Technological Forecasting & Social Change 107 (2016) 97–111

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Consumer information search behavior and purchasing decisions: Empirical evidence from Korea

Seung-Pyo Jun [a,*,1], Do-Hyung Park [b,*,1]

[a] Department of Industrial Information Analysis, Korea Institute of Science and Technology Information, 66, Hoegi-ro, Dongdaemun-gu, Seoul 130-741, Republic of Korea
[b] School of Management Information Systems, Kookmin University, 77, Jeongneung-ro, Seongbuk-gu, Seoul 136-702, Republic of Korea

ARTICLE INFO

Article history:
Received 24 April 2015
Received in revised form 18 March 2016
Accepted 23 March 2016
Available online 6 April 2016

Keywords:
Information search behavior
Purchase behavior
Search traffic
Product classification
Mobile search behavior

ABSTRACT

Recently, Internet activity by consumers adopting innovation or purchasing products has increased markedly. To understand this phenomenon, our study focuses on the correlation between purchase behavior and search activity. Utilizing the product classifications established in previous studies, we classify physical products into durable, nondurable, and industrial goods. We then empirically analyze case studies to determine the correlation between Internet searches and product purchases. Our research results show that the correlation between sales and search traffic is more significant for consumer goods than for industrial goods; furthermore, in the consumer goods category, search traffic is a particularly strong predictor of sales in the case of consumer durable goods. These results may be self-evident, implicit in the definition of each product category. However, the presented findings confirm that even among nondurable goods, search traffic can be a significant predictor of purchases, depending on both price and frequency of purchases. In contrast, for durable goods, search traffic may not be strongly indicative of actual purchases for new products, for which traffic simply reflects rising interest. We also show that PC searches are a stronger predictor of sales than mobile searches. The conclusions drawn from this study provide an important foundation for effectively using search statistics in technology business management to formulate marketing strategies as well as to forecast and analyze the adoption of new technology based on real-time monitoring of the changing involvement with each product.

1. Introduction

Recently, many researchers have analyzed so-called big data in order to empirically verify the correlation between series of behaviors exhibited by consumers and their subsequent purchases. In particular, researchers have focused on performing various types of analyses on the basis of consumers' Internet search activity, as reflected in data such as search traffic. Although this approach has yielded many useful empirical findings, various methodological challenges nonetheless hamper it. For example, critics maintain that some researchers have failed to adequately understand and classify the objects of these searches.

We add to the literature by investigating whether the relationship between search behavior and purchases depends on the product type. We comparatively analyze the relationship between search and purchase behaviors by differentiating between product types, using case studies of products identified as representative of each product category. In this study, we use data provided by Statistics Korea (KOSTAT) and leading search engine websites in Korea, which allow us to perform a more rigorous empirical analysis of differences previously assumed only to exist when consumer search activities are distinguished by product type.

Based on whether a product is tangible or intangible, products can first be divided into goods and services (McDermott et al., 2001; Walsh and Linton, 2011).[2] Goods (physical products) are then further divisible into durable and nondurable goods, according to their duration and form of use. Durable goods generally refer to tangible products that can be used multiple times, such as refrigerators, and machines, whereas nondurable goods refer to tangible products that are generally expended after two or three times of use, such as food and cleaning products. Services are intangible products that are inseparable and are generally highly variable and perishable. Products can also be classified

* Corresponding authors.
E-mail addresses: spjun@kisti.re.kr (S.-P. Jun), dohyungpark@kookmin.ac.kr (D.-H. Park).

[1] These authors have equally contributed to this work, so you can contact anyone about the work.

[2] Alternatively, in innovation studies, tangible products are sometimes referred to as physical products in contrast with services, and the differences are explained in terms of innovation.

http://dx.doi.org/10.1016/j.techfore.2016.03.021
0040-1625/

on the basis of the purpose of use into consumer goods and industrial goods (Kotler et al., 2014; Kotler and Keller, 2007).[3]

In this study, we first classify products into consumer and industrial goods on the basis of their intended use. Then, we classify consumer goods into durable and nondurable goods on the basis of their durability and utility. Next, we select a representative product for each of these categories and comparatively analyze the relationship between search activity and purchasing behavior in Korea. In addition, we account for the rising interest in or the effects of concentrated marketing efforts that occur when a new product is launched. In the case of consumer durables, which exhibit a strong correlation between searches and purchases, we further examine the possible effects of a "hype cycle" that may exist when a new product is launched.

Our study is of use to those that wish to utilize consumer search-activity data since our conclusions offer a framework for directing various activities in technology management, and marketing. A company's efficacy in market information processing—involving the gathering, sharing, and use of consumer information related to the market—plays an important role in determining the success or failure of its products (Ottum and Moore, 1997). In general, companies implement marketing strategies with the aim of transforming a low-involvement product into a high-involvement product (Kotler et al., 2014). This study thereby offers a method for monitoring the involvement of individual products in real time and tracking how products shift from being low-involvement to high-involvement products. The conclusions drawn from this study would have practical value if they are selectively applied to developing effective marketing methods.

The remainder of the paper is organized as follows. Chapter 2 introduces the theoretical background of our research by briefly explaining how products are classified. We also outline the respective theories pertaining to purchase behavior and information searches and review previous studies in this field. Chapter 3 explains our research model, the cases we examine, and the methods used to collect data. Chapter 4 presents our research results, and Chapter 5 addresses issues related to this study that require further discussion, including mobile search traffic and forecasting using the vector error correlation model (VECM).

2. Theoretical background and hypotheses

2.1. Product type depending on product characteristics

Marketing managers typically classify products (consumer and industrial goods) according to a range of criteria, such as durability, tangibility, and purpose, in order to establish an optimal marketing mix strategy (Murphy and Enis, 1986). At the broadest level, products can be divided into consumer and industrial goods according to their intended use. The term "consumer goods" refers to products purchased by general consumers for their own specific purposes, whereas the term "industrial goods" refers to intermediate goods such as raw materials and parts that are input into the manufacturing process to make final products. Industrial goods can be further classified according to the manner of their input and relative cost into subcategories, such as materials and parts, capital goods, supplies, and services.

Consumer goods can also be classified into durable and nondurable goods. Durable goods refer to physical (tangible) goods expended over the long term as the convenience obtainable from their use gradually decreases; in contrast, nondurable goods refer to goods consumed after short-term use. This classification applies to consumer goods: sewing machines and electric refrigerators for home or family use belong to the former category, while groceries, soap, and cigarettes are examples of the latter. The boundaries of the classification are somewhat fluid; for example, clothing and books may be kept and used by some consumers for several years, but they are usually classified as nondurable goods. Durable goods usually involve a large amount of human resources for sales and service. Their price is set to include a high margin of profit, and a strong warranty may be offered to appeal to the buyer. In comparison, nondurable goods, such as beer or soap, are purchased more frequently; therefore, these types of goods ought to be made widely available for purchase, priced with a smaller profit margin, and marketed through large-scale advertising campaigns to induce use, with communication aimed to increase consumer preference.

Lastly, goods can be classified as tangible (goods) or nontangible (services); the term "services" refers to the act of providing convenience to people as a commodity. Services do not result in ownership over things and they cannot be transacted in separation from their production, since the production of a service is realized only at the moment its benefits are delivered to a consumer. Services consist in heterogeneous outputs produced upon order, typically realized as the activities of the producer in response to consumer demand. In summary, services are intangible, inseparable, and have a high level of variability and perishability, and thus they require a high level of quality control and demand reliability as well as adaptability on the part of the supplier (Kotler et al., 2014; Sousa and Wallace, 2006). Furthermore, as observed in innovation studies, the process of innovating and adopting services can unfold in a manner very different from the innovation of physical products. For example, while the innovation of physical products proceeds sequentially from product innovation to process innovation (Abernathy and Clark, 1985) in the service sector, this sequence may be reversed (Barras, 1986). We thus limit the focus of this study to physical products (goods).[4] This study aims to verify how well web search traffic data can predict product sales based on distinctions among product types (i.e., durable and nondurable consumer goods).

2.2. Web search traffic information

Can the keywords people use in Internet searches forecast their economic activities? This depends upon what we mean by "forecast." Google Trends recently released real-time statistics on search keywords and many researchers have conducted studies based on these data. The advantage of using search traffic tools, such as Google Trends, is that it makes it possible to analyze trends close to real time: the advantages of using such data have been demonstrated mainly in applications, making near-future, rather than long-term, forecasts.

Researchers have used various indices to measure product sales and innovation adoption. Watts and Porter argued that the technology lifecycle can be measured based on the science citation index, newspaper abstracts, and patents (Watts and Porter, 1997), and there are many ongoing attempts to analyze new and emerging technology based on various information such as research publications and patents (Robinson et al., 2013). By comparison, forecasting or analysis utilizing web search data has become prominent only relatively more recently.

Ginsberg et al. (2009) presented the results of their analysis performed on data obtained from an early version of the Google Flu Trends search engine, used to forecast the current flu level. They presented a computer model able to convert the unprocessed search data into a real-time monitoring system capable of accurately forecasting flu virus activity one to two weeks in advance of the forecasts issued in conventional reports by the Centers for Disease Control and Prevention in the United States. This study earned broad recognition for the wide-ranging potential of using search traffic to make forecasts.

[3] In addition, products have also been classified as high-involvement or low-involvement based on the varying levels of importance or interest perceived by individual consumers who are affected by marketing stimuli. We chose not to adopt a classification based on the level of involvement in our comparative analysis since we analyze search traffic for individual products, in which case the involvement is self-evident by definition.

[4] In the case of services, the name of a service may encompass a broad range of content (e.g., consulting, restaurants, and financial services) and consumers may perform searches for a wide variety of purposes; therefore, it is difficult to consider these searches as representative. Moreover, it is difficult to determine any direct causality between search intent and the purchase. For these reasons, we exclude the category of services from our analysis.

Although the reliability of Google Flu Trends used by Ginsberg et al. (2009) has since been challenged by Butler (2013), the key significance of their work on the utilization of search traffic is its demonstration of the immense speed with which search traffic can detect phenomena and its superior capability for monitoring these phenomena. It also demonstrated the representativeness of the population obtained from searches performed daily by millions of users, the existence of a strong correlation between social phenomena and search traffic, and the possibility of using these data for forecasting.

The research by Choi and Varian (2012) further demonstrated the strengths of demand forecasting made possible by search traffic, which is also one of the goals of the present study. Choi and Varian found that Google Trends was able to improve forecasts of current economic activity over time. Their explanation of economic activities covered categories such as car sales, house sales, retail, and travel. The authors claimed that, in some cases, it is useful to forecast conditions in the present rather than in the distant future, since this helps identify the "turning point" in economic time. For example, an increase in searches for "real estate agencies" in a particular location can indicate a rise in house sales in this location in the near future. This research was therefore an important contribution to the development of a method for using search traffic in making short-term economic forecasts.

The study by Lui et al. (2011) offered an additional example of other potential uses for search traffic. Their research examined the possibility of forecasting candidates' success in the 2008 and 2010 U.S. congressional elections by using search traffic provided by Google Trends. The authors concluded, however, that search traffic is not yet a significant method for forecasting election outcomes. Recently, Vosen and Schmidt (2011) compared the search traffic results provided by Google Trends with the University of Michigan Consumer Sentiment Index and the Consumer Confidence Index (these indices are used to identify consumption trends in the United States) and concluded that Google search traffic can exhibit superior forecasting ability compared with the survey-based indices used previously. This study demonstrated empirically that search traffic can yield forecasts of individual consumption that rival forecasts made through microscopic analysis. Although one limitation of this research was the authors' reliance on category search data, as in Choi and Varian (2012), it was significant that the study classified products into durable goods, nondurable goods, and services when analyzing data from Google Trends; in this regard, it is similar to the approach used in our study. A key difference, however, is that Vosen and Schmidt (2011) analyzed products based on a single index, rather than classifying products more specifically to distinguish between product characteristics.

Kaesbauer et al. (2012) also used category search data from Google Insights to argue that search traffic is useful in forecasting the real estate market. Guzman (2011) used the Google Inflation Search Index and concluded that search traffic is useful for forecasting inflation rates; furthermore, such forecasts can be made 12 months in advance based on Granger causality. This work is specifically relevant to our research because it utilizes indices for durable goods.

Jun (2012a, 2012b) explained technology hype cycles in the case of hybrid cars by comparing web search traffic data with macroeconomic and bibliometric indices. The author argued that search traffic can be used to observe the hype cycles exhibited by users (consumers) or information distributors (news). In addition, Jun et al. (2014a) argued that search traffic for product brands and their attributes can provide superior predictions of demand compared with conventional bibliometric variables such as patents or news.

Although many recent studies have utilized search traffic, the majority has focused on indices designed for a macroscopic view of collective consumer behavior, rather than focusing on individual products. For instance, Jun et al. (2014a)) attempted to analyze a specific product, the hybrid car, but their study focused only on a single product and did not address the issue of product characteristics. Our study will contribute to the literature by taking account of product characteristics before attempting to analyze the sales performance of specific products or the level of interest generated by products, thereby setting the foundation for a more comprehensive understanding of search characteristics.

2.3. Predictability of web search traffic: consumer vs. industrial products

2.3.1. Consumer goods

The study of consumer behavior examines the ways in which individuals, groups, and organizations select, purchase, use, and process certain products, services, ideas, or experiences in order to satisfy their primary and secondary needs. Consumer behavior is affected by cultural, social, and individual factors, and various models have been proposed for understanding consumer behavior. One such model, the stimuli–response model, posits that marketing and environmental stimuli first enter consumer consciousness. This psychological process interacts with the individual attributes of consumers, as he or she enters the decision-making process and is induced into making the final purchase decision. This consumer behavior model seeks to elucidate the process that unfolds within consumer consciousness from the time he or she receives external marketing stimuli to the point of making the ultimate decision to purchase (Hawkins and Mothersbaugh, 2009).

When the purchase decision-making process is examined based on this consumer behavior model, it can be broadly divided into five stages: 1) problem awareness, 2) information search, 3) evaluation of alternatives, 4) decision to purchase, and 5) actions following the purchase. This model emphasizes that the process leading to the purchase is initiated long before the actual act of purchasing and that its results linger long after the purchase has been made. However, not all consumers who purchase a product pass through all five stages in all cases. Some consumers may simply skip over a certain stage, and some move through the stages in the reverse order. As in the consumer behavior model, differences in cultural, social, and personal factors inevitably result in diversifying the forms of adoption exhibited by consumers, especially in terms of their consumption of new products (Kotler et al., 2014).

According to the five-stage purchasing process model described in Fig. 1, the purchasing process begins when the buyer first becomes aware of a problem or need. This need can be generated by internal or external stimuli. In the second stage—information searching, which is the focus of our study—consumers often search for a surprisingly limited range of information. With regard to these searches, we can distinguish between two levels of involvement (Kotler et al., 2014). First, individuals who perform information searches with relatively weak intensity are in a state of heightened attention. Consumers at this level simply become more adoptive of information on a specific product. Second, some individuals engage in information searches more actively, for example by reading articles, discussing information with friends telephonically, using the information available on the web, and even visiting a store to learn more about a product. Such individuals typically belong to the high-involvement learning state.

Meanwhile, consumers generally obtain information from four types of sources: (1) personal sources (e.g., family, friends, and acquaintances), (2) commercial sources (advertisements, websites), (3) public sources (e.g., the mass media), and (4) experience sources (i.e., experience gained by using or researching a product). Recently, consumers have increasingly performed searches based on the opinions or recommendations available online, which tend to combine the characteristics of all four information sources listed above (Kotler et al., 2014; Niranjanamurthy and Kavyashree, 2013).

2.3.2. Industrial goods

As the industrial structure continues to become more complex and segmented, the scale and importance of the industrial goods market has increased. Among scholars, this has increased the necessity of researching industrial goods; more specifically, led to greater emphasis on researching industrial goods purchase behavior (Anderson et al.,

Fig. 1. Five-stage purchasing process model.
Source: Kotler et al. (2014).

1987). Despite the recognition of its importance, however, relatively few scholars have studied the purchase behavior of the buyers of industrial goods. Compared with studies of consumer goods purchase behavior, such studies remain at the conceptualization stage, with inadequate attempts at empirical verification (Anderson et al., 1987; Anderson and Chambers, 1985; Johnston, 1981; McQuiston, 1989; Shaikh and Hansotia, 1985).

The behavior of industrial-good buyers refers to buyers representing an organization, who make purchases to produce certain products or services intended to be supplied, lent, or sold to other people. Such behavior has been found to differ from that of the buyers of consumer goods, and the "theory of buyclasses" developed by Robinson et al. (Robinson et al., 1967) provides one conceptual framework in this respect. Although relatively simple, this theoretical model offers a useful perspective for understanding the buyers of industrial goods and it can be credited with proposing various hypotheses that can be subject to empirical testing (Anderson et al., 1987).

According to the theory of buyclasses, purchase situations can be classified into the three types shown in Table 1, namely a new task, a modified rebuy, and a straight rebuy. This taxonomy is based on three dimensions: information needs, consideration of alternatives, and newness of the task. The theory of buyclasses describes the characteristics of each purchase situation as follows. In a new buy situation, buyers lack experience from past purchases, and therefore organization members involved in the purchase process perceive a high level of risk. Accordingly, organizational members will actively search for a large amount of information and carefully evaluate multiple alternatives before deciding to buy. In this new buy situation, a large number of organizational members are involved in the purchase decision, including technological experts, who have professional knowledge or information that can reduce the perceived risk and thus have a relatively strong influence compared with purchasing officers. Hence, making the purchase decision becomes a lengthy and complex process.

In comparison to the more rare incidence of new purchases, a straight rebuy occurs most frequently in an organization. Here, since buyers have accumulated experience from past purchases and developed a routinized purchasing procedure, organizational members tend to have a lower perception of risk. Rather than searching for new information, they tend to buy from their current supplier; hence, the organizational members who make the purchase decision are limited to the purchasing department or purchasing officers. The decision-making process also becomes more simplified, with the omission of several stages. During this period, price is usually considered to be important for evaluating purchases, as long as buyers are assured of timely delivery and consistent performance.

Finally, modified rebuy represents a midrange purchase situation, somewhere between a new task and a straight rebuy. A product purchased as a straight rebuy may have been enhanced to perform with better precision or technological characteristics and thus be regarded as a somewhat new product; alternatively, buyers may have accumulated more buying experience and thus gained more familiarity about a product that belongs to a new product purchase.

A buyer of industrial goods first determines the types of products and services that need to be purchased, finds and evaluates multiple suppliers or brands to meet this need, and, finally, makes the optimal selection. Such a purchase decision is thus typically complex, despite having a routinized buying procedure. It should be noted that, when purchasing industrial goods, the buyer tends to cooperate closely with the seller in a long-term relationship (Kotler et al., 2014). Indeed, industrial goods are often repurchased, in which case less effort is allocated to information collection. In the case of new purchases, however, the buyer does collect a large amount of information. Since the decision regarding the purchase of industrial goods can be affected by a complex mix of factors (e.g., organizational, interpersonal, and personal), interest in a product may not explain a purchase as well as in the case of consumer goods.

2.3.3. Product sales and web search traffic

Many studies have demonstrated that using web searches is an effective way to forecast macroscopic social phenomena. In this study, we predict that search traffic information will explain the sales of products from a specific brand from a microscopic perspective. The studies by Shim et al. (2001) and To et al. (2007) indicated that the intent of web searches plays an important role in the intent to purchase a product. In this study, we build on the framework provided by the buyer behavior model explained above. In particular, we accept the basic proposition that consumers who show interest proceed to information searches and then to the purchase of a product (technology adoption) and that information searches motivated by consumer interest can be measured based on consumer Internet search activity (Vosen and Schmidt, 2011; Jun et al., 2014a; Jepsen, 2007).

This study aims empirically to verify the relationship between actual sales data and actual search data. Although web search traffic is positively correlated with product sales, we predict that the degree to which web search traffic is able to forecast sales depends on the type of product (see Fig. 2).

As indicated above, consumer goods can be more effectively searched for under online shopping categories; therefore, search traffic in terms of shopping categories tends to better explain product sales than industrial goods. By contrast, in the case of industrial goods, search traffic is typically less able to explain product sales.

Regarding consumer goods, as evident from the definitions of durable and nondurable goods, it is important to examine how much product-related information is obtainable. Obtaining abundant information about a product enables a consumer better to predict product quality and reduces the risk inherent in the purchase, allowing consumers to have greater confidence in the product. The web search activities of consumers thus depend upon whether the goods are durable or nondurable as well as on how much information can be obtained and on what kinds of strategies can be used to search for information. From a quantitative aspect, durable goods offer more formalized information and require consumers to deduce the quality of the product from various sources of information; therefore, we can predict that consumers will use a wider variety of information sources in their search activities compared with the case of nondurable goods (Currim et al., 2015). Indeed, it is easy to deduce the quality of nondurable goods through a small number of information searches (such searches may even not be very important when searching for the product). Therefore, we propose the following four hypotheses:

Hypothesis 1. Sales of durable goods are positively correlated with web search traffic on durable goods.

Hypothesis 2. Sales of nondurable goods are positively correlated with web search traffic on nondurable goods.

Table 1
Buyclasses model.

Types of purchase situations	Newness of the problem	Information requirements	Consideration of new alternatives
New purchase	High	Maximum	Important
Modified rebuy	Medium	Moderate	Limited
Straight rebuy	Low	Minimal	Unnecessary

Hypothesis 3. Sales of industrial goods are positively correlated with web search traffic on industrial goods.

Hypothesis 4. Effects of web search traffic on product sales are stronger for consumer goods than for industrial goods.

2.4. Hype cycle predictability of web search traffic

In addition to the product's characteristics, its lifecycle might also affect information research activities reflected in search traffic. The technology hype cycle forms the theoretical basis of this argument.

In general, when a new technology is introduced, the technology hype cycle model is used to explain the process by which the expectations of that technology evolve and it becomes established in the market and utilized by companies. The phase-by-phase technology hype cycle presented in Fig. 3 offers the following observations. First, in the Technology Trigger phase (i.e., the technology generation or incipient phase), the technology commodity emerges based on its potential. In this phase, however, although the technology receives attention from the media, it may appear to have limited merchandising potential or may fail to become commercialized.

Second, the Peak of Inflated Expectations phase (bubble phase) is the period of heightening interest when numerous initial success stories are publicized but few companies participate. In this phase, the media report unrealistic and excessive market forecasts regarding the technology's success.

Third, the Trough of Disillusionment (disillusionment phase) is the phase during which the hype rapidly declines owing to falling interest in the experiment results or the failure of commercialization. During this phase, the technology must be developed into a commodity that can satisfy early adopters if it is to be able to secure continued investment. This is a period of realistic readjustment marked by a rapidly declining curve, and the media lose interest aside from expressing suspicions regarding the technology.

Fourth, the Slope of Enlightenment (stabilization phase) occurs when a wider understanding can be gained regarding the specific means by which the technology in question will generate a profit, and a second or third generation version that represents an improvement over the initial commodity may make an appearance.

Finally, in the Plateau of Productivity (growth phase), the technology's commercial viability is recognized and its advancement into a broad market can take place (Jun, 2012a; Gartner, 2011).

The hype cycle can be distinguished from the lifecycle in that when a new technology emerges and is evaluated to have potential for applicability (Technology Trigger), expectations of both the market and consumers rapidly rise and reach a peak (Peak of Inflated Expectation). However, after the majority of new technologies peak, they begin to be disseminated more broadly. Hence, a gap arises between expectations and level of actual satisfaction, resulting in the bubble's collapse (Trough of Disillusionment). This subsidence of the bubble and return of expectation level to its original point are attributed to technological problems in the new technology itself and deficiencies in the related infrastructure required for its implementation. In this way, the hype cycle originates in the "marketing hype" that explains the negative effects of excessive marketing, or, in other words, excessive exposure. Therefore, the visibility of the hype cycle brings about the rapid bubble phase

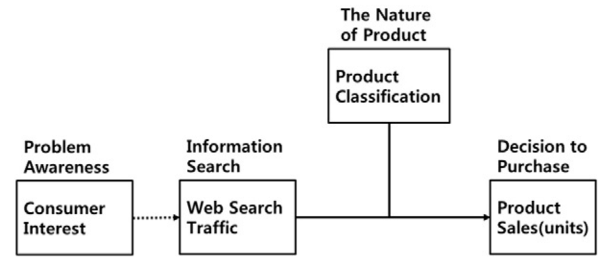

Fig. 2. Research model.

arising from the technological vision or the media, and such visibility becomes hyped according to the content and amount of exposure (Jun, 2012a; Gartner, 2011).

This hype cycle model provides the theoretical basis that we use to argue that the correlations between search traffic and sales for each type of product observed specifically during the introductory phase (when consumer interest is unusually heightened) may differ from the correlations observed generally. Taking this into account, when examining the correlation between search traffic and sales with distinctions between types of products, we also examine the possibility of a hype cycle occurring during the introduction phase for specific types of products for which search traffic was found to be highly indicative of sales. Hence,

Hypothesis 5. In the case of new durable goods that undergo a hype cycle, web search traffic cannot be used to predict product sales.

3. Research methodology and case study

3.1. Data collection and case study

For our case studies of durable consumer goods, we first selected digital video recorders (DVRs) and kimchi refrigerators. These two products were chosen because of their relatively high price and duration (i.e., they continue to be used for two to three years or even longer once purchased). We also picked beds as a typical example of durable goods, as well as fishing rods, which are slightly different in character from the other products but have the common characteristic of being an infrequently repurchased item.

First, a DVR is a device widely used to monitor security, such as in a parking lot or areas inside or outside a building. It is similar to CCTV in its intended use, but one feature that distinguishes a DVR is that it stores videos on a computer or in the hardware of a separate terminal. In response to increased concern over security issues, more consumers have recently begun to purchase DVRs, and we chose this product as a useful case study for comparison with other durable goods. Kimchi refrigerators, which have become a widely popular appliance in Korea, are designed especially for storing kimchi, a staple of Korean cuisine, with the function of controlling the rate at which kimchi ferments. Both these electronic products were selected because they have only recently become widely disseminated compared with more conventional electronic products (such as ordinary refrigerators and televisions). Second, beds are durable goods distinct in character from electronic products. Beds are included in our study because demand for beds is not anticipated to show particular seasonal fluctuations. Lastly, fishing rods were also selected because of the recent increase in leisure activities among Koreans.

Nondurable goods are tangible products usually consumed after a couple of uses and frequently repurchased. As nondurable goods for our analysis, we selected several types of alcoholic drinks and bottled water. Soju, a type of Korean distilled spirit made by boiling alcohol from grain or other starches such as sweet potatoes, is one of the most widely consumed alcoholic drinks in Korea. Bokbunjaju (hereafter

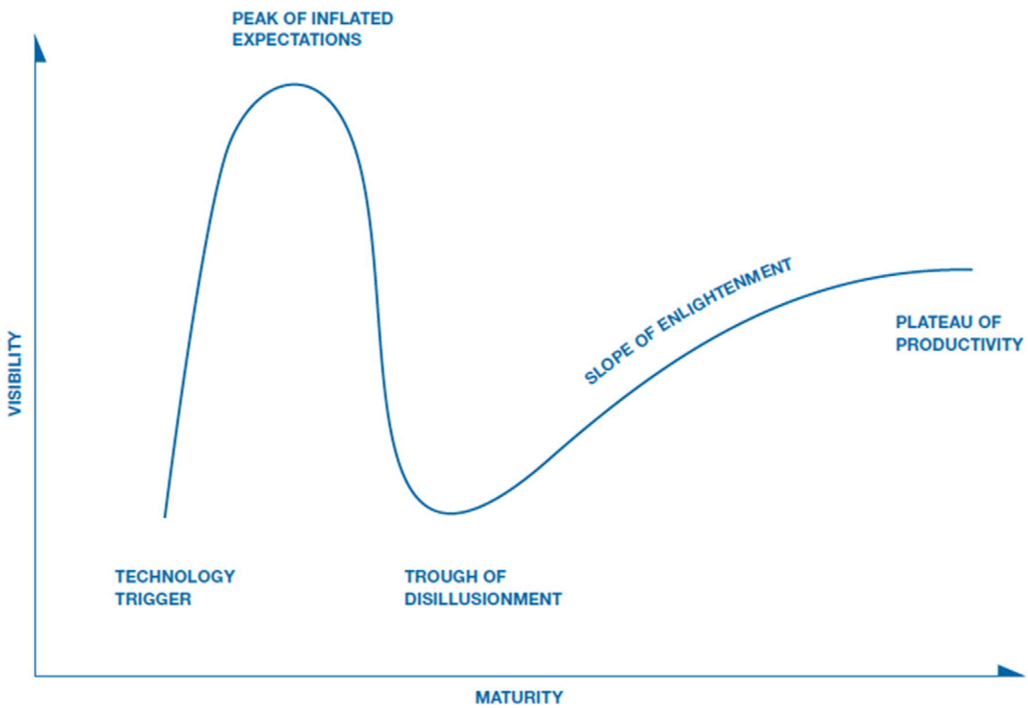

Fig. 3. Gartner technology hype cycle.
Source: Gartner (2011).

referred to as raspberry wine) is a fruit wine made by fermenting *Rubus coreanus* (a type of raspberry native to Korea). This wine has recently gained popularity as a result of the rising interest in personal health among Koreans and has also seen a successful increase in export sales. Lastly, whiskey is a relatively high-priced distilled alcoholic beverage that is not native to Korea. We included whiskey in our analysis because it is less frequently purchased and consumed compared with the other alcoholic beverages. In addition, we also chose bottled water, which we regard as highly typical of nondurable goods.

Lastly, industrial goods refer to products used for industrial or commercial purposes. For our case studies, we chose uninterruptible power supplies (UPS),[5] inverters, machining centers, and synthetic rubber. UPS provide continuous power supply in the event of a power outage by supplying electric energy from storage batteries via a transformer. They are widely used in places that require a stable and reliable power supply, such as communications companies, hospitals, and broadcasting companies. An inverter is a device that changes direct current (DC) into alternating current (AC). A machining center is a single, combined processing machine that includes the functions of a boring machine, milling machine, and drilling machine. With a single setting, the machining center can perform multi-axes machining and multi-process machining. It is therefore widely used for automating the processing of parts manufactured in small quantities and large varieties. We included the machining center in our case studies because it can be more aptly characterized as capital goods in comparison to UPS and inverters, which have more of the character of parts or supplies. Lastly, synthetic rubber refers to any type of artificial elastomer mainly synthesized from petroleum byproducts. We chose to examine synthetic rubber because it is closer in character to raw materials than the other industrial goods included.

In addition, we also selected hybrid electric vehicles (HEVs) as a case of (new) durable consumer goods. Our purpose here was to observe the possibility of hype, but we were limited to analyzing a single product because of the challenges inherent in collecting sales statistics for new durable consumer goods. HEVs are next-generation eco-friendly cars that have significantly reduced emissions of harmful gases and improved mileage compared with conventional cars through innovations, most notably the inclusion of both an internal combustion engine and an electric car battery engine. In Korea, Hyundai and Kia Motors released the country's first HEVs in July 2009, launching LPG HEVs named the Avante Lpi (Hyundai) and Forte Lpi (Kia). With the help of governmental support and aggressive marketing, these cars rapidly gained market success. As of September 2012, their market share in terms of monthly new car sales grew to 2.5% and cumulative sales (for cars made in Korea) reached 50,000 cars (Jun et al., 2014a).

Our source of sales data for all these products (with the exception of the HEV, a new product) was the monthly sales data made available by the Korean government through Statistics Korea. These data are highly credible and made available quickly. For new products, however, research data were often not yet available; we were thus somewhat limited by the fact that sales data become available only after the market adoption of the products has proceeded to some degree. For this reason, we did not have access to sales data for recently launched products such as HEVs. Sales data for HEVs were therefore obtained from the statistics on sales of domestically manufactured and imported cars collected by individual associations. For search traffic data, we used the search traffic on Naver, as this search engine dominates the Korean search market with a share of nearly 80% (Koreanclick, 2012). We accept Naver search traffic as an index of consumer behavior because Naver has already gained a monopolistic position in the market. Although producers and researchers may also use Naver for searches, the majority of Naver users can be assumed to consist of consumers without access to a specialized database (Jun et al., 2014b). Table 2 summarizes the information on our data sources.

The conventional method of analyzing time series data is the time series decomposition method, which decomposes and analyzes each factor that influences any changes in time series data. Factors that

[5] Uninterruptible power supplies are also commonly referred to as UPS (note that among Koreans, UPS is not associated with the name of a postal delivery service as in the United States). The use of the abbreviation UPS is already common practice, and for this reason, we simultaneously included both forms of the name for our comparison of search traffic and sales.

Table 2
Data sources for the major variables.

Variables	Site	Explanation
Search traffic (KOR)	trend.naver.com	Weekly and monthly search traffic in the United States (2007–present)
Product sales (excluding HEV)	kostat.go.kr (Statistics Korea)	Monthly sales for each product (2000–present)
HEV sales	kama.or.kr kaida.co.kr	Monthly unit sales of HEV (2004–present)

influence changes in time series data include 1) the trend component, 2) the seasonal component, 3) the cycle component, and 4) the irregular (remainder) component (Dodge et al., 2003). Search data and sales data used in this study both consisted of monthly time-series data, which can exhibit seasonality. For this reason, we divided the monthly data into the four components above and then applied trend and cycle factors, which are important for mid- to long-term forecasting, to analyze the correlations among the data. The seasonal component was eliminated partly in order to compare the long-term trends; however, some of the major variables also exhibited a seasonal component, which can create noise when examining the significance among these variables. The trend-cycle estimation performed after excluding the seasonal adjustment component and irregular component (errors) was used to convert the time-series data for all monthly variables (Jun et al., 2014a).

4. Research results

4.1. Relations between search traffic and purchase behavior

4.1.1. Durable goods

As explained above, to analyze the relation between searches for durable consumer goods and the sales of these goods, we selected two electronic products, namely kimchi refrigerators and DVRs, as representative examples of durable goods. In addition, we included fishing rods and beds, which are not electronic products, to perform a comparison across different types of durable consumer goods. The DVR, in particular, is an example of an industrial product relatively new to the market, with data from Statistics Korea dating back only to the mid-2009.

Fig. 4 shows that, regardless of the product characteristics, search traffic corresponded relatively closely to sales trends for the majority of durable consumer goods. Overall, among the four durable consumer goods selected, search traffic was found to be a relatively strong predictor of sales. Moreover, the difference between electronic and non-electronic products was small. In the case of DVRs, we found that search traffic was active even before the collection of statistics began. Here, although the trends in search traffic and sales were somewhat similar, the peak time for searches preceded that for sales. Hype can occur if a product is newly launched, or was launched previously but only recently gained sudden popularity, and in such cases, the search peak can occur in advance of sales.

4.1.2. Non-durable goods

To analyze the correlation between the search traffic and sales performance of nondurable consumer goods, we selected alcoholic beverages (soju, whiskey, and raspberry wine) and bottled water as the subjects of our case studies. As presented in Fig. 5, in the same category of alcoholic beverages, search traffic closely reflected sales. In the case of the relatively high-priced whiskey and newly-popular raspberry wine, search traffic generally corresponded closely to sales. Raspberry wine was particularly distinguishable from the other consumer goods in that its popularity is a comparatively recent phenomenon, with statistics for this product only available since 2004. In the case of raspberry wine, search traffic peaked in advance of the peak in sales, as in the case of DVRs. As for soju, which is relatively cheap and is consumed in quantities ten times greater than the consumption of other alcoholic beverages, sales increased continually but search traffic showed a decline, which could reflect consumer familiarity with the product. In the case of bottled water, the escalating interest in health was reflected in the increase in searches; however, demand remained relatively stable, with no significant increase.

The correlation between search traffic and sales is expected to differ between durable consumer goods, which buyers select with a relatively high degree of care (e.g., collecting a lot of information) and nondurable consumer goods, for which there is generally little need to search for information. For the durable consumer goods discussed above, we found that search traffic provided a relatively strong explanation of sales in the case of electronic products, a category that consists in high-priced items used for a relatively long period. However, in our representative cases of nondurable consumer goods including soju and bottled water, which are repurchased frequently and consumed within a short period, search traffic failed to be a useful predictor of sales. However, among alcoholic beverages, which can be considered to belong to the category of nondurable consumer goods, in the cases of products that are relatively high in price, consumed in smaller quantities, and purchased infrequently (e.g., whiskey), we found that searches can be a predictor of purchase behavior. Furthermore, although a time lag can be observed between searches and purchases in the case of raspberry wine, searches were confirmed to be a significant predictor of purchase behavior. In sum, search traffic exhibits a meaningful correlation with purchases, even for nondurable consumer goods, with variations depending on factors such as the newness of the product, frequency of purchases, and price.

4.1.3. Industrial goods

In contrast to consumer goods, industrial goods are sold in a B2B fashion and thus buyers often already have a relatively large amount of information. Therefore, search traffic is not expected to be a strong predictor of sales compared with consumer goods. To verify this expectation, we examined the correlation between sales and search traffic in the cases of UPS, inverters, machining centers, and synthetic rubber, for which governmental statistics are available from Statistics Korea and which serve as examples of industrial goods that tend to be sold individually. We chose to include synthetic rubber, which is distinct in character from the other three types of industrial goods, because it is a material used to manufacture other products.

As shown in Fig. 6, the total volume of actual searches performed on Naver was ten times higher for the keyword “UPS.” Search traffic, however, was not a strong predictor of sales: although search traffic increased steadily, domestic sales actually declined from 2010. Synthetic rubber and inverters also exhibited similar trends. Search traffic for synthetic rubber was relatively high in 2011, while a decrease in sales was observed after 2010. The correlation was weak in the case of inverters as well. Likewise, we had difficulty finding a relation between search trends and sales in the case of machining centers. Search traffic decreased noticeably after 2009, whereas there was a strong increase in sales from 2010 to 2012. These findings met our expectations.

4.1.4. New durable goods

As examined above, search information tends to be a better predictor for consumer goods than for industrial goods; moreover, among consumer goods, it performs most strongly in the case of durable consumer goods. However, even for the case of durable consumer goods for which search traffic correlates strongly with sales, the results may differ for repurchases compared with the cases of new purchases, as discussed above (Jun and Kim, 2011). In the case of newly launched items, information search activities in response to the newness of the product or service may overlap with search activities intended for purchase, and these searches may occur far in advance of purchases. As in the case of DVRs, discussed above, we can often observe the existence of hype. To verify this probability, we selected HEVs as an example of a recently emerged new durable consumer good. Fig. 7 compares HEV sales in Korea with their search traffic.

In the case of HEVs, an example of a relatively high-priced durable good, we observed that even during low-purchase periods, search traffic was significantly higher than sales; sales and searches also followed a similar pattern. However, after 2012, search traffic noticeably increased. After the initial information searches on HEVs petered out and fewer HEVs were being repurchased, the volume of search traffic dropped considerably.

4.2. Empirical results

4.2.1. Unit root tests

Before examining causality to test our hypotheses, we performed unit root tests to verify that the variables were stationary. We applied the commonly used augmented Dickey–Fuller and Dickey–Fuller GLS tests (Mackinnon, 1996), and the results of these unit root tests are presented in Tables 3 and 4.

First, Table 3 shows the unit root test results for the sales of each product. Unit roots were found to be present in a level variable for the majority of products, based on the results from at least one test method. Therefore, the sales time series of the products are nonstationary and it would require a first- or second-order difference to make the series stationary; thus, using a difference variable to determine significance in regards to sales would improve the accuracy of our findings.

Next, Table 4 presents the unit root test results for the search traffic of each product. In the level variable, we likewise found no cases in which unit roots were not present in the results from either test method. Therefore, for the search traffic time series of the products, we again found it necessary to use a variable with a first- or second-order difference to make the series stationary.

4.2.2. Cointegration analysis

In the unit root test results presented above, sales and search traffic were observed to be nonstationary for the majority of products, and therefore we decided to use a difference variable to examine causality based on the vector autoregressive regression (VAR) model. First, however, we examined whether there was a long-run stationary relationship between the two variables. The studies by Granger (1988) and Bahmani-Oskooee and Alse (1993) showed that when unit roots are present in the data and the two variables are in a long-run stationary relation, it may be invalid to infer causality based on the usual Granger Test using a difference variable. These studies demonstrated that the results have stronger validity if the causality test is performed using an error correction model that takes into account the long-run balance between variables. Therefore, we found it necessary to perform cointegration tests as a second stage to determine whether the two variables have a long-run stationary relationship. The cointegration test methods used in previous studies include those proposed by Engle and Granger (1987), Engle and Yoo (1987), and Johansen (1988); among these, the Johansen cointegration test is a representative method that is widely accepted and used. If times series that have unit roots are cointegrated, the obtained regression coefficients' estimations will have consistency, and therefore we can use time series that have unit roots to build a forecasting model supported by econometric theory. Compared with other test methods, the advantage of the Johansen cointegration test is that it estimates the number of cointegration relations and parameters of the model by using the maximum likelihood estimation. Since all variables are regarded as endogenous, we do not need to select a dependent

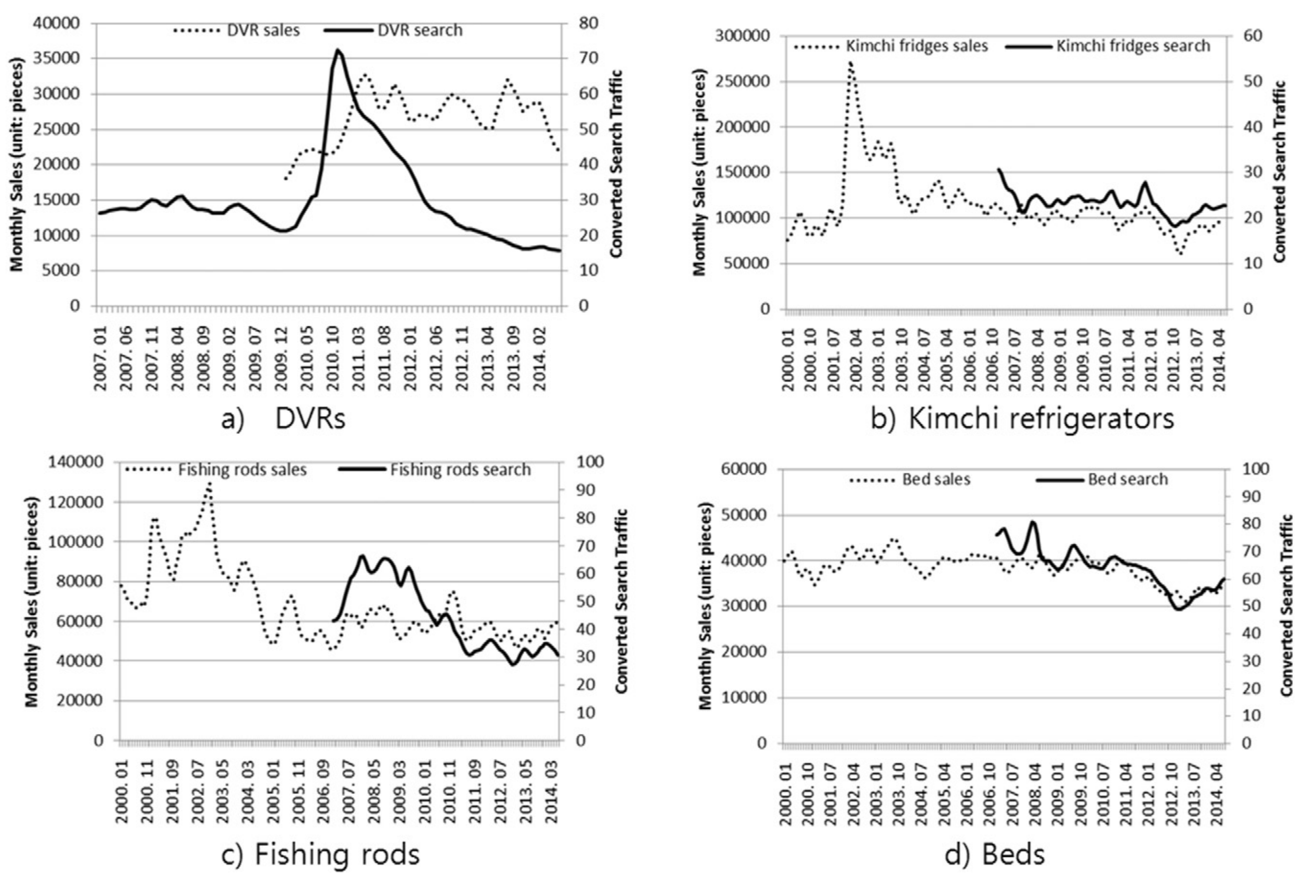

Fig. 4. Comparison of search traffic and monthly sales of durable consumer goods.

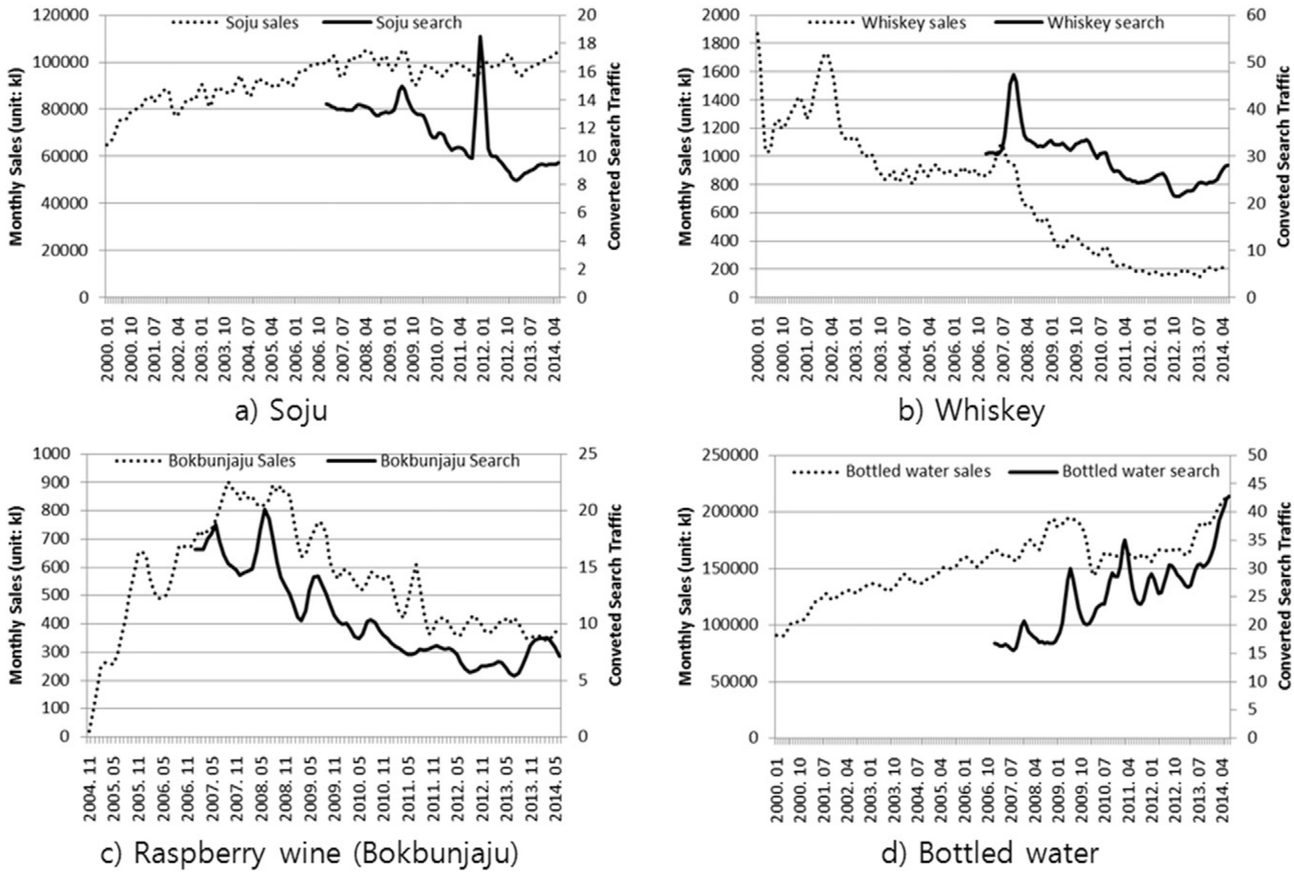

Fig. 5. Comparison of search traffic and monthly sales of nondurable consumer goods.

variable and are thus able to identify multiple cointegration relations. In addition to using the maximum likelihood estimation method to estimate the cointegration relationship based on the VAR model, we also used the likelihood ratio test as the basis for determining the cointegration coefficient. The advantage of this approach is that we can do more than simply test cointegration; when cointegration is found to be present, we can also estimate the cointegration parameters as well as perform various other tests of hypotheses that are relevant to establishing the model (MacKinnon et al., 1999). In this study, we also used the Johansen cointegration test to test whether there were long-run stationary relationships between each of the variables; when cointegration was observed, we applied the Johansen cointegration test to verify our hypotheses.

In this study, in preparation for performing the Johansen cointegration test, we first set up a VAR model in the levels (not the differences) of the data and then determined the appropriate maximum lag length for the variables in the VAR. If we set up an overly extended lag (order) length, this would reduce the serial correlation of the error term but it would also result in a somewhat lower efficiency. Taking account of this trade-off, this study sets the maximum lag (order) length for the VAR model to below 4. Table 5 shows the cointegration test results obtained by applying this lag length. In Table 5, if cointegration is present, we also show the normalized cointegrating coefficient. This information allows us to determine whether the effects of two variables are positive or negative when cointegration is present in a relationship.

As shown in Table 5, we found cointegration relationships for some of the hypotheses, but not for others. For all durable goods products, with the exception of HEVs, we found one or more cointegration equations, confirming that searches and sales have a long-run relationship. However, the majority of industrial goods and nondurable goods did not show cointegration relationships and therefore we were unable to affirm the existence of long-run relationships between the two variables. We were able to verify the existence of a long-run relationship only in the case of raspberry wine in the category of nondurable goods and for UPS in the industrial goods category.

4.2.3. Hypothesis testing

If a cointegration equation was found to be present in the cointegration analysis, then differencing the raw data would result in a problematic loss of significant data. For this reason, we performed the Granger causality test without differencing. In other words, we performed the Granger causality test to confirm the lead/lag structure in regards to the causality verified in the cointegration analysis. If cointegration had not been confirmed, we took account of the unit root test results and used differenced data when performing the Granger causality test.

Table 6 presents the Granger causality test results for cases in which a cointegration equation was found. If cointegration exists, then Granger causality is also expected to exist. It is important to note, however, that Granger causality cannot inform us if a relation is positive or negative. Therefore, to determine the direction of the effects in Granger causality or in the cointegration equations, we also examined the normalized cointegrating coefficients, as presented in Table 5. The results in Tables 5 and 6 show that while Granger causality was observed in UPS, when search traffic increased by one unit, there was a decrease amounting to 4.6 units of UPS. Therefore, we had to reject our research Hypothesis 3.

Table 7 presents the VAR Granger causality test results for searches and sales pertaining to products that do not have

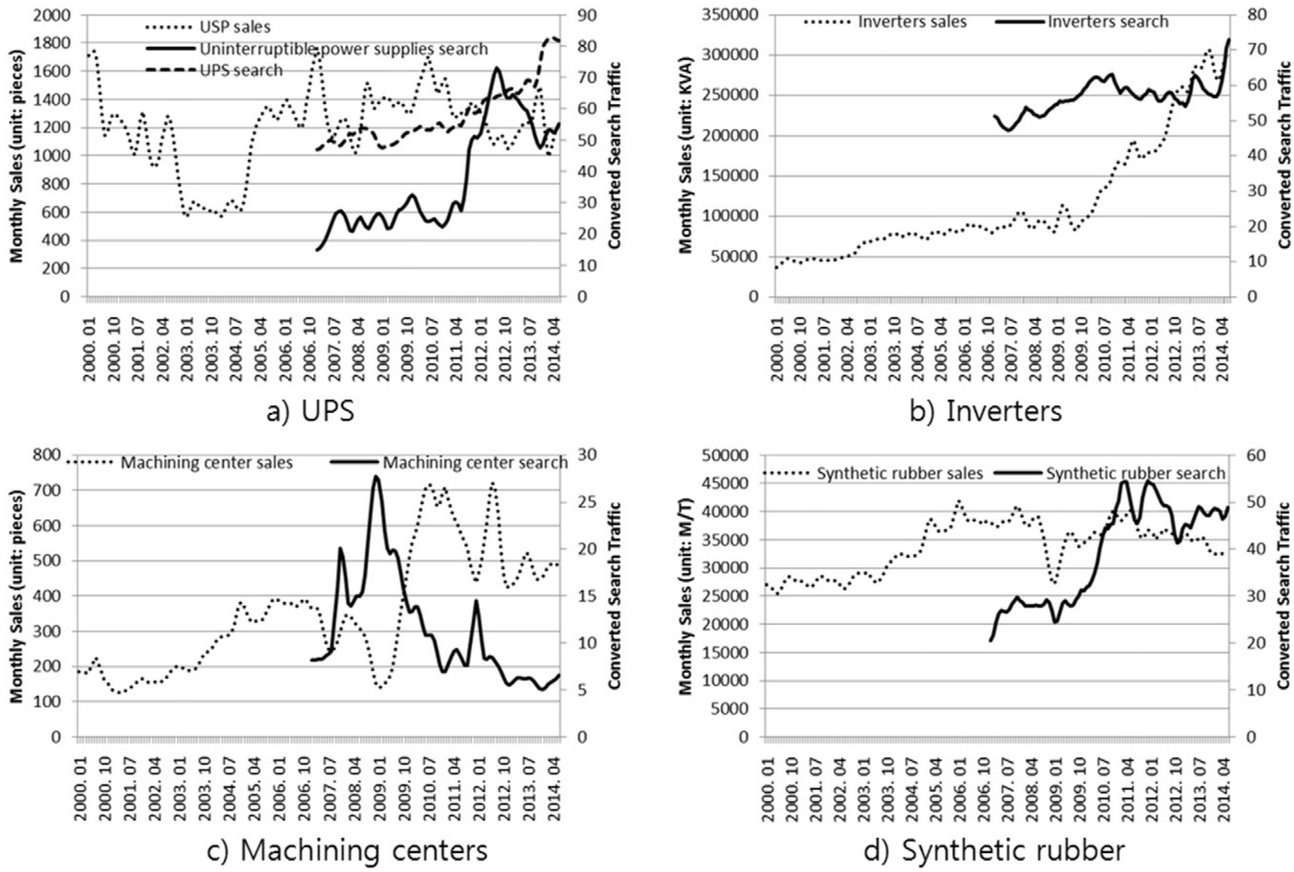

Fig. 6. Comparison of the search traffic and monthly sales of industrial goods.

cointegration. Because the sales and search variables were observed to be nonstationary, as discussed on our analysis of Tables 3 and 4, we determined the VAR model by using differenced variables and then performed VAR Granger causality/block exogeneity (GCBEW) tests by using this model. Our results showed that in the optimal VAR model, whiskey was the only case in which we can reject the null hypothesis of excluding the search traffic variables from the sales product equation. In the case of whiskey, search traffic can have a significant effect relevant to the explanation of sales. In the case of the other products, we failed to reject the null hypothesis of excluding the search traffic variables from the sales products equation at the 0.100 significance level.

Let us now summarize the Granger and GCBEW test results described above in relation to our study's hypotheses. We found that in the category of durable goods, search traffic and sales exhibited statistical significance in the majority of our case studies, with the exception of HEVs, a new product. For nondurable goods, although we found that search traffic had statistically significant explanatory power in the case of some products, such as raspberry wine and whiskey, it was difficult to find any statistical significance in the

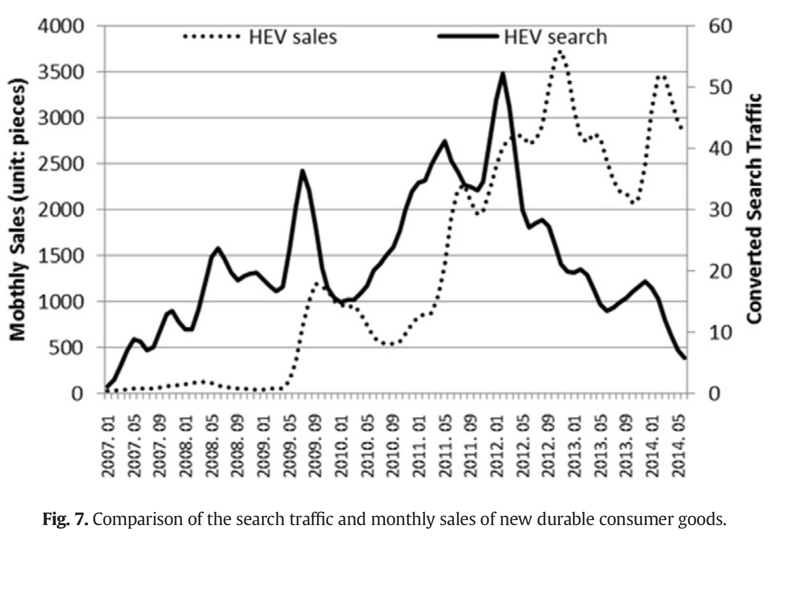

Fig. 7. Comparison of the search traffic and monthly sales of new durable consumer goods.

Table 3
Results of the unit root tests for sales data.

Test type	Augmented Dickey–Fuller (ADF)			Dickey–Fuller GLS (ERS)		
Test for unit root	level	1st difference	Test for unit root	level	1st difference	Test for unit root
Series						
Kimchi fridges	−1.913	−4.210***	−5.433***	−0.971	−2.836***	−11.717***
Fish rods	−2.697*	−4.955***	−7.220***	−0.892	−4.985***	−1.233
Bed	−1.201	−2.792*	−9.631***	−0.640	−1.184	−8.785***
DVR	−3.649***	−1.754	−4.333***	−0.763	−1.156	−0.137
Whisky	−4.280***	−3.492**	−6.089***	−0.531	−3.102***	−1.297
Raspberry wine	−0.872	−4.860***	−5.720***	−0.820	−0.624	−0.112
Soju	−2.461	−3.497**	−5.952***	−2.479**	−1.269	−5.621***
Bottled water	−1.576	−1.867	−5.354***	−1.615*	−1.411	−5.376***
UPS	−1.809	−5.821***	−5.965***	−0.621	−1.001	−0.653
Inverters	0.562	−2.992**	−7.377***	0.841	−2.827***	−2.864***
Synthetic rubber	−1.901	−4.307***	−5.841***	−1.159	−3.310***	−2.368**
Machining centers	−1.499	−2.905**	−3.286**	−1.193	−2.917***	−3.117***
HEVs	−0.700	−4.142***	−5.837***	0.1723	−3.912***	−5.195***

* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

relation between search traffic and sales for more typical nondurable goods such as soju and bottled water. Likewise, in the category of industrial goods, we had difficulty finding any statistically significant correlation between search traffic and sales in any of our case studies. This finding supported our fourth hypothesis, which proposed that search traffic has a more statistically significant relationship to sales in the case of consumer goods than industrial goods. Lastly, with regard to our fifth hypothesis, we confirmed that although search traffic has a strong explanatory relationship with sales in the cases of many durable goods, the explanatory power of search traffic may be weaker for newly-released products.

5. Discussion

Although various challenges may require us to qualify the results of this analysis further, we first address the issue raised by the recent growth in mobile search activities. In our primary research, we limited our source data to search traffic from PC users; however, in this section, we discuss the issue of mobile searches based on the empirical observations made in this study. In addition, we use the VECM to perform a causality analysis to analyze the case in which search traffic can be used to forecast sales. Our purpose is to demonstrate the potential for the utilization of search traffic (Ozkan et al., 2012).

5.1. Mobile vs. PC search traffic

One recent development in this area is that consumers are increasingly performing information searches on their mobile devices. We analyzed whether such mobile searches also serve as a strong predictor of sales. Specifically, we compare mobile and PC search traffic for the four products selected previously, namely DVRs, kimchi refrigerators, fishing rods, and beds.

The results shown in Fig. 8 indicate that, overall, mobile search traffic does not have a significant correlation with sales. In the cases of kimchi refrigerators, fishing rods, and beds, search traffic continued to increase steadily, and thus diverged from the sales trend as well as that for PC search traffic. Although the results for DVRs were slightly more similar, the peak in mobile traffic lagged behind the peaks of PC search traffic and sales.

In the cases of the durable goods examined above, we can thus reasonably presume that the desire to conduct new information searches is low since the products have become somewhat familiar and that more searches for information relate specifically to repurchases. However, the question remains regarding how the results would differ in the case of new products in the market. Fig. 9 illustrates the results from comparing mobile/PC search traffic with sales in the case of HEVs, showing that mobile search traffic exhibited similar trends to PC search traffic. Indeed, the two types of search traffic for information on HEVs

Table 4
Results of the unit root tests for search traffic data.

Test type	Augmented Dickey–Fuller (ADF)			Dickey–Fuller GLS (ERS)		
Test for unit root	Level	1st difference	2nd difference	Level	1st difference	2nd difference
Series						
Kimchi fridges	−1.327	−5.722***	−5.122***	−0.155	−2.131	−0.633
Fish rods	−1.690	−3.212**	−6.184***	−1.925	−2.473	−2.486
Bed	−1.395	−3.840***	−6.611***	−0.383	−2.552	−3.215***
DVR	−1.962	−1.918	−7.295***	−3.153***	−1.903	−7.331***
Whisky	−2.367	−6.542***	−7.300***	−2.007**	−5.995***	−7.275***
Raspberry wine	−1.327	−2.845*	−6.920***	0.095	−2.869***	−6.651***
Soju	−1.751	−3.828***	−8.158***	−1.242	−3.855***	−7.930***
Bottled water	0.492	−3.959***	−7.577***	1.605	−3.090***	−6.910***
UPS	−1.240	−1.994	−6.960***	−0.677	−1.958**	−2.041**
Inverters	−0.547	−3.019**	−5.885***	0.340	−1.650*	−1.409
Synthetic rubber	−1.349	−3.416**	−11.272***	−0.252	−2.628***	−1.152
Machining centers	−1.762	−3.233**	−11.674***	−0.237	−3.224***	−10.893***
HEVs	−1.759	−3.689***	−6.039***	−0.712	−1.669*	−2.138**

* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Table 5
Johansen system cointegration test results.

Cointegration parameters	Unrestricted cointegration rank test (trace)					Normalized cointegrating coefficients (standard error in parentheses)
Series	Hypothesized no. of CE(s)	Eigenvalue	Trace statistic	0.05 critical value	p-Values	
Kimchi fridges	None[a]	0.407	56.729	20.262	0.000	5684.15
	At most 1[a]	0.116	10.801	9.165	0.024	(513.19)
Fish rods	None[a]	0.248	26.983	20.262	0.005	165.87
	At most 1	0.021	1.898	9.165	0.798	(80.57)
Bed	None[a]	0.439	55.024	20.262	0.000	445.10
	At most 1	0.040	3.632	9.165	0.470	(54.51)
DVR	None[a]	0.353	27.040	20.262	0.005	41.26
	At most 1	0.072	3.951	9.165	0.420	(41.78)
Whisky	None	0.134	15.429	18.398	0.124	
	At most 1	0.035	3.105	3.841	0.078	
Raspberry wine	None[a]	0.279	30.448	12.321	0.000	54.07
	At most 1	0.018	1.638	4.130	0.236	(1.43)
Soju	None	0.050	6.434	20.262	0.930	
	At most 1	0.023	2.034	9.165	0.771	
Bottled water	None	0.056	7.166	20.262	0.886	
	At most 1	0.024	2.135	9.165	0.751	
UPS (uninterrupted power supplies)	None[a]	0.306	34.338	20.262	0.000	−4.58
	At most 1	0.025	2.223	9.165	0.733	(1.31)
Inverters	None	0.039	5.144	15.495	0.793	
	At most 1	0.020	1.706	3.841	0.192	
Synthetic rubber	None	0.075	9.223	20.262	0.715	
	At most 1	0.028	2.456	9.165	0.686	
Machining centers	None	0.158	18.215	20.262	0.093	
	At most 1	0.037	3.270	9.165	0.531	
HEVs	None	0.071	8.473	20.262	0.784	
	At most 1	0.023	2.035	9.165	0.771	

[a] Denotes rejection of the hypothesis at the 0.05 level.

exhibited similar trends, and mobile search traffic corresponded closely to sales.

According to our research results, searches for existing (durable) products are often related to purchases, while searches for a newly launched product may reflect not only purchases but also the hype regarding the product. Given that mobile and PC search traffic were similar in our case studies of HEVs and DVRs, we can deduce that PC searches still tend to be generally related to purchases, while mobile traffic is more likely to reflect curiosity or hype regarding the product. Thus, we may argue that PC-based searches remain a better predictor of purchases. Our research also confirms, however, that for new durable goods, mobile searches should also be considered when seeking to explain sales.

5.2. Forecasting using the VECM

The presence of cointegration suggests a long-run relationship among the variables under consideration and thus that the VECM can be applied. The long-run relationship between search traffic and sales for one cointegrating vector for the Korean bed market in the period

Table 6
Results of the pairwise Granger causality tests.

Granger causality test null hypothesis:	Lag = 1		Lag = 2		Research null hypothesis testing
	F-Statistic	p-Value	F-Statistic	p-Value	
Kimchi fridges search does not Granger Cause Kimchi fridges sales	0.513	0.476	4.748**	0.011	Reject (H_0 1)
DVR search does not Granger Cause DVR sales	4.490**	0.039	3.335**	0.044	Reject(H_0 1)
Fish rods search does not Granger Cause Fish rods sales	1.533	0.219	3.081*	0.051	Reject(H_0 1)
Bed search does not Granger Cause Bed sales	5.781**	0.018	12.464***	0.000	Reject(H_0 1)
Raspberry wine search does not Granger Cause Raspberry wine sales	17.673***	0.000	8.005***	0.001	Reject(H_0 2)
UPS search does not Granger Cause UPS sales	0.202	0.654	4.433**	0.015	Do not reject(H_0 3)

* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Table 7
Results of the GCBEW tests.

Dependent variable	Excluded variable	Chi-sq	df	Probability	Research null hypothesis testing
Whisky sales	Whisky search	24.508	4	0.000	Reject (H_0 2)
Soju sales	Soju search	9.007	6	0.173	Do not reject(H_0 2)
Bottled water sales	Bottled water search	7.813	6	0.252	Do not reject(H_0 2)
Inverters sales	Inverters search	2.870	4	0.580	Do not reject(H_0 3)
Synthetic rubber sales	Synthetic rubber search	3.241	4	0.518	Do not reject(H_0 3)
Machining centers sales	Machining centers search	7.565	5	0.182	Do not reject(H_0 3)
HEV sales	HEV search	2.922	6	0.819	Do not reject(H_0 5)

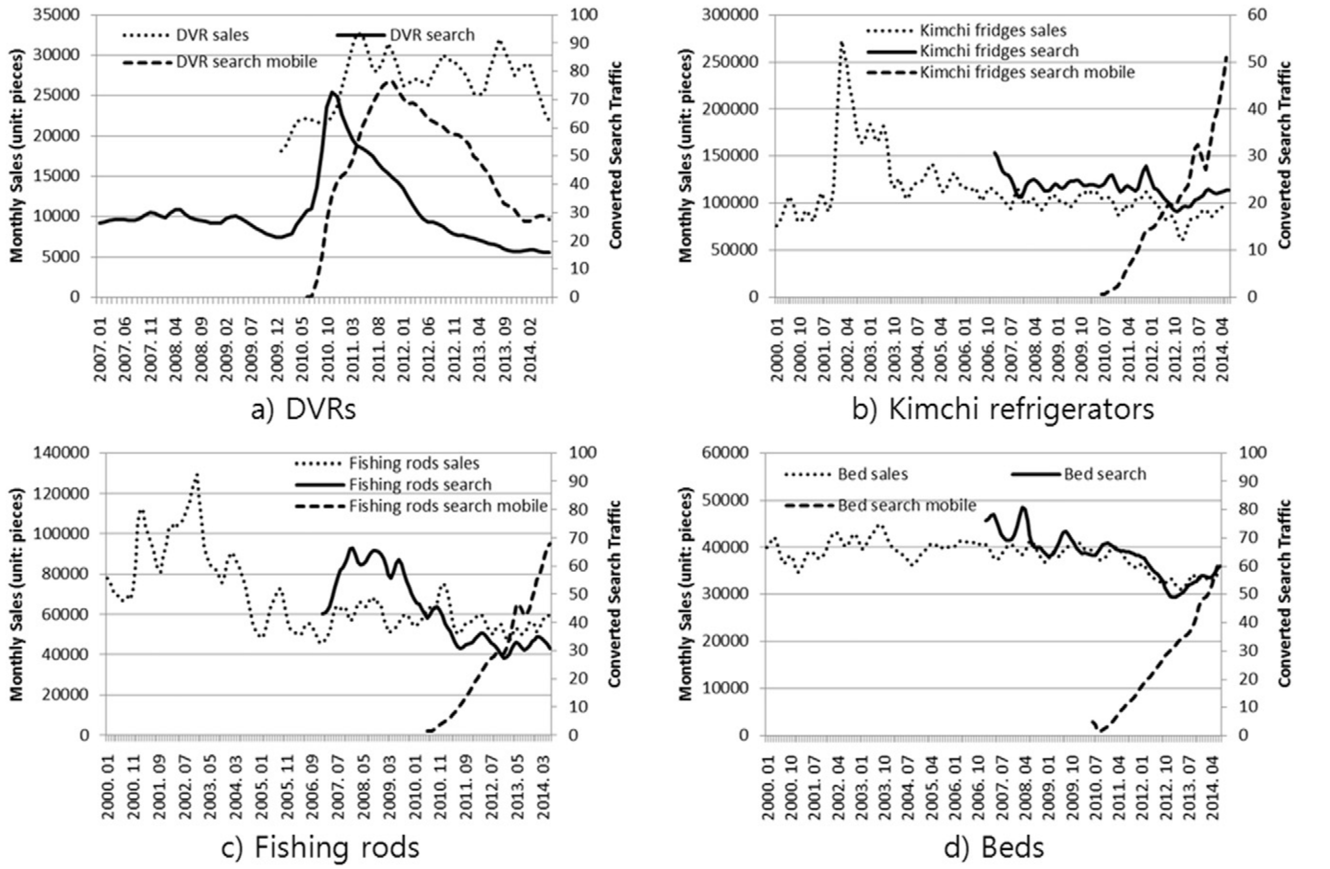

Fig. 8. Comparison of mobile and PC search traffic (DVRs, kimchi refrigerators, fishing rods, and beds).

2007–2014 (July) is displayed below (standard errors are displayed in parentheses):

$$\underset{(52.51)}{\text{Bed Sales} = 445.10} \times \text{Bed Search Traffic} + 8525.18 \tag{1}$$

When the variables are in logarithms and one cointegrating vector is estimated, the coefficients can be interpreted as long-run elasticities. Because the increase in bed sales is related to rising search traffic, the estimated model can produce a consistent result. Thus, a one-unit increase in search traffic resulted in a sales increase of 445 beds, and this estimate was significant (t-statistics: 8.165). These results are consistent with the results illustrated in Fig. 10, clearly demonstrating that information searches for durable goods tend be followed by purchases.

When cointegration is found to be present, we can use the VECM not only to analyze causality but also to perform forecasts. In this study, we performed tests sequentially from Lag 1 to Lag 8 to analyze the maximum lag of the VECM, and the results showed that over the entire time period, the maximum lag for the VECM had a minimum value at Lag 7, according to the Schwarz information criterion. Note that if we overextend the lag length, we may reduce the serial correlation of the

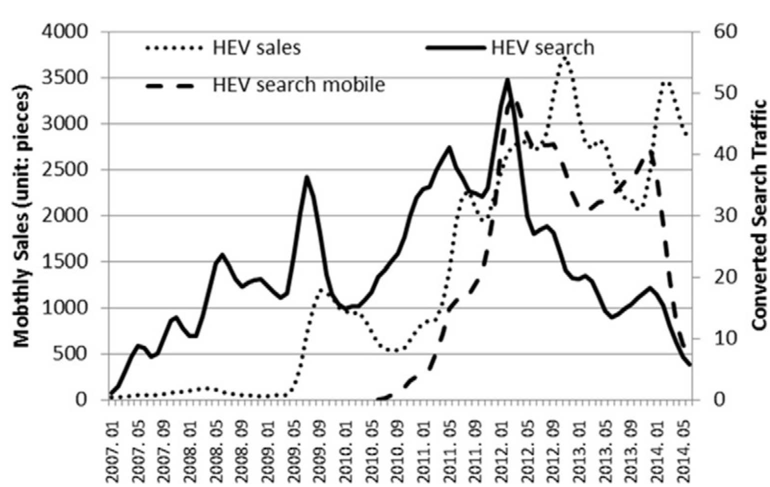

Fig. 9. Comparison of mobile and PC search traffic related to HEVs.

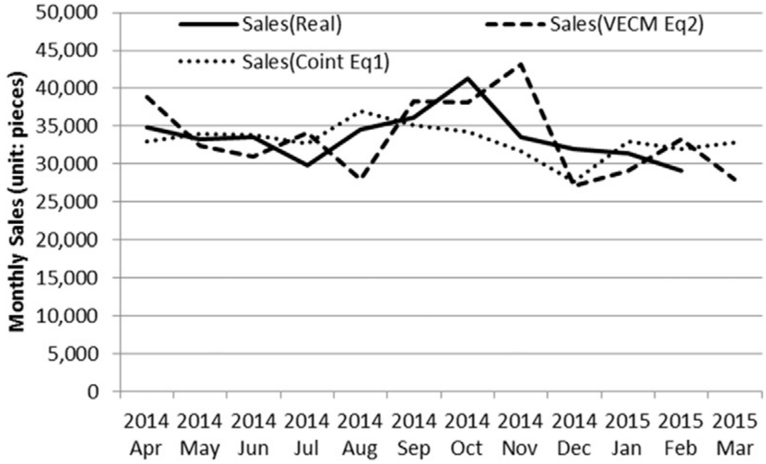

Fig. 10. Comparison of bed sales forecasts in the Korean market based on the VECM and actual bed sales.

error term and thereby reduce efficiency. Taking this trade-off into account, we set the maximum lag for the VECM to 3. The VECM estimation results for bed sales are as follows:

$$\begin{aligned} \mathrm{d(bed\_Sales)} = & -0.067 \times (\mathrm{bed\_sales}(-1) - 445.096 \\ & \times \mathrm{bed\_search}(-1) - 8545.183) + 0.716 \\ & \times \mathrm{d(bed\_sales}(-1)) - 0.025 \times \mathrm{d(bed\_sales}(-2)) \\ & - 0.302 \times \mathrm{d(bed\_sales}(-3)) + 15.736 \\ & \times \mathrm{d(bed\_search}(-1)) + 10.541 \times \mathrm{d(bed\_search}(-2)) \\ & + 25.096 \times \mathrm{d(bed\_search}(-3)) - 10.104. \end{aligned} \tag{2}$$

By applying the above estimation equations to real data from January 2014, we estimated bed sales in the Korean market from April 2014 to March 2015. The forecasting equation (Eq. (2)) that uses the VECM shows the estimations based on search traffic and sales for the past three months, while the forecasting equation (Eq. (1)) that uses the cointegration equation shows the sales forecasts for the current month based on search traffic from the current month. Because search traffic might be measured relatively more easily and quickly than sales information, we applied the current search traffic to Eq. (1). When we compared the forecast results for each equation to the actual values, Eqs. (1) and (2) showed differences of 0.9% and 1.4%, respectively. Therefore, both methods strongly demonstrated the “Nowcasting” capability of search traffic.

6. Conclusion

In this study, search traffic correlated more strongly with sales of consumer goods than that of industrial goods, while among consumer goods, search traffic was a stronger predictor of durable goods or high-priced consumer goods sales, which are bought infrequently compared with low-priced consumer goods, which are repurchased frequently. Furthermore, the products that demonstrated significance in both the VECM and the Granger causality test were mostly durable consumer goods. This finding confirms our stand that we should consider product type when analyzing and forecasting sales on the basis of search traffic. Moreover, even within the same product group, we should consider whether the product has relatively low sales and whether the purchase cycle is long term. In addition, if the product is new to the market, we should also consider that searches might exhibit a time lead over sales, which can be explained in terms of the hype surrounding a newly launched product. In cases where such hype is observed or where searches lead in advance of sales, search traffic tends to perform more poorly as a predictor of sales. These results strongly support the theory that consumers' responses vary by product type and depend on product innovation. In particular, for cases such as HEVs characterized by discontinuous innovation, the hype or lead time of search traffic can be explained by the rapidly increased involvement (Hawkins and Mothersbaugh, 2009).

The research approach and results presented in this study should help those interested in conducting technology management and formulating marketing strategies using search traffic. In the field of technology management, search traffic serves as a powerful tool for analyzing and forecasting the adoption of new technology. Our observation that search traffic may occur ahead of purchases, especially in the case of new consumer products, may be particularly useful for forecasting. Even in the case of industrial goods, search traffic can still be useful for analyzing interest in new products, albeit to a more limited degree. Marketing scholars have, thus far, been forced to rely on surveys to analyze whether experience goods can be converted into search goods and whether involvement with the product has increased (Niranjanamurthy and Kavyashree, 2013; Nakayama et al., 2010). However, our research findings demonstrate that search traffic can be used to track and analyze changes among consumers more easily and in real time. Although PC searches remain a stronger predictor of sales compared with mobile searches, the difference is small for new products, which is an issue with significant implications for future research.

The significance of this study is that we used a large sample of objectively measured data to perform a rigorous empirical analysis of the relationship between people's Internet activity and product sales. We also contribute to the literature by demonstrating the difference between PC and mobile searches.

However, we did experience some limitations in our research, including the limitations imposed by the source of our search traffic data as well as those in our sales data. First, the search traffic data provided by Naver consisted of processed rather than raw data, which imposed limits on our interpretation and utilization. Second, our sales data consisted of production data for Korean companies; thus, if the import volumes suddenly changed, we faced the risk of distortions in data interpretation. Although we carefully selected products that did not involve a large volume of imports, this potential limitation should be recognized. Indeed, when performing trend analyses of search traffic, future research should consider exogenous variables as well as extend our comparative analysis to include data from other search sites such as Google.

Acknowledgments

Seung-Pyo Jun, the first and corresponding author, acknowledge support from the Research Program (no. K16L04C02S04) at Korea

Institute of Science and Technology Information. Furthermore, he thanks Sunhee Jung and Chanhyeok Jun for the support of the research. Do-Hyung Park, the corresponding author, thanks YH Hwang and two CH Park for the financial and mental support. Finally, these authors thank anonymous three reviewers for their constructive comments.

References

Abernathy, W.J., Clark, K.B., 1985. Innovation: mapping the winds of creative destruction. Res. Policy 14, 3–22.

Anderson, P.F., Chambers, T.M., 1985. A reward/measurement model of organizational buying behavior. J. Mark. 7–23.

Anderson, E., Chu, W., Weitz, B., 1987. Industrial purchasing: an empirical exploration of the buyclass framework. J. Mark. 71–86.

Bahmani-Oskooee, M., Alse, J., 1993. Export growth and economic growth: an application of cointegration and error-correction modeling. J. Develop. Areas 535–542.

Barras, R., 1986. Towards a theory of innovation in services. Res. Policy 15, 161–173.

Butler, D., 2013. When Google got flu wrong. Nature 494, 155.

Choi, H., Varian, H., 2012. Predicting the present with Google trends. Economic Record 88, 2–9.

Currim, I.S., Mintz, O., Siddarth, S., 2015. Information accessed or information available? The impact on consumer preferences inferred at a durable product E-commerce website. J. Interact. Mark. 29, 11–25.

Dodge, Y., Cox, D., Commenges, D., Solomon, P.J., Wilson, S., 2003. The Oxford Dictionary of Statistical Terms. Oxford University Press.

Engle, R.F., Granger, C.W., 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica J. Econ. Soc. 251–276.

Engle, R.F., Yoo, B.S., 1987. Forecasting and testing in co-integrated systems. J. Econ. 35, 143–159.

Gartner, 2011. Interpreting Technology Hype. Gartner (www.gartner.com).

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014.

Granger, C.W., 1988. Some recent development in a concept of causality. J. Econ. 39, 199–211.

Guzman, G., 2011. Internet search behavior as an economic forecasting tool: the case of inflation expectations. J. Econ. Soc. Meas. 36, 119–167.

Hawkins, D.I., Mothersbaugh, D.L., 2009. Consumer Behavior: Building Marketing Strategy. 11 ed. McGraw Hill Higher Education.

Jepsen, A.L., 2007. Factors affecting consumer use of the Internet for information search. J. Interact. Mark. 21, 21–34.

Johansen, S., 1988. Statistical analysis of cointegration vectors. J. Econ. Dyn. Control. 12, 231–254.

Johnston, W.J., 1981. Industrial buying behavior: a state of the art review. Rev. Mark. 2, 75–89.

Jun, S.-P., 2012a. A comparative study of hype cycles among actors within the socio-technical system: with a focus on the case study of hybrid cars. Technol. Forecast. Soc. Chang. 79, 1413–1430.

Jun, S.-P., 2012b. An empirical study of users' hype cycle based on search traffic: the case study on hybrid cars. Scientometrics 91, 81–99.

Jun, D.B., Kim, J.I., 2011. A choice-based multi-product diffusion model incorporating replacement demand. Technol. Forecast. Soc. Chang. 78, 674–689.

Jun, S.-P., Yeom, J., Son, J.-K., 2014a. A study of the method using search traffic to analyze new technology adoption. Technol. Forecast. Soc. Chang. 81, 82–95.

Jun, S.-P., Park, D.-H., Yeom, J., 2014b. The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference. Technol. Forecast. Soc. Chang. 86, 237–253.

Kaesbauer, M., Hohenstatt, R., Reed, R., 2012. Direct versus search engine traffic: an innovative approach to demand analysis in the property market. Int. J. Housing Markets Anal. 5, 392–413.

Koreanclick, 2012. The Highest Growth Site (The First Half of 2012 in Korea). Koreanclick (www.koreanclick.com).

Kotler, P., Keller, K.L., 2007. A Framework for Marketing Management. Pearson.

Kotler, P., Keller, K.L., Ancarani, F., Costabile, M., 2014. Marketing Management 14/e. Pearson.

Lui, C., Metaxas, P.T., Mustafaraj, E., 2011. On the Predictability of the US Elections through Search Volume Activity, in: Proceedings of the IADIS International Conference on e-Society. Citeseer.

Mackinnon, J.G., 1996. Numerical distribution functions for unit root and cointegration tests. J. Appl. Econ. 11, 601–618.

MacKinnon, J.G., Haug, A.A., Michelis, L., 1999. Numerical distributions of likelihood ratio tests of cointegration. J. Appl. Econ. 14, n577.

McDermott, C.M., Kang, H., Walsh, S., 2001. A framework for technology management in services. Engineering Management, IEEE Transactions on 48, 333–341.

McQuiston, D.H., 1989. Novelty, complexity, and importance as causal determinants of industrial buyer behavior. J. Mark. 66–79.

Murphy, P.E., Enis, B.M., 1986. Classifying products strategically. J. Mark. 50, 24–42.

Nakayama, M., Sutcliffe, N., Wan, Y., 2010. Has the web transformed experience goods into search goods? Electron. Mark. 20, 251–262.

Niranjanamurthy, M., Kavyashree, N., 2013. Analysis of e-commerce and m-commerce: advantages, limitations and security issues. Int. J. Adv. Res. Comput. Commun. Eng. 2, 2360–2370.

Ottum, B.D., Moore, W.L., 1997. The role of market information in new product success/failure. J. Prod. Innov. Manag. 14, 258–273.

Ozkan, F., Ozkan, O., Gunduz, M., 2012. Causal relationship between construction investment policy and economic growth in Turkey. Technol. Forecast. Soc. Chang. 79, 362–370.

Robinson, P.J., Faris, C.W., Wind, Y., 1967. Industrial Buying and Creative Marketing. Allyn & Bacon Boston, MA.

Robinson, D.K., Huang, L., Guo, Y., Porter, A.L., 2013. Forecasting innovation pathways (FIP) for new and emerging science and technologies. Technol. Forecast. Soc. Chang. 80, 267–285.

Shaikh, M.A., Hansotia, B.J., 1985. Historical and new perspectives on industrial buyer behavior. In: Sheth, J.N. (Ed.), Research in Consumer Behavior, Greenwich, Conn.JAI Press, pp. 233–283

Shim, S., Eastlick, M.A., Lotz, S.L., Warrington, P., 2001. An online prepurchase intentions model: the role of intention to search. J. Retail. 77, 397–416.

Sousa, I., Wallace, D., 2006. Product classification to support approximate life-cycle assessment of design concepts. Technol. Forecast. Soc. Chang. 73, 228–249.

To, P.-L., Liao, C., Lin, T.-H., 2007. Shopping motivations on Internet: a study based on utilitarian and hedonic value. Technovation 27, 774–787.

Vosen, S., Schmidt, T., 2011. Forecasting private consumption: survey-based indicators vs. Google trends. J. Forecast. 30, 565–578.

Walsh, S.T., Linton, J.D., 2011. The strategy-technology firm fit audit: a guide to opportunity assessment and selection. Technol. Forecast. Soc. Chang. 78, 199–216.

Watts, R.J., Porter, A.L., 1997. Innovation forecasting. Technol. Forecast. Soc. Chang. 56, 25–47.

Seung-Pyo Jun is a principal researcher at the Korea Institute of Science and Technology Information (KISTI). He is also an associate professor of Science and Technology Policy at the University of Science & Technology (UST). He received his Ph.D. from the Science and Technology Studies Program at Korea University. His area of interest includes demand forecasting, emerging technology detection, technology evaluation and new technology adoption analysis. Recently, he is focusing on new technology adoption analysis using big data.

Do-Hyung Park is an assistant professor at the School of Management Information Systems, Kookmin University, Seoul 136-702, Korea. He received his Ph.D. in management engineering at KAIST (Korea Advanced Institute of Science and Technology). His research covers online user behavior, such as online feedback, online community and social network services, new technology adoption and innovation process, and model development using big data.