Summer 6-19-2015

The Dynamic Impact of Web Search Volume on Product Sales — An Empirical Study Based on Box Office Revenues

Rui Miao
School of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian, 116025, China,
ruimiaodufe@163.com

Yueyue Ma
School of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian, 116025, China,
mayueyue812@126.com

The Dynamic Impact of Web Search Volume on Product Sales — An Empirical Study Based on Box Office Revenues

Rui Miao[1*], *Yueyue Ma*[1]
[1]School of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian, 116025, China

Abstract: In order to explore how Web search volume dynamically influences product sales during the whole product life cycle, this paper collects Web search volume and sales data of movies and does an empirical analysis using econometric models. The empirical results show that Web search volume before the launch of a new product has a positive impact on the product sales in the initial period of introduction stage. During the whole product life cycle, Web search volume has a positive and significant impact on product sales, but the impact declines gradually across the life cycle. The impact of Web search volume on sales is larger in the early stage of the product life cycle than in the late stage of the product life cycle.

Keywords: Web search volume, product sales, product life cycle, empirical analysis

1. INTRODUCTION

With the development of Internet, more and more people turn to the Internet for news and information. Web search volume measures the counts of search queries about a certain keyword aggregated by the search engine, and it reflects the collective interests, concerns and needs of the global population. Recent works have demonstrated that what people are searching for online can predict their future behavior, and there are relations between Web search volume and offline outcomes, such as disease prevalence, home sales, stock returns and consumer price index (CPI) in near time.

According to the consumer behavior theory, a consumer's purchase process includes the stages of need recognition, information search and decision-making. After recognizing their needs for some products, many consumers will use search engines to search for relevant information, evaluate the products using these information and make a decision. As a result, Web search volume is predictive of what consumers will buy in the near future, and has a positive impact on product sales. It is of great significance for enterprises to use Web search volume to improve the accuracy of prediction for the product sales and make better plans for operation. This paper aims to explore how Web search volume dynamically influences product sales during the whole product life cycle by collecting the panel data of movies and making an empirical analysis using econometric models.

2. LITERATURE REVIEW AND HYPOTHESES DEVELOPMENT

Web search volume recorded by search engines has received some attention for their usefulness in explaining and predicting a variety of economic and social events. Ginsberg et al. used Google search query data to build an early detection system for influenza epidemics [1], and Li Xiuting et al also used Google search query data to detect China influenza [2]. Wu and Brynjolfsson utilized Google search data to predict future house sales and price indices [3]. In the field of macro economies, some studies showed that there are relations between Web search data and unemployment [4] as well as private consumption [5], [6]. Some other studies used Web search volume to measure investor attention for stocks, and suggested the correlation between Web search volume and stock returns [7]-[9]. To explore the predictive capability of Web search volume in product sales, Choi and Varian

* Corresponding author. Email: ruimiaodufe@163.com (Rui Miao), mayueyue812@126.com (Yueyue Ma)

used Web search data to predict the sales of motor vehicle parts, travel and automotive sales [10]. Du and Kamakura as well as Seebach et al. used Google data to predict automotive sales [11], [12]. Goel et al. used Yahoo!'s search engine data to predict various outcomes, including weekend box office revenues for feature films, video game sales and song ranks [13]. Lv benfu et al. used search data to predict the automotive sales and e-commerce trading volume [14], [15]. An explanation of why web search data are useful in predicting future product sales is provided by Wu and Brynjolfsson, who suggest that web search logs constitute honest signals of decision-makers' intentions [3]. Although the correlation between Web search volume and product sales have been widely explored in various fields, there is a lack of research on how Web search volume dynamically influences the product sales in the whole product life cycle. Hence, this paper aims to fill this gap.

According to the consumer behavior theory, a consumer's purchase process includes the stages of need recognition, information search and decision-making. After recognizing the need for one product, the consumer will use search engine to search for relevant information such as price, product attributes, and where to buy it. If he is satisfied with the product, he will buy it immediately or in the near future. The counts of search queries on one product recorded by the search engine measure the collective demands for the product on the current or next period. Hence, this paper proposes the following hypothesis:

Hypothesis 1: Web search volume on the current period has a positive impact on the product sales on the current and next period.

Every product has its own product life cycle, including stages of introduction, growth, maturity and decline. Before a new product is launched into the market, search engine is a very important channel for consumers to acquire the information about the new product in the Internet age. Hence, the behaviors of searching in the search engine reflect the collective concerns of consumers, and these concerns can be turned into the demands and sales of the new product after launching. After the introduction stage of the product, more channels can be used by consumers to acquire product information, such as acquaintance referrals, online product reviews and so on. The substitute effects between different information channels will make the impact of Web search volume on product sales decrease. Hence, this paper proposes the following hypotheses:

Hypothesis 2: Web search volume before the launch of a new product has a positive impact on the sales of the new product in the initial period of the introduction stage.

Hypothesis 3: The impact of Web search volume on product sales is stronger in the early stage of product life cycle than in the late stage of product life cycle.

3. EMPIRICAL ANALYSIS

3.1 Data collection

In order to test the hypotheses proposed above, this paper focuses on the impact of Web search volume on movie box office revenues, because movie revenues are publicly available during the whole life cycle. This paper takes 38 movies that are widely released between October 2013 and August 2014 in China as examples, and collects the daily box office revenues of these movies from www.entgroup.cn. The data of Web search volume are collected from Baidu Index(http://index.baidu.com). This paper uses each movie's name as the keyword and collects the daily Web search volume of each movie from the week before it is released.

3.2 Impact of pre-release Web search volume on the opening day revenues

This paper first uses all the 38 movies and builds a linear regression model (1) to explore the correlation between the Web search volume at the week before a movie is released and the box office revenues of the movie on its opening day.

$$\ln firstdayrevenue_i = \alpha_0 + \alpha_1 \ln avgsearch_i + \alpha_2 series_i + \alpha_3 weekend_i + u_i \qquad (1)$$

Where the dependent variable $firstdayrevenue_i$ is the box office revenues of movie i on its opening day, and the independent variable $avgsearch_i$ is the average of the Web search volume for movie i over the week before it is released. In order to account for the highly skewed distributions of popularity, both revenue and Web search volume are log-transformed. $series_i$ and $weekend_i$ are control variables and dummy variables. $series_i$ represents whether movie i is a sequel, and $weekend_i$ represents whether movie i is released at weekend (including Friday, Saturday and Sunday).

Table 1 shows the descriptive statistics of $firstdayrevenue_i$ and $avgsearch_i$.

Table 1. Descriptive statistics of key variables for all movies

Variables	No. of observations	Mean	Standard deviation	Minimum	Maxmum
$firstdayrevenue_i$	38	3388.263	2186.316	720	10400
$avgsearch_i$	38	35399.36	33632.09	9209.286	203480.9

The estimates of model (1) are reported on Column 2 in Table 2. It can be seen that the coefficients of $lnavgsearch_i$ and $series_i$ are all positive and significant, which suggests that the average of the Web search volume over the week before a movie is released as well as sequel all have positive and significant effects on the opening day revenues. The coefficient of $weekend_i$ is not significant. One possible explanation for this result is that whether the opening day is weekend is not important for the audience who really like the movie. Then this paper re-estimates a new model without the variable $weekend_i$, and the estimates of the new model are shown on Column 3 in Table 2. The coefficients of $lnavgsearch_i$ and $series_i$ are still positive and significant, and just change a little compared to model (1). But after $weekend_i$ is deleted from model (1), both F value and the adjusted R^2 are improved, and this suggests that the performance of the new model is better than model (1). The estimates of both the two models show that the average of the Web search volume over the week before a movie is released has a positive impact on the opening day revenues, which supports Hypothesis 2 that Web search volume before the launch of a new product has a positive impact on the sales of the new product in the initial period of the introduction stage.

Table 2. Impact of pre-release Web search volume on the opening day revenues

	Model (1)	New model without the variable $weekend_i$
Variables	Coefficient (t-statistics)	Coefficient (t-statistics)
constant	1.856 (1.347)	2.020 (1.496)
ln$avgsearch_i$	0.577 (4.358)***	0.567 (4.338)***
$series_i$	0.328 (1.917)*	0.333 (1.962)*
$weekend_i$	0.118 (0.719)	
R^2	0.384	0.375
Adjusted R^2	0.330	0.339
F value	7.072***	10.496***

Note: *** $p<0.01$, * $p<0.1$

3.3 Impact of post-release Web search volume on the revenues

To verify Hypothesis 1 and 3, this paper chooses the first three weeks after a movie is released to approximately represent the product life cycle of the movie, because on average the box office revenues of the movies during the first three weeks account for 95% of their total revenues. This paper eliminates 4 movies whose daily revenues are incomplete from the original samples, and uses the panel data of the remaining 34 movies to estimate the following panel data model (2) with fixed effects.

$$\ln revenue_{it} = \alpha_0 + \alpha_1 \ln search_{it} + \alpha_2 weekend_{it} + \alpha_3 trend_{it} + \mu_i + \varepsilon_{it} \tag{2}$$

Where $revenue_{it}$ and $search_{it}$ are box office revenues and Web search volume for movie i on day t respectively, and they are also log-transformed. $weekend_{it}$ and $trend_{it}$ are control variables. $weekend_{it}$ represents whether the day t after movie i is released is weekend, and $trend_{it}$ is a variable to control the trend that the revenues change over time. There are many other variables that will affect movie revenues, such as movie genre, director and actors, MPAA rating, and production budget. However, because these variables are individual-specific and do not vary over time, they appear in μ_i. ε_{it} is the error term. As μ_i is not independent to ε_{it}, the fixed effects model is more appropriate than the random effects model, and the results of Hausman test also support this. Before estimating the model, this paper conducts a unit root test, and the results of LLC test and Fisher-ADF test all show that the series of all the key variables do not have unit roots.

The estimates of model (2) are reported on Column 2 in Table 3. The control variables in the model all have significant effects on the dependent variable. The coefficient of $weekend_{it}$ is positive and significant, and this suggests that the revenues at weekends are higher than at weekdays. The coefficient of $trend_{it}$ is negative and significant, which means that the revenues decrease over time. The coefficient of $\ln search_{it}$ is positive and significant, and a 1% increase in Web search volume results in a 0.832% increase in the revenues on the current period, which suggests that Web search volume on the current period has a positive impact on the product sales on the same period.

Table 3. Impact of post-release Web search volume on the revenues

Model (2)		Model (3)	
Variables	Coefficient (t-statistics)	Variables	Coefficient (t-statistics)
constant	-1.820 (-2.141)**	constant	2.353 (2.681)***
$\ln search_{it}$	0.832 (11.670)***	$\ln search_{it-1}$	0.480 (6.604)***
$weekend_{it}$	0.293 (8.334)***	$weekend_{it}$	0.443 (12.625)***
$trend_{it}$	-0.088 (-19.665)***	$trend_{it}$	-0.109 (-22.547)***
R^2	0.859	R^2	0.845
Adjusted R^2	0.851	Adjusted R^2	0.836
F value	114.5608***	F value	97.319***

Note: *** p<0.01, ** p<0.05

Then this paper takes the one-period lagged Web search volume as the independent variable to estimate the fixed effects model (3) as follows:

$$\ln revenue_{it} = \alpha_0 + \alpha_1 \ln search_{i,t-1} + \alpha_2 weekend_{it} + \alpha_3 trend_{it} + \mu_i + \varepsilon_{it} \tag{3}$$

The estimates of model (3) can be seen on Column 4 in Table 3. The coefficient of the one-period lagged Web search volume is also positive and significant at 1% level, but the coefficient drops to 0.480. This suggests that the one-period lagged Web search volume has a smaller impact on the current revenues than Web search volume on the current period. In conclusion, Hypothesis 1 is supported that Web search volume on the current period has a positive impact on product sales on the current and next period.

3.4 Dynamic Impact of Web search volume on the revenues during the whole product life cycle

In order to explore the dynamic impact of Web search volume on the revenues during the whole product life cycle, this paper divides the panel data of the 34 movies into two datasets. One dataset includes the panel data of the 34 movies for the first two weeks after being released to represent the early stage of the product life cycle, and the other includes the panel data of the 34 movies for the third week after being released to represent the late stage of the product life cycle. Then this paper builds the same fixed effects model as model (2) on the two datasets respectively, and the estimates are shown in Table 4. It can be seen that the coefficients of the independent variables are still positive and significant. However, in the late stage of the product life cycle, the impact of Web search volume on the revenues decreases. When in the early stage of the product life cycle, a 1% increase in Web search volume results in a 0.922% increase in the revenues, but When in the late stage of the product life cycle, a 1% increase in Web search volume results in a 0.664% increase in the revenues. This suggests that the impact of Web search volume on product sales is bigger in the early stage of product life cycle than in the late stage of product life cycle, and Hypothesis 3 is supported.

Table 4. Dynamic Impact of Web search volume on the revenues

	Early stage of the product life cycle (The first two weeks after released)	Late stage of the product life cycle (The third weeks after released)
Variables	Coefficient (t-statistics)	Coefficient (t-statistics)
constant	-3.014 (-3.948)***	-1.207 (-0.694)
$lnsearch_{it}$	0.922 (14.401)***	0.664 (4.267)***
$weekend_{it}$	0.296 (10.030)***	0.339 (5.658)***
$trend_t$	-0.062 (-13.589)***	-0.109 (-7.772)***
R^2	0.887	0.897
Adjusted R^2	0.878	0.879
F value	95.944***	48.712***

Note: *** p<0.01

4. CONCLUSIONS

The goal of this paper is to explore how Web search volume dynamically influences product sales during the whole product life cycle. This paper collects Web search volume and sales data of 38 movies and makes an empirical analysis using econometric models. The empirical results show that the movie revenues are higher when the Web search volume about the movie is larger before and after the movie is released, and the impact of Web search volume on movie revenues is bigger in the early stage of product life cycle than in the late stage of product life cycle. This paper adds several contributions to the literature. First, before a new product is launched into market, the higher the Web search volume is, and the higher the product sales in the initial period of the introduction stage are. Second, Web search volume has a positive impact on product sales during the whole product life cycle. But because of the substitute effects between different information channels, the impact of Web search volume on product sales decreases over time. The impact of Web search volume on product sales is stronger in the early stage of product life cycle than in the late stage of product life cycle. In terms of business implications, Web search volume can be used by enterprises to improve the accuracy of prediction for product sales. And enterprises should take search engine as an important marketing tool to maximizing product sales especially in the early stage of the product life cycle. This paper also has some limitations. First, this paper only takes the movie (experience product) as example, and whether the findings can be apply to other products and services deserves further exploration. Second, future study should explore the interaction effects between Web search and other information channels, such as blogs, online reviews, social networks and so on.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71301021 and Dongbei University of Finance and Economics under Grant DUFE2014Q43.

REFERENCES

[1] Ginsberg J, Mohebbi M H, Patel R S, et al. (2008). Detecting influenza epidemics using search engine query data. Nature, 457(7232): 1012-1014

[2] Li Xiuting, Liu Fan, Dong Jicang, et al. (2013). Detecting China influenza using search engine data. Systems Engineering — Theory & Practice, 33(12): 3028-3034 (in Chinese)

[3] Wu L, Brynjolfsson E. (2013). The future of prediction: How Google searches foreshadow housing prices and sales. Economics of Digitization. Chicago: University of Chicago Press

[4] Askitas N, Zimmermann K F. (2009). Google Econometrics and Unemployment Forecasting. Applied Economics Quarterly, 55(2): 107-120

[5] Vosen S, Schmidt T. (2011). Forecasting private consumption: survey - based indicators vs. Google trends. Journal of Forecasting, 30(6): 565-578

[6] Zhang Chong, Lv Benfu, Peng Geng, et al. (2012). A study on the correlation between web search data and CPI. Journal of Management Sciences in China, 15(7): 50-59 (in Chinese)

[7] Da Z, Engelberg J, Gao P. (2011). In search of attention. The Journal of Finance, 66(5): 1461-1499

[8] Yu Qingjin, Zhang Bing. (2012). The investor limited attention and stock performance - an empirical study using Baidu Index as a direct measure of investor attention. Journal of Financial Research, 2012 (8): 152-165 (in Chinese)

[9] Liu Feng, Ye Qiang, Li Yijun. (2014). Impacts of interactions between news attention and investor attention on stock returns: Empirical investigation on financial shares in China. Journal of Management Sciences in China, 17(1): 72-85(in Chinese)

[10] Choi H, Varian H. (2012). Predicting the present with google trends. Economic Record, 88(s1): 2-9

[11] Du R Y, and Kamakura W A. (2012). Quantitative Trendspotting. Journal of Marketing Research 49(4): 514-536

[12] Seebach C, Pahlke I, Beck R. (2011). Tracking the Digital Footprints of Customers: How Firms Can Improve Their Sensing Abilities to Achieve Business Agility. In ECIS 2011

[13] Goel S, Hofman J M, Lahaie S, et al. (2010). Predicting Consumer Behavior with Web Search. Proceedings of the National Academy of Sciences, 107(41): 17486-17490

[14] Yuan Qingyu, Peng Geng, Liu Ying, et al. (2011). A Prediction Study on the Amount of Automotive Sales Based on Web Search Data. An Academic Edition of ManaMaga, 2011(1): 12-24 (in Chinese)

[15] Peng Geng, Li Na, Lv Benfu. (2014). A study on the prediction of e-commerce trading volume - the prediction model based on web search data and product classification. Modernization of Management, 2014 (2): 30-32 (in Chinese)