Traditional Marketing, Online Communication and Market Outcomes

Essay 1: Marketing Activity, Blogging and Sales
Essay 2: Consumers' Social Learning about Videogame Consoles through Multi-Website Browsing
Essay 3: Co-evolution of Network Growth and Group Formation

by

Hiroshi Onishi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in The University of Michigan
2011

Doctoral Committee:

Professor Puneet Manchanda, Chair
Professor Gerald F. Davis
Professor Fred M. Feinberg
Associate Professor Lada A. Adamic
Associate Professor Anocha Aribarg

Acknowledgements

I would like to thank all the members of the doctoral dissertation committee, participants at the 2009 YCCI Conference at Yale SOM, at the Michigan Marketing PhD Camp 2008-2010, at the 2008 Marketing Science conference in Vancouver, at the 2009 Marketing Science conference in Ann Arbor and the 2010 Marketing Dynamics conference in Istanbul for valuable feedback, and Hotlink Inc., Dentsu Inc. and Video Research International Inc. for providing the data. I would also like to thank the Ross School of Business for research support via Phelps Doctoral Fellowship, Leo Burnett Fellowship and Spivey/Hall Fellowship.

Table of Contents

Lists of Figures

Lists of Tables

Abstract

The scope of my dissertation is to understand how consumers' online communication affects consumers' behavior and market outcomes. The recent growth of consumer generated media (CGM), also known as "new" media, has provided consumers easier interaction, both to search for product information and to obtain evaluations and opinions from peers. This dissertation addresses two key issues – (a) Does the new media affect consumer behavior and market outcomes? (b) Is there any relationship between traditional marketing activity and new media?

In the first essay, we assemble a unique data set from Japan containing market outcomes (sales), new media (blogs) and traditional media (TV ads) for three product categories. We specify a simultaneous equation log-log system for market outcomes and the blogging volume. Results suggest that blogs are predictive of market outcomes, traditional and new media have a synergistic effect, and pre-launch TV ads lead blogging activity but become less effective in inducing blogging activity post-launch. A novel text mining analysis provides "process" explanation for these results.

In the second essay, we examine the micro-level correlation between traditional media (TV ads and public relations) and consumers' social learning about new videogame consoles (Wii and PS3). We propose consumers' learning processes via perusal of information in online communities by using "pageview" data of multiple websites from a clickstream panel as indicators. We propose a bivariate Bayesian learning model combined with complementary purchase choices. Results show that companies' traditional media have positive impact on social learning. This suggests that by optimizing marketing actions, firms can enhance consumers' learning and promote higher engagement of the products.

In the third essay, we investigate whether the introduction of a “group” function in a social networking website changes the engagement level of current website users. We expect that the "stickiness" effect of the group function motivates group members to increase networking activity (numbers of visits, connections and messages). In addition, group growth may contribute to enlarging the entire network and vice versa (“co-evolution” effect). We use data from a professional job-search social network website and find that introducing the group function can promote both stickiness and co-evolution effects.

Chapter I

Introduction

The scope of my dissertation is to understand how consumer generated online communication (*e.g.*, blogging, consumers' learning and social networking) affects consumer behavior and market outcomes (*e.g.*, sales). The recent development of consumer generated media (CGM) such as blogging and networking community has changed the nature of interaction between consumers and firms. CGM has enabled consumers to interact easily, both to search for information about products and to obtain evaluations and opinions from other consumers. Essentially, communication between firms and consumers has become bidirectional rather than unidirectional. The broad research questions that, therefore, emerge are whether this bidirectional communication impacts managerial action and whether it affects consumer behavior and market outcomes such as sales or profits (*e.g.*, Godes and Mayzlin 2004; Liu 2006). On the other hand, little is known about the relationship between firm controlled traditional communication (*e.g.*, advertisings and public relations) and CGM. Thus, this dissertation addresses two key issues – (a) do CGM affect consumer behavior and market outcomes and, (b) is there relationship between managerial communication and CGM, especially for the cases of new product launches? We, therefore, will focus on capturing the three-way relationship among traditional media, online communication (CGM) and market outcomes in three essays (Figure I-1).

The first essay examines synergistic relationship between traditional media (TV ads) and CGM (blogging). The majority of the existing research has focused on online chatters (*e.g.*, newsgroup postings, reviews and ratings) and their effect on market outcomes. There were some evidences that volume of online user ratings was positively correlated with sales (Liu 2006; Dellarocas, Zhang, and Awad 2007; Duan, Gu and Whinston 2008; Chintagunta, Gopinath and Venkataraman 2009). Given the goal-driven nature of product ratings and reviews, however, these findings may not be that unexpected. On the other hand, blogs, the most prevalent form of new media, encompass a much larger domain of discussion and therefore are less likely to affect market outcomes. There was limited empirical research that examined the role of blogs in forecasting sales volumes (Gruhl, Kumar, Novak and Tomkins 2005; Dhar and Chang 2007). In this study, we also investigate (a) whether blogging activity leads to differential market outcomes. In addition, we investigate if firms' traditional marketing actions act as antecedents of the blogging activity, in other words, (b) do traditional ads lead to changes in blogging behavior? Our expectation is that there is a positive correlation between the quantity of TV ads and blog posts. We also anticipate that this effect of ads on blogs becomes weaker after the product launch. This is because while TV ads can independently increase blogging pre-release by providing product information and content, but once the product is available, consumers may rely less on traditional media, leading to a weaker relationship between new and old media at that point. We assemble a unique data set from Japan for three different product categories of green tea drinks, movies and cellular phone service. Each category has at least one new product launch during the duration of our sample periods. We specify a simultaneous equation log-log system for market outcomes and the volume of blogs. Our results suggest that blogs are predictive of market outcomes, new and traditional media act synergistically, pre-launch TV advertising spurs blogging activity but become less effective post-launch and that market outcomes have some effect on blogging. We find detailed support for some of these findings via a unique and novel text mining analysis. From a managerial perspective, sales forecast becomes accurate by considering the synergy and indirect relationships. Managers can also leverage these relationships to carry out "better" resource allocation (to traditional media) as they can take the spillover effects (from traditional to new media) into account.

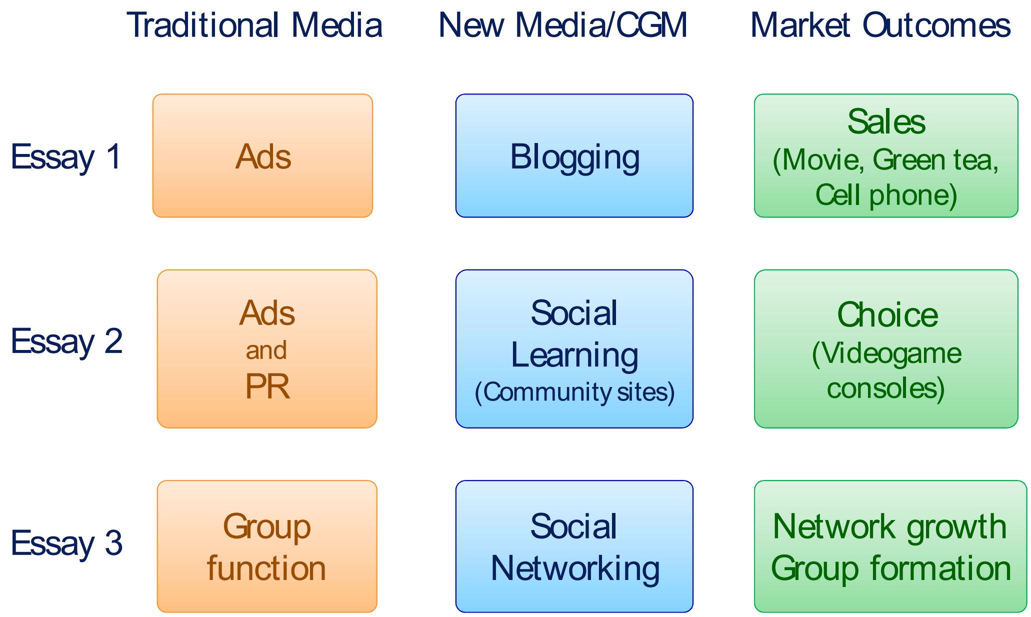
The second essay explores micro-level correlation between traditional marketing activity (TV ads and public relations) and online communication (consumers' social learning) about newly launched videogame consoles (Wii and PS3 in 2006) via browsing on product community websites. We postulate that consumers learn about these products in two ways - from other consumers ("social" learning) and from entire product review websites ("personal" learning). Luan and Neslin (2009) and Erdem et al. (2005) investigated product learning processes through word-of-mouth communication, but their studies only considered a single learning process and did not differentiate social learning from personal learning. Social learning is the learning process that is promoted by exchanging information among diverse consumers in (online) communities (Jayanti and Singh 2010; Calder, Malthouse and Schaedel 2009). In practice, firms have conducted engagement marketing to enhance social learning which leads to providing long-term customer loyalty and maximizing purchase conversions (see ARF Defining Engagement Initiative). As the industry practice, consumer's engagement is assessed by measurements such as duration, frequency and/or recency of visit at online community websites. Therefore, we assume that pageviews of community-based websites indicate a level of consumer's engagement via social learning. We also assume that, at the same time, consumers conduct another mode of learning, personal learning by browsing all videogame related websites, and that the two processes are correlated with each other. We also anticipate that those two learning processes affect consumers' purchase decisions differentially. We propose a bivariate Bayesian learning model combined with complementary purchase choices (Ackerberg 2003; Gentzkow 2007). The proposed model enables a simpler estimation of parameters and allows for detailed information about the interaction between the social and the personal learning processes. In summary, our research questions are (a) do traditional ads and PR campaigns enhance consumer learning and, (b) what is the relative importance of the two types of learning on consumer purchase choice. From empirical results we find that companies' traditional marketing actions have larger impacts on the social learning than on the regular personal learning, during the pre-launch period. When consumers make purchase decision, they weight social beliefs of product quality at least three times more than personal beliefs. Policy simulations suggest that by optimizing the marketing actions, firms can manage to enhance consumers' learning and promote higher engagement of the products.

In the third essay, we investigate whether the introduction of the social function of "groups" in a social networking site changes the engagement and visitation levels of the current website users. The group function allows users to share information on a special topic with self-selected peers who have common interests (*e.g.*, music, sports, fashion, campus, career, etc.). Although most groups are initiated by users, website managers can still manipulate group structures and topics as well as develop its functionality to improve the group function. Therefore, we expect that the "stickiness" effect of the group function motivates group members to increase their networking activity. We investigate the stickiness effect of the group function in terms of three networking activity measurements, such as a number of visits, connections and messages. In addition, this group growth may contribute to enlarging the entire network. Similarly, the growing network would lead to enhancing group formations and increasing group membership. Virtually no research has taken into account this “co-evolution” aspect. Our research plans to make an initial attempt to examine this inter-dependence of the network and the group as the stickiness effect and the co-evolution effect in the marketing context. We use data from a professional social network website in which users build professional connections, exchange job search information and get recommendations. We respectively modeled the stickiness and the co-evolution at an individual user level. We apply a hierarchical Bayes regression model for examining each of the three stickiness measures by comparing between-sample and within-sample of pre- and post group joining behaviors. Our results indicate the presence of a significant effect of group formation on the websites’ stickiness. The group function is proved to be more effective in creating stickiness. We find that some group topics facilitate users’ activity. Results also suggest the presence of a co-evolution effect in the sense that group membership tends to increase primary membership on the website.

The stream of my dissertation research contributes managerially both to forecasting sales of new products and evaluating effectiveness of new forms of consumer generated media, especially in the presence of traditional marketing communication. The addition of new media (CGM) seems to improve the prediction of market outcomes. Moreover, the documentation of a positive and synergistic relationship between new and traditional media is helpful for managers to optimize traditional marketing activity.

Tables and Figures

Figure I-1 Framework of the dissertation



References

ARF Defining Engagement Initiative (2006); http://www.thearf.org/assets/research-arf-initiatives-defining-engagement?fbid=bB_008jPX7a

Ackerberg, D. A. (2003), “Advertising, Learning and Consumer Choice in Experience Good Markets: An Empirical Examination,” *International Economic Review*, 44 (3), 1007-1040.

Calder, B.J., E.C. Malthouse and U. Schaedel (2009), "An Experimental Study of the Relationship between Online Engagement and Advertising Effectiveness," *Journal of Interactive Marketing*, 23 (4), 321-331.

Chintagunta, P. K., S. Gopinath and S. Venkataraman (2009), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation across Local Markets," *Marketing Science*, 29(5), 944-57.

Dellarocas, C., X. Q. Zhang, and N. F. Awad (2007), "Exploring the value of online product reviews in forecasting sales: The case of motion pictures," *Journal of Interactive Marketing*, 21 (4), 23-45.

Dhar, V. and E. Chang (2009), "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Journal of Interactive Marketing*, 23 (4), 300-307.

Duan, W., B. Gu and A. Whinston (2008), "Do Online Reviews Matter? An Empirical Investigation of Panel Data," *Decision Support Systems*, 45 (4), 1007-1016.

Erdem, T., M. Keane, T. Öncü and J. Strebel (2005), “Learning About Computers: An Analysis of Information Search and Technology Choice,” *Quantitative Marketing and Economics*,3 (3), 207-247.

Gentzkow, M. (2007), “Valuing New Goods in a Model with Complementarity: Online Newspapers,” *American Economic Review*, 97 (3), 713-744.

Godes, D. and D. Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545-560.

Gruhl, D, R. Guha, R. Kumar, J. Novak and A. Tomkins (2005), "The Predictive Power of Online Chatter," KDD 2005, Chicago, IL, August 2005.

Jayanti, R. K. and J. Singh (2010), ”Pragmatic Learning Theory: An Inquiry-Action Framework for Distributed Consumer Learning in Online Communities,” *Journal of Consumer Research*, 36 (April), 1058-1081.

Liu, Y. (2006), "Word of mouth for movies: Its dynamics and impact on box office revenue," *Journal of Marketing*, 70 (3), 74-89.

Luan, Y. J. and S. Neslin (2010), “The Development and Impact of Consumer Word of Mouth in New Product Diffusion," Working Paper, Tuck School of Business, SSRN 1462336.

Chapter II

Marketing Activity, Blogging and Sales

Abstract

The recent growth of consumer generated media (CGM), also known as “new” media, has changed the nature of interaction between consumers and firms from unidirectional to bidirectional. This research addresses two key issues - does CGM affect consumer behavior and market outcomes and, is there any relationship between managerial communication and CGM?

The most prevalent form of new media is blogs. Thus, we first investigate whether blogging activity leads to (differential) market outcomes. We then examine whether managerial communication (TV advertising) and blogging are synergistic.

We assemble a unique data set from Japan containing market outcomes (sales), new media (blogs) and traditional media (TV advertising) for green tea drinks, movies and cellular phone service. Each category has at least one product launch during the duration of our sample periods.

We specify a simultaneous equation log-linear system for market outcomes and the volume of blogs. Our results suggest that blogs are predictive of market outcomes, new and traditional media act synergistically, pre-launch TV advertising spurs blogging activity but become less effective post-launch and that market outcomes have some effect on blogging. We find detailed support for some of these findings via a unique and novel text mining analysis. We discuss the managerial implications of our findings.

II-1 Introduction

Consumer generated media (CGM) such as blogs (a contraction of the term "Web logs") have witnessed explosive growth in the last few years. For example, the number of blogs worldwide is estimated to be 184 million with a readership of 346 million (Technorati/Universal McCann, March 2008). In contrast, in March 2003, the number of blogs was essentially zero. Other types of CGM have also seen similar growth patterns, *e.g.*, Facebook, which started in February 2004, now has about 400 million members worldwide (February 2010 Comscore Media Metrix traffic estimates). There are also indications that blogs are now being seen as similar to mainstream media sites – the number of blog sites in the top 100 most popular sites (blogs and mainstream media) worldwide was twenty-two in 2007 and blogs were being viewed by consumers as "sites for news, information, gossip etc." (Technorati 2007). In 2008, four of the top ten entertainment sites were blogs (Technorati/Universal McCann, March 2008).

It is clear from these statistics that there is considerable activity (multi-media posting, blogging, visits, traffic etc.) on the part of consumers. However, an important question, from a managerial perspective, is whether this activity leads to (differential) business outcomes such as sales or profits. In addition, little is known about the relationship between traditional or old media (where the company creates content and delivers it to consumers) and consumer generated, or new, media (where consumers create content and there in an exchange of this content between other consumers and potentially, the company). That is, are there any synergies between new media and old media? In this research, we take the first step towards answering these questions.

Blogging is perhaps the most established and largest form of consumer generated media at this point in time. The total worldwide viewership of blogs is estimated to be about 346 million (Technorati/Universal McCann, March 2008). Wikipedia defines as a blog as "a Web site, usually maintained by an individual with regular entries of commentary, descriptions of events, or other material such as graphics or video. Entries are commonly displayed in reverse-chronological order." Blogging is a worldwide phenomenon with the two biggest blogging markets being the United States and Japan. The number of blogs in the United States is about 23 million (about 12% of all US Internet users) and about 8 million in Japan (about 5% of all Japanese Internet users) in 2007 (Technorati/Universal McCann, March 2008). However, if one examines the total number of posts by language, Japanese language posts account for 37% of all posts worldwide followed closely by English language posts at 36% (Technorati 2007). Finally, readership of blogs in these two markets is very high - about half of all Internet users in the US and about one-fifth of all Japanese Internet users have read a blog in the past year.

While there are many informal opinions on the effectiveness of CGM in general (and blogs in particular) vis-à-vis market outcomes, there is limited empirical research that sheds light on this issue, especially for the launch of new products. The majority of the existing research has focused on online chatter (newsgroup postings, reviews and ratings) and its effect on market outcomes. There is some evidence that volume of online user ratings is positively correlated to sales (Liu 2006; Dellarocas, Zhang, and Awad 2007; Duan, Gu and Whinston 2008; Chintagunta, Gopinath and Venkataraman 2009). In addition, Dellarocas et al. (2007) found that the average valence (positive/negative) of user ratings also explained future movie sales as well in their diffusion model. On the other hand, Godes and Mayzlin (2004) found that only the dispersion of Usenet conversations among different newsgroups was significant but the volume and the valence of online conversations were not correlated with TV shows' viewership.

Given the goal-driven nature of product ratings (and reviews), these findings may not be that unexpected. Blogging, on the other hand, has been seen as a unique type of user generated content as being a highly personal, non-directed communication tool. As Kumar et al. (2005) note, blogs are unique for sociological reasons – they comprise a "highly dynamic, temporal community structure" that "focuses heavily on local community interactions" - and for technical reasons – blogs "offer us a ready-made view of evolution (of content) in continuous time." In addition, blogging activity was probably the most pervasive CGM activity on the web during the time of our data. Given these unique characteristics of blogs as opposed to reviews, it is not obvious that bloggers' activity should affect market outcomes.

Surprisingly, there is very little research that has tried to quantify the effect of blogs on market outcomes, especially in the presence of traditional media and/or an examination of pre- and post-launch changes in the role of old and new media (we provide the detailed list of past studies and comparisons in Table II-1). Two recent empirical papers have focused on blogs and market outcomes.[1] Dhar and Chang (2009) explore the relationship between music album sales (imputed via sales ranks on Amazon.com) and online chatter (as seen in blogs and on social networks). Using 108 music albums in early 2007 (before four weeks and after four weeks of their release), they find a positive correlation between both the number of blogs and Myspace member intensity with future music sales. Gruhl et al. (2005) propose a new methodology to automatically generate a query of blog keywords to detect spikes in Amazon.com's book sales rank. They conclude that their new algorithm could adequately predict the changes and spikes of future sales ranks. Thus, while these two studies suggest that there may be a correlation between blogging activity and market outcomes, they do not use actual sales data but only sales ranks from Amazon.com.[2]

To the best of our knowledge, the second issue that we outline above - the positive relationship between traditional media and new media - has not been investigated in the literature. Our expectation is that there will be a positive correlation between the quantity of traditional media and new media as traditional media is likely to provide discussion materials for bloggers. From a managerial perspective this issue is crucial, as managers have no direct control over CGM (blogs in our case). However, if there is indeed a synergistic relationship between traditional media, which are under managerial control, and new media, which are outside managerial control, then managers can leverage this relationship. Specifically, they can carry out "better" resource allocation and media planning (to

---

[1] There is also emerging work on other (non market outcome) aspects of blogs e.g., linking decisions between and across blogs (see Mayzlin and Yoganarasimhan 2008).

[2] This is an important issue as (a) there is a potential for the rank-sales relationship to be very noisy (especially if Amazon.com is not a representative seller of the good in question), (b) ranking data are very idiosyncratic and usually not available for a wide variety of products (e.g., movie box-office sales, packaged goods, services etc.) and, (c) it is not clear if managers can use the imputed sales data to obtain actual effect sizes in order to set policy.

traditional media) as they can take the spillover effect (from traditional to new media) into consideration.

We examine the role of new media with respect to market outcomes as well as the relationship between new media and traditional media using data of new product launches from three diverse product markets – green tea drinks, movies and cellular phone service – in Japan. We consider the number of units sold, customers or subscribers (all a proxy for demand) as market outcomes, blogs as representations of consumer generated media and TV advertising as traditional media. We specify a simultaneous equation model that links sales to advertising and blogs as well as a model that links blogs to advertising.

Our results, after controlling for many temporal and cross-sectional factors, suggest that first, the volume of Blogstock (cumulative sum of past blog posts) is positively correlated with market outcomes (volume of green tea drinks sold, the number of movie goers and cellular phone subscribers) post launch. Second, the interaction between blogs and TV advertising has a positive effect on market outcomes. Third, we also find that traditional media (TV advertising) positively affects new media (the volume of blogs) pre launch. In other words, bloggers consume advertising, independent of the product, and this increases their blogging activity. Finally, we find that the effect of blogs varies between pre and post launch. In general, the positive relationship between TV advertising and the volume of blogs pre-launch becomes weaker after launch. This result suggests that while TV advertising can independently increase blogging pre-launch via the provision of information and content, post-launch (*i.e.*, once the product is available), consumers may rely less on traditional media, leading to a much weaker relationship between new and old media at that point. These last three sets of results shed light on the possibility that, broadly speaking, advertising and blogs act synergistically (with the relationship changing somewhat post-launch). To the best of our knowledge, the documentation of this synergistic relationship between traditional and new media is a novel finding.

Given the aggregate nature of our data, the process explanations for our findings is not obvious. We take the first step in eliciting process explanations by carrying out a novel text mining analysis of the blog posts for two of the product markets for which we have access to the textual content data. The findings from the text mining analysis suggest that blogs may affect market outcomes as they represent a rich source of product information and consumer opinion for other consumers. Also, bloggers do use advertising as a subject for blogging pre-launch but turn their attention to product attributes post-launch. Finally, we carry out some simple simulations to compute the economic significance of the relationship between market outcomes, blogging volume and TV advertising.

The rest of the chapter is organized as follows. Section II-2 describes the institutional setting and the data. We discuss the model next in Section II-3. Section II-4 contains the results, details on text mining and managerial implications. We conclude in Section II-5.

II-2 Data

Our data come from the Japanese market. As mentioned earlier, Japan is in the top two markets worldwide in blog participation and the number of blog posts. We consider data from three Japanese product markets – green tea drinks, movies and cellular phone service. We first describe the market outcome data for each product market and then we describe the measurement of traditional and new media.

II-2.1 Market Outcomes

The daily sales of green tea drinks were made available to us for the total Japanese market based on a nationally representative consumer panel maintained by Video Research, Ltd. The data include daily sales of five new green tea drink brands introduced in the period from May 2006 to August 2006 (Table II-2 provides more details on the five brands and their market positions). Details on daily sales post-launch are provided in Table II-5.

For movies, the outcome variable we use is the size of a movie's audience for each day after the launch of the movie. These data were obtained from Kyogyo Tsushinsha – an industry periodical. We have data of twelve major movies that were released (launched) in the period from January 2007 to August 2007. The movie titles were chosen to reflect sufficient variation in movie genre (action, animation etc.) across American, European and Japanese movies – the list of titles is available in Table II-3 while Table II-5 describes the daily sales patterns post-launch.

Finally, for cellular phone service, we obtained the monthly number of subscribers from the TCA (the Japanese Telecommunication Carriers Association) website. As listed in Table II-4, we use data on monthly subscribers from five Japanese cell phone companies for the eighteen-month period from November 2006 to April 2008. In contrast to the other two datasets, we observe only one launch in this product market (EMOBILE entered the Japanese market as a new cellular phone carrier in May 2007). We therefore use that data to define the pre and post launch period. Table II-5 details the sales data for this product market.

II-2.2 Traditional Media

The traditional marketing variable we use is TV advertising. This was measured in units of daily or monthly Gross Rating Points (GRPs). The measure of national GRPs delivered to Japanese households for all the brands across the three product markets was provided to us by Dentsu Inc. (the largest advertising agency in Japan). Details on the TV advertising across the three product markets are provided in Table II-6.

As can be seen from the tables, there are some differences in the patterns of TV advertising pre and post launch across the three product markets. For green tea drinks, most of the advertising is post launch. Typically, TV commercials in this market begin to air about five days pre launch and then the heavier advertising kicks in post launch. In contrast, for movies, pre-release TV GRPs are larger (on average) than the post-release TV GRPs. Specifically, peak advertising for a movie was, not surprisingly, a week before its launch date in order to generate high demand at the time of the opening. Finally, for the cellular phone service market, there is a little difference in TV advertising on average pre and post launch. This is probably due to the fact that our data contains only one brand that was launched during this period.

II-2.3 New Media

We obtain blogging data from Kizasi Co., Inc (www.kizasi.jp) for green tea drink data and Dentsu Buzz Research Inc (www.dbuzz.jp) for the movie and cellular phone service data. Both the companies scan and index the major blogging sites in Japan on a daily basis using keywords with coverage of about 64% of all blog articles in 2008.[3] They then aggregate the data and provide the count of the daily number of blogs that mention a particular keyword on a specific temporal period such as day or month (multiple mentions in the same temporal unit are counted as one). A typical blog that includes a discussion about green tea drinks can be seen in Figure II-1. This is a blog focused on the movie title "Spider-man 3" on its release date of May 1st, 2007. As is typical for most blogs, its contents appear in a reversal chronological order and also include the blogger's profile, "trackbacks" (links showing other websites, typically other blogs, that a blog is linked to), and comments.

Buzz Research archives the contents of all blog posts. It also carries out lexical analysis of the contents of each tracked blog by using a proprietary text-mining method and classifies each blog as positive, negative and/or neutral with respect to a given keyword.[4] We therefore have access to the actual content of all posts as well as the daily percentage of positive, negative and neutral blogs for the movies and cellular phone service markets. Tables II-7 and II-8 contain details on the number of blogs pre and post launch as well as the number of positive, negative and neutral blogs for the movies and cellular phone service markets.

As can be seen from the tables, there is big increase in the average number of blogs per period post launch in all three product markets. Interestingly, for the two product markets where we have valence data, the biggest growth is in the percentage of neutral blogs post launch.

To illustrate the relationship between marketing outcomes and both traditional and new media, we pick a product across our three product markets.

---

[3] Note that none of the brands in our analysis had their own "corporate" blog.

[4] The algorithm used for this lexical analysis is proprietary and thus, unfortunately, we have no access to its details.

Figure II-2 contains the data series from the movie "Spider-man 3." The figure suggests that TV advertising, blog volume and movie viewership are temporally correlated. Dividing the data temporally at the date of release (May 1, 2007), we see that TV GRPs and the number of blogs exhibit an increasing trend pre-release, but a decreasing one post-release. While we illustrate a typical data pattern through this example, the pattern is not identical for all brands across product markets.

In conclusion, these data are novel in the sense that they combine marketing data for both traditional and new media along with market outcomes from a market where new media have proven to be important (at least in terms of activity). Our data are also novel in the sense that they enable us to focus on new product launches. In addition, the fact that we have data from three different product markets (frequently purchased consumer packaged goods, experiential goods and services) with varying characteristics (*e.g.*, more versus fewer new product launches) will help us determine if the relationship between market outcomes and new media as well as the relationship between new media and traditional media generalizes across product markets. Finally, the availability of the actual blog post text (for two categories) opens up the possibility to conduct a deeper text-mining analysis (details are in Section II-4.3). We also reports correlation coefficient matrices of the variables used in our empirical in Tables II-9, II-10 and II-11.

II-3 Model

Our model contains two dependent variables – market outcomes (daily volume of green tea drinks sold, daily movie viewers and monthly cell phone subscribers) and the volume (number) of blogs. As mentioned earlier, it is quite feasible that market outcomes and the volume of blogs are determined simultaneously. Our model therefore takes the form of a system of equations (with correlated errors) in order to account for this potential simultaneity.

Given our data, the market outcomes are a function of past market outcomes, TV advertising, blog activity (volume and valence), interaction terms and a set of control variables. Specifically (to take the case of movies), the number of customers (the daily audience) is a function of the past number of customers (both in terms of the number of customers in the last time period and the cumulative number of customers up to the last time period) who are now presumably out of the market.

In the case of new products, where quality is difficult to evaluate before purchase, consumers look to other consumers to seek information and obtain opinions (Rogers 1983, Harrison-Walker 2001). One such source of information and opinions is the blogosphere (Dhar and Chang 2009). Blogs are unique, relative to other user generated content, in the generation and transmission of information (Kumar et al. 2005). In addition, 75 percent of blog readers consider information in blogs compared to 43 percent for online newspapers, magazines etc. (Johnson and Kyle 2004). Thus, we expect that the number of current and past blogs (the current number of blogs and a number of cumulative blogs) that discuss a product will have a positive effect on market outcomes post launch. Besides the volume of information available from others, previous research has also shown that the valence of this information impacts product performance (Richins 1983, Nam, Manchanda and Chintagunta 2006) and especially when both positive and negative information is present (Rozin and Royzman 2001). Thus, we expect that the valence of current blogs (number of positive blogs and number of negative blogs), will also affect market outcomes over and above the volume of blogs.

Besides opinions of others, consumers may also respond to marketing actions by the firm, primarily communication in the form of TV advertising. The relationship between sales (or share) and advertising is well documented in the marketing (*e.g.*, Lodish et al. 1995) and economics (*e.g.*, Bagwell 2007) literatures. Thus, we expect to see a direct effect of current and past advertising on market outcomes.

Based on the above, it is likely that that consumers may not only find information from other consumers and from the firm directly useful, but the two sources of information (and persuasion) could also potentially reinforce (or weaken) each other. We therefore include an interaction term between current and past blogging volume and current and past TV advertising. To the best of our knowledge, this is the first time that this interaction has been studied and hence there is no clear theory that predicts the sign of the interaction terms.

In addition to these variables, we include a set of control variables. These control variables are temporal (trend from launch and day of week fixed effects) as well as cross-sectional (movie-specific fixed effects). These controls for diffusion are driven purely by temporal phenomena and movie heterogeneity respectively. In the case of movies, for example, the movie fixed effect controls for all (non-varying) idiosyncratic aspects of the movie such as stars, director, awards, genre, MPAA rating etc. (all parts that contribute to its overall quality). Other factors that could affect the size of the daily audience (*e.g.*, distribution represented by the number of movie screens) are also controlled for by the lagged audience variable and the movie fixed effects. For empirical analysis, we transform these variables by taking their natural log, resulting in a log linear structure. The specification is similar for both the other product markets.

There are two issues that need to be addressed regarding this specification. First, it is possible that advertising is endogenous. In other words, the level of advertising is related to the error term as a result of some factors that are not observed by the researcher (but are by the manager). When we discussed this with the advertising agency that provided us the data, we were informed that for product launches, advertising plans (level of GRPs and schedules) were set far in advance of the launch and were usually not varied after launch. Thus, this institutional feature of the setting allows us to assume that advertising is exogenous. In order to make sure that this was indeed the case in the data, we also estimate an alternate version of our model where we allow advertising to be an additional dependent variable. In other words, we estimate a system of equations with three (market outcome, volume of blogs and advertising GRPs) dependent variables and allow the errors to be correlated. We find that our results are unchanged (details on these results are reported in Appendix II-D). Given these two reasons, we specify our model with two dependent variables – market outcomes and volume of blogs.

Second, we do not include price in the specification. This is because for all the product markets, there was no variation in prices over the time period of our data.

This is not surprising as movie audiences are used to fixed prices. This is also true for green tea drinks as the vast majority of sales of these drinks are at vending machines and convenience stores (where prices are fixed) in our data set. Finally, for cellular phone service, EMOBILE launched a series of calling plans but did not change the prices and the attributes of these plans for the duration of our data.

Thus, the market outcome equation specification is given in Equation (II-1). In this and subsequent equations, j $(=1,\ldots, J)$ represents a product while t $(=1,\ldots,T_j)$ represents time (daily/ monthly) for each brands.

(II-1)

$$\begin{aligned}\ln(\#\text{Customers}_{jt}) &= \alpha_0 + \alpha_1 \text{Trend}_t + \alpha_2 \text{Season}_t + \alpha_{3j} \text{Brand Dummies}_j \\ &+ \alpha_4 \ln(\#\text{Customers}_{jt-1}) + \alpha_5 \ln(\#\text{Blogs}_{jt}) + \alpha_6 \ln(\#\text{PositiveBlogs}_{jt}) + \alpha_7 \ln(\#\text{NegativeBlogs}_{jt}) \\ &+ \alpha_8 \ln(\text{Salestock}_{jt}) + \alpha_9 \ln(\text{Blogstock}_{jt}) + \alpha_{10} \ln(\text{Adstock}_{jt}) \\ &+ \alpha_{11} \ln(\text{Blogstock}_{jt}) \times \ln(\text{Adstock}_{jt}) + \varepsilon_{1jt},\end{aligned}$$

As mentioned earlier, our second dependent variable is the count of the daily number of blogs that mention a particular keyword on a specific temporal period such as day or month (multiple mentions in the same temporal unit on the same blog are counted as one).

We hypothesize that as the number of customers who consume a product increases, the more these customers are likely to disseminate information (Rogers 1983), especially if some bloggers have an asymmetric relationship with their followers in the sense that their opinions are more valued and follow them (Christakis and Fowler 2007 and Nair, Manchanda and Bhatia 2006). Thus, we include the number of current and past customers (the current number of customers and the number of cumulative customers, *i.e.*, Salestock, up to the previous period) as independent variables.

Second, as noted earlier, prior to launch, consumers (including bloggers) cannot try out the product. However, they are exposed to marketing efforts of the firm – TV advertising in our case. There is a very long history of research in marketing that has documented that consumers not only extract information and form attitudes about the product advertised, but also form attitudes directly towards the advertisement. The attitude toward the ads, has been seen to mediate the relationship between beliefs and attitude to the brand (Mitchell and Olson 1981, Brown and Stayman 1992), which in turn influences purchase intentions. We therefore expect bloggers to be influenced by advertising, via the development of attitudes both about the ads and the brand, and then to disseminate opinions about the advertising and the brand in their posts. We therefore include the current and past advertising as an explanatory variable.[5] Relative to pre-launch however, we expect that bloggers, especially influential ones, will have more access to actual products, and will therefore base their posts less on advertising content and more on product experiences and attributes.

Third, in order to control for the "viral" nature of user generated content and the unique information network of blogs (Kumar et al. 2005), we also include the number of past blogs (the number of blogs at the previous time period and the number of cumulative blogs, *i.e.*, Blogstock, up to the previous time period) as explanatory variables. It is possible that the stock of blog posts and advertising acts synergistically to produce more blog posts. We therefore include the interaction term of blogs and advertising. Finally, we also include an instrumental variable as an exclusion restriction to assure the identification of our system of two equations.[6] As the standard methodology in the econometrics literature, each equation in a system should have at least one independent variable which is correlated with the dependent variables of the equation but not correlated with the other dependent variables in the other equations in the system. For this purpose, we pick the sum of the number of blog posts of other products in the category – that is, the number of current blog posts of the category excluding the number of blog posts of the focal brand (*i.e.,* the value of the left-hand-side of the equation). Since the number of category blog posts is correlated only with the number of current blogs but not with

---

[5] We also confirmed with our data provider(s) that the firms' expectation was that their advertising would have a direct effect on sales. Thus, firms were not setting advertising with the objective of influencing the blogosphere directly through their advertising. We thank an anonymous reviewer for bring this possibility to our notice.

[6] We also expected that the brand fixed effects and trends may provide resources for identification since these values are estimated differently in the both equations.

the market outcomes, this instrument term should be work as the exclusion restriction.

As noted above, we expect the actual launch of the product to represent a structural change in the environment. We therefore specify different coefficients for the pre- and post-release periods for each of the independent variables. Finally, as in Equation (1), we also include the control variables. The resulting specification can be seen in Equation (II-2).

$$
\begin{aligned}
\ln(\#\text{Blogs}_{jt}) = {} & \beta_0 + \beta_1 \text{Trend}_t + \beta_2 \text{Season}_t + \beta_{3j} \text{Brand Dummies}_j \\
& + \beta_4 \ln(\#\text{pre Blogs}_{jt-1}) + \beta_5 \ln(\text{pre Blogstock}_{jt}) + \beta_6 \ln(\text{pre Adstock}_{jt}) \\
& + \beta_7 \ln(\text{pre Blogstock}_{jt}) \times \ln(\text{pre Adstock}_{jt}) \\
& + \beta_8 \ln(\#\text{post Customers}_{jt}) + \beta_9 \ln(\text{post Salestock}_{jt}) \\
& + \beta_{10} \ln(\#\text{post Blogs}_{jt-1}) + \beta_{11} \ln(\text{post Blogstock}_{jt}) + \beta_{12} \ln(\text{post Adstock}_{jt}) \\
& + \beta_{13} \ln(\text{post Blogstock}_{jt}) \times \ln(\text{post Adstock}_{jt}) \\
& + \beta_{14} \ln(\#\text{post other products' Blogs}_{jt-1}) + \varepsilon_{2jt}.
\end{aligned} \tag{II-2}
$$

As discussed earlier, there exists the possibility of simultaneity between the market outcomes (green tea volume/ movie viewers/ cell phone subscribers) and the number of blogs. We therefore allow the errors in Equations (II-1) and (II-2) to be jointly distributed bivariate normal as in Equation (II-3).

$$
\begin{bmatrix} \varepsilon_{1jt} \\ \varepsilon_{2jt} \end{bmatrix} \sim MVN(0, \Sigma), \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{II-3}
$$

Note that all the stock (cumulative) variables are defined using the standard distributed lag structure (Rao and Miller 1975) as below in Equation (II-4).[7]

$$
\begin{cases}
\text{Salestock}_{jt} = \#\text{Customers}_{jt-1} + \gamma_1 \text{Salestock}_{jt-1} \\
\text{Blogstock}_{jt} = \#\text{Blogs}_{jt-1} + \gamma_2 \text{Blogstock}_{jt-1} \\
\text{Adstock}_{jt} = \text{TV GRP}_{jt} + \gamma_3 \text{Adstock}_{jt-1}
\end{cases} \tag{II-4}
$$

[7] The exact value of the carryover constant is estimated via a grid search with best fit as the objective function. Details on this analysis are in Appendix II-C.

We use feasible generalized least squares (FGLS) to estimate the model (Wooldridge 2001, p.157).

II-4 Results

II-4.1 Parameter Estimates

Results of the sales volume equation

We first focus on the sales volume equation (Tables II-12, II-13 and II-14). We find somewhat mixed results for the coefficient of volume of current blogs – this is not significant for all three product markets. Interestingly, however, the coefficient of the volume of current blogs that are classified as having a positive valence is positive and significant for movies and cellular phone services. This result is in contrast to previous findings for online reviews for movies that have found that the valence of the reviews is not significantly correlated with outcomes (*e.g.*, Liu 2006 and Duan, Gu and Whinston 2008 found that the number of positive reviews was not predictive of movie sales). More importantly, we find that the volume of Blogstock is positive and significant for all three product markets. This suggests that the volume of blogs, especially cumulative volume, is a predictor of market outcomes. In general, we expect that these effects are driven by the fact that blog posts represent information and opinion about service experiences and product attributes by other consumers – this makes these posts a rich source of insights to consumers at large.

The coefficient for GRP Adstock is positive and significant for green tea drinks and movies (consistent with Elberse and Anand 2007) but negative and significant for cellular phone service. While the latter result may seem counterintuitive, a closer look at the data patterns suggests the reason for this finding. In order to defend against the new launch, the incumbents raised their total advertising spend by about thirty percent and twenty percent respectively in the first and second period after launch. However, their sales remained virtually flat after launch (the new brand stole a very insignificant amount of market share). This coupled with the use of the Adstock formulation, results in a situation of flat sales with increased advertising. Thus, the estimated coefficient ends up being negative.

There is a variation in the estimated values of the self lag of sales volume, #Consumer(-1), across the product categories. These are negative and significant for green tea drinks and cellular phone service but not for movies. In contrast, the coefficients of the Salestock (cumulative number of customers) are positive and significant for all the three categories. Especially in movies, the audience size during the first opening weekend generally has the most significant impact on market outcomes (Liu 2006; Elberse and Eliashberg 2003). Thus, it is consistent that the current audience size has the strong immediate correlation with the previous audience size. Whereas, in green tea drinks and cellular phone service, there is only a long-term carry-over impact from the past values of sales volume but not an immediate impact.

Continuing to the interaction term, we find that the interaction term between the cumulative number of blogs and the Adstock is positive and significant for all three product markets. This suggests that the two media (new and traditional) act synergistically – an interesting and hitherto undocumented finding. As expected, the lagged outcome terms are significant for all three product markets.

Results of the blog equation

Moving on to the equation with the number of blogs as the dependent variable, we examine the pattern of results (Tables II-15, II-16 and II-17) by looking at the pre- and post- launch estimates. We first look at the effect of advertising on blog production. Across all three product markets, we find that pre-launch, advertising leads to a positive and significant increase in the number of blogs. Interestingly, this pattern changes at post-launch. In general, across the three product markets, the effect of advertising on blogs becomes weaker – the effect is dramatically lower for green tea drinks, goes to zero for movies and remains about the same for cellular phone service.[8] As more and more consumers are able to sample the product (post-launch), it is likely that bloggers are less affected by traditional advertising as they can rely on their own service experiences and/or information about product

[8] Recall again that for cellular phone service, the post-launch effect is calibrated off only one brand (out of five).

attributes and performance. Again, these findings are novel and shed some light on the relationship between traditional and new media.

We also find that the past Blogstock is predictive (positive and significant) of the current number of blogs in a given period both pre- and post-launch (except for green tea drinks). This finding also seems to be reasonable as an anecdotal evidence suggesting that the blogosphere feeds off of itself.

Turning to the effects of marketing outcomes on blogging, we find that in general, there is some relationship between the two. For movies, the contemporaneous audience size has a positive and significant effect on the number of blogs (the effect is insignificant for green tea drinks and cellular phone service). The Salestock in the case of movies has a negative and significant effect on the number of blogs. This perhaps arises from the fact that movie audiences decline rapidly (*e.g.*, as shown in Figure II-2) but the discussion of movies persists in the blogosphere. Also, movies are typically a one-time consumption good. In contrast, for green tea drinks, the cumulative number of customers has a positive and significant effect on the number of blogs, reflecting perhaps the repeated experiences associated with consuming these drinks.

In conclusion, we find patterns across the three categories that show clear linkages between traditional media, new media and market outcomes. The most salient results can be summarized as follows – (a) the volume of blogs that exist about a product is predictive of market outcomes, (b) the valence of recent (current) blogs could also be predictive of market outcomes, (c) the effect of blogs and advertising is synergistic on market outcomes (d) pre-launch advertising spurs blogging activity (that is predictive of marketing activity) but becomes less effective in inducing blogging activity post-launch and (e) market outcomes do have some effect on blogging activity.

II-4.2 Model Fit

We also examine our model fit both in and out of sample, relative to two baseline models. In the first baseline model (Model 1), we retain only the fixed effects, the time trend and seasonality dummies. In the second baseline model (Model 2), we include the covariates in Model 1 and add the lagged dependent variables. For the in sample data, as can be seen from Table II-18, the addition of the lagged dependent variable improves the fit (based on the Akaike Information Criterion or AIC, R-square and the Adjusted R-Square) considerably (Model 2 versus Model 1) for both equations across all three product markets. The addition of the traditional and new media variables increases the fit of the outcome equation significantly (Full Model versus Model 2) for green tea drinks and for cellular phone service while the increase in fit for the movie product market is very modest. For the blog equation, the inclusion of the additional variables (in the Full Model) shows a reasonable improvement for Movies but not for the other two product markets.

We also examined the predictive ability of the model relative to the baseline models by carrying out two holdout tests. In the first test, we hold out the last observation for each product or movie. In the second test, we hold out the data for one of the products chosen at random in each of the product markets (Kiki-Cha in green tea drinks, Shrek the Third in movies and EMOBILE in cellular phone service). Tables II-19 and II-20 show that the full model performs much better (based on the AIC) than Model 1 baseline model in all cases. The same patterns can be seen for Model 2 with the exception of the blog equation for the movie product market where the full model does almost as well as Model 2.

II-4.3 Text Mining and Analysis

For two of the three product categories that we consider – movie and cellular phone service – we have the textual contents of the blogs (all posts) available to us. We therefore decided to carry out a text mining analysis of these blogs with the objective of providing more insights into the results (that are based on counts and valences) discussed above.

Our procedure was as follows. We first looked at the number of posts across all blogs in each category. For movies, the number of posts was about 200,000 and for cellular phones, the number of posts was approximately 1,560,000. Using a lexical software program (KH Coder) for Japanese text, we carried out the following steps:

1. Parse each sentence in the blog posts into words. This is an important step as text is not naturally delimited into words in Japanese. Then classify words into lexical categories (nouns, verbs etc.).
2. Assess words for tense (*e.g.*, the same word in three different tenses counts as three unique words).
3. Carry out a frequency count of all unique words.
4. Choose words relevant to our analysis.
5. Compute the number of times a relevant word appeared across all blog posts for the temporal unit of analysis (*e.g.*, day or week).

This was quite a laborious task as it involved frequent human intervention to see that the parsing, classification and word frequencies were correct. We were able to carry out this process for all the blog posts (200,000) for movies. However, given the nature of the task, we drew a five percent random sample (260,000 posts) for cellular phone service. The top 30 words by frequency are given in Tables II-21 and II-22. We also show the frequencies for words that were germane to our analysis, *e.g.*, "advertising" (see discussion below).

Our objective was to provide some "process" explanations for our findings above. We focus on two of our main findings – blogs are predictive of market outcomes and the role of traditional media and new media pre- and post-launch. Our explanation of the first finding (above) was that blog posts represent useful information to consumers in terms of their decisions to evaluate products and services, leading to an eventual purchase decision. Our explanation of the second finding was that bloggers use firm actions such as advertising much more pre-launch than post-launch (where the product experience was likely to dominate). Thus we focused on the content of blogs, both pre- and post-launch, to help us shed light on both these findings.

We chose four words for each product market. For movies, we chose "advertising," "award,"[9] "interesting" and "viewed" as we expect that consumers would find posts that contained these words (especially "award," "interesting" and "viewed") useful in terms of making their own decision to see the movie. We also expect "advertising" and "award" to represent a higher proportion of blog posts pre-

---

[9] All the non-Japanese movies in our sample had been released previously in their home countries. They had been candidates for awards or had won awards already.

release as they reflect informational attributes that are available pre-launch and "viewed" to represent a higher proportion of posts post-launch as it would represent discussions based on consumers' movie viewing experience. For cellular phones we chose "advertising," "interesting," "subscribed," and "price" – again as these words (especially the latter three) can be seen as diagnostic about the service. Also, as before, we expect "advertising" and "price" (prices are usually not announced in this product market pre-launch leading to considerable interest in pricing) to account for a higher proportion of blog posts pre-launch and "subscribed" to have a higher proportion post-launch. We included the word "interesting" as it had a high rank for movies. However, "interesting" could pertain to movies and cellular phone service pre- or post-launch. Thus, we had no prediction about the proportion of blog posts that use the word pre- or post-launch.[10] Tables II-23 and II-24 show the proportion of blogs that use each of the chosen words. The raw proportion of mentions (to shed light on our first findings) is not informative in itself (our statistical analysis below speaks to that), but it is useful to look at the pre- and post-launch proportions. As can be seen from the tables, the pre- and post-launch proportions are generally in line with our expectations (except for "price"). Specifically, bloggers mention advertising in their blog posts much more than pre-launch and mention viewing and subscribing (along with price) much more after launch. This suggests that traditional advertising is probably used to enrich blog posts pre-launch while service experiences and knowledge of product attributes matter to bloggers (and presumably, to their readers) much more post-launch.

To validate these data patterns statistically, we ran our full model and included the number of blogs mentioning these four words as covariates. For the sales equations (see Tables II-13 and II-14), we find that "viewed," "subscribed" and "price" are significant (with the expected signs) while "award" is negative and significant. "Interesting" is not significant for both product markets. This suggests that blogs affect market outcomes because they represent a direct and valid source of information of consumer opinion for other consumers. Turning to the blog

[10] The chosen words had to be general enough (e.g., no proper nouns), have the possibility to be used differently pre- and post-launch and ranked reasonably high in (word) frequency. We arrived at these words using these criteria.

equations (Tables II-16 and II-17), we find evidence that the mention of "advertising" is positive and significant pre-launch but not post-launch for both product markets. Similarly, for movies, "award" is positive and significant pre-launch but not post-launch. While "viewed" is significant both pre- and post-launch, the size of post-launch coefficient is twenty times as large post-launch. For cellular phones, price is positive and significant pre-launch but not post-launch and "subscribed" is positive and significant post-launch. Thus, the content of blog posts based much more on traditional media (relative to post-launch), and experience and/or product attribute mentions affect blogging activity post-launch. Taken together, the text mining (and the associated statistical analysis) helps to shed light on our econometric findings.

II-4.4 Managerial Implications

So far, we have discussed our findings purely from a statistical point of view. However, it may be useful to translate these findings in a manner that quantifies the effect sizes from a managerial point of view. We therefore ran two experiments – the first to get a sense of how managers could change resource allocation and the second to see how managers could use blog data to improve sales forecasts.

In the first experiment, we use the estimates from the green tea drink market data. To illustrate short-term effects, in the experiment, we assumed there were only three periods, two in the pre-release and one in the post-release. Recall that blogging is outside the control of managers. We therefore used the marketing instrument under managerial control in our data set – traditional TV advertising.[11] In the experiment, we increased the Adstock by one percent in the first pre-release period. The output we measured was the percentage increase in the size of the daily volume sold in the post-release period.

As shown in Figure II-3, a ten percent increase in the Adstock results in a 3.3 percent increase in the number of blogs at the second pre-release period. As a result

[11] Note that our analysis is essentially correlational as it is hard to prove causality with our data. Our experiment must therefore be seen as an attempt to document the economic impact of firm actions, rather than a prescriptive guideline for managerial action. We thank an anonymous reviewer for bringing this to our notice.

of this increase in the Adstock, we find that the net increase in the sales volume is 2.1 percent. A decomposition of this overall increase due to traditional media versus new media suggested that the increase in the Adstock directly enhances the sales by 0.13 percent while the interaction between blogging and advertising increases the sales by 0.1 percent. Furthermore, the largest and most significant increase in the sales volume at post-launch is led by the indirect impact from advertising via blogging activity, which accounts for 1.9 percent. Similar experiments for the other two product markets also support these findings with the overall effect being slightly higher for cellular phone service (2.7%) and slightly smaller for movies (0.84%) as in Figures II-4 and II-5.

In addition to simulating the short-term effects of advertising, Figures II-6, II-7 and II-8 illustrate the evaluations of long-term effects of increases in Adstock. We use a simulation setting similar to the above experiments and expand the time horizon from one period to ten periods. The largest indirect effect of the ten percent increase in Adstock decays slower than do the other two effects across three product categories. The peaks of the indirect effects are located at the third period for the green tea drinks and at the second period for the cellular phone services. These are resulted from the larger estimates of the carry-over constants of Adstock and Blogstock (see details in Appendix II-C) at post-launch in the blog equations (Figures II-7 and II-8).

In the second experiment, we hold out the last observation from each brand and re-estimated the model. We then use the model estimates for prediction and computed the difference in the predicted value and the actual data across all the held out observations. We do this for the full model and a restricted version of the full model where the response coefficients for the number of blogs and the cumulative number of blogs were set to zero. Thus, the difference in prediction (based on the Root Mean Square Deviation) between these two models shows the extent to which the use of blog data can improve sales forecasts. As can be seen from Table II-25, the improvement in RMSD is very high for cellular phone service, high for movies and modest for green tea drinks.

II-5 Conclusion, Limitations and Directions for Future Research

This paper adds to the very limited, but rapidly growing field of research into the effectiveness of new media, especially in the case of new product launches. Using a unique dataset from three product markets from Japan (a major new media market), we are able to combine into a single source, data on market outcomes, traditional media (TV advertising) and new media (volume and content of blogs). Given the newness of this area, we take a more pragmatic and empirical approach to our analysis. While we motivate the use of our specification by leveraging theoretical and empirical findings from previous research, we hope that our efforts will lead to the development of a unifying theoretical framework for understanding user generated content in future research.

We used a simultaneous equation model to capture the effect of new media on market outcomes and the effect of market outcomes on new media. While this in itself is somewhat novel, we were also able to include the major marketing activity (TV advertising) in both equations, both directly and via interactions. In addition, we also include a set of cross-sectional and temporal fixed effect variables as controls and an exclusion restriction in the blog equation. Thus this allows us to investigate two open questions in this domain - (a) whether new media (blogging activity in our case) leads to (differential) market outcomes and (b) whether traditional marketing actions (*i.e.*, TV advertising) and new media act synergistically. We also make a first attempt, to the best of our knowledge, to use the content of the blog posts to shed “process” light on our econometric findings via a careful and methodical text mining analysis.

Using data from green tea drinks, movies and cellular phone service, we find that patterns across the three categories showing clear linkages between traditional media, new media and market outcomes. In general, we find that cumulative blogs (Blogstock) are predictive of market outcomes, blogs and TV advertising act synergistically, pre-launch advertising spurs blogging activity (that is predictive of marketing activity) but becomes less effective in inducing blogging activity post-launch and market outcomes also do have some effect on blogging activity. Our text mining results provide additional support for some of these findings. From a managerial point of view, in the experiment using green tea drink estimation results, we find that a one percent increase in the traditional marketing instrument (TV advertising) leads to a median increase in market outcomes of 0.2%, with a majority of the increase coming from the increase in blogging activity generated by the advertising pre-launch. In another experiment , we also show that using blog data can lead to better sales forecasts for all three product markets.

Although our analyses are descriptive and primarily leverage correlational patterns, our results (and the second experiment above) suggest that blogs can be a good predictor of market outcomes and managers would do well to consider including them in sales forecasting models. While this is a "passive" implication, our results also suggest a more "active" implication. Essentially, by understanding the specific relationship between traditional and new media, managers can allocate resources much better to traditional media as they can exploit the "multiplier" effect of traditional media on new media (as we show in the first experiment above). Another useful implication that emerges from our findings is that blogging activity can be a surrogate measure of advertising effectiveness – something that is typically hard to gauge directly using traditional methods. Finally, managers may also find that rigorous text mining analyses can shed light on consumer evaluation and adoption of new products.

Our analyses do also have a few limitations (driven mostly by the nature of the data). First, as noted earlier, the aggregate nature of our data makes it very hard to offer micro-level causal explanations of the effectiveness of new media and the synergistic relationship between new and traditional media. While our text mining analyses shed some light on our findings, it would be very beneficial to obtain datasets that link individual activity to market outcomes for a larger variety of new media. Second, our measures of new media are at present limited to blog content - volume – and in two of three product markets, keywords and valence. We do not have any knowledge of the identity and demographics of the bloggers and the extent of linkage to and from a given blog (*e.g.*, trackbacks). Third, our model could be improved with the potential use of non-parametric models to model the effects of both old and new media and the associated interactions. Finally, our data do not contain information on all marketing instruments and hence we use proxies (such as lagged sales in the case of distribution). We hope that with better data, future research will be able to address these limitations.

Tables and Figures

Table II-1 List of Past Studies and Comparisons

Paper	Product category	Online WOM	New & old media interaction	Pre vs. Post release	WOM measures	Outcome measures	Data type
This paper	Green tea drinks, Movies, Cell phone subscriptions	Blog	Yes (TV GRPs)	Yes	Volume, Valence, (Text mining)	Sales volume, # audience, # customers	Aggregate (daily/monthly panel)
Dhar and Chang (2009)	Music	Online review, Blog, SNS intensity	No	No	Volume, Rating, Intensity	Sales rank	Aggregate (weekly panel)
Gruhl et al. (2005)	Books	Blog	No	No	Volume	Sales rank	Aggregate (daily panel)
Chintagunta et al. (2010)	Movies	Online rating	No	No	Valence	Opening-day revenue	Aggregate (cross-sectional)
Tirunillai and Tellis (2010)	Multiple	Product reviews	No	No	Volume, Valence	Stock Prices	Aggregate (daily panel)
Goel et al. (2010)	Movies, Videogames, Music Flu	Web search	No	Yes	Volume	Revenue Flu-caseloads	Aggregated (cross-sectional)
Trusov et al. (2009)	SNS signups	Referral	No (but proxy measure for PR - media appearances)	No	Volume	# signup customers	Aggregate (weekly panel)
Chakravarty et al. (2008)	Movies	Online review	No	Yes	Volume	7-point scale evaluation	Lab experiment
Duan et al. (2008)	Movies	Online review	No	No	Volume, Valence	Box-office revenue	Aggregate (daily panel)

Table II-1 (Cont.) List of Past Studies and Comparisons

Paper	Product category	Online WOM	New & old media interaction	Pre vs. Post release	WOM measures	Outcome measures	Data type
Dellarocas et al. (2007)	Movies	Online rating	No	No	Volume, Valence, Variance	Box-office revenue	Aggregate (weekly panel)
Moul (2007)	Movies	Latent quality learning	No	No	Latent quality	Box-office revenue	Aggregate (panel weekly)
Liu (2006)	Movies	Online review	No	Yes	Volume, Valence	Box-office revenue	Aggregate (weekly panel)
Chevalier and Mayzlin (2006)	Books	Online rating	No	No	Volume, Valence	Sales rank	Aggregate (weekly panel)
Godes and Mayzlin (2004)	TV shows	Online review	No	No	Volume, Valence, Variance	Household rating	Aggregate panel (episode)

Table II-2 Brand Market Share (Green Tea Drinks)

Brand	Volume share
Karada Meguri Cha	1.99%
Koi Nama Cha	1.35%
Iemon Koime	1.28%
Seiryu Nana Cha	0.73%
Kiki Cha	0.26%

Table II-3 Movie List

Title	Description
Spider-man 3	US
300	US
Apocalypto	US
Die Hard 4.0	US
The Prestige	US & British
Volver	European
The Magic Flute	European
Confession of Pain	Hong Kong
Maiko Haaaan!!!	Japanese
Saiyuki	Japanese
Dai Nipponjin	Japanese
Shrek The Third	Animation

Table II-4 Number of Subscribers (Cellular Phone Service)

Brand	Subscribers (Apr 2008)
NTT DoCoMo	53 mm
SoftBank Mobile	19 mm
Au	30 mm
Willcom	4.6 mm
EMOBILE	0.5 mm

Table II-5 Market Outcomes

Product category	Description	Mean	SD
Green Tea Drinks	Purchase volume per day	5305.78	12960.64
Movie	Number of audiences per day	1,298.95	4,716.47
Cellular Phone Service	Number of subscribers per month	110,145.45	120,691.66

Table II-6 Traditional Media (TV GRPs)

Product category	Period	Mean	SD
Green Tea Drinks	All periods	14.59	44.78
	Pre-release	0.58	9.60
	Post-release	59.30	74.14
Movie	All periods	3,047.60	9,621.23
	Pre-release	3,471.31	10,553.11
	Post-release	1,992.98	6,657.42
Cellular Phone Service	All periods	5,804.03	4,288.31
	Pre-release	5,823.75	4,623.10
	Post-release	5,791.48	4,105.04

Table II-7 New media (Number of blogs)

Product category	Period	Mean	SD
Green Tea Drinks	All periods	2.86	20.54
	Pre-release	0.17	1.11
	Post-release	11.45	40.89
Movie	All periods	99.43	184.93
	Pre-release	43.78	77.35
	Post-release	237.95	278.61
Cellular Phone Service	All periods	59,468.79	61,409.34
	Pre-release	24,332.89	19,451.93
	Post-release	81,827.99	68,320.97

Table II-8 Blog Valence Percentage (Movies and Cellular Phone Service)

Valence of blogs	Product category	Period	Mean	SD
Positive blogs	Movie	All periods	37.9%	40.8%
		Pre-release	32.4%	44.1%
		Post-release	50.6%	28.3%
	Cellular Phone Service	All periods	62.4%	15.1%
		Pre-release	58.9%	20.7%
		Post-release	64.6%	9.8%
Neutral blogs	Movie	All periods	14.1%	23.5%
		Pre-release	8.0%	20.9%
		Post-release	28.3%	23.0%
	Cellular Phone Service	All periods	18.4%	10.1%
		Pre-release	16.4%	9.8%
		Post-release	19.6%	10.3%
Negative blogs	Movie	All periods	6.3%	16.7%
		Pre-release	4.3%	15.9%
		Post-release	10.9%	17.5%
	Cellular Phone Service	All periods	15.9%	5.4%
		Pre-release	16.1%	7.4%
		Post-release	15.8%	3.6%

Table II-9 Correlation Matrix (Green Tea Drinks)

	Mean	s.d.	Cust omer	Blog	Blog stock	Sale stock	Ad-stock	Bst* Adst
#Customer	1265.5	6715.7	1.00					
#Blog	2.9	20.5	0.13	1.00				
Blogstock	1.9	11.1	0.15	0.49	1.00			
Salestock	10026.7	23400.8	0.29	0.13	0.20	1.00		
Adstock	9.7	28.1	0.24	0.35	0.42	0.54	1.00	
Blogstock*Adstock	148.7	1578.3	0.11	0.46	0.77	0.10	0.35	1.00

Table II-10 Correlation Matrix (Movies)

	Mean	s.d.	Audi ence	Blog	Posi tive	Nega tive	View ed	Awa rd	Inter est	Ad	Blog stock	Sale stock	Ad-stock	Bst* Adst
#Audience	1336.1	4203.1	1.00											
#Blog	108.3	195.7	0.73	1.00										
#Blog Positive	54.5	119.1	0.69	0.54	1.00									
#Blog Negative	11.5	55.7	0.28	0.47	0.25	1.00								
#Blog Viewed	33.3	63.5	0.77	0.72	0.50	0.42	1.00							
#Blog Award	11.3	30.3	0.59	0.59	0.48	0.24	0.56	1.00						
#Blog Interesting	13.1	27.6	0.65	0.59	0.66	0.44	0.65	0.35	1.00					
#Blog Advertising	7.7	27.5	0.52	0.40	0.37	0.15	0.41	0.53	0.17	1.00				
Blogstock	984.7	1573.7	0.59	0.79	0.60	0.39	0.65	0.43	0.73	0.29	1.00			
Salestock	1662.5	5055.2	0.81	0.68	0.67	0.22	0.72	0.57	0.61	0.50	0.65	1.00		
Adstock	8269.0	20424.7	0.36	0.42	0.41	0.17	0.41	0.25	0.36	0.21	0.19	0.27	1.00	
Blogstock*Adstock	14375090.0	56181540.0	0.68	0.71	0.67	0.28	0.63	0.30	0.61	0.16	0.54	0.67	0.65	1.00

Table II-11 Correlation Matrix (Cellular Phone Service)

	Mean	s.d.	Cust omer	Blog	Posi tive	Nega tive	Sub scr	Price	Inter est	Ad	Blog stock	Sale stock	Ad-stock	Bst* Adst
#Customer	121987.8	144242.7	1.00											
#Blog	59359.3	61246.7	0.23	1.00										
#Blog Positive	39979.7	42524.2	0.24	0.80	1.00									
#Blog Negative	9592.5	9434.4	0.19	0.78	0.86	1.00								
#Blog Subscribed	2106.3	1516.8	0.47	0.74	0.73	0.72	1.00							
#Blog Price	2744.1	1840.2	0.33	0.80	0.79	0.78	0.95	1.00						
#Blog Interesting	318.6	188.1	0.40	0.72	0.70	0.71	0.95	0.93	1.00					
#Blog Advertising	1856.5	1378.4	0.42	0.75	0.74	0.73	0.97	0.95	0.91	1.00				
Blogstock	6005.4	6669.0	0.23	0.53	0.44	0.59	0.68	0.73	0.63	0.72	1.00			
Salestock	348779.8	391010.9	0.61	0.44	0.44	0.41	0.59	0.54	0.50	0.62	0.43	1.00		
Adstock	26984.4	23352.4	0.37	0.75	0.85	0.83	0.75	0.76	0.68	0.79	0.35	0.32	1.00	
Blogstock*Adstock	293270700.0	411715700.0	0.19	0.51	0.62	0.86	0.67	0.70	0.60	0.70	0.78	0.42	0.86	1.00

Table II-12 Sales Volume Equation (Green Tea Drinks)

	Estimate	SE	t value	Pr(>|t|)	
Intercept	-1.71	4.20	-0.41	0.683	
Seasonal					
Tue	0.41	0.96	0.43	0.667	
Wed	0.63	0.79	0.79	0.431	
Thu	0.75	0.75	1.01	0.314	
Fri	0.06	0.72	0.08	0.934	
Sat	0.65	0.73	0.90	0.371	
Sun	-2.34	0.72	-3.26	0.001	**
Holiday	-2.55	2.42	-1.06	0.292	
Trend	-0.03	0.01	-2.28	0.023	*
Brand					
Brand1	6.87	1.06	6.47	0.000	***
Brand2	8.52	0.77	11.02	0.000	***
Brand3	8.23	0.83	9.97	0.000	***
Brand4	6.10	0.65	9.34	0.000	***
#Consumer(-1)	-0.09	0.04	-2.12	0.035	*
#Blog	0.01	0.05	0.31	0.760	
Blogstock	0.56	0.18	3.05	0.002	**
Salestock	0.25	0.03	9.46	0.000	***
Adstock	0.08	0.04	2.16	0.031	*
Blogstock *Adstock	0.04	0.02	1.81	0.071	.

Notes: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-13 Sales Volume (Audience) Equation (Movies)

	Estimate	Std. Error	t value	Pr(>|t|)	
Intercept	-1.06	0.61	-1.74	0.083	.
Seasonal					
Tue	0.40	0.07	5.73	0.000	***
Wed	0.93	0.05	16.85	0.000	***
Thu	0.38	0.05	7.29	0.000	***
Fri	0.49	0.05	10.58	0.000	***
Sat	1.07	0.05	22.78	0.000	***
Sun	0.98	0.05	20.33	0.000	***
Holiday	0.92	0.10	9.61	0.000	***
Trend	-0.01	0.00	-4.05	0.000	***
Brand					
Movie1	0.26	0.10	2.57	0.010	*
Movie2	0.39	0.10	3.88	0.000	***
Movie3	0.42	0.07	6.06	0.000	***
Movie4	-0.07	0.08	-0.80	0.425	
Movie5	0.45	0.11	4.18	0.000	***
Movie6	0.00	0.07	0.04	0.965	
Movie7	0.14	0.06	2.25	0.025	*
Movie8	0.28	0.06	4.52	0.000	***
Movie9	-0.26	0.10	-2.58	0.010	*
Movie10	0.31	0.05	6.29	0.000	***
Movie11	-0.22	0.07	-3.11	0.002	**
#Audience(-1)	0.28	0.11	2.63	0.009	**
#Blog	-0.03	0.03	-1.00	0.318	
#Blog Positive	0.01	0.00	4.28	0.000	***
#Blog Negative	0.00	0.00	0.47	0.635	
#Blog Viewed	0.30	0.04	7.57	0.000	***
#Blog Award	-0.02	0.01	-2.20	0.028	*
#Blog Interesting	-0.01	0.01	-1.39	0.166	
#Blog Advertising	-0.01	0.00	-1.50	0.134	
Blogstock	0.40	0.02	21.02	0.000	***
Salestock	0.26	0.00	74.03	0.000	***
Adstock	0.08	0.02	4.50	0.000	***
Blogstock *Adstock	0.01	0.00	44.21	0.000	***

Notes: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-14 Sales Volume (Subscribers) Equation (Cellular Phone Service)

	Estimate	Std. Error	t value	Pr(>|t|)	
Intercept	3.79	5.17	0.73	0.469	
Seasonal					
March	0.32	0.08	4.08	0.000	***
April	0.11	0.10	1.15	0.258	
Trend	0.00	0.01	0.23	0.820	
Brand					
Brand1	0.71	0.24	2.93	0.007	**
Brand2	0.14	0.10	1.42	0.165	
Brand3	0.32	0.11	2.87	0.008	**
Brand4	0.99	0.15	6.50	0.000	***
#Customer(-1)	-0.34	0.11	-3.00	0.006	**
#Blog	-0.50	0.52	-0.96	0.343	
#Blog Positive	0.45	0.09	5.21	0.000	***
#Blog Negative	-0.04	0.05	-0.86	0.394	
#Blog Subscribed	1.04	0.18	5.68	0.000	***
#Blog Price	-0.62	0.11	-5.45	0.000	***
#Blog Interesting	0.02	0.09	0.20	0.844	
#Blog Advertising	0.25	0.04	6.03	0.000	***
Blogstock	1.22	0.11	11.44	0.000	***
Salestock	0.20	0.02	9.81	0.000	***
Adstock	-0.12	0.01	-10.31	0.000	***
Blogstock *Adstock	0.11	0.00	629.82	0.000	***

Notes: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-15 Blog Equation (Green Tea Drinks)

	Estimate	SE	t value	Pr(>|t|)	
Intercept	-4.37	1.98	-2.21	0.028	*
Seasonal					
Tue	-0.70	0.88	-0.80	0.423	
Wed	1.40	0.76	1.85	0.065	.
Thu	-0.53	0.72	-0.73	0.464	
Fri	0.09	0.68	0.14	0.890	
Sat	-0.50	0.69	-0.72	0.472	
Sun	0.28	0.68	0.41	0.681	
Holiday	2.12	1.85	1.15	0.251	
Trend	-0.02	0.01	-2.55	0.011	*
Brand					
Brand1	3.02	0.90	3.36	0.001	**
Brand2	1.04	0.80	1.29	0.197	
Brand3	0.90	0.72	1.24	0.214	
Brand4	-2.24	0.64	-3.53	0.000	**
Pre					
#Blog(-1)	0.10	0.14	0.77	0.442	
Blogstock	0.03	0.99	0.03	0.977	
Adstock	0.60	0.08	7.10	0.000	***
Blogstock *Adstock	-0.03	0.01	-6.06	0.000	***
Post					
#Blog(-1)	-0.07	0.06	-1.31	0.189	
#Consumer	0.01	0.04	0.22	0.825	
Blogstock	-0.45	0.51	-0.88	0.378	
Salestock	0.36	0.10	3.41	0.001	***
Adstock	-0.26	0.11	-2.33	0.020	*
Blogstock *Adstock	0.27	0.02	13.59	0.000	***
Other products' blog	0.14	0.06	2.17	0.031	***

Notes: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-16 Blog Equation (Movies)

	Estimate	Std. Error	t value	Pr(>|t|)	
Intercept	-7.69	4.94	-1.56	0.120	
Seasonal: Tue	-0.01	0.23	-0.05	0.961	
Wed	-0.03	0.20	-0.17	0.864	
Thu	0.09	0.19	0.46	0.648	
Fri	0.12	0.18	0.66	0.508	
Sat	-0.09	0.18	-0.53	0.597	
Sun	-0.23	0.18	-1.27	0.204	
Holiday	0.50	0.34	1.50	0.134	
Trend	0.03	0.00	9.49	0.000	***
Brand					
Movie1	-0.44	0.36	-1.23	0.217	
Movie2	-1.09	0.30	-3.64	0.000	**
Movie3	-4.82	0.25	-19.13	0.000	***
Movie4	1.12	0.27	4.11	0.000	***
Movie5	0.10	0.31	0.30	0.761	
Movie6	0.10	0.25	0.39	0.698	
Movie7	-0.19	0.24	-0.77	0.441	
Movie8	-0.03	0.25	-0.10	0.918	
Movie9	0.60	0.27	2.22	0.027	*
Movie10	0.22	0.22	0.98	0.325	
Movie11	-1.66	0.20	-8.11	0.000	***
Pre					
#Blog(-1)	0.20	0.03	6.78	0.000	***
#Blog Positive	0.00	0.01	0.35	0.728	
#Blog Negative	0.00	0.02	0.19	0.851	
#Blog Award	0.04	0.01	2.63	0.009	**
#Blog Interesting	0.08	0.01	5.58	0.000	***
#Blog Advertising	0.03	0.01	1.85	0.064	.
Blogstock	0.08	0.04	1.70	0.089	.
Adstock	0.12	0.01	12.23	0.000	***
Blogstock *Adstock	-0.03	0.00	-13.89	0.000	***
Post					
#Blog(-1)	-0.34	0.93	-0.37	0.710	
#Blog Positive	-0.04	0.02	-1.78	0.076	.
#Blog Negative	0.00	0.02	0.18	0.859	
#Blog Viewed	0.66	0.31	2.17	0.030	*
#Blog Award	-0.03	0.08	-0.34	0.733	
#Blog Interesting	-0.05	0.06	-0.87	0.383	
#Blog Advertising	-0.02	0.07	-0.20	0.839	
#Audience	0.83	0.16	5.31	0.000	***
Blogstock	0.07	0.18	0.36	0.717	
Salestock	-0.77	0.02	-38.28	0.000	***
Adstock	-0.05	0.10	-0.49	0.621	
Blogstock *Adstock	0.02	0.00	7.18	0.000	***
Other products' blog	0.90	0.01	157.45	0.000	***

Notes: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-17 Blog Equation (Cellular Phone Service)

	Estimate	Std. Error	t value	Pr(>|t|)	
Intercept	-19.61	7.68	-2.55	0.013	*
Seasonal					
March	-0.04	0.12	-0.30	0.761	
April	-0.04	0.13	-0.33	0.740	
Trend	0.02	0.02	0.79	0.434	
Brand					
Brand1	1.82	0.34	5.30	0.000	***
Brand2	2.14	0.13	16.10	0.000	***
Brand3	1.95	0.13	15.29	0.000	***
Brand4	-0.65	0.27	-2.39	0.020	*
Pre					
#Blog(-1)	0.09	0.19	0.45	0.657	
#Blog Positive	0.20	0.03	6.01	0.000	***
#Blog Negative	-0.07	0.07	-0.95	0.344	
#Blog Price	0.44	0.22	2.02	0.048	*
#Blog Interesting	0.15	0.20	0.73	0.466	
#Blog Advertising	0.48	0.11	4.40	0.000	***
Blogstock	0.24	0.08	2.95	0.004	**
Adstock	0.20	0.03	6.56	0.000	***
Blogstock *Adstock	-0.03	0.00	-27.41	0.000	***
Post					
#Blog(-1)	2.80	1.26	2.22	0.030	*
#Blog Positive	1.76	0.32	5.53	0.000	***
#Blog Negative	-0.58	0.17	-3.51	0.001	**
#Blog Subscribed	0.71	0.42	1.69	0.096	.
#Blog Price	0.54	0.28	1.94	0.057	.
#Blog Interesting	0.12	0.20	0.63	0.530	
#Blog Advertising	-0.06	0.11	-0.59	0.556	
#Customer	0.10	0.18	0.57	0.573	
Blogstock	0.45	0.20	2.32	0.023	*
Salestock	-0.09	0.06	-1.46	0.148	
Adstock	-0.04	0.03	-1.25	0.216	
Blogstock *Adstock	0.02	0.00	31.86	0.000	***
Other products' blog	1.84	0.00	891.40	0.000	***

Notes: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-18 Model Fit

			Model 1 (Fixed effects & trend)	Model 2 (Model 1 & lagged DVs)	Full Model
Green Tea Drinks	1st equation (Sales Volume)	AIC	14159.38	2914.88	2914.15
		R-square	0.1575	0.6648	0.9788
		Adj. R-square	0.1524	0.6567	0.9772
	2nd equation (Blog)	AIC	3106.06	3070.10	2987.78
		R-square	0.2794	0.6057	0.6766
		Adj. R-square	0.2612	0.5949	0.6673
Movie	1st equation (Sales Volume)	AIC	49.28	-1078.11	-1577.18
		R-square	0.8572	0.9917	0.9963
		Adj. R-square	0.8514	0.9813	0.9961
	2nd equation (Blog)	AIC	9985.75	7895.21	7799.26
		R-square	0.5345	0.6399	0.6637
		Adj. R-square	0.5294	0.6356	0.6582
Cellular Phone Service	1st equation (Sales Volume)	AIC	391.38	390.78	-133.08
		R-square	0.5990	0.6054	0.9992
		Adj. R-square	0.5321	0.5284	0.9989
	2nd equation (Blog)	AIC	149.48	88.15	72.81
		R-square	0.8898	0.9867	0.9889
		Adj. R-square	0.8797	0.9853	0.9867

Table II-19 Holdout (last observation) test AIC

		Model 1 (Fixed effects & time trend)	Model 2 (Model 1 & lag of DVs)	Model 3 (All variables)
Green Tea Drinks	1st equation	74.66	63.05	60.05
	2nd equation	75.53	63.91	60.62
	Overall	150.19	126.96	120.68
Movie	1st equation	144.91	138.04	89.60
	2nd equation	149.48	106.35	98.11
	Overall	294.39	244.39	187.71
Cellular Phone Service	1st equation	171.77	150.63	116.52
	2nd equation	84.43	73.82	59.71
	Overall	256.20	224.45	177.23

Table II-20 Holdout (chosen at random) test AIC

		Model 1 (Fixed effects & time trend)	Model 2 (Model 1 & lag of DVs)	Model 3 (All variables)
Green Tea Drinks	1st equation	1266.63	978.61	984.04
	2nd equation	2065.22	1089.36	922.21
	Overall	3331.85	2067.97	1906.25
Movie	1st equation	45.52	-17.23	-62.55
	2nd equation	553.49	554.71	526.74
	Overall	599.01	537.48	464.19
Cellular Phone Service	1st equation	223.71	116.45	115.91
	2nd equation	55.80	53.39	43.94
	Overall	279.51	169.84	159.85

Table II-21 Word Counts (Movies)

Word	Rank	Frequency	Lexical Category
Movie	1	19362	noun
Dainihonjin	2	10284	proper noun
View	**3**	**16150**	**verb**
Matsumoto Hitoshi	4	9259	proper noun
Think	5	8508	verb
Director	6	8402	noun
Go	7	4701	verb
Say	8	4631	verb
Interesting	**9**	**4372**	**adjective**
Film	10	4281	noun
Award	**11**	**3228**	**noun**
Release	12	3152	verb
Warai	13	2723	proper noun
Feeling	14	2677	noun
Theater	15	2509	noun
Out	16	2380	verb
Story	17	2352	noun
Japan	18	2287	noun
Fun	19	2174	noun
Love	20	2002	verb
Self	21	1900	noun
Interview	22	1891	noun
News	23	1882	noun
Content	24	1825	noun
TV	25	1811	noun
Matsu	26	1782	proper noun
Down Town	27	1733	proper noun
Laugh	28	1718	verb
Kitano Takeshi	29	1603	proper noun
Write	30	1586	verb
Advertising	**80**	**686**	**noun**

Table II-22 Word Counts (Cellular Phone Service)

Word	Rank	Frequency	Lexical Category
Cellphone	1	546400	noun
DoCoMo	2	264091	proper noun
Think	3	248589	verb
Phone	4	188871	noun
SoftBank	5	154878	proper noun
Model	6	125798	noun
au	7	112657	proper noun
Use	8	109641	verb
See	9	106815	verb
Say	10	103783	verb
Email	11	98065	noun
Service	12	87017	noun
Information	13	84075	noun
Laugh	14	82486	verb
Willcom	15	79317	proper noun
Love	16	78906	verb
Self	17	77589	noun
Handset	18	77158	noun
Release	19	72199	verb
Go	20	71589	verb
Sell	21	70151	verb
Buy	22	69973	verb
Compatible	23	67030	adjective
Free	24	64081	noun
Function	25	63338	noun
Read	26	63280	verb
Website	27	62801	noun
Plan	28	61587	noun
Series	29	59334	noun
Price	**30**	**58573**	**noun**
Subscribe	**53**	**44116**	**verb**
Advertising	**108**	**29666**	**noun**
Interesting	**208**	**15560**	**adjective**

Table II-23 Word Percentage (Movies)

Word	Period	Mean	SD
Viewed	All periods	54.2%	27.1%
	Pre-release	43.3%	30.7%
	Post-release	64.8%	10.3%
Award	All periods	21.9%	28.4%
	Pre-release	22.8%	30.8%
	Post-release	19.9%	21.9%
Interesting	All periods	20.3%	19.3%
	Pre-release	17.6%	21.2%
	Post-release	26.4%	11.8%
Advertising	All periods	14.0%	26.5%
	Pre-release	15.0%	27.8%
	Post-release	11.7%	23.0%

Table II-24 Word Percentage (Cellular Phone Service)

Word	Period	Mean	SD
Subscribed	All periods	30.6%	9.4%
	Pre-release	27.6%	5.7%
	Post-release	42.0%	11.6%
Price	All periods	42.7%	13.6%
	Pre-release	40.8%	11.4%
	Post-release	49.4%	18.6%
Interesting	All periods	10.4%	3.0%
	Pre-release	10.6%	2.5%
	Post-release	9.7%	4.3%
Advertising	All periods	23.5%	6.4%
	Pre-release	24.7%	6.0%
	Post-release	19.4%	6.3%

Table II-25 Root Mean Square Deviation Comparison

	Model without Blog information	Model with Full information
Green Tea Drinks	7.91	7.73
Movie	8.42	2.67
Cellular Phone Service	94.18	9.04

Figure II-1 An Japanese Blog focused on “Spider-man 3” (May 1, 2007)

Figure II-2 "Spider-man 3" Pre- and Post-Release Patterns (2007)

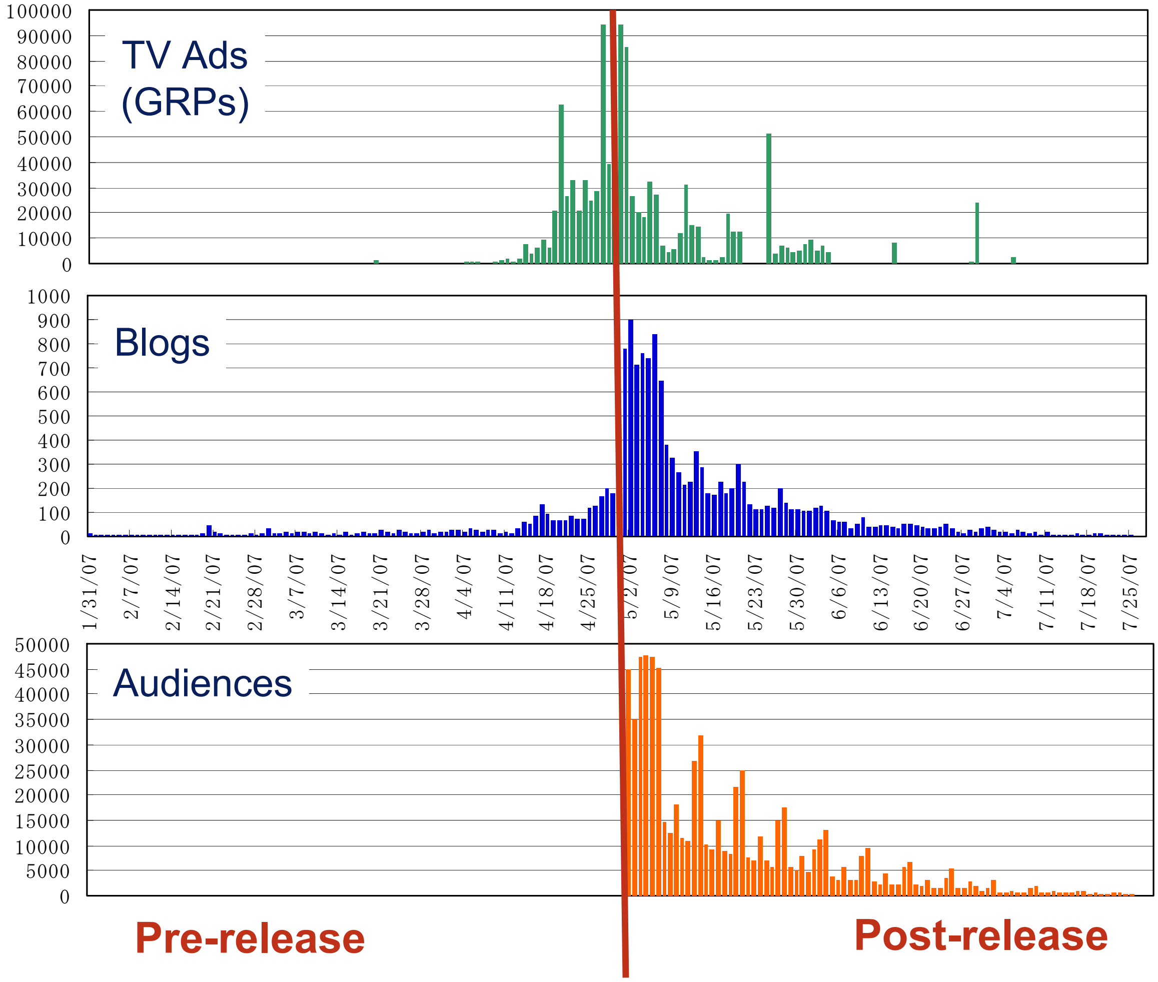


Figure II-3 Effect of 10% Increase in Adstock (Green Tea Drinks)

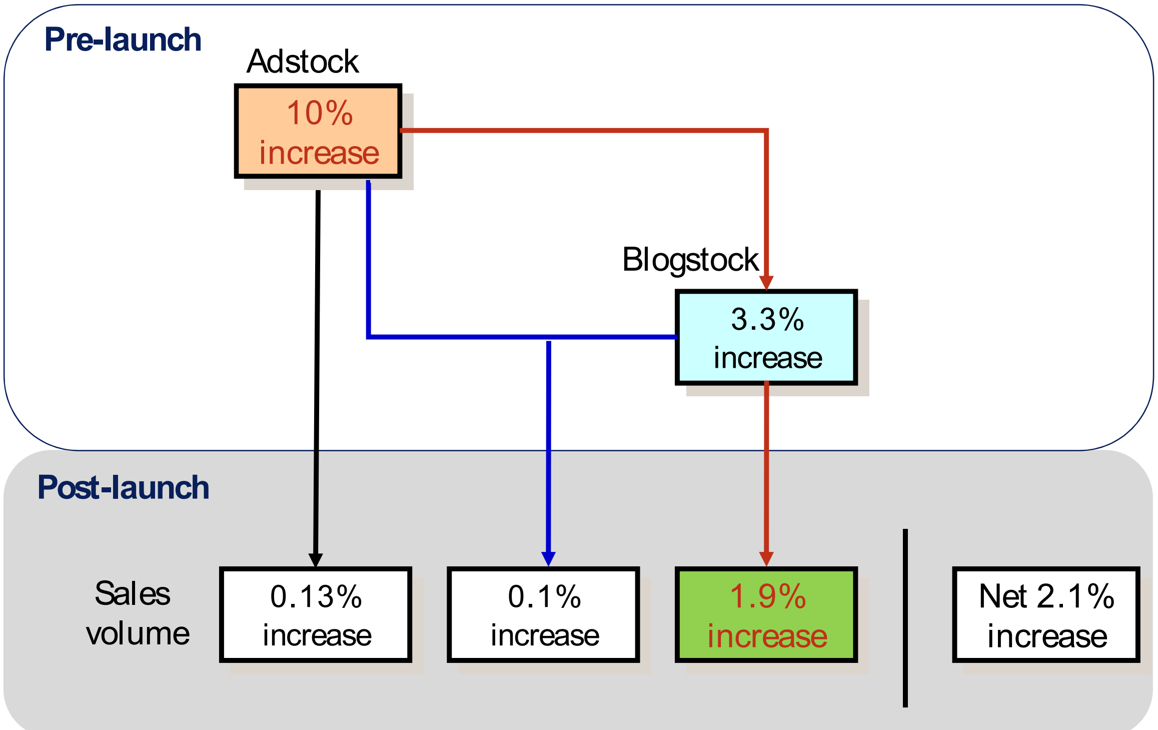


Figure II-4 Effect of 10% Increase in Adstock (Movies)

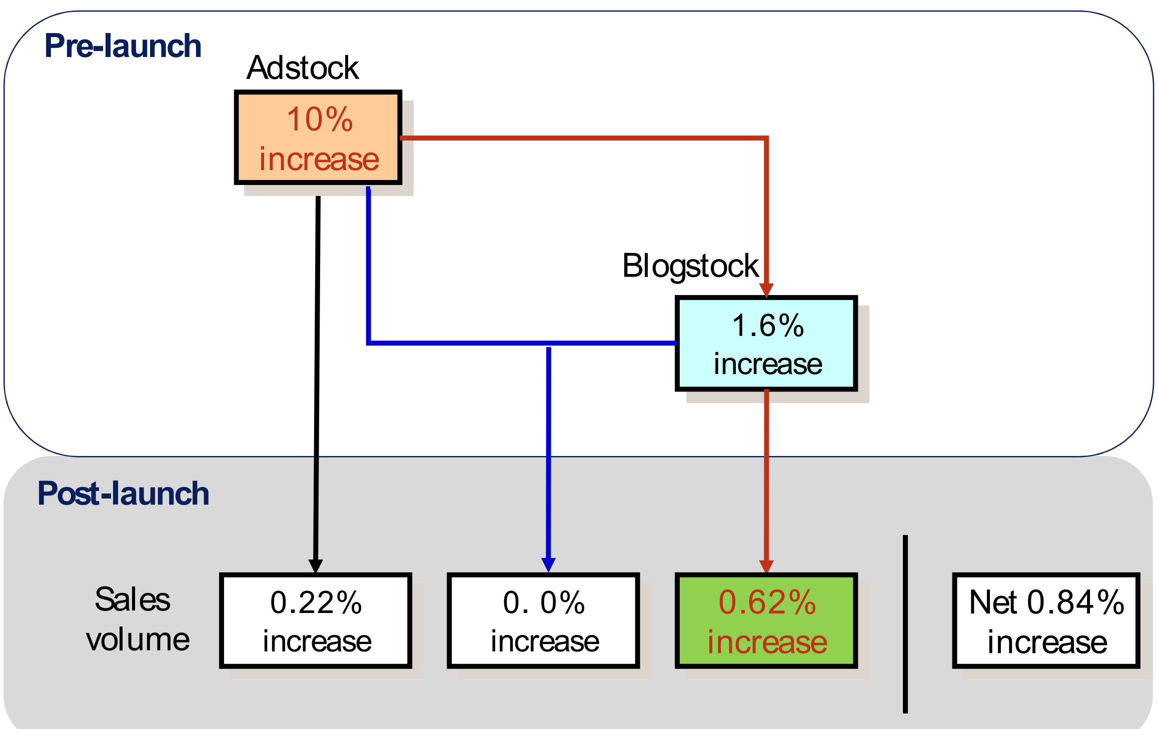


Figure II-5 Effect of 10% Increase in Adstock (Cellular Phone Service)

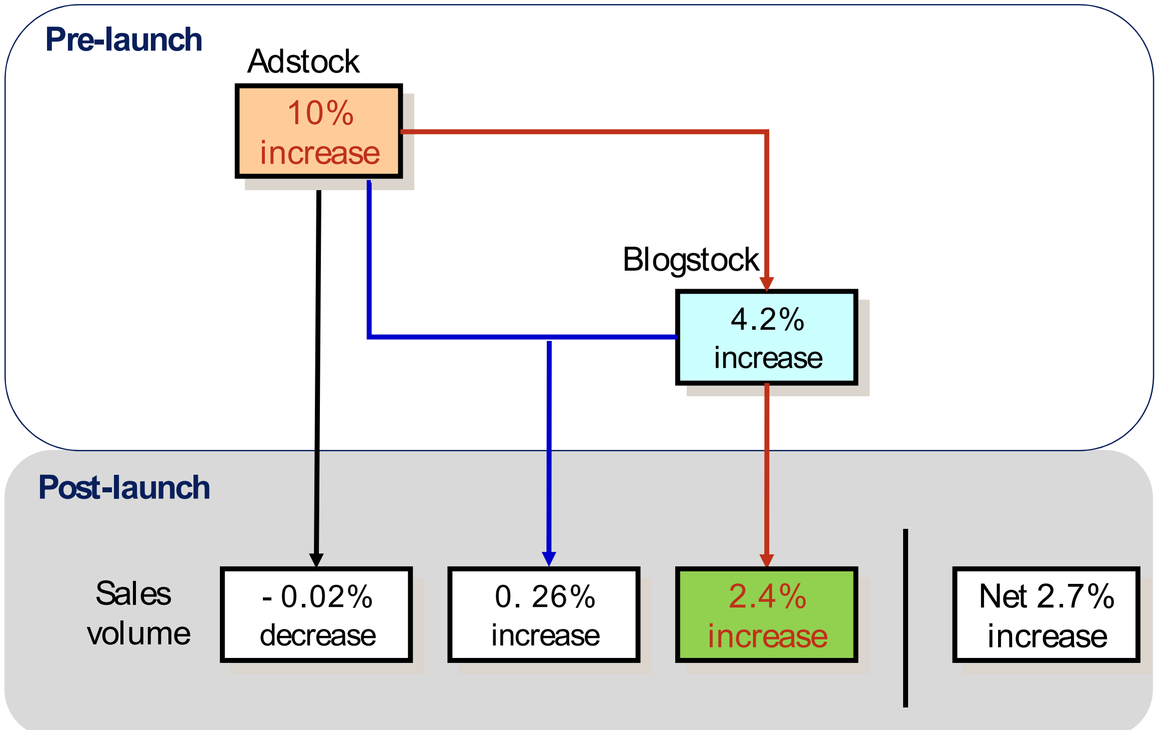


Figure II-6 Long-term Effects of 10% Increase in Adstock (Green Tea Drinks)

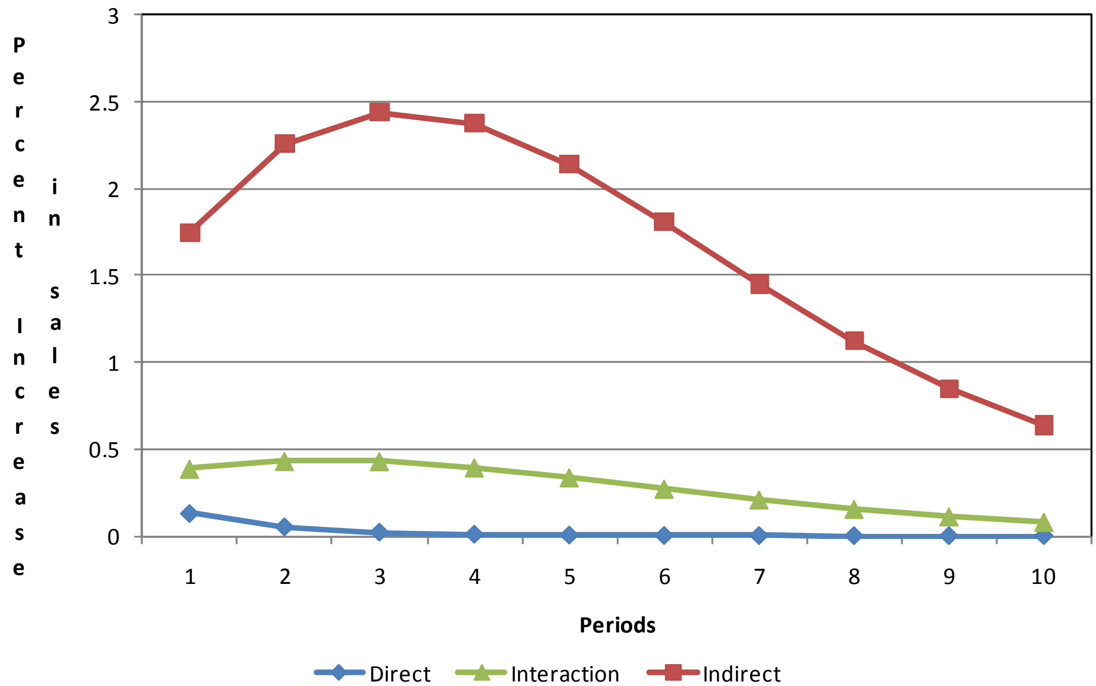


Figure II-7 Long-term Effects of 10% Increase in Adstock (Movies)

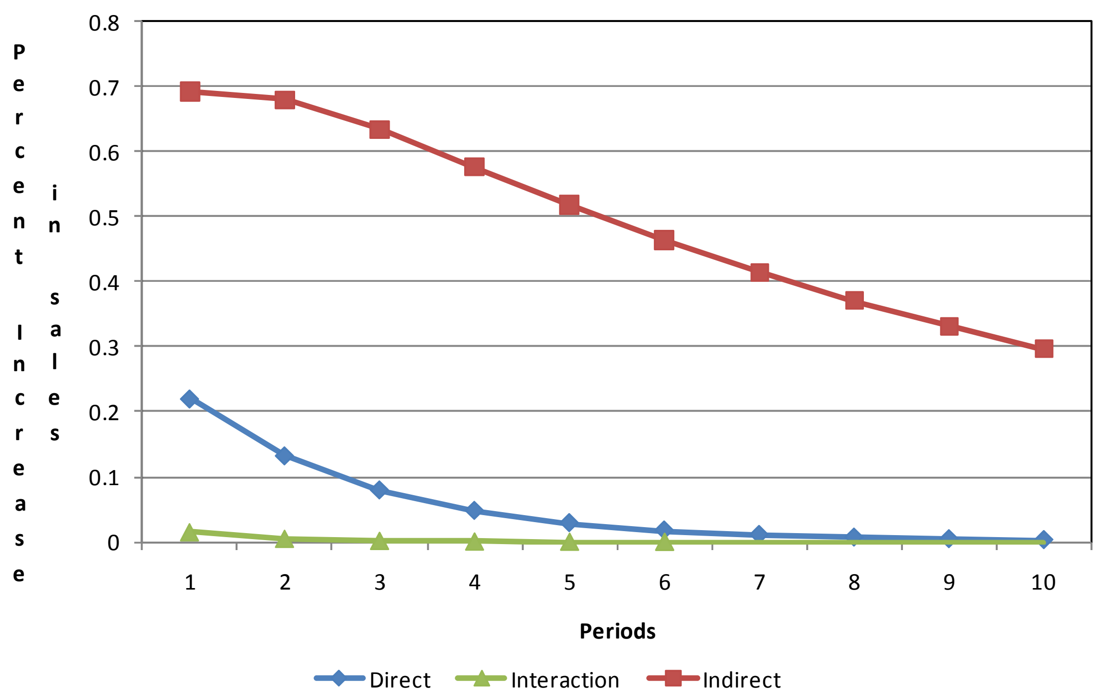
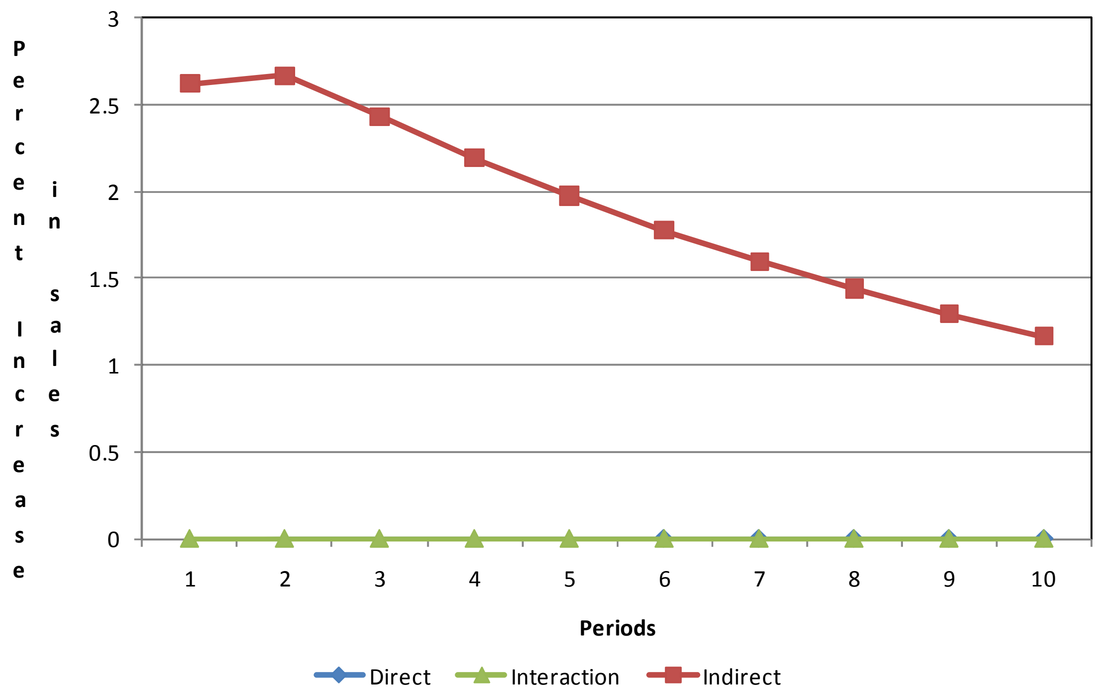


Figure II-8 Long-term Effects of 10% Increase in Adstock (Cellular Phone Service)



Appendix

Appendix II-A Testing data series stationarity

We conducted a Augmented Dickey-Fuller test for checking stationarity of variables and a Durbin-Watson test for autocorrelation of disturbances in equation by equation regressions.

1. We found that all the variables (sales, blog and TV GRPs) were stationary.

2. Results from DW tests for equation by equation regressions showed that disturbances of baseline models were not stationary in movie and cellular phone service. But these do not cause problems, since FGLS accommodates serial correlation by estimating weighted variance matrix (see Wooldridge 2001, Ch. 10.5.4 Serial Correlation and the Robust Variance Matrix Estimator, p. 274-276).

Detailed results are reported in the following tables.

Table II-26 Results of Augmented Dickey-Fuller (ADF) test for a unit root (stationarity of data series)

	ADF	p-value
Green Tea Drinks		
#Customer	-4.2171	0.000
#Blog	-4.2250	0.000
TV GRPs	-5.0784	0.000
Movie		
#Customer	-5.8757	0.000
#Blog	-5.0609	0.000
TV GRPs	-6.1035	0.000
Cellular Phone Service		
#Customer	-4.6193	0.000
#Blog	-4.9432	0.000
TV GRPs	-5.1807	0.000

Table II-27 Results of Durbin-Watson test for autocorrelation of disturbances

	Sales equation		Blog equation	
	DW	p-value	DW	p-value
Green Tea Drinks				
Full model	1.9440	0.0916	2.0053	0.4548
Baseline model	0.5145	0.0000	1.0386	0.0000
Movie				
Full model	1.8787	0.0017	2.0758	0.7604
Full model with text-mining	1.8867	0.0027	2.0495	0.7604
Baseline model	0.1304	0.0000	1.1776	0.7604
Cellular Phone Service				
Full model	2.0382	0.2814	1.6289	0.0016
Full model with text-mining	2.0634	0.2860	1.8539	0.0189
Baseline model	1.9930	0.3064	0.4906	0.0000

II-6Appendix II-B Testing Granger causality between dependent variables

As we noted in the text, all our results were based on correlational relationship because we employed a system of regression models. Thus we cannot say that the results were based on causal relationship. However, the model included sets of controls (seasonality, trends, fixed effects and exclusion restrictions) and captured correlated errors between the two equations. Those parameters should take into account the unobserved variables.

Furthermore, we also conducted Granger causality test for main three variables (# customers, # blogs and TV GRPs). We found mutual Granger causality among every two sets of the dependent variables – sales, blogs and TV GRPs – except for one case (from blogs to TV GRPs).

Table II-28 Results of Granger causality test

	→		←	
	F -value	p-value	F-value	p-value
Green Tea Drinks				
#Customer vs. #Blog	213.46	0.0000	558.94	0.0000
#Customer vs. TV GRPs	207.36	0.0000	79.644	0.0000
#Blog vs. TV GRPs	337.67	0.0000	35.817	0.0000
Movie				
#Customer vs. #Blog	8.5027	0.0036	29.850	0.0000
#Customer vs. TV GRPs	10.5120	0.0000	83.160	0.0000
#Blog vs. TV GRPs	30.882	0.0000	18.809	0.0000
Cellular Phone Service				
#Customer vs. #Blog	8.2394	0.0049	4.9230	0.0287
#Customer vs. TV GRPs	6.3666	0.0131	4.5891	0.0345
#Blog vs. TV GRPs	1.0729	0.3027	2.7105	0.0346

Appendix II-C Results of parameter estimates of carry-over constants

As stated in the text, the estimated value of the carryover constant is estimated via a grid search with best fit as the objective function. The estimated gammas used in the model are given in the table below. It is difficult to find any particular patterns across the three categories.

However, as noted in the text, the estimated values of the carry-over constants of Adstock and Blogstock for the green tea drinks and the cellular phone services are relatively lager than those for the movies at the post-launch in the blog equations. This fact leads the difference in managerial implications in simulating the long-term effects by increased Adstock in the sales volume, resulting in the peaks of the indirect effects of advertisings which are located at the third period for the green tea drinks and at the second period for the cellular phone services. But these effects for the movies decay constantly.

Table II-29 Results of parameter estimates of carry-over constants

		1st equation (Outcome)	2nd equation (Blog)	
			Pre	Post
Green Tea Drinks	Salestock	0.9	-	0.1
	Blogstock	0.4	0.1	0.6
	Adstock	0.4	0.1	0.9
Movie	Salestock	0.2	-	0.9
	Blogstock	0.9	0.3	0.4
	Adstock	0.6	0.6	0.6
Cellular Phone Service	Salestock	0.8	-	0.1
	Blogstock	0.1	0.9	0.4
	Adstock	0.9	0.1	0.9

Appendix II-D Robustness check with a three-equation system

In order to allow for a complete system, as mentioned in the text, we also estimated a simultaneous equation model where we added a third equation to the system. This equation is given below:

$$
\begin{aligned}
\ln(\text{TV GRP}_{jt}) &= \gamma_0 + \gamma_1 \text{Trend}_{jt} + \gamma_2 \text{Season}_{jt} + \gamma_{3j} \text{Brand}_j \\
&\quad + \gamma_4 \ln(\text{TV GRP}_{jt-1}) + \gamma_5 \ln(\#\text{Blogs}_{jt}) + \gamma_6 \ln(\#\text{Customers}_{jt}) \\
&\quad + \gamma_7 \ln(\text{CumCustomers}_{jt}) + \gamma_8 \ln(\text{CumBlogs}_{jt}) + \varepsilon_{3jt},
\end{aligned}
\tag{II-B.1}
$$

We found that the results from sales volume equations and blog equations were the same between the two equation models and the three equation specifications. The advertising equation results (Tables II-33, II-3, and II-35) showed that the current TV GRPs were significantly correlated with the previous TV GRPs for all the three product categories. The results also suggested that the Blogstock and Salestock had a relation with TV GRPs especially at post-launch period.

Table II-30 Sales Volume Equation Comparison (Green Tea Drinks)

	Full model (Proposed)				Three-equation model			
	Est.	Std. Error	t value	Pr(>|t|)	Est.	Std. Error	t value	Pr(>|t|)
Intercept	-1.71	4.20	-0.41	0.683	-1.72	4.30	-0.40	0.690
Seasonal								
Tue	0.41	0.96	0.43	0.667	0.41	0.99	0.42	0.676
Wed	0.63	0.79	0.79	0.431	0.63	0.81	0.77	0.443
Thu	0.75	0.75	1.01	0.314	0.75	0.77	0.99	0.325
Fri	0.06	0.72	0.08	0.934	0.06	0.74	0.08	0.937
Sat	0.65	0.73	0.90	0.371	0.65	0.75	0.87	0.384
Sun	-2.34	0.72	-3.26	0.001	-2.34	0.74	-3.18	0.002
Holiday	-2.55	2.42	-1.06	0.292	-2.55	2.48	-1.03	0.304
Trend	-0.03	0.01	-2.28	0.023	-0.03	0.01	-2.22	0.027
Brand								
Brand1	6.87	1.06	6.47	0.000	6.85	1.09	6.29	0.000
Brand2	8.52	0.77	11.02	0.000	8.51	0.79	10.73	0.000
Brand3	8.23	0.83	9.97	0.000	8.22	0.85	9.71	0.000
Brand4	6.10	0.65	9.34	0.000	6.10	0.67	9.10	0.000
#Audience(-1)	-0.09	0.04	-2.12	0.035	-0.09	0.05	-2.06	0.040
#Blog	0.01	0.05	0.31	0.760	0.02	0.05	0.32	0.748
Blogstock	0.56	0.18	3.05	0.002	0.56	0.19	2.98	0.003
Salestock	0.25	0.03	9.46	0.000	0.25	0.03	9.27	0.000
Adstock	0.08	0.04	2.16	0.031	0.08	0.04	2.12	0.035
Blogstock *Adstock	0.04	0.02	1.81	0.071	0.04	0.02	1.77	0.078

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-31 Sales Volume (Audience) Equation Comparison (Movies)

	Full model (Proposed)				Three-equation model			
	Est.	Std. Error	t value	Pr(>|t|)	Est.	Std. Error	t value	Pr(>|t|)
Intercept	-1.06	0.61	-1.74	0.083	-1.00	0.33	-3.03	0.003
Seasonal								
Tue	0.40	0.07	5.73	0.000	0.40	0.04	10.63	0.000
Wed	0.93	0.05	16.85	0.000	0.92	0.03	30.95	0.000
Thu	0.38	0.05	7.29	0.000	0.38	0.03	13.49	0.000
Fri	0.49	0.05	10.58	0.000	0.50	0.03	19.67	0.000
Sat	1.07	0.05	22.78	0.000	1.07	0.03	41.89	0.000
Sun	0.98	0.05	20.33	0.000	0.99	0.03	37.53	0.000
Holiday	0.92	0.10	9.61	0.000	0.92	0.05	17.73	0.000
Trend	-0.01	0.00	-4.05	0.000	-0.01	0.00	-7.80	0.000
Brand								
Movie1	0.26	0.10	2.57	0.010	0.26	0.05	4.73	0.000
Movie2	0.39	0.10	3.88	0.000	0.39	0.05	7.03	0.000
Movie3	0.42	0.07	6.06	0.000	0.42	0.04	11.04	0.000
Movie4	-0.07	0.08	-0.80	0.425	-0.07	0.04	-1.50	0.135
Movie5	0.45	0.11	4.18	0.000	0.45	0.06	7.71	0.000
Movie6	0.00	0.07	0.04	0.965	0.00	0.04	0.09	0.931
Movie7	0.14	0.06	2.25	0.025	0.14	0.03	4.26	0.000
Movie8	0.28	0.06	4.52	0.000	0.28	0.03	8.32	0.000
Movie9	-0.26	0.10	-2.58	0.010	-0.26	0.05	-4.84	0.000
Movie10	0.31	0.05	6.29	0.000	0.32	0.03	12.19	0.000
Movie11	-0.22	0.07	-3.11	0.002	-0.22	0.04	-5.53	0.000
#Audience(-1)	0.28	0.11	2.63	0.009	0.28	0.06	4.80	0.000
#Blog	-0.03	0.03	-1.00	0.318	-0.03	0.02	-1.89	0.060
#PositiveBlog	0.01	0.00	4.28	0.000	0.01	0.00	7.88	0.000
#NegativeBlog	0.00	0.00	0.47	0.635	0.00	0.00	0.91	0.363
#Blog Viewed	0.30	0.04	7.57	0.000	0.30	0.02	13.93	0.000
#Blog Award	-0.02	0.01	-2.20	0.028	-0.02	0.01	-4.08	0.000
#Blog Interesting	-0.01	0.01	-1.39	0.166	-0.01	0.00	-2.56	0.011
#Blog Advertising	-0.01	0.00	-1.50	0.134	-0.01	0.00	-2.66	0.008
Blogstock	0.40	0.02	21.02	0.000	0.40	0.01	38.36	0.000
Salestock	0.26	0.00	74.03	0.000	0.26	0.00	136.37	0.000
Adstock	0.08	0.02	4.50	0.000	0.08	0.01	8.16	0.000
Blogstock *Adstock	0.01	0.00	44.21	0.000	0.01	0.00	80.62	0.000

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-32 Sales Volume (Subscribers) Equation Comparison (Cellular Phone Service)

	Full model (Proposed)				Three-equation model			
	Est.	Std. Error	t value	Pr(>|t|)	Est.	Std. Error	t value	Pr(>|t|)
Intercept	3.79	5.17	0.73	0.469	4.84	7.14	0.68	0.503
Seasonal								
Tue	0.32	0.08	4.08	0.000	0.32	0.11	2.93	0.007
Wed	0.11	0.10	1.15	0.258	0.14	0.13	1.06	0.296
Trend	0.00	0.01	0.23	0.820	0.00	0.02	-0.12	0.906
Brand								
Brand1	0.71	0.24	2.93	0.007	0.69	0.34	2.03	0.052
Brand2	0.14	0.10	1.42	0.165	0.05	0.14	0.36	0.721
Brand3	0.32	0.11	2.87	0.008	0.27	0.16	1.70	0.099
Brand4	0.99	0.15	6.50	0.000	1.10	0.21	5.18	0.000
#Customer(-1)	-0.34	0.11	-3.00	0.006	-0.38	0.16	-2.43	0.021
#Blog	-0.50	0.52	-0.96	0.343	-0.55	0.72	-0.77	0.449
#PositiveBlog	0.45	0.09	5.21	0.000	0.50	0.12	4.19	0.000
#NegativeBlog	-0.04	0.05	-0.86	0.394	-0.03	0.07	-0.49	0.627
#Blog Subscribed	1.04	0.18	5.68	0.000	1.10	0.26	4.31	0.000
#Blog Price	-0.62	0.11	-5.45	0.000	-0.66	0.16	-4.18	0.000
#Blog Interesting	0.02	0.09	0.20	0.844	0.02	0.12	0.16	0.872
#Blog Advertising	0.25	0.04	6.03	0.000	0.26	0.06	4.57	0.000
Blogstock	1.22	0.11	11.44	0.000	1.13	0.15	7.48	0.000
Salestock	0.20	0.02	9.81	0.000	0.17	0.03	5.84	0.000
Adstock	-0.12	0.01	-10.31	0.000	0.91	0.02	55.06	0.000
Blogstock *Adstock	0.11	0.00	629.82	0.000	0.10	0.00	409.66	0.000

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-33 Blog Equation Comparison (Green Tea Drinks)

	Full model (Proposed)				Three-equation model			
	Est.	Std. Error	t value	Pr(>|t|)	Est.	Std. Error	t value	Pr(>|t|)
Intercept	-4.37	1.98	-2.21	0.028	-4.41	2.03	-2.18	0.030
Seasonal								
Tue	-0.70	0.88	-0.80	0.423	-0.70	0.90	-0.78	0.435
Wed	1.40	0.76	1.85	0.065	1.40	0.78	1.81	0.071
Thu	-0.53	0.72	-0.73	0.464	-0.53	0.74	-0.71	0.475
Fri	0.09	0.68	0.14	0.890	0.09	0.70	0.13	0.894
Sat	-0.50	0.69	-0.72	0.472	-0.50	0.70	-0.70	0.483
Sun	0.28	0.68	0.41	0.681	0.28	0.70	0.40	0.689
Holiday	2.12	1.85	1.15	0.251	2.12	1.89	1.12	0.263
Trend	-0.02	0.01	-2.55	0.011	-0.02	0.01	-2.47	0.014
Brand								
Movie1	3.02	0.90	3.36	0.001	3.01	0.92	3.27	0.001
Movie2	1.04	0.80	1.29	0.197	1.05	0.82	1.27	0.204
Movie3	0.90	0.72	1.24	0.214	0.90	0.74	1.22	0.223
Movie4	-2.24	0.64	-3.53	0.000	-2.23	0.65	-3.42	0.001
Pre								
#Blog(-1)	0.10	0.14	0.77	0.442	0.11	0.14	0.76	0.447
Blogstock	0.03	0.99	0.03	0.977	0.00	1.01	0.00	0.997
Adstock	0.60	0.08	7.10	0.000	0.60	0.09	6.90	0.000
Blogstock *Adstock	-0.03	0.01	-6.06	0.000	-0.03	0.01	-5.56	0.000
Post								
#Blog(-1)	-0.07	0.06	-1.31	0.189	-0.07	0.06	-1.30	0.195
#Consumer	0.01	0.04	0.22	0.825	0.01	0.05	0.21	0.832
Blogstock	-0.45	0.51	-0.88	0.378	-0.44	0.53	-0.84	0.400
Salestock	0.36	0.10	3.41	0.001	0.36	0.11	3.33	0.001
Adstock	-0.26	0.11	-2.33	0.020	0.26	0.12	2.28	0.023
Blogstock *Adstock	0.27	0.02	13.59	0.000	0.27	0.02	13.29	0.000
Other products' blog	0.14	0.06	2.17	0.031	0.14	0.06	2.14	0.033

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-34 Blog Equation Comparison (Movies)

	Full model (Proposed)				Three-equation model			
	Est.	StdErr	t val.	Pr(>|t|)	Est.	StdErr	t val.	Pr(>|t|)
Intercept	-7.69	4.94	-1.56	0.120	-7.17	4.97	-1.44	0.149
Seasonal: Tue	-0.01	0.23	-0.05	0.961	0.23	-0.06	0.952	-0.01
Wed	-0.03	0.20	-0.17	0.864	0.20	-0.18	0.857	-0.04
Thu	0.09	0.19	0.46	0.648	0.19	0.45	0.651	0.09
Fri	0.12	0.18	0.66	0.508	0.18	0.65	0.518	0.12
Sat	-0.09	0.18	-0.53	0.597	0.18	-0.54	0.590	-0.10
Sun	-0.23	0.18	-1.27	0.204	0.18	-1.25	0.211	-0.22
Holiday	0.50	0.34	1.50	0.134	0.34	1.47	0.141	0.50
Trend	0.03	0.00	9.49	0.000	0.00	9.42	0.000	0.03
Brand: Movie1	-0.44	0.36	-1.23	0.217	-0.44	0.36	-1.22	0.223
Movie2	-1.09	0.30	-3.64	0.000	-1.08	0.30	-3.58	0.000
Movie3	-4.82	0.25	-19.13	0.000	-4.82	0.25	-18.95	0.000
Movie4	1.12	0.27	4.11	0.000	1.12	0.27	4.09	0.000
Movie5	0.10	0.31	0.30	0.761	0.10	0.32	0.32	0.752
Movie6	0.10	0.25	0.39	0.698	0.10	0.25	0.40	0.690
Movie7	-0.19	0.24	-0.77	0.441	-0.19	0.25	-0.76	0.447
Movie8	-0.03	0.25	-0.10	0.918	-0.03	0.25	-0.10	0.921
Movie9	0.60	0.27	2.22	0.027	0.61	0.28	2.20	0.028
Movie10	0.22	0.22	0.98	0.325	0.21	0.22	0.95	0.342
Movie11	-1.66	0.20	-8.11	0.000	-1.66	0.21	-8.04	0.000
Pre								
#Blog(-1)	0.20	0.03	6.78	0.000	0.20	0.03	6.73	0.000
#Blog Positive	0.00	0.01	0.35	0.728	0.00	0.01	0.35	0.725
#Blog Negative	0.00	0.02	0.19	0.851	0.00	0.02	0.20	0.843
#Blog Award	0.04	0.01	2.63	0.009	0.04	0.01	2.60	0.009
#Blog Interesting	0.08	0.01	5.58	0.000	0.08	0.01	5.53	0.000
#Blog Advertising	0.03	0.01	1.85	0.064	0.03	0.01	1.84	0.066
Blogstock	0.08	0.04	1.70	0.089	0.07	0.05	1.64	0.102
Adstock	0.12	0.01	12.23	0.000	0.12	0.01	12.05	0.000
Blogstock *Adstock	-0.03	0.00	-13.89	0.000	-0.03	0.00	-13.61	0.000
Post								
#Blog(-1)	-0.34	0.93	-0.37	0.710	-0.35	0.93	-0.38	0.705
#Blog Positive	-0.04	0.02	-1.78	0.076	-0.04	0.02	-1.74	0.082
#Blog Negative	0.00	0.02	0.18	0.859	0.00	0.02	0.16	0.873
#Blog Viewed	0.66	0.31	2.17	0.030	0.67	0.31	2.16	0.031
#Blog Award	-0.03	0.08	-0.34	0.733	-0.03	0.08	-0.36	0.720
#Blog Interesting	-0.05	0.06	-0.87	0.383	-0.05	0.06	-0.87	0.383
#Blog Advertising	-0.02	0.07	-0.20	0.839	-0.02	0.08	-0.21	0.837
#Audience	0.83	0.16	5.31	0.000	0.83	0.16	5.25	0.000
Blogstock	0.07	0.18	0.36	0.717	0.07	0.18	0.41	0.680
Salestock	-0.77	0.02	-38.28	0.000	-0.77	0.02	-37.83	0.000
Adstock	-0.05	0.10	-0.49	0.621	-0.05	0.10	-0.52	0.601
Blogstock *Adstock	0.02	0.00	7.18	0.000	0.02	0.00	7.38	0.000
Other products' blog	0.90	0.01	157.45	0.000	0.85	0.01	147.21	0.000

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-35 Blog Equation Comparison (Cellular Phone Service)

	Full model (Proposed)				Three-equation model			
	Est.	Std. Error	t value	Pr(>|t|)	Est.	Std. Error	t value	Pr(>|t|)
Intercept	-19.61	7.68	-2.55	0.013	-9.76	11.31	-0.86	0.392
Seasonal								
March	-0.04	0.12	-0.30	0.761	-0.04	0.18	-0.23	0.817
April	-0.04	0.13	-0.33	0.740	-0.13	0.20	-0.66	0.510
Trend	0.02	0.02	0.79	0.434	0.04	0.03	1.16	0.250
Brand								
Brand1	1.82	0.34	5.30	0.000	1.93	0.52	3.68	0.000
Brand2	2.14	0.13	16.10	0.000	2.26	0.20	11.23	0.000
Brand3	1.95	0.13	15.29	0.000	2.05	0.19	10.65	0.000
Brand4	-0.65	0.27	-2.39	0.020	-0.57	0.41	-1.38	0.173
Pre								
#Blog(-1)	0.09	0.19	0.45	0.657	0.03	0.29	0.11	0.911
#Blog Positive	0.20	0.03	6.01	0.000	0.20	0.05	3.97	0.000
#Blog Negative	-0.07	0.07	-0.95	0.344	-0.05	0.11	-0.46	0.648
#Blog Price	0.44	0.22	2.02	0.048	0.58	0.33	1.77	0.082
#Blog Interesting	0.15	0.20	0.73	0.466	0.40	0.31	1.30	0.199
#Blog Advertising	0.48	0.11	4.40	0.000	0.42	0.16	2.55	0.013
Blogstock	0.24	0.08	2.95	0.004	0.21	0.12	1.78	0.080
Adstock	0.20	0.03	6.56	0.000	0.18	0.05	3.89	0.000
Blogstock *Adstock	-0.03	0.00	-27.41	0.000	-0.03	0.00	-15.84	0.000
Post								
#Blog(-1)	2.80	1.26	2.22	0.030	2.61	1.91	1.37	0.176
#Blog Positive	1.76	0.32	5.53	0.000	1.50	0.48	3.13	0.003
#Blog Negative	-0.58	0.17	-3.51	0.001	-0.47	0.25	-1.89	0.063
#Blog Subscribed	0.71	0.42	1.69	0.096	0.68	0.64	1.06	0.293
#Blog Price	0.54	0.28	1.94	0.057	0.51	0.42	1.21	0.230
#Blog Interesting	0.12	0.20	0.63	0.530	0.18	0.30	0.61	0.546
#Blog Advertising	-0.06	0.11	-0.59	0.556	-0.12	0.16	-0.72	0.477
#Customer	0.10	0.18	0.57	0.573	-0.04	0.28	-0.16	0.874
Blogstock	0.45	0.20	2.32	0.023	0.52	0.30	1.77	0.081
Salestock	-0.09	0.06	-1.46	0.148	0.00	0.09	-0.03	0.979
Adstock	-0.04	0.03	-1.25	0.216	-0.05	0.05	-0.87	0.386
Blogstock *Adstock	0.02	0.00	31.86	0.000	0.01	0.00	10.80	0.000
Other products' blog	1.84	0.00	891.40	0.000	1.14	0.00	361.35	0.000

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-36 Advertising Equation (Green Tea Drinks)

	Three-equation model				
	Estimate	Std. Error	t value	Pr(>|t|)	
Intercept	2.37	1.43	1.66	0.098	.
Seasonal					
Tue	-1.44	0.67	-2.14	0.033	*
Wed	-1.24	0.58	-2.13	0.034	*
Thu	-0.11	0.55	-0.20	0.840	
Fri	-1.25	0.52	-2.39	0.018	*
Sat	-0.71	0.53	-1.34	0.180	
Sun	0.95	0.53	1.81	0.071	.
Holiday	-0.90	1.35	-0.67	0.505	
Trend	-0.05	0.01	-5.78	0.000	***
Brand					
Movie1	1.16	0.74	1.58	0.115	
Movie2	-1.41	0.71	-2.00	0.047	*
Movie3	-0.13	0.56	-0.22	0.825	
Movie4	-1.82	0.52	-3.49	0.001	**
Pre					
TV GRP(-1)	0.86	0.07	12.70	0.000	***
#Blog	0.15	0.05	3.31	0.001	**
CumBlog	-0.33	0.25	-1.30	0.195	
Post					
TV GRP(-1)	0.65	0.03	21.49	0.000	***
#Blog	0.15	0.04	4.03	0.000	***
#Consumer	0.04	0.03	1.26	0.210	
Blogstock	-0.77	0.22	-3.52	0.000	***
Salestock	0.44	0.02	19.34	0.000	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-37 Advertising Equation (Movies)

	Three-equation model				
	Estimate	Std. Error	t value	Pr(>|t|)	
Intercept	-5.44	0.78	-6.95	0.000	***
Seasonal					
Tue	0.92	0.52	1.78	0.075	.
Wed	-0.29	0.47	-0.63	0.531	
Thu	1.20	0.45	2.69	0.007	**
Fri	1.54	0.44	3.54	0.000	***
Sat	-0.78	0.42	-1.83	0.067	.
Sun	-0.58	0.43	-1.36	0.176	
Holiday	-1.43	0.81	-1.78	0.076	.
Trend	0.05	0.01	6.79	0.000	***
Brand					
Movie1	3.95	0.65	6.08	0.000	***
Movie2	-0.31	0.61	-0.50	0.616	
Movie3	-0.61	0.51	-1.19	0.234	
Movie4	0.16	0.63	0.25	0.805	
Movie5	1.08	0.59	1.83	0.067	.
Movie6	-0.65	0.56	-1.14	0.253	
Movie7	2.60	0.53	4.90	0.000	***
Movie8	1.98	0.59	3.33	0.001	**
Movie9	-0.20	0.62	-0.32	0.749	
Movie10	1.29	0.51	2.51	0.012	**
Movie11	-1.72	0.46	-3.75	0.000	***
Pre					
TV GRP(-1)	0.63	0.02	37.03	0.000	***
#Blog	0.09	0.05	1.99	0.047	
CumBlog	0.14	0.04	3.12	0.002	**
Post					
TV GRP(-1)	0.51	0.04	14.41	0.000	***
#Blog	-1.50	0.94	-1.60	0.111	
#Consumer	1.65	0.34	4.80	0.000	***
Blogstock	2.00	0.59	3.38	0.001	**
Salestock	-3.23	0.16	-19.80	0.000	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table II-38 Advertising Equation (Cellular Phone Service)

	Three-equation model				
	Estimate	Std. Error	t value	Pr(>|t|)	
Intercept	-25.88	19.01	-1.36	0.178	
Seasonal					
March	1.30	2.74	0.48	0.636	
April	-1.59	2.48	-0.64	0.523	
Trend	-0.43	0.44	-0.98	0.329	
Brand					
Brand1	-5.70	4.24	-1.34	0.184	
Brand2	-5.79	2.56	-2.26	0.028	*
Brand3	-5.69	2.36	-2.41	0.019	*
Brand4	0.88	2.83	0.31	0.758	
Pre					
TV GRP(-1)	0.29	0.44	0.66	0.513	
#Blog	4.71	1.02	4.62	0.000	***
CumBlog	-1.15	0.15	-7.73	0.000	***
Post					
TV GRP(-1)	0.69	0.19	3.63	0.001	**
#Blog	-0.73	2.52	-0.29	0.771	
#Consumer	-1.86	4.54	-0.41	0.683	
Blogstock	2.54	3.25	0.78	0.438	
Salestock	1.31	0.62	2.11	0.039	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

References

Bagwell, K. (2007), “The Economic Analysis of Advertising,” in M. Armstrong and R. Porter (eds.), *Handbook of Industrial Organization*, Vol. 3, North-Holland: Amsterdam, 1701-1844.

Brown, S. P. and D. M Stayman (1992), “Antecedents and Consequences of Attitude toward the Ad: A Meta-Analysis,” *Journal of Consumer Research*, 19 (June), 34-51.

Chintagunta, P. K., S. Gopinath and S. Venkataraman (2009), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation across Local Markets," *Marketing Science*, 29(5), 944-57.

Christakis, N. A. and J. H. Fowler (2007), “The Spread of Obesity in a Large Social Network over 32 Years,” *The New England Journal of Medicine*, 357, 370-379.

Dellarocas, C., X. Q. Zhang, and N. F. Awad (2007), "Exploring the value of online product reviews in forecasting sales: The case of motion pictures," *Journal of Interactive Marketing*, 21 (4), 23-45.

Dhar, V. and E. Chang (2009), "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Journal of Interactive Marketing*, 23 (4), 300-307.

Duan, W., B. Gu and A. Whinston (2008), "Do Online Reviews Matter? An Empirical Investigation of Panel Data," *Decision Support Systems*, 45 (4), 1007-16.

Elberse, A. and B. N. Anand (2007), “The Effectiveness of Pre-Release Advertising for Motion Pictures: An Empirical Investigation Using a Simulated Market,” *Information Economics and Policy*, 19, nos. 3-4 (October), 319-343.

Goel; S., J. M. Hofman, S. Lahaie, D. M. Pennock and D. J. Watts (2010), “Predicting consumer behavior with Web search,” Proceedings of the *National Academy of Sciences*.

Godes, D. and D. Mayzlin (2004), "Using online conversations to study word-of-mouth communication," *Marketing Science*, 23 (4), 545-60.

Gruhl, D, R. Guha, R. Kumar, J. Novak and A. Tomkins (2005), "The Predictive Power of Online Chatter," KDD 2005, Chicago, IL, August 2005.

Harrison-Walker, L. J. (2001), “The Measurement of Word-of-Mouth Communication and an Investigation of Service Quality and Customer Commitment as Potential Antecedents,” *Journal of Service Research*, 4 (1), 60-75.

Johnson, T. J. and B. K. Kaye (2004), “Wag the blog: How reliance on traditional media and the Internet influence credibility perceptions of Weblogs among blog users,” *Journalism and Mass Communication Quarterly*, 81 (3), 622-642.

Kumar, R., J. Novak, P. Raghavan and A. Tomkins (2005), “On the Bursty Evolution of Blogspace,” *Internet and Web Information Systems*, 8, 159-178.

Liu, Y. (2006), "Word of mouth for movies: Its dynamics and impact on box office revenue," *Journal of Marketing*, 70 (3), 74-89.

Lodish, L. M., M. Abraham, S. Kalmenson, J. Livelsberger, B. Lubetkin, B. Richardson and M. E. Stevens (1995), “How TV advertising works: A meta-analysis of 389 real world split cable TV advertising experiments,” *Journal of Marketing Research*, 32 (May), 125–139.

Mayzlin, D. and H. Yoganarasimhan (2008) “Link to Success: How Blogs Build an Audience by Promoting Rivals,” Working Paper, Yale University

Mitchell, A. A. and J. C. Olson (1981), “Are Product Attribute Beliefs the Only Mediator of Advertising Effects on Brand Attitude,” *Journal of Marketing Research*, 18 (August), 318-332.

Moul, C. (2007), “Measuring Word of Mouth’s Impact on Theatrical Movie Admissions.” *Journal of Economics and Management Strategy*, 16 (4), 859-892.

Nair, H., P. Manchanda and T. Bhatia (2006), “Asymmetric Social Interaction in Physician Prescription Behavior: The Role of Opinion Leaders,” *Journal of Marketing Research*, 47 (October), 883-895.

Nam, S., P. Manchanda and P. K. Chintagunta (2006), “The Effect of Signal Quality of Contiguous Word of Mouth on Customer Acquisition for a Video on Demand Service,” *Marketing Science*, 29 (4), 690-700.

Rao, A. G. and P. B. Miller (1975), “Advertising/sales response functions,” *Journal of Advertising Research*, **15** (2), 259-267.

Richins, M. L. (1983), “Negative Word of Mouth by Dissatisfied Consumers: A Pilot Study,” *Journal of Marketing*, 47, 68-78.

Rogers, E. (1983), *Diffusion of Innovations*, 3rd ed., Free Press: New York.

Rozin, P. and E. B. Royzman (2001), “Negativity Bias, Negativity Dominance, and Contagion,” *Personality and Social Psychology Review*, 5(4), 296-320.

Technorati (2007), http://www.sifry.com/alerts/archives/000493.html, accessed on July 30, 2009.

Technorati/ Universal McCann (2008), http://technorati.com/blogging/state-of-the-blogosphere//, accessed on May 24, 2010.

Tirunillai, S. and G. J. Tellis (2010), "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance," Working Paper, University of Southern California.

Trusov, M., A. V. Bodapati and R. E. Bucklin (2009) "Your Members are Also Your Customers: Marketing for Internet Social Networks," *Journal of Marketing*, 73 (September), 90–102.

Wooldridge, J. M. (2001), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press: Cambridge.

Chapter III

Consumers' Social Learning about Videogame Consoles through Multiple Website Browsing

Abstract

This essay examines the micro-level correlation between traditional marketing actions (TV ads and public relations) and consumers' social learning about newly launched videogame consoles (Wii and PS3 in 2006). We propose consumers' learning processes via perusal of information in online communities, by using "pageview" data of multiple websites from a clickstream panel as indicators. We propose a bivariate Bayesian learning model combined with complementary purchase choices. The proposed model enables simpler estimation of parameters and allows for detailed information about interaction between social and personal learning processes. From results we find empirical evidence that companies' traditional marketing actions have a larger impact on social learning than on regular personal learning, during the pre-launch period. When consumers make purchase decisions, their social beliefs of product quality are weighted at least three times more than their personal beliefs. Policy simulations suggest that by optimizing marketing actions, firms can enhance consumers' learning and promote higher engagement of the products.

III-1 Introduction

Recent advances in internet communication tools and online social networks enable consumers to conduct "social learning" about products more easily. Social learning is the learning process that is promoted by exchanging information among diverse consumers in (online) communities (Jayanti and Singh 2010; Calder, Malthouse and Schaedel 2009). A concept related to social learning is "consumers' engagement" – a new keyword among advertising circles. Since 2006, the Association of National Advertisers (ANA), American Association of Advertising Agencies (AAAA) and the Advertising Research Foundation (ARF), have worked together to develop a definition and metric for consumers' engagement (details on ARF Defining Engagement Initiative website). Although there is no established definition yet, the concept of consumers' engagement is described as the consumers' prospect of a brand idea, which is enhanced and stimulated by online interaction with other consumers, and not only through the offline one-way marketing communications. In practice, consumer's engagement is measured as, for example, duration, frequency and/or recency of visiting, viewing high-value or medium-value content, providing personal information, and posting customer reviews and comments in online communities. Firms conduct engagement marketing to enhance social learning which leads to long-term customer loyalty, resulting in maximizing purchase conversions.

This study explores individual level correlation between traditional marketing activity (TV ads and public relations (PRs)) and online communication (consumer's social learning) for newly launched videogame consoles (Wii and PS3 in 2006), measured by consumers browsing on product community websites. An anecdotal story about an advertising campaign for the introduction of the Wii videogame console, in the Japanese market, illustrates an example of conducting consumers' engagement marketing and a study of social learning in the videogame industry. Nintendo initially showed only the silhouette of the console at its introduction, and then provided additional information over time until Wii released in December 2006. By releasing limiting information at the beginning, Nintendo forced consumers to search product information and discuss their expectations for the console in online communities. Throughout the learning process, if consumers were highly engaged, Nintendo's strategy may result in higher sales.

According to the recent development of consumers' learning literature (Jayanti and Singh 2010; Calder, Malthouse and Schaedel 2009), it can be assumed that consumers learn about products in two ways (as outlined in Figure III-1). The first is by traditional personal learning. This type of learning can be also called cognitive learning and reason-based learning. In this process, consumers access all the possible information about products and cognitively evaluate them. The second source of learning product quality comes from other consumers (social learning). In the process of social learning, consumers interact with other consumers to acquire product information. The social learning is expected to stimulate the consumers' engagement, leading to higher product purchase probability.

We also assume that pageviews of community-based websites act as an indicator of the level of the consumer's engagement via social learning. At the same time, the personal learning is assumed to be indicated by browsing all videogame related websites including both community-based and non community-based websites. Furthermore, we anticipate that the two processes are correlated with each other. To best of our knowledge, no past studies have this assumption. That is, they assume that dual or multiple learning processes were conducted individually (*e.g.* Erdem and Keane 1996).

In summary, our working research questions are (a) do traditional ads and PR campaigns enhance consumer learning and, (b) what is the relative importance of the two types of learning on consumer purchase choice? By quantifying the impact of traditional marketing media, firms can manage to enhance consumers' learning and promote higher engagement with the product, potentially leading to increase in product purchase.

The rest of the chapter is organized as follows. Section III-2 reviews the literature and discusses our position in the literature. In section III-3, we describe the data. The model is discussed in Section III-4. Section III-5 contains results and Section III-6 is the conclusion.

III-2 Literature review

In the marketing literature, many studies explored how consumers learn about quality of products from information in the market, and these studies suggested that the consumer's learning occurs through dual or multiple processes. Petty and Cacciopo (1986) conceptualized the dual learning processes of systematic and heuristic routes. Other papers introduced emotional or experiential response in addition to cognitive processing (*e.g.*, Meyers-Levy and Malaviya 1999; Forgas 1995; MacInnis and Jaworski 1989; Edell and Burke 1987; Batra and Ray 1986).

On top of these existing well-known learning processes, social learning may also play an important role in consumers' purchase decisions. Social learning is the learning process that is promoted by exchanging information among diverse consumers in problem-solving communities. Jayanti and Singh (2010) defined and examined the social learning process – is generated by interactive cycles among communities and is motivated by actions for solving problems – about health care via online BBSs (Bulletin Board Systems). Calder, Malthouse and Schaedel (2009) discussed the effectiveness of advertising on consumer engagement by using experiments with eight different online experiences on websites. They examined two types of engagement with online media - Personal and Social-Interactive Engagement. They found that both types were positively associated with advertising effectiveness. Moreover, Social-Interactive Engagement was strongly correlated with advertising after controlling for Personal Engagement. Therefore, this study follows their definitions of social learning and personal learning processes.

The personal learning is conducted via consumers' cognitive route based on logical, deliberate and systematic information processing (*e.g.*, Petty and Cacciopo 1986). This route is self-acquired in nature so that it becomes personal learning. This leads to the concept of personal learning which encompasses the aspects of the cognitive learning. On the other hand, as we already outlined our definition of social learning, we assume that this route of learning is conducted based on information from interactions with other consumers. However, the past two studies measured the social learning by either directly surveying the subjects or observing consumers' behavior in field experiment settings. In contrast, we did not observe the learning processes, but instead estimated them from the consumers' browsing behaviors on community-based websites.

Bucklin (2008) reviewed past clickstream data studies in detail. In his article, the literature was categorized as providing three types of insights: how to (1) attract visitors to the site, (2) understand site usage behavior, predict purchase, and manage the site, and (3) assess activity across multiple sites and multiple channels. Our research intends to explore the issues in categories 1 and 2. In the category 1, Ilfeld and Winer (2002) examined the impact of online and offline advertising expenditures on website visits using aggregated data. They found that website visits were positively correlated with online ad spending but negatively correlated with offline ad spending. They also found that website visits, as independent variables, were positively correlated with awareness and brand equity measures in return. Compared to their study, we examine the dynamic aspect of individual's learning process using the individual panel of clickstream data. Other studies considered the impact of only online marketing activities, such as banner ads (Chatterjee, Hoffman and Novak 2003; Manchanda, Dube, Yong Goh and Chintagunta 2006) and email ads (Ansari and Mela 2003).

Many studies in the category 2 – understanding site usage behavior, predicting purchase and managing the site – have investigated the predictive power of website browsing behavior. Moe and Fader (2004) found that past visit behavior at the Amazon website increased the future probability of purchase conversion. Sismeiro and Bucklin (2004) and Montgomery, Li, Srinivasan and Liechty (2004) explored predictions of consumers' browsing path or completion of successive tasks to purchase. All those studies were limited to examining consumers' browsing behavior within a single e-commerce website. In contrast, our research is interested in browsing behavior among a large number of multiple websites, since consumers try to learn about product quality from various websites and compare informative contents.

Luan and Neslin (2010) and Erdem et al. (2005) closely shared our research interests. They investigated product learning processes through word-of-mouth communication. But these studies only considered a single learning process and did not differentiate social learning from personal learning. In addition, they used aggregate product-level datasets and assumed a structure to replicate unobserved individual consumer's learning process. In contrast, this study uses disaggregate data to directly infer individual learning processes.

III-3 Data

The date evaluated in this study was user-centric internet clickstream data collected by Video Research Interactive, Inc., which maintained a panel of approximately 12,000 Japanese panel members, whose website browsing behaviors were recorded over time by a firm's proprietary software, installed on their computers at home. The collected data contain visited sites' URLs and when they visited.

III-3.1 Videogame console ownership

The company also conducted annual written surveys to a randomly selected part of its existing panelists. There were 7,053 subjects who responded to the annual survey in November 2007 (around one year after releasing Wii and PS3 in December 2006).

As shown in Table III-2, 24 percent of the subjects owned one of the available videogame consoles at the date of the survey conducted in 2007. Wii gained 25.5% share among all the videogame users. In contrast, the share of PS3 was 4.5% and 1.6% of the videogame users owned both consoles.

III-3.2 Pre-purchase browsing behavior of videogame websites and classification of community-based websites

The study centered on the pre-purchase website browsing behavior of the subjects who bought the newly released videogame consoles, Wii or PS3 or both. For this purpose, we translated their website browsing records into the daily number of pageviews of videogame-related websites from the date that product information was first announced (April 28, 2006) to the date of the product launch (December 2, 2006). We selected 49 major videogame related Japanese websites (Table III-1).

Three individual raters classified the part of these videogame related websites as community-based, when the website had community features such as

BBS (Bulletin Board System) or users' review posting systems. [12] Final classification was determined by the majority votes of the three raters' responses for each videogame related website. Consequently, we chose nine community-based websites as marked in Table III-1. Note that in order to validate the classifications, we examined the inter-rater agreement and found "substantial agreement" among the raters by using Fleiss's kappa (Fleiss 1971). The validation details are in Appendix III-A.

Next, we counted pageviews of the community-based websites and used them as the indicator of the social learning process. For the empirical analysis, we selected and used the observations by the panelists who owned any of the available videogame consoles and visited the videogame-related websites more than twice during the analysis period from April to December 2006. This condition resulted in 1,078 panelists remaining for the analysis. The average numbers of daily pageviews per subject are reported in Table III 3. It shows that subjects who bought new videogame consoles (Wii and PS3) visited videogame-related websites almost two times more often during the period before the product was released. Moreover, Wii and PS3 owners viewed community-based videogame websites three times more often than the other videogame console owners.

In order to show videogame owners' cross-browsing behavior among multiple websites, we calculated Pearson's correlation coefficients for individual owner's pageviews between every two videogame-related websites over the analysis period. Table III-4 reports the averages of those correlation coefficients across (a) only the community-based websites, across (b) only the other videogame-related websites, and across (c) both the community-based and the other websites. Wii owners browsed across videogame-related websites more than did PS3 owners. PS3 owners were likely to focus on visiting particular websites. The other interesting finding was that Wii owners cross-browsed more on the other videogame websites than on the

[12] We used this ex-ante classification of the community-based websites rather than latent classifications which are estimated in a model. First, since both the social and personal beliefs are also latent, the identification of the two latent beliefs is based on the ex-ante classification from the observed data. Second, if it makes the group classifications latent, the model cannot figure out which characteristics of the websites derive to differentiate those groups.

community websites. In contrast, PS3 owners cross-browsed more on the community-based videogame websites than on the other websites.

III-3.3 TV GRPs and PRs (Public Relations)

Video Research Interactive, Inc. has also reported GRPs of all TV commercials which were aired in major Japanese geographic markets. We used the aggregated TV GRPs of videogame ads, segmented in genders of male/female and ages of teen/20's/30's/40's/50's/60's, and matched the segmented GRPs with the pageview data of the subjects who were in the same demographic segment according to the gender and the age groups. In addition, we classified the types of the videogame TV ads into console ads and software ads by Nintendo, Inc. and SCE (Sony Computer Entertainment, Inc.). Table III-5 shows the mean of the daily videogame TV GRPs by each type.

III-3.4 Limitation of the data: gap between the dates of launch and survey

It should be noted that, due to the limitation of the data set, we only observed which video game consoles the panel of consumers owned on the date of responding to the annual survey. The survey was conducted around one year after the new products had launched in the Japanese market. It is unknown when the panel subjects actually bought a new videogame console. This leads to the problem that we cannot directly match the consumers' beliefs about product quality with their purchase decision on their buying date.

To deal with this problem, in our model we examined the correlation between the cumulative quality beliefs on the date of the product release, and panelists' product ownership choices one year after the release. This could be a substantial issue when evaluating the empirical implications from estimated results by our model. But we may be able to regard our study as a conservative test of our proposed model if it reasonably fits to the observations of the videogame ownership and if their correlations of cumulative quality beliefs and choices are estimated to be significant. This study could provide evidence that quality beliefs have effect on purchase decisions, even with a one-year gap between the date of purchases and the estimated constructions of the quality beliefs.

This assumption relies on consumers' beliefs staying fairly constant after the product is released and until the date of their purchase. After the products are available, the quality beliefs could be updated by other possible factors such as product experiences (if consumers actually play the videogames somewhere) or reputation and word-of-mouth from other consumers who already own and play videogames. If these confounding factors have significant effects on the choice of the videogame consoles, estimated parameters for our model could be inconsistent. Due to the possible confounding factors of those unobservables, we cannot strongly conclude significant correlations between videogame purchase choices and quality beliefs, if any, in our empirical analysis.

However, CESA (Computer Entertainment Supplier's Association) reported that one third of unit sales of Wii were achieved by the first four months after the product released in Japan (http://report.cesa.or.jp). Thus, we can expect that the gap may be shorter for many of the subjects in the panel, who may have bought the products in the early stage, which would result in reducing the impact of the confounding factors.

III-4 Model

Data was estimated using a model consisting of two components – a bivariate learning process about product quality over time before the product launch, and individual purchase decisions at post launch.

III-4.1 A learning process of product quality

We assume that consumers have two different modes of learning processes. As a first case, we will describe a simpler single process model which only considers the personal learning process. Note that our unique formulation of signal information enables a simpler estimation procedure. Then, in a second dual learning processes case, the model is extended from the bivariate Bayesian learning model by Ackerberg (2003). Our proposed model considers the possible correlation between the two learning processes.

Basic case: a single learning process

Personal learning is expected to capture cognitive aspects of the products and construct cognitive beliefs about the quality of products over the time. Thus, $\widetilde{C}_{ijt}$ denotes consumer *i*'s personal (cognitive) belief about the product *j* at date *t*. C_{ij} denotes the true quality of the product *j* for the consumer *i*. Following the standard context of the Bayesian learning process (*e.g.* Erdem and Keane 1996), the consumers have uncertainty about the true quality of the product but have belief about its value.

At the date of *t*=0, the consumer *i*'s initial prior belief about the quality of the product *j* is assumed to be normally distributed. We also assume that the mean of the initial belief c_j^0 is also normally distributed across consumers.

(III-1) $\widetilde{C}_{ij0} \sim N(c_j^0, \sigma_{C0}^2),\ c_j^0 \sim N(0, \sigma_{c0}^2)$.

We expected that the videogame related website browsing behaviors of consumers can indicate their level of engagement or strength of their interest in products, that are driven by the beliefs about the product quality. Therefore, we consider the number of pageviews indicating a signal of the consumers' personal belief. The signal, S_{ijt}^C, is assumed to follow normal distribution.

(III-2) $S_{ijt}^C \sim N(C_{ij}, \sigma_{\widetilde{C}}^2)$.

Pageviews can be biased indicators of the signal of the product quality, influenced by other factors that may also lead to website browsing, such as seasonal factors like weekends and holidays. In addition, we are interested in the impact of the firms' marketing actions such as TV ads and PRs. Thus, we formulate the log number of pageviews as the additive combination of the signal of the product belief, seasonal factors, the company's marketing actions and a white noise as Equation (III-3).

(III-3) $\ln(n_{it}) = S_{ijt}^C + \beta_0\, Seasonal_t + \beta_1\, AD_{jkt} + \beta_2\, PR_{jt} + \eta_{ijt}$.

This equation can be rewritten as shown below by substituting Equation (III-2).

(III-4) $\ln(n_{it}) = C_{ij} + \beta_0 Seasonal_t + \beta_1 AD_{jkt} + \beta_2 PR_{jt} + \nu_{ijt}$, $\nu_{ijt} \sim N(0,\sigma_C^2)$.

It leads to a consistent estimate of the consumer's product belief which can be expressed as Equation (III-5) given the available information by the date *t*. $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ denote consistent estimates of the covariates via linear regression and $\overline{\ln(n_{i(t)})}$, $\overline{Seasonal_{(t)}}$, $\overline{AD_{jk(t)}}$ and $\overline{PR_{j(t)}}$ are means of log number of pageviews, seasonal factors, *k* type of TV ads (console and software) and PRs up to the date *t*.

(III-5) $\hat{C}_{ij(t)} = \overline{\ln(n_{i(t)})} - \left(\hat{\beta}_0 \overline{Seasonal_{(t)}} + \hat{\beta}_1 \overline{AD_{jk(t)}} + \hat{\beta}_2 \overline{PR_{j(t)}}\right) \sim N(C_{ij}, \frac{\sigma_C^2}{t})$.

Finally, by combining the prior belief and the information from the signals, the posterior belief about the product quality is normally distributed with an updating formulation for dates *t*=1,…, *T*.

(III-6) $\tilde{C}_{ijt} \sim N(\overline{C}_{ijt}, \Sigma_{Cijt})$,

where $\overline{C}_{ijt} = \Sigma_{Cijt}\left(\Sigma_{Cij(t-1)}^{-1}\tilde{C}_{ij(t-1)} + \frac{t}{\sigma_C^2}\hat{C}_{ij(t)}\right)$ and $\Sigma_{Cijt} = \left[\Sigma_{Cij(t-1)}^{-1} + \frac{t}{\sigma_C^2}\right]^{-1}$.

The advantage of this formulation is that it does not require simulations in its estimation procedure. The standard methods of the learning model need to infer the mean value of the signal information by simulating to recover its distribution. In our model, the consistent estimates of the signals are simply provided from the linear regression of the pageviews as in Equation (III-5).

Extended case: bivariate learning processes

As discussed, we assume that there is another mode of learning in addition to the personal learning. The social learning occurs through interactions with other consumers in community websites and helps consumers construct social beliefs about the product quality. The social belief is assumed to be correlated with the individual's personal belief. Ackerberg (2003) proposed bivariate learning processes.

However, the two beliefs were considered to be correlated only at the initial belief in his model. In contrast, our model takes the correlation into account both at the initial belief and throughout the entire updating processes as below.

(III-7) $D_{ij} = \gamma C_{ij} + d_j$

Equation (III-7) assumes that consumers have initial social belief d_j. In other words, during the entire updating process, social beliefs are systematically biased against the personal belief by the amount of d_j and by proportional to the value of γ.

According to the formulation of the social belief of Equation (III-7), we can rewrite Equation (III-1), consumer *i*'s initial personal and social beliefs about the quality of product *j* on the date *t*=0 and the distribution of different signals.

(III-8) $$\begin{cases} \tilde{C}_{ij0} \sim N(c_j^0, \sigma_{C0}^2), & c_j^0 \sim N(0, \sigma_{c0}^2), \\ \tilde{D}_{ij0} = \gamma c_j^0 + d_j^0, & d_j^0 \sim N(0, \gamma^2 \sigma_{c0}^2) \end{cases}$$

(III-9) $$\begin{cases} S_{ijt}^C \sim N(C_{ij}, \sigma_{\tilde{C}}^2), \\ S_{ijt}^D \sim N(D_{ij}, \sigma_{\tilde{D}}^2) \end{cases}$$

As explained in the data section, we classified the pageviews of the community-based websites from the videogame related websites. We assume that the pageviews of the community-based websites are led by the signal from the social belief and the other confounding factors, as by the similar discussion in the single learning process case. Then, we can show the consistent estimates of the personal and the social beliefs as follows.

(III-50)

$$\begin{cases} \ln(n_{it}) = C_{ij} + \beta_{0j} Seasonal_t + \beta_{1j} AD_{jkt} + \beta_{2j} PR_{jt} + \nu_{ijt}, & \nu_{ijt} \sim N(0, \sigma_C^2) \\ \ln(n_{it}^{COM}) = D_{ij} + \beta_{0j}^{COM} Seasonal_t + \beta_{1j}^{COM} AD_{jkt} + \beta_{2j}^{COM} PR_{jt} + \nu_{ijt}^{COM}, & \nu_{ijt}^{COM} \sim N(0, \sigma_D^2) \end{cases}$$

(III-11) $$\begin{cases} \hat{C}_{ij(t)} = \overline{\ln(n_{i(t)})} - \left(\hat{\beta}_{0j} \overline{Seasonal_{(t)}} + \hat{\beta}_{1j} \overline{AD_{jk(t)}} + \hat{\beta}_{2j} \overline{PR_{j(t)}}\right), \\ \hat{D}_{ij(t)} = \overline{\ln(n_{i(t)}^{COM})} - \left(\hat{\beta}_{0j}^{COM} \overline{Seasonal_{(t)}} + \hat{\beta}_{1j}^{COM} \overline{AD_{jk(t)}} + \hat{\beta}_{2j}^{COM} \overline{PR_{j(t)}}\right) \end{cases}$$

where $\begin{bmatrix} \hat{C}_{ij(t)} \\ \hat{D}_{ij(t)} \end{bmatrix} \sim NVN\left(\begin{bmatrix} C_{ij} \\ D_{ij} \end{bmatrix}, \begin{bmatrix} \frac{\sigma_C^2}{t} & \frac{\gamma\sigma_C^2}{t} \\ \frac{\gamma\sigma_C^2}{t} & \frac{\sigma_D^2}{t} \end{bmatrix} \right)$

Finally, similar to Equation (III-6) of the single learning case, the posterior beliefs of the product quality follow a bivariate Bayesian learning process during t=1,…, T based on the initial beliefs and the signal information updates

(III-12) $$\begin{bmatrix} \widetilde{C}_{ijt} \\ \widetilde{D}_{ijt} \end{bmatrix} \sim NVN(m_{ijt}, \Sigma_{ijt}),$$

$$m_{ijt} = \begin{bmatrix} \overline{C}_{ijt} \\ \overline{D}_{ijt} \end{bmatrix} = \Sigma_{ijt}\left(\Sigma_0^{-1} m_0 + t\Phi^{-1}\hat{Z}_{ijt}\right), \quad m_0 = \begin{bmatrix} 0 \\ d_j^0 \end{bmatrix}, \quad \Sigma_0 = \begin{bmatrix} \sigma_{C0}^2 + \sigma_{c0}^2 & \gamma\sigma_{c0}^2 \\ \gamma\sigma_{c0}^2 & \gamma^2\sigma_{c0}^2 \end{bmatrix},$$

$$\Sigma_{ijt} = \left(\Sigma_{Cij(t-1)}^{-1} + t\Phi^{-1}\right)^{-1} \quad \hat{Z}_{ijt} = \begin{bmatrix} \hat{C}_{ij(t)} \\ \hat{D}_{ij(t)} \end{bmatrix}, \quad \Phi = \begin{bmatrix} \sigma_C^2 & \gamma\sigma_C^2 \\ \gamma\sigma_C^2 & \sigma_P^2 \end{bmatrix}$$

III-4.2 Purchase choice based on the cumulative product beliefs

As we stated previously, there could be an issue with our data. Since we do not observe the date of consumers' product purchases, we cannot directly match the consumers' beliefs about product quality with their purchase decision on their buying date. To deal with this problem, we examined the correlation between product choices and the cumulative quality beliefs on the date of the product release.

Complementary bundle choice model

When there are two new products available in the market, consumers choose one of the following options {0, 1, 2, 1&2}: where 0 denotes buying neither product and 1&2 means buying both products. By using the complementary bundle choice model of Gentzkow (2007), the expected mean utility functions are assumed as follows.

(III-13)

$$\begin{cases} E\overline{u}_i(0) = 0, \\ E\overline{u}_i(j) = \theta_1 \overline{Q}_{ijT} + \theta_2 (\overline{Q}_{ijT})^2 + \alpha_1 \mathrm{Pr}ice_j + \alpha_2 CumAd_{jkT} + \alpha_3 CumPR_{jkT} + \xi_{1i} \qquad j = 1,2, \\ E\overline{u}_i(1\,\&\,2) = u_i(1) + u_i(2) + \Gamma. \end{cases}$$

where the mean of overall quality beliefs, $\overline{Q}_{ijT}$, are assumed to be a convex combination of the mean of the personal beliefs and the mean of the social beliefs on the date of the products launched, *T*, as $\overline{Q}_{ijT} = \lambda \overline{C}_{ijT} + (1-\lambda)\overline{D}_{ijT}$, $0 < \lambda < 1$. $CumAd_{jT}$ and $CumPR_{jT}$ denote cumulative summations of the all TV ads and all PRs from the dates *t*=1 to *T*. The parameter Γ indicates the complementarity (if it is positive) between two products. The error terms, ξ_{ji}, indicate the consumers' persistent taste about the products are distributed as bivariate normal.

(III-14) $$\begin{bmatrix} \xi_{1i} \\ \xi_{2i} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right).$$

More specifically, denoting *k*=0,1,2,3 to indicate the element of the choice options {0, 1, 2, 1&2}, the indirect expected utility can be defined over the choice options as follows.

(III-15) $$EU_{ik} = E\overline{u}_i\ (k) + \varepsilon_{ik}.$$

Assuming that the error term, ε_{ik}, follows the type-I extreme distribution, then the probability that the consumers choose to purchase the products given the consumers' persistent taste vector ξ_i can be written as the multinomial logit formulation.

(III-16) $$\Pr_i\left[y_i = k, Data_{iT}, \xi_i, \Omega \mid \Theta\right] = \frac{\exp\left[EU_{ik}\right]}{\sum_{k'=1}^{3} \exp\left[EU_{ik'}\right]}.$$

where Θ is the set of parameters includes $\{\lambda, \theta_1, \theta_2, \alpha_1, \alpha_2, \alpha_3, \Gamma\}$ and Ω is the set of parameters which are determined through the dual learning processes denotes $\{d_j, \gamma, \sigma_{C0}, \sigma_{c0}, \sigma_C, \sigma_D, \vec{\beta}\}$.

III-4.3 Identification and estimation

The main identification issue is laid on differentiating the latent social belief from the personal belief. The first source of identification comes from the formula assumption in Equation (III-7). The size parameter, d_j, accounts for shifting bias of the social belief from the personal belief. The proportional parameter, γ, is considered proportional difference. In addition, data may be able to support the identification of these two parameters. When the difference between two beliefs varies over time, those changes are taken account by the proportional parameter. On the other hand, if one time shock causes a stationary difference across the entire period, it is accommodated by the size parameter.

In the estimation, we need to normalize the first element of the variance of the initial beliefs as $\sigma_{C0}^2 + \sigma_{c0}^2$ to 1 for identification purpose in Equation (III-12). Finally, we can integrate out error terms, ξ_i, from Equation (III-16) and from the conditional likelihood function as in Equation (III-17). To replicate the error distribution, we apply simulated maximum likelihood method in our empirical parameter estimation.

(III-17) $$L = \prod_{i=1}^{N} \int \Pr_i \left[y_i = k, Data_{iT}, \xi_i \mid \Theta \right] dF(\xi_i \mid \sigma_{12}, \sigma_2).$$

III-5 Results and discussions

In this section, the estimated results are discussed in five subsections. The first subsection illustrates the results of estimated parameters in the bivariate learning process. The second subsection shows results from two pageview equations – social (community-based) and personal – as in Equation (III-10). Then, estimated results of consumers' purchase choice by the complementary bundle choice model are discussed in the third subsection. The fourth subsection reports results of testing the proposed model compared with two benchmark models. Finally, for an example of managerial implications, we conducted policy simulations based on different advertising schedule plans.

III-5.1 Results of the bivariate learning process

Table III-6 presents the parameter estimates of the bivariate learning process. The variance of posterior beliefs in personal learning (σ_C) and social learning (σ_D) are estimated to be larger than the variance of initial belief (σ_{c0}). The results suggest that the updates of the product beliefs by the social learning are more precise than the updates by the personal learning because its estimated variance is smaller. In other words, the information signal from the social learning is more informative for updating consumers' quality perceptions.

In Table III-6, we also find that initial social beliefs (d_j^0) are small compared to the zero means of initial personal beliefs (c_j^0). According to Equation (III-7), this means that the biases of the social beliefs from the personal beliefs are significant but small at the initial learning periods. However, the proportional relationship of beliefs (γ) is reasonably large, suggesting that as the number of updates increases, the biases of the social beliefs become larger by proportional to the value of the correlation between beliefs.

Figures III-2 and III-3 report the updated paths of means of the personal beliefs and the social beliefs about the two videogame consoles across the analysis period. Major differences between the two learning processes are (i) the personal learning occurs slightly faster than does the social learning. The inflection points of the personal learning are around the 20th day. In contrast, for the social learning they are around the 40th day or later. We also find that (ii) the updates of the personal beliefs follow a similar path for both Wii and PS3, but those of the social beliefs deviate from one another, resulting in higher values of PS3's quality belief. Furthermore, (iii) the terminal values of the social beliefs are larger than those of the personal beliefs, two times larger for Wii, and three times larger for PS3. As in Figure III-3, (iv) the updates of the social learning fluctuate largely over time. We anticipate that those spikes may result from the major exogenous shocks which the proposed model is not able to fully control for (e.g. large conventions and major advertisements, which would seem fitting to the large spikes).

III-5.2 Results of the pageview equations

The estimated results from the pageview equations as in Equation (III-10) are illustrated in Tables III-7 and III-8. Overall, we find that parameters for the constant, daily trend and all pageviews are positive and significant, in all cases. In addition, the results suggest that two types of TV ads and public relations have positive and significant correlations with the social pageviews. In contrast, only TV GRPs of Wii software ads are marginally significant in the personal pageview equations.

The estimated constants of the social pageview equations are larger than those of the personal pageview equations. They are two times larger for Wii and three times larger for PS3. This is consistent with the above results of the bivariate learning process. Since, according to Equation (III-11), the estimated constants should be the consistent estimators of the personal and the social beliefs. The updates of the two learning process should approach the values of the estimated constants.

III-5.3 Results of the purchase choice

Parameter estimates and elasticity

From the purchase choice model results in Table III-9, we find positive and significant estimates of consumers' overall beliefs of the product quality. Furthermore, the coefficient of quadratic overall beliefs is estimated to be positive and significant, suggesting that the relationship between consumers' overall beliefs of product quality and choice probability is U-shaped. Consequently, consumers tend to be risk-taking in terms of the product quality. This means that consumers are likely to buy videogame consoles in cases where their social belief and personal belief are at either high or low levels, but not moderate level.

The weight of personal belief (λ) is about .4, which means that the composition ratio of the overall belief is 4 to 6 between the personal belief and the social belief. In other words, combining the results of the dual learning process, which shows that the average of social belief is at least two times larger than that of personal belief, we can conclude that consumers weight social beliefs from 3 to 4.5 times more than personal beliefs in their decision in purchase choice.

Table III-9 also presents that the price coefficient is negative and significant. The average of price elasticity was calculated based on the parameter estimates, defined as the mean value of percentage increase in choice probability relative to a one percent change in console price. The results suggest that the own price elasticity has the largest effect size among all variables. In addition, the cumulative values of a companies' public relations and cumulative TV GRPs of console and software ads, are positively and significantly correlated with purchase decisions. Among all, the cumulative software ads have the largest elasticity.

Complimentarity and substitution

One of the advantages of the proposed model allows us to investigate the complimentary pattern between purchase decisions in Wii and PS3. Table III-9 indicates that complementarity parameter (Γ) is not significant and almost zero[13], which means that there is no incremental utility of buying both Wii and PS3 consoles instead of purchasing one of them. This result seems consistent with the common perception that the characteristics of the two products are very different and the two products' targeted segments are also varied (our data also suggested that only 1.6% of videogame users owned both consoles).

From the estimated values of the variance and covariance of the persistent taste in Table III-9, we can see that there is much less heterogeneity in consumer preference in the product utility of Wii than that of PS3. Also, we do not find significant preference correlation between the persistent tastes of the two products.

In addition, we examined substitution patterns between changes in the independent variables of two products. Table III-10 shows cross elasticity indicators of purchase probabilities, calculated as a one percent change in independent variables. The results suggest that the substitution patterns are asymmetric and the cross-elasticity of PS3 variables leads to a larger increase in Wii's purchase probability in general. Exceptionally, a change in Wii's price has a larger increase in PS3's purchase probability. We believe that it is because Wii's price was cheaper

[13] The relationship between Wii and PS3 is estimated to be no-complimentary ($\Gamma = 0$) but not substitutive ($\Gamma < 0$). The substitutive relationship means that, for instance, increase in Wii's price would lead to a direct switch from buying Wii to buying PS3 but no impact on buying both. However, the no-complimentary relationship indicates that increase in Wii's price results in a decrease in buying Wii and both products, but purchasing PS3 at the same level, since it is offsetted by switching from buying both to buying only PS3 (see detailed discussion in Gentzkow 2007, page 718).

than PS3's price, and a decrease in Wii's price would hurt PS3's demand more than the opposite scenario would hurt Wii's.

III-5.4 Model comparisons

Two benchmark models were tested to assess the performance of the proposed model. The first benchmark model was a complementary bundle choice model of Gentzkow (2007) without a product quality learning process. Instead of including consumers' social and personal beliefs about product quality, this model regarded exponential sums of the social pageviews and the entire personal (videogame-related) pageviews as proxies of the two beliefs. The second baseline model allowed only a single learning process instead of the bivariate learning process in the proposed model.

Estimated parameters and model fit measures are reported in Table III-11. The proposed model outperformed the two benchmark models according to three measures: AIC (in-sample), AIC (holdout-sample) and fitting ratio (true positive). The no learning baseline model fits better than the single learning model in terms of in-sample AIC and fitting ratio. The single learning model overestimated most of the response parameters. This might be because the model excluded the important differences between social and personal learning processes on purchase decision, compared to the other two models with two quality measures.

III-5.5 Policy simulations: advertising schedule plans

Planning advertising schedules is an important part of real marketing practices. The results showed that advertising enhanced consumers' engagement through social learning via browsing community-based websites. Therefore, changing advertising scheduling is likely to have long-term effects on consumers' purchase decisions.

To assess such effects, two different simulated advertising plans for Wii were examined. The total amount of TV GRPs remained the same during the analysis periods as that of the original plan. For the first simulated plan, we distributed the total TV GRPs evenly across the period. The second advertising plan decreasingly allocated the TV GRPs. [14] Then we calculated estimated purchase probabilities of

[14] We excluded a plan which increasing allocated TV GRPs, since the actual advertising plan was this case and the simulated results were also the same as the actual updates of estimated quality beliefs.

Wii in our sample by using the above estimated parameters and the actual datasets except for the different advertising schedule policies.

Figures III-4 and III-5 report the simulated updated paths of average quality beliefs of Wii's for the flat advertising plan, the decreasing advertising plan and the actual plan. While the simulated updates of the average personal beliefs of the three plans converged to almost the same value at the end, the terminal values of the average social beliefs differed among the three advertising schedules. The decreasing and flat advertising plans grew faster for both quality beliefs but also decayed faster than does the actual plan. The actual plan increasingly allocated TV GRPs across the period and the increasing amount of advertising prevented the quality beliefs from decaying at the last periods. However, the decreasing advertising plan developed higher values of the social learning at the beginning of the updates. Even though those values decayed later, they are high enough to stay substantially larger at the terminal values, than those of the actual adverting plan. Consequently, the purchase probability of Wii for the decreasing advertising plan is estimated to be 26.9%, which is 2.3% larger than the estimated buying probability for the actual plan.

III-6 Conclusion

Many marketers have struggled to manage consumers' engagement with their products and have missed opportunities to enhance social learning. We undertook this research to answer the empirical questions: (a) Do traditional advertisings and public relations campaigns enhance consumers' social and personal learning? (b) What is the relative importance of the two types of learning on consumer purchase choice?

In this research, we used disaggregate data of Japanese panelists' individual multiple-website browsing behavior in order to link consumers' social learning and personal learning about newly launched videogame consoles to their purchase decisions. Methodologically, we proposed a bivariate Bayesian learning model with a simpler formulation of estimating signal information and an assumption about correlation between two learning processes during the entire updating processes. Daily pageviews of videogame-related websites are used as indicators of the personal learning process. Social learning is assumed to be indicated by the pageviews of the community-based websites, that include community features.

The empirical analyses yielded substantive insights for marketing managers to implement engagement marketing by enhancing consumers' social learning. (i) When consumers consider their purchase of new videogame consoles, they weight social beliefs of product quality at least three times more than personal beliefs. (ii) Consumers are found to be risk-taking in terms of product quality, so that they are likely to buy videogame consoles when their social belief and personal belief are at either high or low levels, but not a moderate levels. (iii) While consumers update their quality beliefs, the information signal from the social learning is perceived to be more informative and becomes larger over time than the personal learning signal. (iv) Two types of TV commercials (console ads and software ads) and public relations have positive effects on the social learning. In contrast, only Wii's software ads are marginally significant for the personal learning. Finally, (v) the counterfactual advertising plans result in slightly higher estimated purchase probability in our policy simulation. Marketers are able to optimize their advertising schedule by taking the dual learning processes into account.

Further study should consider other factors in the analysis. First, this model does not directly consider the roles of social network structure and opinion leadership (Watts and Dodds 2007; Godes and Mayzlin 2009; Iyengar, Valente, and Van den Bulte 2009). However, the social learning process is expected to accommodate those factors implicitly. Since the social beliefs are indicated by consumers' multiple browsing behavior on community-based websites, the structure of browsing patterns can be considered a representative of the consumers' social network. Second, while we classified the videogame-related websites from the view of the community feature, we did not consider detailed contents of the websites. It may be useful to apply further classifications of the websites and use text-mining approaches to incorporate websites' contents and their WOM messages.

Last but not least, due to the limitation of the data, there is a time-gap between the date of product purchases and the estimated constructions of the quality beliefs. As we discussed above, the gap may be much shorter and this fact could support our study as a conservative test, based on the fact that the proposed model fits the observed choices well. Additional information of the gap period could enable improvements in the evaluations of the empirical results.

Tables and Figures

Table III-1 A list of videogame related websites

Domain name	Mean daily pageviews	Community-based
nintendo.co.jp	31.52	
konami.jp	13.66	
itmedia.co.jp	9.74	x
playstation.com	6.82	
nintendo.jp	6.76	
gamecity.ne.jp	4.94	x
sega.jp	4.17	
pokemon.co.jp	3.64	
capcom.co.jp	3.61	
famitsu.com	2.81	x
4gamer.net	1.51	x
segadirect.jp	1.25	
square-enix.com	1.17	
mediaworks.co.jp	0.84	
broccoli.co.jp	0.76	
nintendo-inside.jp	0.75	x
d3p.co.jp	0.69	
gpara.com	0.61	
xbox.com	0.59	
gameiroiro.com	0.52	
taito.co.jp	0.50	
bandainamcogames.co.jp	0.39	
girls-style.jp	0.35	
alchemist-net.co.jp	0.34	
mmv.co.jp	0.34	
g-res.com	0.32	
banpresto-game.com	0.29	
tecmo.co.jp	0.19	
capcom-fc.com	0.13	x
xbox-news.com	0.12	
aa-game.com	0.12	
ge-iro1.com	0.08	
gamers-review.net	0.08	x
koei.co.jp	0.07	
dorasu.com	0.06	

Table III—1 (cont.) A list of videogame related websites

Domain name	Mean daily pageviews	Community-based
games-capture.com	0.05	x
nintendo-m.com	0.04	
n-wii.net	0.03	
aae.co.jp	0.02	
em-game.net	0.01	
nindori.com	0.01	
onlineplayer.jp	0.01	
ps3-fan.net	0.01	x
gizmodo.co.jp	0.01	
webbee.net	0.01	
wii-ds.com	0.00	
nintendo-search.com	0.00	
nintendo-e.com	0.00	
wiifan.jp	0.00	

Table III-2 Data description: Videogame console possessions

	# Users	Percentage	Share
Wii owners	434	6.2%	25.5%
PS3 owners	77	1.1%	4.5%
Both (Wii & PS3) owners	28	0.4%	1.6%
Other console owners	1,160	16.4%	68.3%
All videogame owners	1,699	24.1%	100.0%

Table III-3 Mean daily pageviews of videogame-related websites

	All videogame sites		Community-based sites	
	Mean	s.d.	Mean	s.d.
Wii owners	1.14	4.02	0.32	1.72
PS3 owners	1.47	4.02	0.34	1.29
Both (Wii & PS3) owners	1.36	1.54	0.16	0.33
Other console owners	0.70	2.21	0.11	0.45

Table III-4 Average of Pearson's correlation coefficients across videogame-related websites

	(a) Community sites	(b) Other videogame sites	(c) Community & Other sites
Wii owners	0.081	0.140	0.072
PS3 owners	0.058	0.026	0.013
Both (Wii & PS3) owners	0.026	0.031	0.052
Other console owners	0.096	0.046	0.054
All videogame owners	0.095	0.068	0.059

Table III-5 Mean daily GRPs of videogame ads and PRs

	Wii (Nintendo)		PS3 (SCE)	
	Mean	s.d.	Mean	s.d.
Console TV ads				
Wii users	10.80	27.07	2.85	15.70
PS3 users	10.71	26.66	2.83	15.55
Both (Wii & PS3) users	9.41	23.22	2.53	13.40
Other console users	10.97	27.51	2.88	15.85
Software TV ads				
Wii users	10.90	40.59	6.67	21.97
PS3 users	10.92	40.26	6.77	22.26
Both (Wii & PS3) users	9.95	36.30	6.70	21.45
Other console users	11.09	41.46	6.77	22.22
Other consoles & software TV ads				
Wii users	103.25	56.44	99.69	57.32
PS3 users	104.97	55.86	101.19	57.38
Both (Wii & PS3) users	90.39	46.74	96.46	52.48
Other console users	103.78	56.31	99.96	57.83
PRs	0.07	0.25	0.08	0.28

Table III-6 Estimates of the bivariate learning process

	Estimate	Std.Err.	
Initial social belief (d_j^0)			
Wii	0.937	0.550	*
PS3	0.865	0.418	*
Correlation of beliefs (γ)	2.131	0.034	**
Variance of initial belief (σ_{c0})	0.443	0.089	**
Variance of posterior beliefs			
Personal learning (σ_C)	10.708	0.883	**
Social learning (σ_D)	8.200	0.713	**

Significant code: 0.01 '**', 0.05 '*', 0.1 '.'

Table III-7 Estimates of (personal) pageview equation: $\ln(n_{it})$

	Wii			PS3		
	Estimate	Std.Err.		Estimate	Std.Err.	
Constant	2.180	0.271	**	2.130	0.336	**
Trend	0.031	0.002	**	0.034	0.002	**
Tuesday	-0.017	0.305		-0.029	0.389	
Wednesday	0.096	0.312		0.001	0.390	
Thursday	0.074	0.310		0.039	0.393	
Friday	0.011	0.310		-0.003	0.390	
Saturday	-0.017	0.308		-0.054	0.389	
Sunday	0.010	0.305		-0.035	0.387	
Holiday	0.442	0.473		-0.139	0.598	
All pageviews	0.003	0.000	**	0.001	0.000	.
PRs	0.164	0.334		-0.043	0.390	
TV GRPs Console	0.004	0.004		-0.003	0.007	
TV GRPs Software	0.003	0.002	.	0.004	0.005	

Significant code: 0.01 '**', 0.05 '*', 0.1 '.'

Table III-8 Estimates of social (community-based) pageview equation: $\ln(n_{it}^{COM})$

	Wii			PS3		
	Estimate	Std.Err.		Estimate	Std.Err.	
Constant	5.461	0.235	**	6.935	0.269	**
Trend	0.058	0.001	**	0.049	0.001	**
Tuesday	-0.100	0.265		0.104	0.311	
Wednesday	0.229	0.270		0.217	0.312	
Thursday	0.158	0.269		0.179	0.314	
Friday	-0.006	0.269		-0.041	0.312	
Saturday	0.009	0.267		-0.107	0.312	
Sunday	0.056	0.265		-0.147	0.310	
Holiday	1.094	0.410	**	1.259	0.479	**
All pageviews	0.003	0.000	**	0.004	0.000	**
PRs	0.428	0.290	.	0.665	0.312	*
TV GRPs Console	0.018	0.003	**	0.011	0.006	*
TV GRPs Software	0.015	0.002	**	0.022	0.004	**

Significant code: 0.01 '**', 0.05 '*', 0.1 '.'

Table III-9 Estimates of purchase choice

	Elasticity	Estimate	Std.Err.	
Weight of personal belief (λ)	-	0.410	0.045	**
Complementarity (Γ)	-	-0.001	7.567	
Variance of persistent taste (σ_2)	-	1.910	0.010	**
Covariance of persistent taste (σ_{12})	-	0.000	0.070	
Overall belief	0.051	0.064	0.000	**
Quadratic of overall belief	3.659	0.465	0.000	**
Price ($)	-910.768	-0.443	0.000	**
Cumulative PRs	2.355	0.284	0.000	**
Cumulative TV GRPs Console	243.384	0.340	0.000	**
Cumulative TV GRPs Software	754.314	0.796	0.000	**

Significant code: 0.01 '**', 0.05 '*', 0.1 '.'

Table III-10 Cross elasticity

	Cross elasticity Wii to PS3	Cross elasticity PS3 to Wii
Overall belief	-0.026	-0.085
Quadratic of overall belief	-1.717	-6.329
Price ($)	1129.102	594.685
Cumulative PRs	-2.364	-2.283
Cumulative TV GRPs Console	-98.935	-434.092
Cumulative TV GRPs Software	-544.172	-1024.622

Table III-11 Model comparisons

	Model 1: No learning		Model 2: Single learning		Proposed model	
Weight of personal belief (λ)	0.404	**	-		0.410	**
Complementarity (Γ)	-0.001		0.003		-0.001	
Variance of persistent taste (σ_2)	2.303	**	2.013	**	1.910	**
Covariance of persistent taste (σ_{12})	0.000		0.000		0.000	
Overall belief	0.020	**	0.642	**	0.064	**
Quadratic of overall belief	0.211	**	2.469	**	0.465	**
Price ($)	-0.404	**	-1.973	**	-0.443	**
Cumulative PRs	0.271	**	1.082	**	0.284	**
Cumulative TV GRPs Console	0.381	**	0.707	**	0.340	**
Cumulative TV GRPs Software	0.801	**	1.308	**	0.796	**
AIC (in-sample)	3442000		3632578		2884920	
AIC (holdout-sample)	1781125		1730389		1700486	
Fitting ratio (true positive)	0.496		0.358		0.768	

Significant code: 0.01 ‘**’, 0.05 ‘*’, 0.1 ‘.’

* We randomly set aside 100 subjects to be used as holdout sample.

Figure III-1 Outline of the research and the proposed model

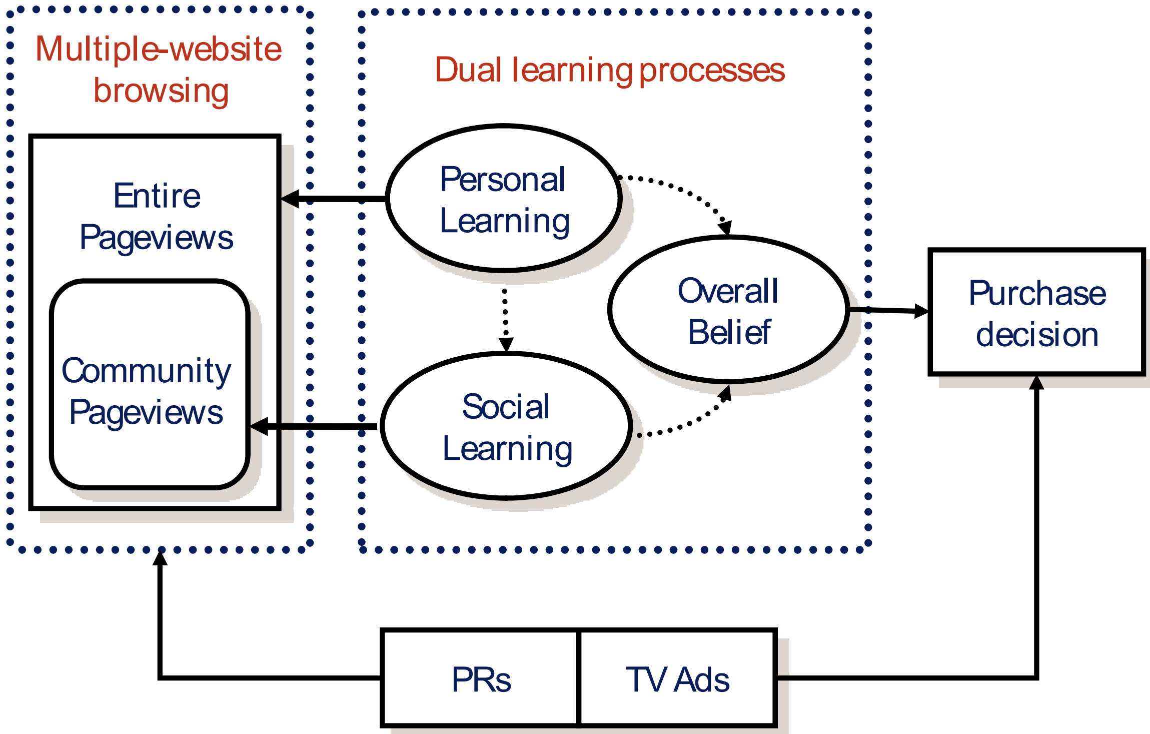
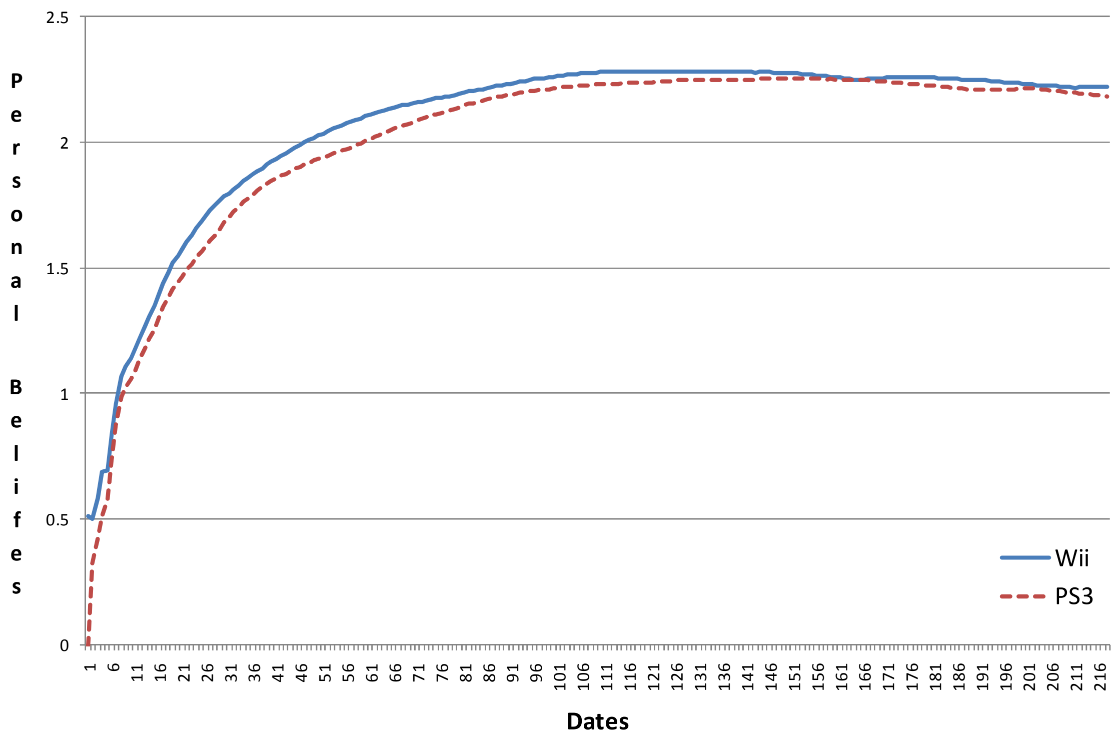


Figure III-2 Updates of mean personal beliefs



FigureIII-3 Updates of mean social beliefs

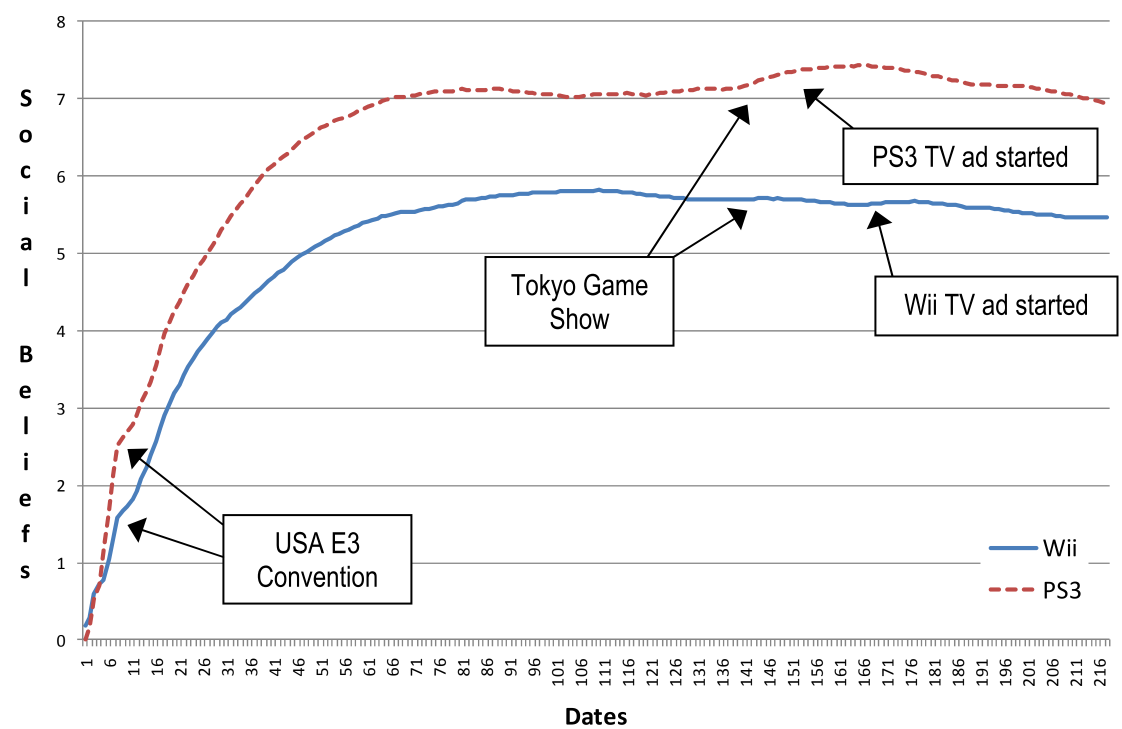


Figure III-4 Simulated updates of mean personal beliefs of Wii

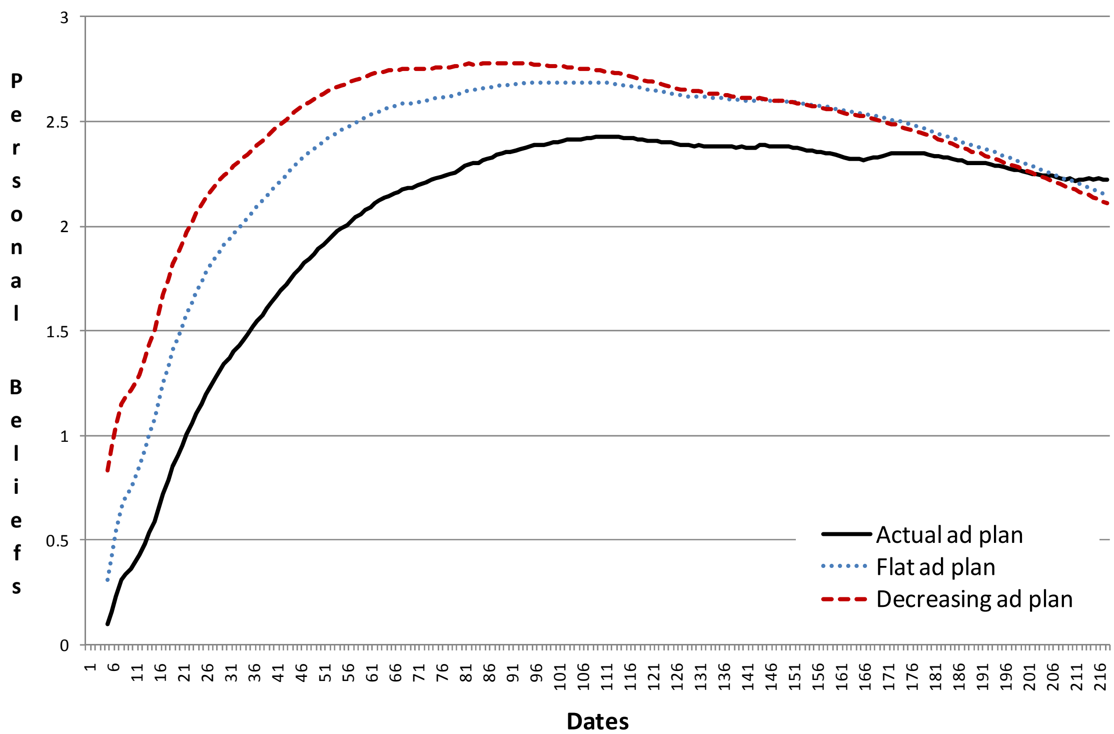
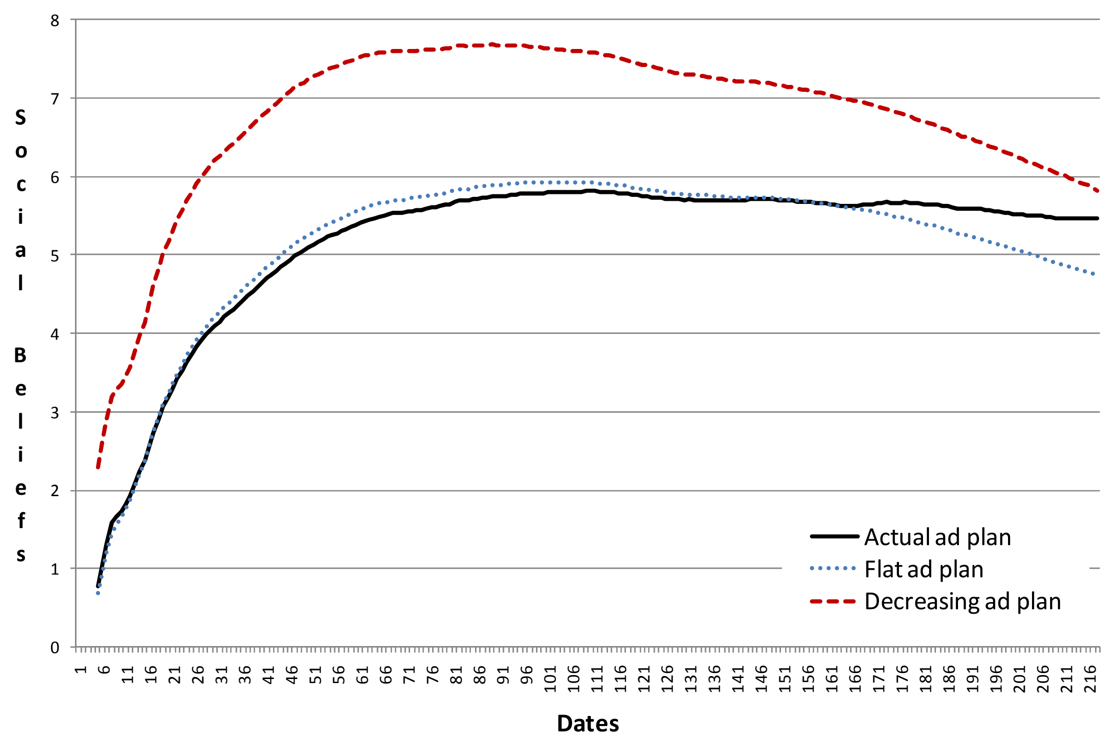


Figure III-5 Simulated updates of mean social beliefs of Wii



Appendix

Appendix III-A Inter-rater agreement in classifying community-based websites

Agreement among raters

To classify the community-based websites, the three raters rated them on a categorical scale (yes/no). Thus, we calculated inter-rater agreement using coefficient of Fleiss's kappa (Fleiss 1971), which measures the agreement among three or more judges beyond that expected by random classification.

The value of kappa was 0.603, indicating "substantial agreement" among raters (Landis and Koch 1977). Based on the test statistics, the null hypothesis, that the kappa is equal to zero was rejected at high significant level (p-value = 1.5e-13).

Agreement with groups by statistical cluster analysis

To study further validation of raters' classification, we also used a statistical cluster analysis to group videogame related websites into two types. Based on the panel members' cross-browsing behaviors of every two websites during the analysis period, the clustering method basically assigns similar websites into the same sub-group in the sense of having similar visitation patterns.

We employed the standard hierarchical cluster analysis with Euclidian distances and Ward's criterion of merging clusters. From the clustering results, we took the last two sub-clusters of the websites as group 1 and group 2. Then we calculated Fleiss's kappa to examine the agreement between the three raters and the cluster analysis. The value of kappa lowered to 0.219 which was evaluated as "fair agreement" according to Landis and Koch's table. The null hypothesis was also rejected at p-value = 0.00085.

From the above two validation results, we may be able to conclude that the inter-rater agreement in the classification of the community-based websites is reasonable and these raters' responses are also consistent with the statistical clustering based on the cross-websites browsing behaviors.

Appendix III-B Parameter Recovery

To assess the accuracy of parameter recovery we simulated a dataset with the same variables as in the model, generated randomly by setting arbitrary values to the parameters, for 150 individuals, 200 time-frames, two products, totaling 60,000 observations.

All estimated values of the parameters were recovered within a 95% confidence interval. The simulated maximum likelihood method took around 3,000 iterations to converge and recover the true parameter values well.

Table III-12 Key parameters with true values and 5%-95% confidence intervals

	True value	Estimate	5%	95%
Initial social belief (d_j^0)				
Wii	1.0	0.965	0.931	1.712
PS3	0.5	0.499	0.484	0.507
Correlation of beliefs (γ)	1.5	1.420	1.251	1.694
Variance of initial belief (σ_{c0})	0.6	0.653	0.273	1.446
Variance of posterior belief				
Personal learning (σ_C)	2.3	2.252	2.001	2.453
Social learning (σ_D)	1.2	1.231	1.138	1.246

References

Ackerberg, D. A. (2003), "Advertising, Learning and Consumer Choice in Experience Good Markets: An Empirical Examination," *International Economic Review*, 44 (3), 1007-1040.

Ansari, A. and C. F. Mela (2003), "E-Customization," *Journal of Marketing Research*, 40 (May), 131–145.

ARF Defining Engagement Initiative (2006); http://www.thearf.org/assets/research-arf-initiatives-defining-engagement?fbid=bB_008jPX7a

Bucklin, R. E. (2008). "Marketing Models for Electronic Commerce," in: Berend, W., ed., Handbook of Marketing Decision Models, Chapter. 10, 327-369.

Calder, B.J., E.C. Malthouse and U. Schaedel (2009) ,"An Experimental Study of the Relationship between Online Engagement and Advertising Effectiveness," *Journal of Interactive Marketing*, 23 (4), 321-331.

Chatterjee, P., D. L. Hoffman and T. P. Novak (2003), "Modeling the Clickstream: Implications for Web-Based Advertising Efforts," *Marketing Science*, 22(4), 520-541.

Erdem, T. and M. Keane (1996), "Decision-Making under Uncertainty: Capturing Choice Dynamics in Turbulent Consumer Goods Markets," *Marketing Science*, 16 (1), 1–21.

Erdem, T., M. Keane, T. Öncü and J. Strebel (2005), "Learning About Computers: An Analysis of Information Search and Technology Choice," *Quantitative Marketing and Economics*, 3 (3), 207-247.

Fleiss, J. L. (1971), "Measuring Nominal Scale Agreement Among Many Raters," *Psychological Bulletin*, 76 (5), 378-382.

Gentzkow, M. (2007), "Valuing New Goods in a Model with Complementarity: Online Newspapers," *American Economic Review*, 97 (3), 713-744.

Godes, D. and D. Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545-60.

Ilfeld, J. S. and R. S. Winer (2002), "Generating Website Traffic," *Journal of Advertising Research*, 42 (September–October), 49-61.

Iyengar, R., T. W. Valente and C. Van den Bulte (2011), "Opinion Leadership and Social Contagion in New Product Diffusion," Marketing Science, 30 (2), 195-212.

Jayanti, R. K. and J. Singh (2010), ”Pragmatic Learning Theory: An Inquiry-Action Framework for Distributed Consumer Learning in Online Communities,” *Journal of Consumer Research*, 36 (April), 1058-1081.

Landis, J. R. and G. G. Koch (1977), “The Measurement of Observer Agreement for Categorical Data,” *Biometrics*, 33, 159-174.

Luan, Y. J. and S. Neslin (2010), “The Development and Impact of Consumer Word of Mouth in New Product Diffusion," Working Paper, Tuck School of Business, SSRN 1462336.

Manchanda, P., J.-P. Dube, K. Yong Goh and P.K. Chintagunta (2006), “The Effect of Banner Advertising on Internet Purchasing,” *Journal of Marketing Research*, 43(February), 98-108.

Moe, W. W. and P. S. Fader (2004), “Dynamic Conversion Behavior at e-Commerce Sites,” *Management Science*, 50(3), 326-335.

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

Watts, D. J. and P. S. Dodds (2007), "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research*, 34 (4), 441-58.

Chapter IV

Co-evolution of Network Growth and Group Formation

Abstract

The use of social networking websites to conduct online marketing is a growing trend in the online advertising industry. To maintain their customer base and increase customer engagement, many social networking websites have developed the "group" function (also known as a forum or a community) as a marketing tool. The group function allows users to share information on a special topic with self-selected peers who have common interests (*e.g.*, music, sports, fashion, campus, career, etc.). In this research, we investigate the "stickiness" effect of the group function by measuring the propensity of group members to increase three networking activity measures (numbers of visits, connections and messages). In addition, we explore "co-evolution" of the network and the groups, since group growth is expected to contribute to enlarge the entire network (and vice versa).

We use a unique data set from a professional social network in which users build professional connections, exchange job search information and get recommendations. We applied a hierarchical Bayes regression model for examining each of the three stickiness measures by comparing the pre- and post group joining behavior relative to a control group of consumers. Our results indicate the presence of a significant effect of group formation on the website stickiness. We find that some group topics facilitate users' activity. Results also suggest the presence of a co-evolution effect – in the sense that group membership tends to increase primary membership on the website. Overall, our research sheds some light on a central concern to the administrators of (directed) social networking websites – what can be done to increase engagement and stickiness at these websites?

IV-1 Introduction

Advertisers and marketers frequently utilize social network websites as a marketing tool. For example, every major social network website has banner ads attached to their pages. As a more advanced example, Myspace has allowed music labels or musicians to set up their "Music page" for promoting their products. To keep websites attractive to network users, website managers need to retain users' engagement, and exhibit high levels of user activity on their websites, such as consumers' visiting the website more frequently, staying longer time, connecting and socializing with other members. A large customer base can result from these efforts, making a website more attractive to marketers who wish to conduct marketing campaign online.

Many social networking websites use a "group" function (also known as a forum or a community) as a mechanism to increase users' engagement. The group function allows users to share information on a special topic with self-selected peers who have common interests (*e.g.*, music, sports, fashion, campus, career, etc.). Although most groups are initiated by users, website managers can still manipulate group structures, their topics and develop its website's functionality to improve the group function. For instance, Facebook started a new feature called the "Fan page", in addition to the current group function, in 2008. By employing the fan page function, both users and companies can initiate a group page for their products. Figure IV-1 is an example of the fan page of the Volks Wagen CC, on which CC customers posts their stories about their own VW CC and share photos. Volks Wagen, in turn, is also allowed to provide information, and respond to questions from customers on its fan page. Therefore, the group function is expected to facilitate consumer-product bonding and attachment.

In this research, we investigate the "stickiness" effect of the group function by measuring the propensity of group members to increase networking activity defined as number of visits, connections and messages. In addition, this group growth may also help to enlarge the entire network. Similarly, the growing network would lead to the formation of groups in the network and an increase in group memberships. Virtually no research in the marketing literature has taken this "co-evolution" aspect into account. Our research makes an initial attempt to examine this inter-dependence of the network and the group by examining the stickiness effect and the co-evolution effect in the marketing context.

We use a unique dataset from a professional social network in which users build professional connections, exchange job search information and get recommendations. We respectively model the stickiness and the co-evolution at individual user level. The user-level stickiness effect is measured by their amount of networking activity such as a number of visits, connections and messages. We apply a hierarchical Bayes regression model to examine each of the three stickiness measures by comparing between-sample and within-sample before and after users join groups.

The rest of the chapter is organized as follows. In Section IV-2 we review the past studies of group functionality in networks. Section IV-3 describes the institutional setting and the data. Then we discuss the model in Section IV-4. Section IV-5 contains the results and we conclude in Section IV-6.

IV-2 Literature review

Despite the long history of network research, only a few studies have considered group functionality in networks. As a marketing study, Dholakiaa, Bagozzi and Pearo (2004) conducted an internet survey to investigate individuals' mental motivation to participate in sub-groups in several online virtual communities. Based on social psychology theories, they constructed a structural equation model of social influence variables associated with participation behaviors. They found two key constructs — group norms and social identity —lead to increases in the participation decisions.

In information technology literature, there are several studies looking at users' motivation to join groups. Backstrom et al. (2006) employed a decision-tree method to classify structural factors and examined their impact on individual motivation for membership, group growth and topic changes over time. Backstrom et al. (2008), in contrast, investigated engaged group users' impacts on the other group users. They found that group users who became heavily engaged were more likely to receive responses from core-users and stay longer after joining. Huffaker et al. (2011) studied diffusion of a virtual product in different groups of the Second Life network. They examined group structure characteristics and found that structure measures, such as cluster coefficient, average degree and largest connected component, had significant correlation with increasing adoption of the product. Another stream of group research in information technology examined user's motivation of contributing to communities. Rashid et al. (2006) and Beenen et al. (2004) conducted field experiments to explore user's willingness to rate movies in terms of stimulating several social settings and structural factors.

IV-3 Data

We used data from a professional social network website company. Like LinkedIn, this website has users that make connections with their business acquaintances and share job search information. The website was started in January 2008 and has grown rapidly, leading to more than 1.2 million registered users by the end of May 2009.

As Table IV-1 shows, the number of total users and newly registered users dramatically increased after October 2008. However, the website is confronted with a problem in retaining its users. Therefore, the number of confirmed users who accepted confirmation emails from the website, after signing up, was less than half of the number of total registered users.

Similar to past studies that evaluated networking activity of website users, we used the number of outgoing connections (*i.e.*, out-degrees) and the numbers of messages which were sent through the messaging system on the website (not including emails sent by regular emailing software). As seen in Table IV-2, the levels of users' network activity were extremely low – the average number of connections and messages of each user were less than one. In addition, we used the number of days each user visited the site to gauge the user's level of engagement. Due to the limitation of the weekly data set, however, we only observed the date of users' last login for a given week. Based on the last login date, we calculated the number of visits per week as the number of days from the first day of the week to the date of the last login if a user logged-in on that week.

As the new feature in the website, a group function was introduced on July 25, 2008, roughly seven months after the website was launched. By May 2009, there were about 2,000 groups initiated by users and 14,000 unique users in those groups. Those group users have joined 3.6 groups on average (Table IV-3). Table IV-4, shows the range of group topics from “Information Technology” to “Travel”.

To evaluate how introducing the group function influenced user behavior, we conducted a couple of simple preliminary analyses. First, we investigated the stickiness effect as a selection effect (between subjects) and a behavioral change effect (within subject). The left tile in Figure IV-2 compares the means of the number of outgoing connections by non-group users and group users. The significant difference between them provides basic support for the selection effect of the stickiness for the group function. Note that the numbers were normalized by users' membership days since the measured values tend to become larger as the users stayed longer in the website. Thus, we denominated the measure divided by the days from the registration date to the last login date. Second, we examined the behavioral change effect of the stickiness. A common problem in examining the selection effect is that the difference might be caused by selection bias. For example, the group users were inherently active users and they may have joined groups because they were inherently more engaged. In order to accommodate the selection bias effects, the right tile in Figure IV-2 compares the mean of the number of connections which the group users made before and after they joined their first group. The chart shows that the group users changed their behavior after joining groups, making more connections with other members in the network compared to the pre-joined period. Figures IV-3 and IV-4 demonstrates similar results for other network activity measures – the number of messages sent and the number days visited. As a minor counter-example, the selection effect in the mean number of visits occurred in the opposite direction (the left tile of Figure IV-4). However, the difference between non-group users and group users was relatively small compared with the difference in the other charts. Moreover, the behavioral change effect within group users was still significant. Overall, by simply comparing the network activity measures (the number of connections, messages and visits), the group users were more active than the non-group users and they became more engaged after joining groups.

Another expected benefit from the group function was that group users tend to invite new users from outside the network. To provide intuitive evidence of the co-evolution between the network and groups, we again employed the same simple analysis used to examine the stickiness effect. In this case, we used the mean number of connections with new users to measure attractiveness. We anticipate that group users have the power to attract new users, leading to a larger number of connections with newer users. We examined this by defining the new users as the users who registered for the website shortly before they had made a connection. For example, Figure IV-5 shows the results by defining the new users as those who registered within one day from the date of making a connection. The numbers in the figure are the mean connections by these users. There were significant differences in both the selection effect and the behavioral change effect. Figures IV-6 and IV-7 show comparisons based on measuring the number of new users by multiple number of days (three and seven days). Consequently, different definitions of the new users provided the same results. Those results suggested that group users attract more new users after joining groups, and that the group function will consequently contribute to the network growth.

In addition, we then conducted regressions to examine the stickiness effect and the co-evolution effect after controlling for other covariates, such as membership length (days since registered, days between registered and last visit, and days between joined and last visit), user characteristics (occupation, state, and number of groups joined) and group characteristics (topic genre, and number of group users). The results were similar to those of the above simple analyses (see Appendix IV-A).

Although we found that the introduction of the group function in the social network websites makes a difference in the stickiness effect and the co-evolution effect by the between-sample and the within-sample comparisons, there still may be some possibility that unobserved factors may confound with those comparisons and lead to fallacious results in the stickiness and the co-evolution effects. Therefore, in Appendix IV-B, we reported two descriptive statistics which comparing the number of dyadic connections of group users between before and after they join the same group, and the number of dyadic links of homophily users (in the same occupation and state) vs. the non-homophily users. Results showed that the number of dyadic links were statistically larger for post-join group users and non-homophily users, which means that the introducing the group function was the major factor to increase the number of connections among group users.

IV-4 Model

The preliminary results, discussed in the Data section, provided some intuitive evidence of the stickiness effect and the co-evolution effect of introducing the group function. In essence, we found that the group users were more active than the non-group users in visiting the website, connecting and messaging with other users and attracting newer users (the selection effect). Moreover, the group users became more engaged after they joined groups (the behavioral change effect). In this section, we focus only on the group users and investigate what types of group structures and topics may induce the users' behavioral change.

As described in the Data section, the network activity was measured by three types of dependent variables and used to examine the stickiness effect. These variables were the number of outgoing connections, messages sent and days visited (Table IV-5). On the other hand, the variable to evaluate the co-evolution effect was the number of connections with new users who registered shortly before making a connection. We took averages of these dependent variables at the two time points, before and after the user joined the group. With this set up, the dummy independent variable indicated the main condition of pre vs. post-joined. As in Equation (IV-1), the network activity measures per user i $(=1,\ldots, N)$ and per group k $(=1,\ldots, G_i)$ at occasion t $(=[Pre, Post])$ are regressed on the dummy indicator of joining and the other controlling covariates.

(IV-6)

$$
\begin{aligned}
(\text{Network Activity})_{ikt} &= \alpha_{1i} + \alpha_{2i}(\text{Join Dummy})_{ikt} + \beta_{1k} + \beta_{2k}(\text{Join Dummy})_{ikt} \\
&+ \gamma_1(\text{User Characteristics})_i + \gamma_2(\text{Join Dummy})_{ikt} \times (\text{User - Group Characteristics})_{ik} \\
&+ \gamma_3(\text{Join Dummy})_{ikt} \times (\text{Group Characteristics})_k + \varepsilon_{ikt},
\end{aligned}
$$

$$
\varepsilon_{ikt} \sim N(0, \sigma^2),
$$

$$
\begin{bmatrix} \alpha_{1i} \\ \alpha_{2i} \end{bmatrix} \sim MVN(\Delta_\alpha, V_\alpha), \begin{bmatrix} \beta_{1k} \\ \beta_{2k} \end{bmatrix} \sim MVN(\Delta_\beta, V_\beta).
$$

For both the stickiness effect and the co-evolution effect, we employed the same independent variables as listed in Tables IV-6, IV-7 and IV-8. As baseline covariates, the network activity is regressed on factors of the user characteristics, such as users' living state, type of occupation and their membership length. These variables are static per user so that the coefficients accompanied them are time-invariant (Table IV-6). In contrast, the other user characteristics in Table IV-7 – days since joined the group to last log-in and the total number of groups joined – were related to group membership and had an impact only after the users joined the group. Therefore, these coefficients are used only for the post-joined period, as part of an interaction term using the join dummy and the group related user characteristics. For the same reason, the group characteristics are set to have impact only at the post-joined occasions (Table IV-8). Note that, from the results of the past studies, we expected that the group structure characteristics may have a significant effect. Thus, we included three structural measures – average degree, clustering coefficient and largest connected component – which were found to be significant in the past studies (Huffaker et al. 2011).

In addition to the above controlling covariates, the model also included idiosyncratic random-effect intercepts for each user and each group to take into account heterogeneity across individual users and groups. These random effects are expected to accommodate individual variations of users and groups that are not captured by the user characteristics and the group characteristics in the model. Moreover, the heterogeneous constants consist of time-invariant baseline and intercepts of the join dummy. The mean of the mixing distribution of the intercepts of the join dummy represents main effects by the overall group function, if the mean is estimated as positive and significant. The model can be estimated by a MCMC procedure used in the standard hierarchical Bayes regression with two independent heterogeneous distributions across users and groups. The MCMC procedure and the prior settings used are detailed in Appendix IV-C.

IV-5 Results

In Table IV-9, we compared the proposed model with a baseline homogenous model and a heterogeneous model that included only a group random effect. In terms of posterior marginal log likelihood, the full model improved fitting to the observed data compared to the other models for all network activity measurements.

IV-5.1 Stickiness effect

The estimated results of the stickiness effects are shown in Tables IV-10, IV-11 and IV-12. The main effect of the behavioral change by joining the group was only significant for the mean number of visits. The main effects of the other network activity measures – the mean number of connections and messages – disappeared after controlling for the users' and groups' covariates and heterogeneous random effects. However, we found that the interaction effects of the join dummy with the user characteristics were mostly positive and significant. For example, the estimated coefficients of the interaction between the join dummy and the number of groups joined was positive for all the activity measures. In addition, the interaction effect of the join dummy with the number of days since joined was also significant, and the sign was negative, leading us to believe that users become more engaged the longer they are members of groups.

The estimates of the interaction between the join dummy and the number of group users was only significant for the mean number of visits. In terms of three group structure characteristics that we examined, only average degree had significant correlation with the network activity measurements. This result suggests that after a user joined groups, where members are densely connected, his or her network activity became more active.

The behavioral change effect, based on group topics, was marginally significant and varies over the network activity measurements. For example, the group topics of “Customer Service” and “Management” were positively correlated with the mean number of connections. In contrast, the group topics of “Accounting” and “Education” had a negative impact on the connections. More surprisingly, the group topics had no significant impact on the mean number of messages.

IV-5.2 Co-evolution effect

As shown in Table IV-13, we found similar results to the stickiness effects. The main effect of the behavioral change on attracting connections with new users was insignificant. Instead, the interaction effects based on joining the group were found to be positive and significant. The interaction effects of the join dummy with the number of days since joined and the number of groups joined were negative and positive, respectively.

In terms of the group characteristics, the estimated coefficients, between the join dummy and the days since a group was created, were positive for all cases. On the other hand, the estimate for group structure of the average degree was positive and significant. These results could lead to the conclusion that groups with longer history might have the power to attract more new users. In addition, some group topics might appeal to new users. For example, the topics of “Customer Service” and “Management” were estimated to have positive impact on attracting the connections with new users. On the other hand, the topic of “Information Technology” was negatively correlated.

In summary, after controlling for user characteristics, group characteristics and heterogeneous random intercepts of individual users and groups, the estimated results suggest that the group users became significantly more engaged after they joined groups. The main effect of the behavioral change was significant only for the mean number of visits, but interaction effects with the user characteristics were found to be significant. The impact of the group topics was also confirmed to be significant for all the network activity measurements. We found that a couple of group topics positively contributed to enhance stickiness and attracted connections with new users.

IV-6 Discussion and Conclusion

Sub-groups and/or communities within social networking websites are a global trend. Our study sheds some light on a central concern of the administrators of social networking websites to maintain their membership base and increase users' engagement. Specifically, we investigated the "stickiness" effect of the group function by comparing group members to increases in three networking activity measures (numbers of visits, connections and messages) at pre vs. post joining groups. In addition, we explored "co-evolution" of the network and the groups, since group growth may contribute to enlarge the entire network (and vice versa). We used a unique data set from a professional social network which introduced the group functionality to enhance the users' networking activity.

We applied a hierarchical Bayes regression model for examining each of the three stickiness measures by comparing the pre and post group joining behavior relative to a control group of consumers. Our results indicated the presence of a significant effect of group formation on website stickiness. Introduction of the group function, in the website, improved the stickiness effect so that networking activity was increased after network users joined groups compared to their activity immediately before joining the groups. The results also suggested the co-evolution between network growth and sub-group formations. Group users were likely to make connections with members who recently registered the network after they joined groups. Thus, we expect that group users have the ability to attract new members from outside the network.

In addition, we also investigated effects of specific user characteristics and group characteristics. Especially, some group topics were found to facilitate users' engagement and attract new users. We expect that it might make possible for website administrators to create and manage groups of certain topics to optimize the attractiveness of the website which then appeals to advertisers and marketers.

For further research, we may need to consider other factors in analysis. From website managers' perspective, it would be useful to examine the attractiveness of websites' functionality (*e.g.*, website design and page layout) and online events (*e.g.*, web-seminars and web-conference).

Tables and Figures

Table IV-1 Network statistics: Monthly growth

Month	Total # users	Total # out-degree	Mean distance	Longest distance	Clustering coefficient
Aug. 2008	40,463	169,063	2.94	9	0.0336
Sep. 2008	62,290	167,188	2.94	9	0.0333
Oct. 2008	90,789	188,451	3.40	12	0.0327
Nov. 2008	149,348	209,987	3.73	13	0.0324
Dec. 2008	227,934	268,762	3.65	15	0.0264
Jan. 2009	325,404	322,356	3.58	15	0.0222
Feb. 2009	407,461	326,128	3.60	15	0.0209

Table IV-2 Network activity statistics

Measures	Mean	s.d	Min	Max
# Connections	0.67	22.83	0	6387
# Messages (sent)	0.20	11.31	0	5131
# days of visits	7.05	5.18	1	86

Table IV-3 Group statistics

# Groups Joined	Mean	s.d.	Min	Max
All users (N=396,916)	0.11	1.51	0	487
Group users (N=9,774)	3.57	7.96	1	487

Table IV-4 Group topics

Topics	# Groups	# Group users
Jobs and Careers	2,301	9,184
Business and Markets	1,263	3,005
Management	775	1,846
Associations and Organizations	772	2,682
Information Technology	674	1,950
Health Care	659	1,269
Community and Social Issues	627	1,763
Accounting	611	1,200
Sales	604	2,460
Marketing	578	1,289
Admin and Clerical	383	1,056
Education	322	522
Human Resources	279	477
Travel	192	437
Customer Service	169	380
Companies	132	224
Events	100	259
Current Events	94	163
Other	2,030	5,154

Table IV-5 List of dependent variables (Network activity)

Effects	Variable name	Description
Stickiness effect	Connections	Mean number of outgoing connections per user at pre and post-joined
	Messages	Mean number of messages sent per user at pre and post-joined
	Visits	Mean days of weekly visiting the website per user at pre and post-joined
Co-evolution effect	Connections with new users	Mean number of outgoing connections with new users who registered recently per user at pre and post-joined

Table IV-6 List of independent variables: User baseline characteristics

Variable name	Description
State	Dummy indicators [0,1] of user's living state in the USA
Occupation	Dummy indicators [0,1] of user's job category (*e.g.*, Accounting, Consultant, Insurance, Sales, etc.)
Days since registered	Number of days since the user registered for the website to present
Days between registered & last visit	Number of days between the date of the user registered and the date when one visited the website at the last time

Table IV-7 List of independent variables: User-group related characteristics

Variable name	Description
Days since joined	Number of days since the data of the user joined the group to the present
# groups joined	Total number of groups that the user joined

Table IV-8 List of independent variables: Group characteristics

Variable name	Description
# groups users	Total number of unique users joined in the group
Group topic	Dummy indicators [0,1] of the group's topic genre (*e.g.*, Customer Service, Health Care, Marketing, etc.)
Days since group created	Number of days since the group created in the website to the present
Average degree	Mean number of connections among members in the group
Clustering Coefficient	Number of closed triads as a proportion of all connected triples in the group
Largest connected component	Largest subset of the group members within one can reach one another

Table IV-9 Model selection compared by posterior log likelihood

Model	Connection	Message	Visit	New user
Full homogenous model (baseline model)	-917469.2	-799053.4	-247254.7	-1265132.0
+ Group intercept & Join dummy	-552494.0	-620686.5	-137918.0	-676782.0
+ Group/ User intercept & Join dummy (proposed)	-504657.0	-608656.0	-134556.0	-680681.0

Table IV-10 Stickiness effect estimates (DV: Connections)

Parameter	Estimate	s.d.	(5%, 95%)
Join dummy (user-specific)	-6.0510	26.09	(-3048.3, 3042.2)
Join dummy (group-specific)	1.5130	1994.20	(-46.99, 37.23)
Days btw joined & last login	**-0.0023**	**0.0002**	**(-0.0026, -0.0020)**
# groups joined	**0.9034**	**0.0148**	**(0.8744, 0.9324)**
Days since group created	**0.2775**	**0.0184**	**(0.2415, 0.3135)**
# groups users	0.0014	0.0015	(-0.0016, 0.0044)
Average degree	**0.0029**	**0.0014**	**(0.0007, 0.0051)**
Cluster coefficient	0.3069	0.2389	(-0.1132, 0.7903)
Largest connected component	0.0022	0.0328	(-0.0483, 0.0613)
Accounting	**-27.4300**	**9.8000**	**(-46.6380, -8.2220)**
Admin and Clerical	-7.4660	9.6810	(-26.4408, 11.5088)
Associations and Organizations	-3.5720	8.8080	(-20.8357, 13.6917)
Business and Markets	4.3870	8.7050	(-12.6748, 21.4488)
Community and Social Issues	-1.1350	9.2110	(-19.1886, 16.9186)
Companies	12.7800	13.4800	(-13.6408, 39.2008)
Current Events	**-30.8200**	**15.7600**	**(-61.7096, 0.0696)**
Customer Service	**22.7200**	**11.4800**	**(0.2192, 45.2208)**
Education	**-31.4900**	**10.5800**	**(-52.2268, -10.7532)**
Events	-0.3855	13.4300	(-26.7083, 25.9373)
Health Care	-12.7600	9.4330	(-31.2487, 5.7287)
Human Resources	13.2800	10.8000	(-7.8880, 34.4480)
Information Technology	**-22.6300**	**9.1240**	**(-40.5130, -4.7470)**
Jobs and Careers	-6.6980	8.5330	(-23.4227, 10.0267)
Management	**16.0200**	**9.0870**	**(1.7905, 33.8305)**
Marketing	6.0260	9.3700	(-12.3392, 24.3912)
Sales	-7.2310	8.5790	(-24.0458, 9.5838)
Travel	-4.0510	8.8860	(-21.4676, 13.3656)

* The table presents the estimates for the population means and standard deviations of the parameters after 50,000 runs and 25,000 burn-ins.

* The significant posterior means are shown in bold according to the 95% posterior intervals around the mean.

Table IV-11 Stickiness effect estimates (DV: Messages)

Parameter	Estimate	s.d.	(5%, 95%)
Join dummy (user-specific)	0.0230	25.9800	(-43.77, 43.34)
Join dummy (group-specific)	4.3620	3.6660	(-0.1224,9.6235)
Days btw joined & last login	**-0.0010**	**0.0001**	**(-0.0012, -0.0009)**
# groups joined	**0.1729**	**0.0064**	**(0.1604, 0.1854)**
Days since group created	**-0.0228**	**0.0079**	**(-0.0382, -0.0073)**
# groups users	**-0.0037**	**0.0007**	**(-0.0050, -0.0024)**
Average degree	**0.0046**	**0.0027**	**(0.0034, 0.0076)**
Cluster coefficient	-0.0624	0.1770	(-0.3586, 0.2320)
Largest connected component	0.0354	0.0433	(-0.0349, 0.1072)
Accounting	-1.5300	4.2190	(-9.7992, 6.7392)
Admin and Clerical	0.4045	4.1680	(-7.7648, 8.5738)
Associations and Organizations	-1.1230	3.7920	(-8.5553, 6.3093)
Business and Markets	2.0480	3.7480	(-5.2981, 9.3941)
Community and Social Issues	-3.1350	3.9660	(-10.9084, 4.6384)
Companies	4.9570	5.8040	(-6.4188, 16.3328)
Current Events	6.9460	6.7840	(-6.3506, 20.2426)
Customer Service	1.9600	4.9410	(-7.7244, 11.6444)
Education	3.8840	4.5530	(-5.0399, 12.8079)
Events	1.2030	5.7830	(-10.1317, 12.5377)
Health Care	-2.2620	4.0620	(-10.2235, 5.6995)
Human Resources	3.3170	4.6490	(-5.7950, 12.4290)
Information Technology	-1.9550	3.9280	(-9.6539, 5.7439)
Jobs and Careers	2.3570	3.6740	(-4.8440, 9.5580)
Management	-1.7960	3.9130	(-9.4655, 5.8735)
Marketing	-0.1422	4.0330	(-8.0469, 7.7625)
Sales	1.7240	3.6930	(-5.5143, 8.9623)
Travel	1.9750	3.8200	(-5.5122, 9.4622)

* The table presents the estimates for the population means and standard deviations of the parameters after 50,000 runs and 25,000 burn-ins.

* The significant posterior means are shown in bold according to the 95% posterior intervals around the mean.

Table IV-12 Stickiness effect estimates (DV: Visits)

Parameter	Estimate	s.d.	(5%, 95%)
Join dummy (user-specific)	-0.03526	15.7340	(-24.47,24.82)
Join dummy (group-specific)	**0.3526**	**0.0728**	**(0.2099,0.4953)**
Days btw joined & last login	**-0.0001**	**0.0000**	**(-0.0001, -0.0001)**
# groups joined	**0.0004**	**0.0001**	**(0.0002, 0.0006)**
Days since group created	**0.0036**	**0.0002**	**(0.0033, 0.0039)**
# groups users	**0.0001**	**0.0000**	**(0.0000, 0.0001)**
Average degree	**0.0088**	**0.0032**	**(0.0028, 0.0133)**
Cluster coefficient	0.4017	0.4874	(-0.4668, 1.1685)
Largest connected component	0.0577	0.0137	(-0.0088, 0.0807)
Accounting	-0.0890	0.0830	(-0.2517, 0.0738)
Admin and Clerical	-0.1539	0.0821	(-0.3147, 0.0069)
Associations and Organizations	0.0023	0.0747	(-0.1440, 0.1486)
Business and Markets	0.0131	0.0738	(-0.1314, 0.1577)
Community and Social Issues	-0.0520	0.0781	(-0.2050, 0.1010)
Companies	**0.2286**	**0.1143**	**(0.0046, 0.4526)**
Current Events	0.0614	0.1336	(-0.2005, 0.3232)
Customer Service	-0.0813	0.0973	(-0.2719, 0.1093)
Education	**-0.1894**	**0.0896**	**(-0.3651, -0.0137)**
Events	-0.0457	0.1139	(-0.2689, 0.1775)
Health Care	-0.0811	0.0800	(-0.2378, 0.0756)
Human Resources	-0.0651	0.0915	(-0.2445, 0.1142)
Information Technology	-0.0027	0.0773	(-0.1542, 0.1489)
Jobs and Careers	-0.1191	0.0723	(-0.2608, 0.0226)
Management	0.0334	0.0770	(-0.1175, 0.1844)
Marketing	0.0627	0.0794	(-0.0930, 0.2183)
Sales	-0.0512	0.0727	(-0.1937, 0.0913)
Travel	0.0004	0.0752	(-0.1470, 0.1478)

* The table presents the estimates for the population means and standard deviations of the parameters after 50,000 runs and 25,000 burn-ins.

* The significant posterior means are shown in bold according to the 95% posterior intervals around the mean.

Table IV-13 Co-evolution effect estimates (DV: New Users within 1 day)

Parameter	Estimate	s.d.	(5%, 95%)
Join dummy (user-specific)	-57.70	77.09	(-756074.0,87.86)
Join dummy (group-specific)	31.77	4972.00	(-7216.0,7229.0)
Days btw joined & last login	**-0.0261**	**0.0023**	**(-0.0305, -0.0216)**
# groups joined	**13.3100**	**0.1757**	**(12.9656, 13.6544)**
Days since group created	**2.2030**	**0.2167**	**(1.7783, 2.6277)**
# groups users	0.0212	0.0183	(-0.0147, 0.0570)
Average degree	**0.0106**	**0.0035**	**(0.0053, 0.0161)**
Cluster coefficient	0.1471	0.4228	(-0.5719, 0.8605)
Largest connected component	0.0148	0.0881	(-0.1232, 0.1674)
Accounting	**-308.9000**	**116.40**	**(-537.0440, -80.7560)**
Admin and Clerical	-3.7990	115.00	(-229.1990, 221.6010)
Associations and Organizations	-7.1210	104.60	(-212.1370, 197.8950)
Business and Markets	27.0200	103.40	(-175.6440, 229.6840)
Community and Social Issues	-2.8070	109.40	(-217.2310, 211.6170)
Companies	-108.5000	160.20	(-422.4920, 205.4920)
Current Events	-171.1000	187.20	(-538.0120, 195.8120)
Customer Service	**393.5000**	**136.30**	**(126.3520, 660.6480)**
Education	-206.6000	125.60	(-452.7760, 39.5760)
Events	28.2800	159.60	(-284.5360, 341.0960)
Health Care	-139.5000	112.00	(-359.0200, 80.0200)
Human Resources	**331.1000**	**128.20**	**(79.8280, 582.3720)**
Information Technology	**-269.8000**	**108.40**	**(-482.2640, -57.3360)**
Jobs and Careers	27.2200	101.40	(-171.5240, 225.9640)
Management	181.3000	107.90	(-30.1840, 392.7840)
Marketing	116.5000	111.30	(-101.6480, 334.6480)
Sales	-37.2000	101.90	(-236.9240, 162.5240)
Travel	25.2100	105.50	(-181.5700, 231.9900)

* The table presents the estimates for the population means and standard deviations of the parameters after 50,000 runs and 25,000 burn-ins.

* The significant posterior means are shown in bold according to the 95% posterior intervals around the mean.

Figure IV-1 Facebook Fan page (*e.g.*, Volks Wagen CC)

Figure IV-2 Mean # connections (normalized by days)

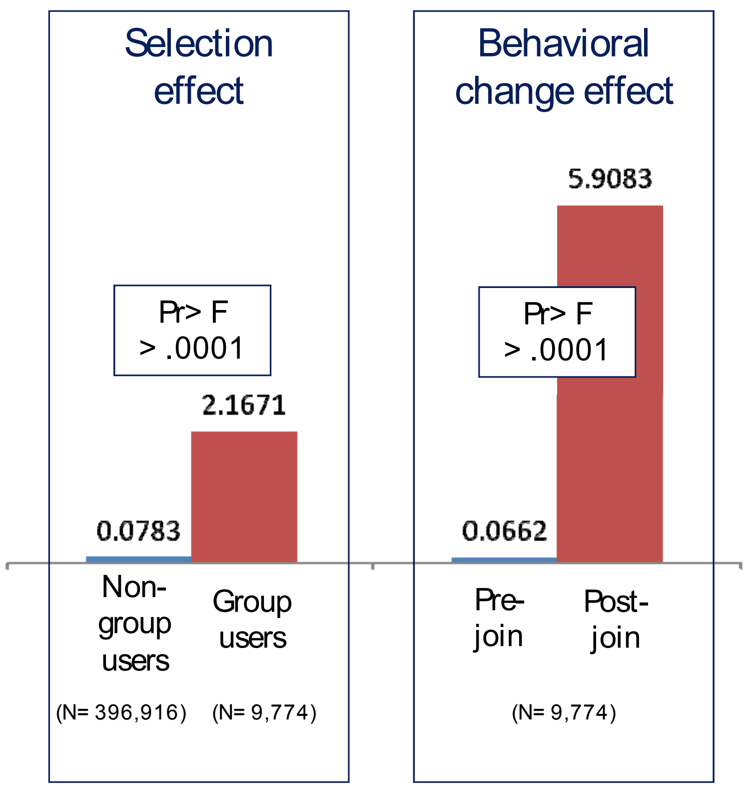


Figure IV-3 Mean # messages (normalized by days)

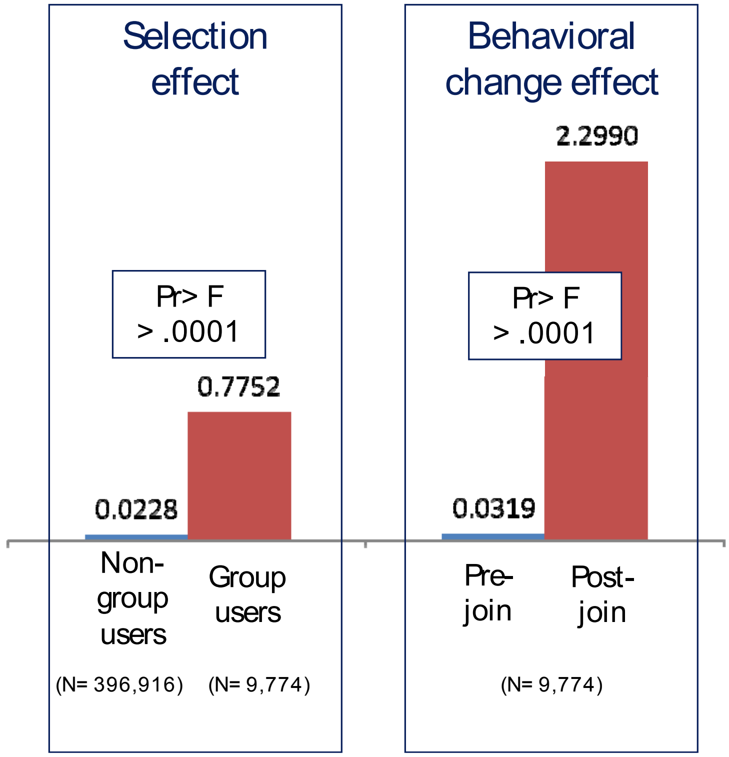


Figure IV-4 Mean # visits (normalized by days)

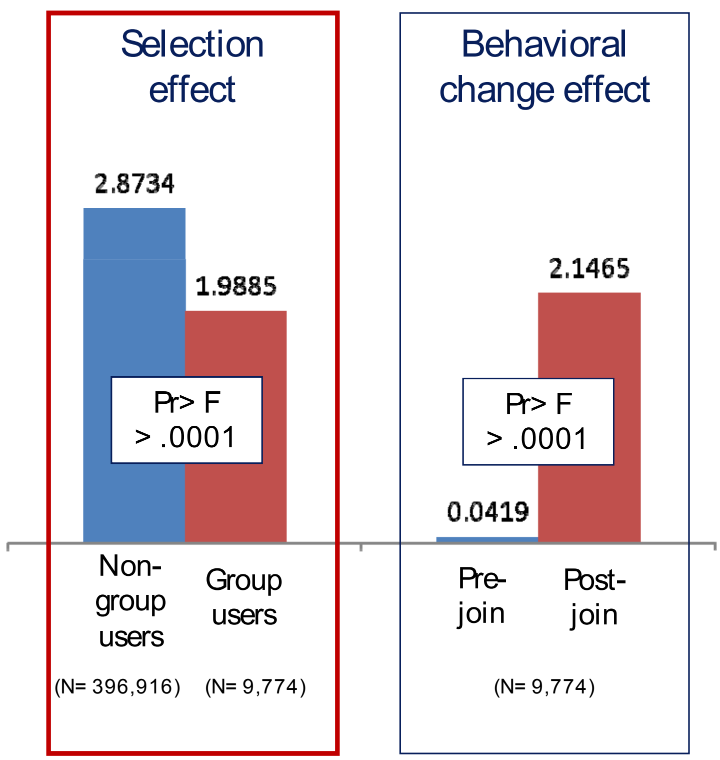


Figure IV-5 Mean # connections with new users who registered within 1 day

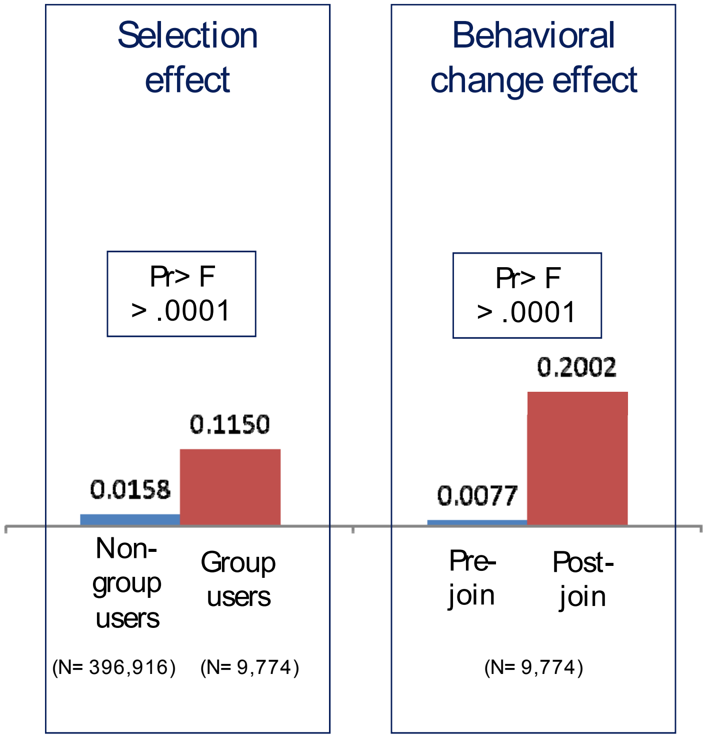


Figure IV-6 Mean # connections with new users who registered within 3 days

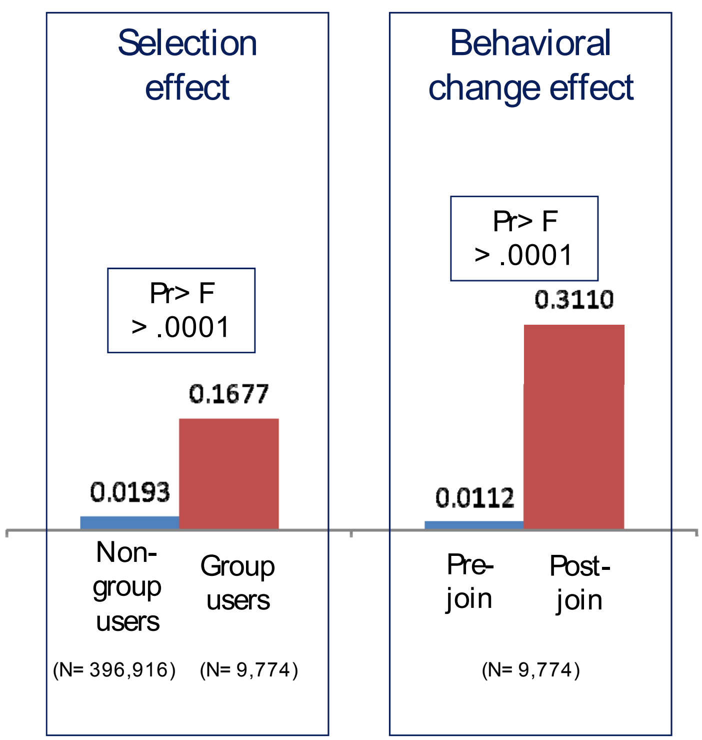
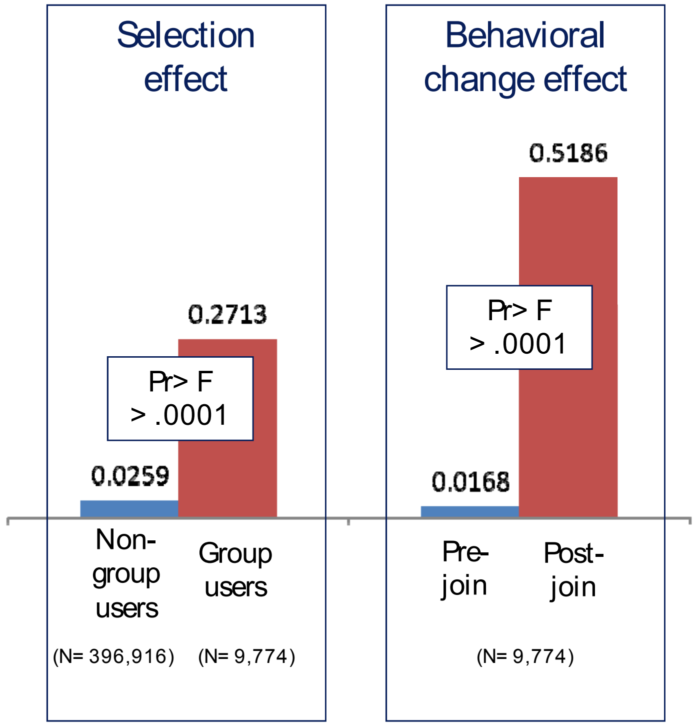


Figure IV-7 Mean # connections with new users who registered within 7 days



Appendix

Appendix IV-A Regression results of stickiness effects and co-evolution effects

In the preliminary analyses, we conducted regressions on the days of visits (days since registered to last login), the number of connections[15], and the number of messages to examine the stickiness effect and the co-evolution effect after controlling the other covariates such as membership length (days since registered, days between registered and last visit, and days between joined and last visit), user characteristics (occupation, state, and a number of groups joined) and group characteristics (topic genre, and a number of group users).

As shown in the tables below, in all cases the coefficients of the number of groups joined, and the interaction of the joined dummy with the number of groups joined, were positive and significant. The results assured that the group users were more active than the non-group users in visiting the website, connecting and messaging with other users, as well as attracting newer users (the selection effect). Moreover, the group users became more engaged after they joined groups (the behavioral change effect).

Table IV-14 Regression results DV: Days of visits (among all users)

Parameter	Estimate	Std Err.	t-value	Pr(>|t|).
Intercept	-1.5379	0.03309	-46.47	<.0001
Since registered	0.61497	0.01769	34.77	<.0001
# Groups joined	**0.05967**	**0.0003**	**195.98**	**<.0001**
# Connections	0.0755	0.00087	87.27	<.0001
# Messages	0.03023	0.00186	16.24	<.0001
# Recommendations	6.20744	0.44693	13.89	<.0001
# Observations	407,461			
Adjusted R-square	0.119			

[15] We indexed the number of connections with new users by taking natural log of days since registered over members connected with these new users.

Table IV-15 Regression results DV: # Connections of each user (among all users)

Parameter	Estimate	Std Err.	t-value	Pr(>|t|).
Intercept	0.026485	0.00264	10.01	<.0001
Dum. Pre/ Post joined	-0.005052	0.00049	-10.21	<.0001
Since registered	-0.000057	0.00000	-17.04	<.0001
Since last login	-0.000060	0.00000	-9.37	<.0001
# Groups joined	**0.000781**	**0.00016**	**4.82**	**<.0001**
Index con. w/ new users	0.001643	0.00054	3.03	0.0025
Groups joined*New users	-0.000096	0.00003	-2.99	0.0028
# Observations		8,064		
Adjusted R-square		0.131		

* Exclude observations of the first 7 days after registering

Table IV-16 Regression results DV: # Connections of each user (among group users)

Parameter	Estimate	Std Err.	t-value	Pr(>|t|).
Intercept	0.041704	0.00428	9.73	<.0001
Dum. Pre/ Post joined	0.001781	0.0054	0.33	0.742
Since registered	0.000065	0.00019	0.34	0.7363
Since joined	-0.000083	0.00000	-19.74	<.0001
Since join to last login	0.000064	0.00001	5.53	<.0001
# Groups joined	**0.000065**	**0.00000**	**7.98**	**<.0001**
Index con. w/ new users	-0.000905	0.00087	-1.03	0.3029
Joined * # Groups joined	**0.000167**	**0.00003**	**4.25**	**<.0001**
Joined * # New users	-0.003428	0.00106	-3.21	0.0013
# Groups joined*# New users	0.000014	0.00003	0.38	0.7051
# Observations		6,260		
Adjusted R-square		0.179		

* Exclude observations of the first 7 days after registering

Table IV-17 Regression results DV: # Messages of each user (among all users)

Parameter	Estimate	Std Err.	t-value	Pr(>|t|).
Intercept	0.080804	0.0243	3.31	0.0009
Dum. Pre/ Post joined	-0.026691	0.00421	-6.33	<.0001
Since registered	-0.000136	0.00003	-3.92	<.0001
Since last login	-0.000217	0.0001	-3.12	0.0019
# Groups joined	**0.001128**	**0.00114**	**0.98**	**0.3248**
Index con. w/ new users	0.005913	0.00495	1.19	0.2333
Groups joined*New users	-0.000197	0.00022	-0.87	0.3863
# Observations		1,165		
Adjusted R-square		0.089		

* Exclude observations of the first 7 days after registering

Table IV-18 Regression results DV: # Messages of each user (among group users)

Parameter	Estimate	Std Err.	t-value	Pr(>|t|).
Intercept	0.072076	0.03445	2.09	0.0367
Dum. Pre/ Post joined	0.092242	0.04135	2.23	0.0259
Since registered	-0.001224	0.00121	-1.01	0.3142
Since joined	-0.000238	0.00004	-6.54	<.0001
Since join to last login	0.000328	0.00009	3.72	0.0002
# Groups joined	**0.000200**	**0.00007**	**2.8**	**0.0052**
Index con. w/ new users	0.005031	0.00707	0.71	0.4772
Joined * # Groups joined	**0.000705**	**0.00027**	**2.6**	**0.0095**
Joined * # New users	-0.030329	0.00826	-3.67	0.0003
# Groups joined*# New users	0.000173	0.00023	0.74	0.4611
# Observations		1,070		
Adjusted R-square		0.196		

* Exclude observations of the first 7 days after registering

Appendix IV-B Checking unobserved factors to make dyadic connections

Although we found that the introduction of the group function in the social network websites makes a difference in the stickiness effect and the co-evolution effect by the between-sample and the within-sample comparisons, there still may be some possibility that unobserved factors may confound with those comparisons and lead to fallacious results in the stickiness and the co-evolution effects.

In order to check those unobserved factors, we conducted two comparisons of the number of dyadic connections among group users. If the unobserved factors exist, website users may make connections irrelevant to joining the same groups. This should lead to the fact that the number of dyadic connections is larger before they join the same groups than after they join. Therefore, the first comparison is based on pre vs. post join groups. Table IV-19 shows that the number of dyadic connections is statistically larger after the group users join the same group. This result supports our findings that joining the group increased the number of connections, which link both with users in the same groups and outside the groups.

On the other hand, in the information study literature, homophily is considered one major factor that makes people flock together. Theory of homophily assumes that individuals seek out others who follow the same interests and self categorization – auto restoration hobbyists, punk rockers – or belong to the same formal or informal groups – ethnicity, age groups (*e.g.*, McPherson et al 2001). Table IV-20 shows the number of connections among the same homophily (occupation and state) vs. the non-homophily. The result indicated that the effect of homophily was very small.

The two comparisons suggested that the introduction of the group function was the major factor that increased the number of connections among group users. Homophily and unobserved factors may only have small impacts on establishing dyadic links between the website users.

Table IV-19 Comparison of # connections at pre vs. post join groups

	# Connections	Percent	Chi-Square	Pr > ChiSq
Post Join	11,409	58.4	4931.0	<.0001
Pre Join	8,139	41.6		
Total	19,548			

Table IV-20 Comparison of # connections among homophily vs. non-homophily

	# Connections	Percent	Chi-Square	Pr > ChiSq
Non-homophily	19,207	98.3	164171.6	<.0001
Homophily	341	1.7		
Total	19,548			

Appendix IV-C MCMC estimation procedure and the prior setting

The proposed model in the text can be estimated by MCMC procedure used in the standard hierarchical Bayes regression with two independent heterogeneous distributions over users and groups. We used rhierLinearModel (bayesm) function in R software to estimate the model.

We assumed the natural conjugate priors of the multivariate regression in the model by specifying the second stage parameter distributions as

$$V_\alpha \sim IW(v_\alpha, \overline{V}_\alpha),\ \Delta_\alpha \mid V_\alpha \sim N(\overline{\Delta}_\alpha, V_\alpha \otimes A_\alpha^{-1}),$$
$$V_\beta \sim IW(v_\beta, \overline{V}_\beta),\ \Delta_\beta \mid V_\beta \sim N(\overline{\Delta}_\beta, V_\beta \otimes A_\beta^{-1}).$$

We also followed the regular way of setting defused priors:

$$v_i = 3, \quad s_{0i}^2 = \mathrm{var}(y_i),$$
$$v_\alpha = K_\alpha + 3, \quad \overline{V}_\alpha = v_\alpha \times 0.1 I_{K_\alpha}, \quad \overline{\Delta}_\alpha = 0, \quad A_\alpha = 0.01,$$
$$v_\beta = K_\beta + 3, \quad \overline{V}_\beta = v_\beta \times 0.1 I_{K_\beta}, \quad \overline{\Delta}_\beta = 0, \quad A_\beta = 0.01.$$

References

Beenen, G., K. Ling, X. Wang, K. Chang, D. Frankowski, P. Resnick and R. E. Kraut (2004), “Using social psychology to motivate contributions to online communities,” Proceedings of *ACM Conference on Computer Supported Cooperative Work*, 212–221.

Backstrom, L., R. Kumar, C. Marlow, J. Novak and A. Tomkins (2008), "Preferential Behavior in Online Groups," *WSDM’08*, February, 11-12, Palo Alto, CA, USA.

Backstrom, L., D. Huttenlocher, J. Kleinberg and X. Lan (2006), "Group Formation in Large Social Networks: Membership, Growth, and Evolution," *KDD’06*, August, 20-23, Philadelphia, PA, USA.

Dholakiaa, U. M., R. P. Bagozzi and L. K. Pearo (2004) “A social influence model of consumer participation in network- and small-group-based virtual communities,” *International Journal of Research in Marketing*, 21, 241–263.

Huffaker, D. A., C. Y. Teng, M. P. Simmons, L. Gong and L. A. Adamic (2011), “Group Membership and Diffusion in Virtual Worlds*,” IEEE’11*, Boston.

McPherson, M., L. Smith-Lovin, and J. Cook (2001), “Birds of a feather: Homophily in social networks,” *Annual review of sociology*, 27, 415–444.

Rashid, A. M., K. Ling, R. D. Tassone, P. Resnick, R. Kraut and J. Riedl (2006), “Motivating participation by displaying the value of contribution,” Proceedings of *ACM Conference on Human Factors in Computing Systems*, 955–958.

Chapter V

Conclusion

The three essays in my dissertation added to the very limited, but rapidly growing field of research into the effectiveness of new media of online communication (CGM). Using very unique datasets of blogging, social learning via web browsing, and social networking, we were able to combine data on market outcomes, traditional media (TV ads, PRs and group functions) and CGM (blogging, social learning and network activity). Thus this allowed us to investigate two open questions in this domain - (a) whether new media lead to differential market outcomes and (b) whether traditional marketing actions and CGM act synergistically.

In the first essay, using data from three product categories of green tea drinks, movies and cellular phone service, we applied a simultaneous equation model to capture the effect of new media on market outcomes and the effect of market outcomes on new media. While this itself is somewhat novel, we were also able to include traditional marketing activity (TV advertising) in both equations, both directly and via interactions. In addition, we also included a set of cross-sectional and temporal fixed effect variables as controls and an exclusion restriction in the blog equation. In general, we found that blogs were predictive of market outcomes, pre-launch TV ads spurred blogging activity (that is predictive of marketing activity) but became less effective in inducing blogging activity post-launch, and market outcomes also had some effect on blogging activity. Our text mining results provided "process" explanation which supported this finding. From a managerial point of view, in the experiment simulations using the estimated values of parameters, we found that a one percent increase in the traditional marketing instrument (TV advertising) led to a median increase in market outcomes of 0.2%, with a majority of the increase coming via the increase in blogging activity generated by the advertising pre-launch. The results suggested that managers would do well to consider including blogging activity in sales forecasting models. Moreover, by understanding the specific relationship between traditional and new media, managers can allocate resources to traditional media much better. Another useful implication that emerged from our findings was that blogging activity can be a surrogate measure of advertising effectiveness – sometimes, which is typically hard to gauge directly using traditional methods. Our analysis did also have a few limitations. The aggregate nature of our data made it very hard to offer micro-level causal explanations of the effectiveness of CGM and the synergistic relationship between CGM and traditional media.

In the second essay, we used disaggregate data of Japanese panelists' individual multiple-website browsing behavior in order to link consumers' social learning and personal learning about new videogame consoles to their purchase decisions. Methodologically, we proposed a bivariate Bayesian learning model with a simpler formulation of estimating signal information and an assumption about the correlation between two learning processes during the entire updating periods. Daily pageviews of videogame related websites were used as the indicators of the personal learning process. Specifically, we assumed that the social learning is indicated by the pageviews of the community-based websites which have community features to offer interactions among consumers. Our empirical analyses yielded substantive insights for marketing managers to implement the engagement marketing by enhancing consumers' social learning. Our results suggested that (i) when consumers considered their purchase of new videogame consoles, they weighted social beliefs of product quality at least three times more than personal beliefs. (ii) Consumers were found to be risk-taking in terms of the product quality, so that they were likely to buy videogame consoles when their social beliefs and personal beliefs were at either high or low levels, but not a moderate level. (iii) While consumers updated their quality beliefs, the information signal from the social learning was perceived to be more informative and becomes larger over time than the signal from the personal learning. (iv) Two types of TV commercials (console ads and software ads) and public relations had positive effects on the social learning. In contrast, only Wii's software ads were marginally significant for personal learning. Finally, (v) the counterfactual advertising plans resulted in the slightly higher estimated purchase probability in our policy simulation. Marketers are able to optimize their advertising schedule by taking the dual learning processes into account. These results suggest that firms can manage to enhance consumers' learning and induce higher engagement with the product, which is likely to result in higher purchase probability. Further research may need to consider the roles of social network structure and opinion leadership and use text-mining approaches to incorporate websites contents and their WOM messages. Due to a data limitation, we should be careful when evaluating the results, since there was a time-gap between the observed date of product purchases and the estimated periods.

In the third essay, we investigated the "stickiness" effect of the group function by a between-samples comparison of group users and non-users as well as a within-sample comparison of group members. We compared increases in three networking activity measurements (numbers of visits, connections and messages) at pre vs. post joining groups. In addition, we explored "co-evolution" of the network and the groups, since group growth may contribute to enlarging the entire network (and vice versa). We used unique data from a professional social network in which users built professional connections, exchanged job search information and got recommendations. We applied a hierarchical Bayes regression model for examining each of the three stickiness measures by comparing the pre- and post group joining behavior relative to a control group of consumers. Our results indicated the presence of a significant effect of group formation on website stickiness. Introduction of the group function in the website improved the stickiness effect so that networking activity was increased after network users joining groups compared to right before. The results also suggested the co-evolution between network growth and sub-group formations. Group users were likely to make connections with members who recently registered the network after they joined groups. Thus, we concluded that group users seemed to have the ability to attract new members from outside the network. In addition, we also investigated effects of specific user characteristics and group characteristics. Specifically, some group topics were found to facilitate users' engagement and attract new users. This result suggested that by creating and managing groups of certain topics, website managers might be able to optimize the attractiveness of the website which appeals to advertisers and marketers.

In summary, my dissertation can managerially contribute to forecasting sales of new products and evaluating effectiveness of traditional marketing communication. In my proposed models, CGM or new media serve as additional information to estimate market outcomes. Thus, they have the potential to make sales forecasts more precise than current methods. In addition, our research also quantifies how online communication is affected by traditional media. Our results are, therefore, likely to help managers to optimize the allocation of traditional marketing activity and CGM to maximize their market outcomes.