EMPIRICAL ADVANCES FOR THE STUDY OF WEBLOGS: RELEVANCE AND TESTING OF RANDOM EFFECTS MODELS

Kai-Lung Hui
Department of Information Systems
City University of Hong Kong
hklung@cityu.edu.hk

Yee-Lin Lai and Shun-Jian Yee
Department of Information Systems
National University of Singapore
laiyeeli@comp.nus.edu.sg

Last revised: January 18, 2007 (edited by Rob Kauffman and Paul Tallon)

Note: Forthcoming in Robert J. Kauffman and Paul P. Tallon (Eds.), *Economics, Information Systems and Electronic Commerce: Empirical Advances*, M. E. Sharpe, Armonk, NY, 2007, in press.

ABSTRACT

The global blogging phenomenon has caught researchers by surprise. While past research has focused on the motivations behind blogging, since the number of weblogs is growing at such an unprecedented rate, the roles played by their visitors should not be overlooked. This chapter presents an empirical analysis of weblogs from a methodological perspective that demonstrates empirical advances in this research area. Owing to the large number of weblogs in our sample and the persistent use of certain technological features in some of these weblogs, accounting for individual blog-specific effects through traditional dummy-variable regression is infeasible. We introduce the use of random effects in empirical weblog research. A random effects model allows researchers to encapsulate blog-specific effects in a regression without significant degree of freedom losses, as in typical dummy variable regressions. The modeling approach also allows the analyst to confront issues and problems with data singularity, which happens because of the linear dependency between blog-specific dummy variables and the persistent use of technological features. By developing and testing a random effects model that accounts for mutual influences between blog visits and comments, we show that blogger-visitor interactions play an important role in affecting blog popularity. Further, the use of technological features may not always raise the number of blog visits, and hence may not add value to bloggers in terms of increasing their blogs' popularity. We discuss some implications of these findings for research and practice.

Keywords: Blogs, blog entries, comments, dyadic relationships, economic analysis, empirical analysis, generalized two-stage least squares, instrumental variables, random effects model, technological features.

Acknowledgements. The authors would like to thank the two editors, and seminar participants at the City University of Hong Kong for their helpful comments and suggestions, and the National University of Singapore for providing the resources needed for this research.

1. INTRODUCTION

The development of Web technology has been booming in recent years. For individuals, there is now a wide spectrum of tools and utilities that they can conveniently use to share their personal thoughts and opinions, and to engage in discussion and collaboration [20]. Among these new tools and utilities, *web-logs* – often simply called *blogs* – are one of the most popular, arousing the interest of people with different backgrounds and demographics. By late 2006, Technorati.com (www.technorati.com), an online blog portal and search engine, was tracking more than 63.2 million blogs, and its statistics indicate that bloggers (people who "blog") post more than 1.6 million messages per day, or an average of 18 updates per second [30]. Clearly, blogging has spawned an unprecedented amount of online traffic and communication exchange.

The majority of blog research to date has focused almost exclusively on the motivations of bloggers, specifically, why would people want to share with others intimate details of their daily lives, personal thoughts, views, and opinions (e.g., [16, 23, 24, 31]). There is very little research, however, on the traffic generated by blogs, and more importantly, the extent of communication exchanges between bloggers and visitors, and factors that stimulate such communication exchanges. It is important to undertake this research for the following reasons. First, it helps us understand the nature of communication between mostly unrelated individuals in an online environment. In particular, recent research has found that the mere existence of audience reading blog entries could be a sufficient motivation for people to maintain a blog [22, 23]. In a blog context, such reading is mostly reflected in a blog's visit statistics and the number of comments that it receives from visitors. Hence, blog visits and comments appear to play an instrumental role in sustaining the ever-expanding *blogosphere*. A better understanding of how this has come about may provide useful and practical insights toward analyzing similar online communities.

Second, the growing popularity of blogs implies that substantial consumer attention is being devoted to them, insinuating that blogs could become an important avenue for advertising. Prior studies have found that the sales of some products spiked after they were mentioned in popular blogs [15]. Sony, as a result, has once paid up to $25,000 per month to sponsor a new blog that focused on electronic gadgets [2].[1] With the increasing popularity of online advertising and the emergence of new concepts such as sponsored links and cost-per-click or cost-per-impression pricing models, even casual bloggers can now earn advertising revenues by simply maintaining a blog. A prerequisite for this, however, is to build a

[1] During the dotcom boom, the values of many start-up companies were determined by the amount of visits that they were able to obtain. Evidently, in the online world traffic can sometimes be translated into a "currency" [7].

substantial, recurring readership and contributor base.[2] How to achieve this recurring stream of readers and contributors has been a recurring research question.

Third, the study of some blog features, such as the categorization of entries, the use of tag boxes, and the posting of a personal profile, allows us to assess the relevance of some possible motivations for people to visit and read blogs (e.g., information seeking, community exchanges, etc.). These assessments could help individuals and enterprise decision-makers determine if a blog could be the right tool for them in facilitating structured information dissemination, the collection of feedback, or collaborative exchanges – or simply, in organizing social functions and activities.

Finally, since the majority of blogs are maintained as venues for expressing personal opinions, visitors will be interacting directly with the bloggers, as opposed to the case of business and firm-centric e-commerce Web sites wherein they interact with a "system," and hence the interaction is less personal. It is interesting to see if the use of the technological features of blogs plays a significant role in affecting visits and communications in the more personal blog context. Further, because of the direct interaction with other people and the public nature of blogs, it is possible for people to give more attention to a popular blog because of its wide readership. In other words, there may be different social network effects in blogs of different popularity, and this will be reflected by the blogs' visit statistics and especially the number of comments that they are able to gather. Hence, it is important to study the relationship between blog visits and comments.

Given the above motivations and our observation that some popular blogs are getting thousands of visits per day, while many others suffer from poor hits and readership, we pose our research questions as follows:

> Are visits to individual blogs influenced by social and technological factors associated with the blogs, such as the number of comments made by visitors, the recency of blog updates, or the provision of content management and navigation tools?
>
> Do people make more comments, and hence, engage in more communication exchanges, in blogs that are more widely visited or that post more entries? Does the number of comments vary with blog entry characteristics?

We conducted an empirical study to address these research questions, and in the process, demonstrate the kinds of empirical advances that can be made in the presence of such weblog data. Specifically, we randomly sampled 100 blogs and recorded their key features, visit rates, and the extent of communication exchanges exhibited in them (in terms of the number of entries made by bloggers and comments made by

[2] Although many blogs are able to generate good publicity and visits, a much larger number of blogs suffer from poor readership and are abandoned because visitors are overloaded by information or simply are not aware of their existence. In general, the popularity of blogs may follow a power law distribution [28], and a *long tail* [4, 8] may well exist in blog traffic.

visitors) over a period of seven days in early 2006. Then, with this set of panel data, we constructed an econometric model to study the dynamics of blog visits and communication exchanges. Our model comprises two endogenous variables, one pertaining to *visit rate*,[3] and the other to the *number of new comments* written by visitors. The latter reflects the extent of communication exchanges exhibited in the blogs under consideration. We will formalize the definitions of both later in this chapter.

In specifying our econometric model, it is important to take account of time-invariant, blog-specific heterogeneity. Such heterogeneity could arise because of the quality of writing, being listed on different search engines or portals, consistent choices of topics by bloggers, etc. A simple and often-employed way to capture such heterogeneity in traditional panel data analysis is to incorporate, in our context, blog-specific dummy variables (often called *fixed effects* to reflect their fixed, time-invariant nature). These dummy variables would extract the variations in the dependent variables due to unobservable blog characteristics, and by including them the effects of other independent variables (often the ones that are of key research interest) can be estimated with better accuracy and precision.

However, the dummy variable approach is costly in terms of degrees of freedom, especially when the number of blogs is large relative to the number of time periods. This is likely to be the case in large-scale empirical blog research. In our case, the sample comprised 100 blogs but only seven time periods. Further, because a subset of the studied blogs may persistently use some technological attributes, there may be linear dependencies in the blog-specific dummy variables and the technological attributes, which prevents estimation of separate effects for the latter.

Hence, instead of using blog-specific dummy variables, we introduce the use of *random effects* in our empirical analysis. The random effects specification assumes that blog-specific heterogeneity follows a statistical distribution, which is reasonable given that the set of blogs that we studied were drawn from a population with a much larger size.[4] Because this approach does not parameterize individual blog effects, it conserves degrees of freedom and, at the same time, allows for estimating the effects of the full set of independent variables, even when the values of a subset of these variables did not vary within some of the sampled blogs during the studied time window.

There was one final estimation challenge that we needed to tackle. Although more comments on blog entries may enrich a discussion and hence lead more people to visit a blog, the number of new comments

[3] We will not distinguish *new visits* and *repeated visits*, as most online tracking sites do not report separate visit and revisit statistics. Hereafter we use the term *visit* to refer to both visits and revisits. We used *visit rate*, which is defined as the number of new visits to a blog divided by its average number of daily visits, as the dependent variable, because visit rate is standardized. The number of new visits may consistently vary across blogs because of unobservable factors (e.g., some blogs may have been listed on more search engines), and hence it is not suitable for our purpose.

[4] This is likely to be the case in future empirical blog research as well. Given the millions of blogs on the Internet, random or some other forms of sampling from the entire blog population is unavoidable.

was correlated with number of visits, as it was likely that more visitors would have written more comments in a blog too. In other words, there was an endogeneity problem in our setting. It is well known that with endogeneity, ordinary least squares (OLS) or generalized least squares (GLS) estimations would produce biased parameter estimates [14].

Given that we simultaneously collected visit and new comment data on the blogs that we studied, and the fact that we were not able to manipulate the number of new comments written in these blogs, the most appropriate solution to this endogeneity problem was to conduct an *instrumental variable regression* [5]. We compiled a set of instrumental variables for the "*Number of New Comments*" variable, and used them in a *generalized two-stage least squares* (G2SLS) procedure to estimate our model and identify its parameters. (See [6].) This gave us information on the effects of the blogs and their technological attributes on visit rates and the number of new comments. We used the two-stage least squares procedure to address the endogeneity issue, while the estimation was "generalized" to incorporate the random effects.

Our results show that blog visit rates and the number of new comments are complementary. There were more comments in blogs that were more widely visited, and these new comments induced more people to visit the blogs. This may have occurred because the comments enriched the overall content contained in the blogs, or because people were interested in following interactive and recurrent blog discussions. These results suggest that blogs could form a new type of online community that features a closely-knit one-to-many dyadic relationship. Further, we found that blogs that employed content management and navigation tools had higher visit rates. This is consistent with the view that some people visit blogs to seek targeted information.

The remainder of this chapter is organized as follows. Section 2 presents a brief history of weblogs and a summary of prior research. Section 3 describes our research model. Section 4 reviews variable definitions and data collection procedures, leading up to Section 5, which presents our main results. Section 6 discusses the implications of our research findings and, finally, Section 7 concludes the chapter with the contributions and limitations of this research.

2. BACKGROUND

A *blog* consists of multiple entries that are typically organized in reverse chronological order [7]. Blogs differ from general Internet Web sites: they are frequently updated and interactive. Visitors or readers of blogs are allowed to leave comments on any particular entries, and bloggers often respond in turn to these comments. Blogs also differ from online communities such as discussion forums in that their exposition and content are primarily governed by the people who contribute content to them – the bloggers. Unlike online communities which empower all of their participants and visitors with almost equal ability to contribute to a community's contents or choices of specific topics, a blog's content is generally con-

structed by the blogger herself. Hence, the relationship between bloggers and visitors is highly asymmetrical. Bloggers determine the topics that they want to write on, and it is then up to visitors to determine whether to respond to an entry. Blog entries are often interconnected to each other. Also, bloggers tend to exhibit intimate self expressions repeatedly over time [16, 31] including penning chronicles of their private lives and exhibiting personal self-portraits, which is not common in online discussion forums.

Figures 1 and 2 show some examples of blog entries. Some of these blog entries were merely exhibits of the bloggers' thoughts, or pictures of their own or their pets (as in Figure 1), whereas others were more informative and contained knowledge on a special topic (as in Figure 2).

Most blogs provide a few common tools or possess some common features, including the following:

Comment tools allow visitors to respond to a blog entry. This facilitates interaction between bloggers and visitors.

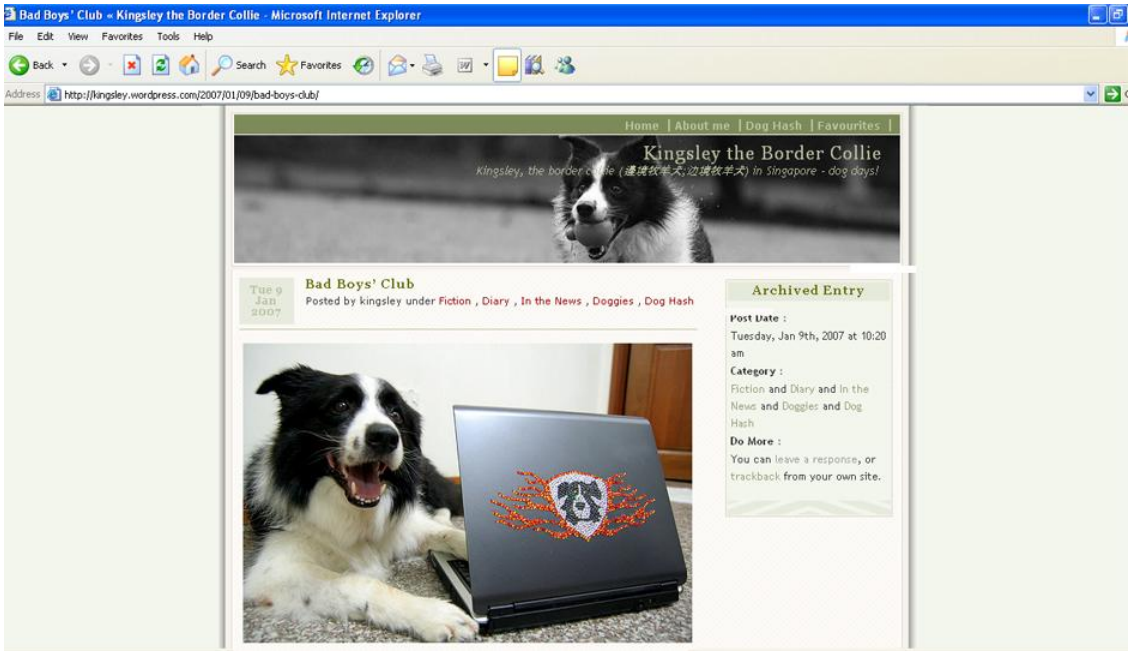
Trackback capabilities inform bloggers about who has referred to their original entries.

Archives allow visitors to conveniently browse old entries.

Personal profiles let visitors view personal data (e.g., age and gender) provided by bloggers.

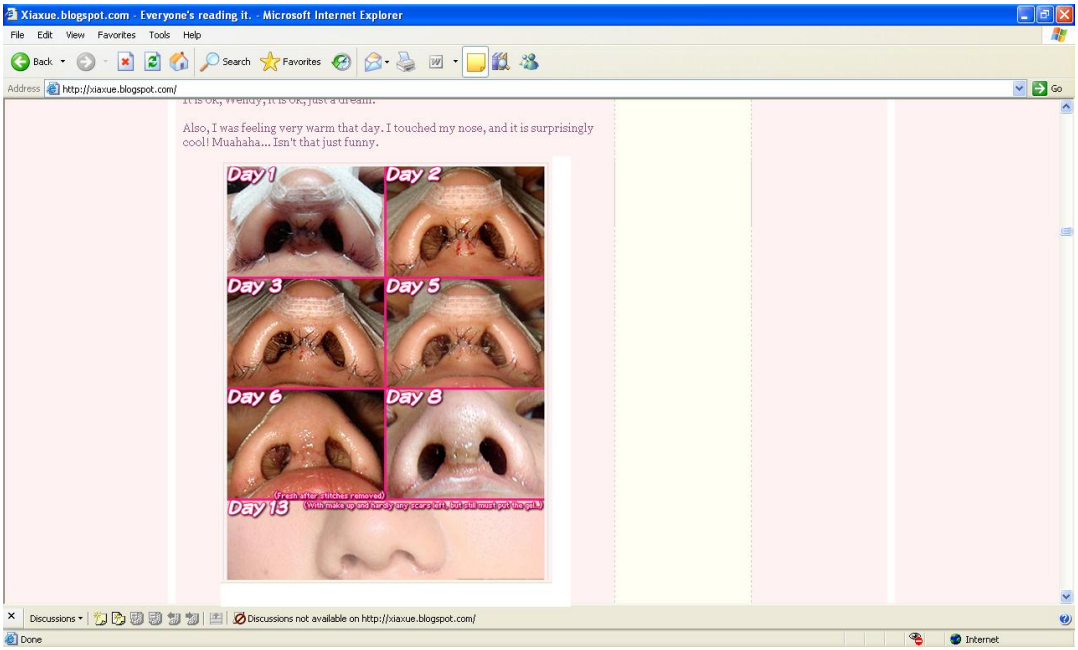
Figure 3 shows a blog that contains various links to some of these features. Most of these features are directly accessible from the home page of the blog with just one mouse click.

Figure 1. Every Dog Has Its Day



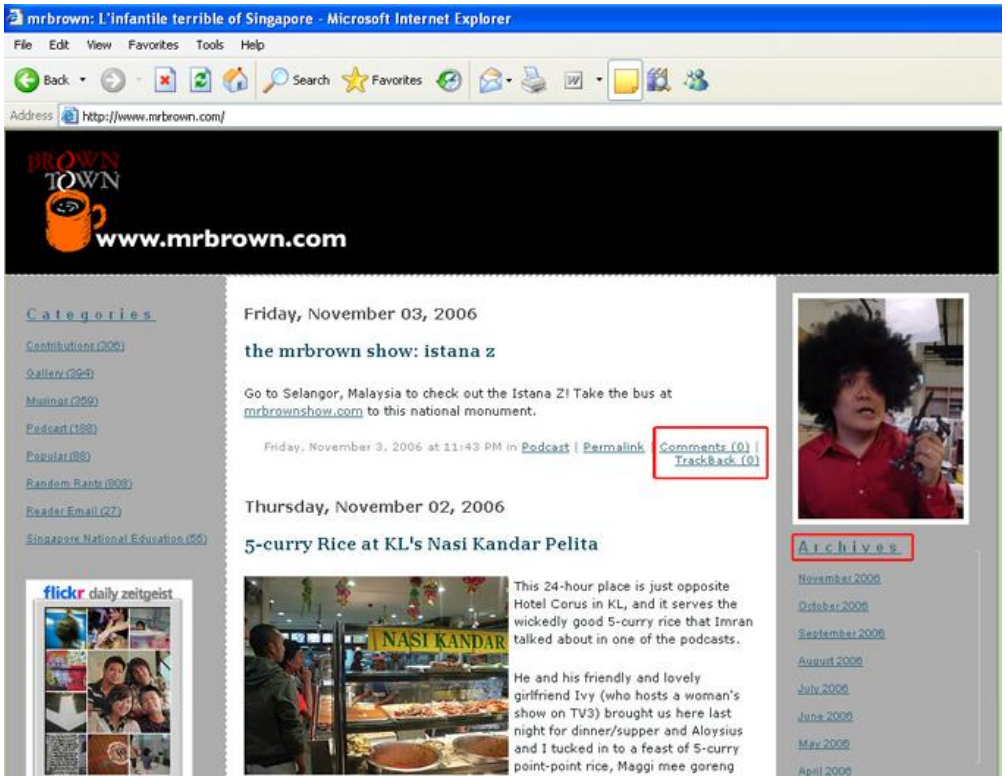
Note: Even canines are keeping up with technology, utilizing blogs as a tool to document their daily adventures and training routines.

Figure 2. No Secrets on the Web?



Note: Plastic surgery, which is conventionally kept surreptitious, is openly disclosed and detailed by this blogger who underwent a rhinoplasty or "nose job" operation.

Figure 3. Components of a Typical Blog



Note: Rectangles added by the authors for emphasis.

Prior research has suggested that the number of unique inbound links from other Web sites is a good predictor of blog traffic [17, 29]. Other than inbound links, however, little research has focused on whether blog features affect popularity in terms of *actual visits*. Based on Technorati.com's ranking, Du and Wagner [12] consider how blogging technologies are used. They considered blogging technologies ranging from basic content editing and linking tools to enhanced integrated applications for social interactions, such as workflow or project management tools. They specifically evaluated how the use of these capabilities affected the popularity of top "A-list" blogs. Technorati.com's ranking, however, was based on the number of inbound links from other *Web sites* rather than the actual number of visits by *people*, which is a more direct measure of popularity. Also, they only studied the use of blogging technologies and tools but did not consider how the dynamics of blogger-visitor communication exchanges might affect blog popularity.

The prior literature on system and Web site quality has focused on usability and design metrics [9, 11, 21, 27]. The suggestions that emerge from these studies, however, pertain more to contexts in which end users or visitors to a Web site have a specific goal in mind. This is often not the case for a user's visit to a blog. Studies on human behavior and technology acceptance have advocated behavioral controls, and ease of use and usefulness of technologies [3, 10]. By contrast, marketing research on "flow" and online user experience is useful, since both are inevitably important in affecting a person's intent to read a blog. The prior research in this area has emphasized factors such as users' skills, motivations, and arousals when browsing a Web site to a greater degree than other issues [18, 26]. Clearly, in a blogging context – namely, a highly personal, selective, and casual environment that enables peer-to-peer communication exchanges – it is useful to consider if design features, including those listed above and other less common ones, affect actual visits.

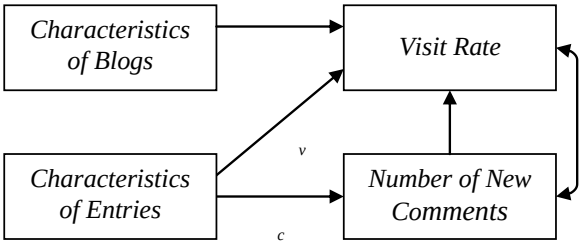
More importantly, previous blog research has mostly ignored the social communications and exchanges exhibited in blogs. As suggested in prior research, one possible motivation for blogging is to seek others' feedback or opinion [22, 23], which may lead to future blogger-visitor communications. It is important to study how visitors respond to blog entries, and if their responses and interactions thereafter could trigger more follow-on blog visits.

3. THE MODEL

The characteristics of blogs, such as their use of content management and navigation tools, may increase the usability of blogs and hence reduce visitors' efforts in navigating and reading their entries [12, 21, 27]. Further, the nature of blog entries, such as their length, their use of multimedia, or their hyperlinks to other Web sites, may affect the content quality of the blogs and lead to different visit rates over time [19]. Hence, we posit that the visit rate of a blog may depend on two groups of factors – *persistent*

blog characteristics (e.g., use of tag boxes or navigation bars) and *dynamic blog entry characteristics* (e.g., number of words or images included in new entries). This leads us to propose a model for the empirical research that we are conducting on blogs (see Figure 4, and in particular the arrows leading to *Visit Rate*).

Figure 4. Research Model



On the other hand, once a visitor has read the blog entries, she may decide to follow up a topic by posting her comments. In general, we expect a visitor's tendency to post comments to be correlated with blog entry characteristics (e.g., a longer entry or one that contains more images may provide more bases for blogger-visitor exchanges and discussions). These comments, in turn, are likely to expand the richness of the blog, thus increasing the likelihood that others will visit the blog also. This relationship is also depicted in Figure 4 (see the arrows related to the *Number of New Comments* variable).

Equations 1 and 2 below present the relationships depicted in Figure 4 in empirical form:

$$v_{it} \quad _i \quad x_{it} \quad _v \quad z_i \quad \quad c_{it} \quad d_t \quad _v \quad \quad _{it}, \tag{1}$$

$$c_{it} \quad _i \quad x_{it} \quad _c \quad \quad v_{it} \quad d_t \quad _c \quad \quad _{it}, \tag{2}$$

where v_{it} denotes the *Visit Rate* to blog *i* in day *t*, c_{it} denotes the *Number of New Comments*, x_{it} is a vector of *temporal new blog entry characteristics*, and z_i is a vector of *persistent blog characteristics* (including technological attributes, such as inclusion of search engines and tag boxes, and other features such as inclusion of advertisements, recency of update, etc.). As we discussed in the previous section, the *Number of New Comments* and *Visit Rate* are inherently correlated (as more visitors will generate more comments), and hence we added an extra parameter in Equation 2 to extract such *a priori* correlations. Overall, $_v$, , , $_c$, and are vectors of parameters, and $_v$ and $_c$ are vectors of period-specific fixed effects that will be estimated from the data; $_i$ and $_i$ are blog-specific effects, d_t is a vector of period-specific dummy variables, and $_{it}$ and $_{it}$ are random error terms.

Realistically, the number of visits to a blog may be serially correlated. A blog may slowly build up its visitor base over time, and hence the number of visits today may be a function of the number of visits in the past. Our use of *Visit Rate* – defined as the number of new visits to a blog divided by its av-

erage number of daily visits – as the dependent variable effectively addresses such serial correlations. This step standardizes the *Visit Rate* variable. It essentially captures the percentage rather than the nominal growth in visits, and should be free from serial correlation since there is no *a priori* reason to expect the percentage change in visits to be correlated over time.

As also mentioned in the previous section, there could be unobserved time-invariant heterogeneity in blogs. Some bloggers may write better than others, some may get more comments because of their choice of topics or stories, or their general style of writing, etc. It is important to control for such time-invariant heterogeneity in our estimation. To do that, we incorporated a blog-specific random effect in each of the two empirical model equations. Random effects could effectively capture the blog-specific heterogeneity without incurring heavy losses in degrees of freedom. Further, it is common for blogs to use some technological attributes continuously over time. It is infeasible to estimate blog-specific dummy variables and the effects due to these technological attributes separately. The use of random effects avoids this identification problem. Specifically, assuming random blog-specific effects, $_i$ and $_i$ can be decomposed as:

$$_i \quad u_i, \tag{3}$$

$$_i \quad v_i. \tag{4}$$

In Equations 3 and 4, and are then the overall constants from Equations 1 and 2, and u_i and v_i capture random heterogeneity specific to blog *i*. By applying standard regression assumptions on u_i and v_i (zero mean, constant variance, etc.), the parameters in Equations 1 and 2 can be efficiently estimated by *feasible generalized least squares* (FGLS) [14].[5]

There is one final caveat though. We have posited that the *Number of New Comments* may have increased the observed *Visit Rate*. These two variables may also correlate simply because a larger number of visitors will likely have posted more comments; in other words, there is a natural and positive association between these two variables. From our data it was impossible to distinguish which additional visitor had written a new comment, and which visits were attracted by the new comments. Therefore, the *Number of New Comments* was endogenous in Equation 1, and so it is important to account for such endogeneity in our estimation.

[5] FGLS is a standard econometric procedure to estimate the parameters with an unknown covariance matrix. In the case of a random effects model, specifically, Equations 3 and 4, the assumption of constant, non-zero variances for u_i and v_i would imply that OLS regressions are inefficient (i.e., the estimated parameters would be less precise if we do not account for these variances). FGLS uses the OLS estimator to compute an unbiased estimator of the covariance matrix of the parameters. This covariance matrix estimator is then entered, in a standard weighted-least squares fashion, into the subsequent estimations to obtain efficient estimates of the parameters of interest (i.e., the effects of the independent variables). See Section 5 below for more details on estimating the random effects model. Up to this point we have not considered the endogeneity of visit rate and the number of new comments. We will come back to that later.

The standard solution to the endogeneity problem is to perform an instrumental variable regression [5]. In our case, for Equation 1, what we need for this are variables that were highly correlated with the *Number of New Comments* (especially on its effect of attracting new visitors), and that have little or no correlation with the parts of *Visit Rate* that were not supposed to be influenced by the *Number of New Comments* (including the inherent positive association between *Visit Rate* and the *Number of New Comments*) [14].

We compiled a set of data on the *Average Time Spent per Visit* and the *Total Pages Viewed* by all visitors during the new visits, and used them as instrumental variables for c_{it}. These two variables were likely to be correlated with the *Number of New Comments*, since all of these three variables were related to the visitors' blog reading or browsing experience. *Time Spent per Visit* and *Total Pages Viewed* should not be related to *Visit Rate*, however, because they captured the depth rather than breadth of blog visits. Hence, they satisfy the conditions for suitable instruments.[6]

4. DATA

We randomly sampled 100 blogs from blo.gs and tracked their visit and comment statistics for seven days in February 2006.[7] The visit statistics were compiled from Site Meter (www.sitemeter.com), which provides the free Site Meter counter to track the visit statistics of a Web page. This tool has been recommended and used in a number of previous blog studies (e.g., [1, 13, 17]). It provides a simple and standardized way to measure blog visits.

Specifically, the 100 blogs that we sampled fulfilled the following two criteria:

They had publicly-accessible visit statistics from Site Meter.

They were not spam blogs that contain no meaningful content but rather an unusually large number of links for the purpose of distorting search engine results;

To focus the analysis on reasonably active blogs, we further restricted our sample to blogs that had at least two entries, and that were updated at least once in February 2006, prior to the data collection. These restrictions eliminated neophyte bloggers or dead blogs that could be expected to get few visits and hence were not of much interest to us. We also removed audio blogs because the nature of interaction in these blogs could be quite different from text blogs which are the dominant type of blog on the Internet.

We collected the data for this research in early February 2006. For each blog, we recorded the following dynamic data on a daily basis for seven consecutive days: cumulative number of entries up to the previous day, number of new entries and comments, visit rate, and total number of images, links, and

[6] For a more detailed discussion on the choices of instrumental variables, the interested reader should see [5].

[7] blo.gs (www.blo.gs) is one of the largest portals that tracks active blogs from different sources. At the time when we composed our sample, blo.gs was tracking close to 68 million blogs, compared with Technorati.com's 30.4 million.

words in the new entries. We also collected some blog characteristics: demographics, inclusion of blogging tools and technologies (e.g., tag boxes and search engines) in the blog sites, and inclusion of advertisements, etc. These characteristics were persistent and did not vary over the period in which the data were collected. All blog and entry characteristics were recorded by directly traversing the blog sites. Table 1 presents definitions of all variables that are included in this study.

INSERT TABLE 1 ABOUT HERE

Generally, the demographics of the blogs that we sampled resemble those studied in previous research (e.g., [16, 32]). Table 2 presents demographic statistics of the blogs used for this study. There were more female bloggers in our sample than male, and the bloggers were primarily adults, with a minimum age of 18. The majority of the blogs were located in the Unites States

Table 2. Blog Demographics

DEMOGRAPHIC VARIABLE	FREQUENCY[+]	PERCENTAGE
Gender (1 = Female; 0 = Male)	55 (n = 93)	59.14
Located in the USA	63 (n = 96)	65.63
Blogger age	Average = 29.64 St. dev. = 7.85	Max = 47 Min = 18

Note: [+] Some blogs do not publish these data, and hence the total number, *n*, varies.

Table 3 presents descriptive statistics of the variables that were recorded daily during the period of study.

Table 3. Descriptive Statistics of the Daily Measures

VARIABLE	MEAN	ST. DEV.	MAX	MIN
Visit Rate	2.69	3.05	24.42	0.09
Number of New Comments	7.27	11.68	97	0
Cumulative Number of Entries	3.67	3.60	29	1
Number of New Entries	0.89	1.19	8	0
Number of New Images	0.63	1.63	14	0
Number of New Links	1.78	4.37	37	0
Number of Words	223.24	383.75	4,458	0
Number of New Visits	704.79	2,615.74	34,475	2
Average Visit Length (*seconds*)	99.76	65.11	356	2
Total Page Views	1,170.18	4,315.96	47,979	2

Table 1. Definitions of Variables

Dependent Variables	
Visit Rate[+]	Total number of new visits to the blog (i.e., visits recorded since the previous day) divided by its average number of daily visits. As defined by sitemeter.com, a "visit" is a series of page views by a person with no more than 30 minutes in between the page views. If a visitor clicks on a link to another site but goes back to the blog within 30 minutes, then she would be considered as remaining in the same visit.
Number of New Comments[+]	Total number of new comments written based on all (old + new) entries since the previous day.
Independent Variables	
Cumulative Number of Entries[+]	Total number of entries in the blog in the previous day.
Number of New Entries[+]	The number of new entries added since the previous day.
Number of New Images[+]	Total number of images contained in the new entries.
Number of New Links[+]	Total number of links contained in the new entries.
Number of Words[+]	Total number of words in the new entries.
Number of New Visits[+]	The number of visits recorded since the previous day.
Blog Age	The age of the blog measured in months.
Recency of Update	Number of days before the first day of data collection when the blog was updated.
Group blog	Binary variable for blog maintained by a group or individual (1 = Group; 0 = Individual).
Top Navigation Bar	Binary variable for whether top navigation bar was used in the blog (1 = Y; 0 = N).
Bottom Navigation Bar	Binary variable for whether bottom navigation bar was used in the blog (1 = Y; 0 = N).
Side Bar	Binary variable for whether a side bar was used in the blog (1 = Y; 0 = N).
Profile	Binary variable for whether a profile of the blogger was posted (1 = Y; 0 = N).
Photo	Binary variable for whether a photo of the blogger was posted (1 = Y; 0 = N).
Email	Binary variable for whether an email address of the blogger was posted (1 = Y; 0 = N).
Archive	Binary variable for whether the blogger allowed visitors to view archives of entries (1 = Y; 0 = N).
Categorized Entries	Binary variable for whether blog entries were categorized by topic (1 = Y; 0 = N).
Tag Box	Binary variable for whether a tag box was used (1 = Y; 0 = N). *Tag box* is an interactive dialog box that allows bloggers and visitors to engage in message exchanges similar to those in online discussion forums.
Trackback	Binary variable indicating whether the *Trackback* feature was enabled (1 = Y; 0 = N). *Trackback* allows a person to keep track of who has referred to an entry.
External Links	Binary variable for whether external links were placed on the blog site, but not within blog entries (e.g., on the top or bottom navigation bars) (1 = Y; 0 = N).
Link to Personal Photo Gallery	Binary variable for whether a link to the blogger's personal photo gallery was provided (1 = Y; 0 = N).
Number of Ads	Number of advertisements placed on the blog site, but not within the blog entries (e.g., on the top or bottom navigation bars).
Search Engine	Binary variable for whether a search engine was used in the blog (1 = Y; 0 = N).
Instrumental Variables	
Average Visit Length[+]	Average time (in seconds) visitors spent reading the blog. *Site Meter* was able to track visit length only when visitors viewed at least two pages in the blog. It is difficult to capture time spent on a Web site if a person only views a page and then leaves.
Total Page View[+]	Number of pages viewed by all visitors during the new visits. The count was increased by one every time a visitor followed a link on the page.
Note: [+] Variables that were measured daily for seven days during the data collection period.	

The statistics in Table 3 indicate that the blogs that we studied were active. On average, the bloggers posted close to one entry per day, and the entries were reasonably long (with more than 200 words on average). The entries also stimulated quite a lot of comments, and the high value of *Visit Rate* suggests that the blogs were getting popular over time. Table 4 presents descriptive statistics of the persistent blog characteristics.

Table 4. Descriptive Statistics of Persistent Blog Characteristics

VARIABLE	MEAN	ST. DEV.	MAX	MIN
Blog Age	21.50	16.31	86	1
Recency of Update	1.89	1.98	10	0
Number of Ads	0.92	1.30	5	0
	Frequency (n = 100)			
Group Blog	7			
Top Navigation Bar	15			
Bottom Navigation Bar	33			
Side Bar	98			
Profile	85			
Photo	58			
Email	81			
Archive	98			
Categorized Entries	45			
Tag Box	9			
Trackback	32			
External Links	96			
Link to Personal Photo Gallery	37			
Search Engine	40			

From Table 4, it is clear that several features were quite commonly used in blogs, such as a *Side Bar*, posting of personal *Profiles* and *Email* addresses, *Archives* of entries, and *External Links*. By contrast, the use of a *Top Navigation Bar* and *Tag Boxes* was not popular. The blogs that we sampled were reasonably mature, with an average age of close to two years, and most had only one identifiable author.

5. DATA ANALYSIS

We performed *generalized two-stage least squares* (G2SLS) estimations to estimate Equations 1 and 2 [6]. In general, the idea is to perform two sets of OLS regressions, one on the pooled data with an overall constant, and the other on mean-adjusted data, and then use the residuals from these regressions to compute estimates for the error variances (i.e., the variances of $_{it}$, $_{it}$, u_i, and v_i). These error variance estimates can then be used to transform the variables in a standard weighted-least squares fashion [14]. The transformed variables would account for the error components in the random effects models. They

can then be used directly in a *two-stage least squares* (2SLS) procedure to obtain consistent parameter estimates.[8]

Table 5 presents the G2SLS estimation results of Equation 1, and the GLS estimation results of Equation 2. For comparison, we also report the GLS estimates of the simple random effects model for Equation 1 (i.e., without instrumenting for the number of new comments).

We first conducted a Lagrange multiplier test to see if the random effects specification fits the data that we collected. According to [14], the test statistic is computed as:

$$\text{LM} \quad \frac{nT}{2(T \quad 1)} \quad \frac{\sum_{i \ 1}^{n} (T\bar{e}_i)^2}{\sum_{i \ 1}^{n} \sum_{t \ 1}^{T} e_{it}^2} \quad 1 \quad ^2 ,$$

where e_{it} are the residuals obtained from an OLS regression with only one overall constant, and $\bar{e}_i$ are the means of the OLS residuals that can be computed from e_{it}. In our case, LM equals 12.27 for the *Visit Rate* Equation 1. This exceeds the 99% critical value of a ^{2} distribution with one degree of freedom, 6.63. Hence, we conclude that the random effects specification is appropriate for the *Visit Rate* equation.[9]

On the other hand, LM equals 0.57 for the *Number of New Comments* Equation 2, which is not statistically significant (p = 0.45). Apparently, allowing for random effects did not significantly improve the fit of the *Number of New Comments* equation. However, given the large number of blogs in this study, it is reasonable to free up the constant of the equation. Thus, we will continue to focus on the results with random effects for the *Number of New Comments*.[10]

Comparing the GLS and G2SLS estimation results for the *Visit Rate* equation, it is clear that after instrumenting for the *Number of New Comments*, all of the coefficients changed considerably, and the coefficient for the *Number of New Comments* almost doubled in G2SLS. This suggests that failing to account for endogeneity could lead to misleading inferences on the effect of comments on blog visits. In the remaining discussion we will focus on the G2SLS results.

The G2SLS results show that the *Number of New Images*, *Blog Age*, use of *Bottom Navigation Bars*,

---

[8] The 2SLS procedure, often called *instrumental variable estimation*, proceeds as follows. First, in the "first stage," the endogenous variable in question (in our case, the *Number of New Comments*) is regressed on all instrumental and other independent variables. The predicted values in this regression are computed, which are then used in the "second stage" (i.e., the original full regression) to replace the original endogenous variable in question. Essentially, the first stage serves to purify and remove unwanted variations in the problematic variable.

[9] In our case it was impossible to estimate a fixed effects model, because some of the blog-specific dummy variables were correlated with the feature variables. The data matrix with blog-specific dummy variables was singular.

[10] Estimation using an overall constant in Equation 2 produced similar results as those in Table 5, except that there was a decrease in the significance of the blog entry characteristics.

provision of *Archives* and *Categorized Entries*, and the *Number of New Comments* had a positive impact on *Visit Rate*. On the other hand, the *Number of Links in New Entries*, use of *Side Bars*, publishing of personal *Profiles* and *External Links*, and *Links to Personal Photo Galleries*, all had a negative impact on *Visit Rate*. The adjusted R^2 was 40.93%, which suggests that our model was able to explain a reasonable amount of the variance in *Visit Rate*.

For the *Number of New Comments* equation, in general, the characteristics of blog entries affected the extent to which visitors submitted comments. The *Number of New Entries* and the *Number of Links* and *Number of Words* in the new entries had a positive effect on the *Number of New Comments* written by visitors. Also, *Visit Rate* was positively correlated with the *Number of New Comments*. The adjusted R^2 was 32.24%.

6. DISCUSSION AND IMPLICATIONS

There are several interesting results in Table 5. First, the visit rate of blogs increased with the number of new comments posted by visitors. This implies that new visitors are interested in the comments written by other people, and blogs may possibly help to facilitate social communications, which may be an important motivation behind the blogging phenomenon as previously noted by Nardi et al. [23, 24]. On the other hand, the increased visits were also coupled with more new comments. We found a systematic complementarity between blog visits and comments, which implies that although blogging is primarily meant for asymmetric sharing of personal stories or opinions, it could also form a new type of online community in which discussion and interaction is anchored on blogger-driven topics.[11] Many blog hosts now allow bloggers to create "friends lists". These "friends" often do not know each other before joining the blogs, but they interact frequently by making comments and responses, which over time may attract other people to visit their blogs and expand their social network.

Clearly, the presence of an audience and feedback could be a key to success of Internet Web sites, including those that aim to share information. Our findings also imply that, similar to personal blogs, it is possible for new knowledge or enterprise systems to enjoy greater success or more widespread adoption if users are encouraged to provide feedback or responses. Feedback or responses may make a new system more interesting and "lively," which may possibly increase the interest of people to "take a look" at the system and possibly post an opinion.

Second, the results in Table 5 show that blogs that were older tended to have a higher visit rate too. This implies that older blogs receive more visits, which could have been due to improved search engine

[11] The possibility for blogs to form such an online community has been recognized by some news agencies. For example, MSNBC.com has used blogs to stimulate interesting debates on stem cell research, the death penalty, democratic nominees for the 2008 U.S. presidential election, etc. The authors thank the editors for making this observation.

indexing over time, the loyalty of visitors, or diffusion through word of mouth since older blogs have a larger base of visitors. In any case, older or more well-established blogs may be a better avenue for companies to reach out to a wider group of audience, or to attract "eyeballs" for their products.

Third, the inclusion of archives and categorization of entries increased visit rate. This finding is consistent with the proposition that some people read blogs to seek specific types of information (e.g., discussion on a social topic, stock market news, etc.), and hence blogs can potentially be used as a means for knowledge sharing and dissemination. Alternatively, since archiving and entry categorization help organize the contents of a blog and enhance the ease of navigation for visitors, it is also possible that such improved usability attracted more people to visit the blogs [25].

Fourth, the use of blogging tools and technologies, and navigation tools had mixed influences on visit rate. Although a bottom navigation bar increased visit rate, a side bar decreased it. One possible reason is that a side bar reduces the available space to display the blog contents, and hence may annoy visitors who want to focus on reading the entries.

The inclusion of links, including *persistent links* (i.e., links that are always displayed on the blog page), *ad hoc links* inside blog entries, or links to personal photo galleries, decreased visit rate. Links provide a convenient way for people to visit other pages, and hence may distract visitors' attention. Although people may visit a blog with the intention of reading its contents, they may not necessarily stay on that site and could easily click away from the blog site to someplace else. Bloggers or companies that sponsor blog advertisements may want to evaluate the inclusion of links carefully if traffic is deemed important to them.

Surprisingly, the posting of a personal profile reduced visit rate. This could perhaps be due to a novelty effect: some people may browse blogs just to learn about other people or to make new friends, and the posting of a personal profile reduces the number of visits needed to achieve such purposes.

Finally, new entries, especially longer ones containing more links, caused visitors to write more comments, but they did not directly lead to more visits. It may be possible for bloggers to stimulate more feedback by enhancing the content or media richness of their entries. By doing this, more people may submit comments, which then indirectly raises the visit rate.[12]

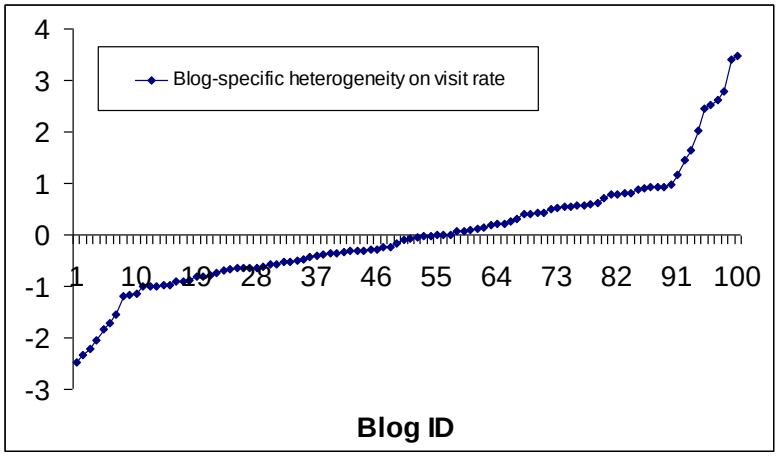
Overall, our findings show that some people may read blogs to seek targeted information, but there are also people who are active in providing comments, which could be one key factor in dictating the success of a blog in terms of gathering more visits and gaining popularity. The use of blogging tools and

[12] We also regressed (by G2SLS with random effects) the *Number of New Entries* on the *Number of New Comments* and the *Cumulative Number of Visits*, and we found both coefficients to be positive and significant ($p < 0.01$). These results are consistent with previous research: one important motivation to blog is to seek feedback and opinions from others [22, 23]. We do not report the detailed statistical results for these findings to maintain brevity in this chapter.

technologies could matter, but only certain tools or features are relevant – we found no evidence that use of top navigation bars, tag boxes and search engine, and inclusion of trackback features, affected blog visits. Most of the features that we now see on blogs are invented to empower people to track a discussion or locate some specific content, but apparently they are not equally useful from all visitors' viewpoints. Some technology features (e.g., side bars) may in fact reduce individual's desire to visit a blog. Obviously, bloggers or blog hosts need to exercise judgment in selecting the right mix of blog site features [12].

Methodologically, the use of a random effects model allowed us to account for blog-specific heterogeneity that persists over time. More importantly, we were able to estimate the effects due to the blogging tools and technologies which otherwise would have been infeasible if we were to employ dummy-variable regressions. The random effects that we estimated from the visit rate equation are plotted in Figure 5. Evidently, the blogs that we studied differed quite significantly in terms of their ability to generate visits.

Figure 5. The Random Effects Computed from the *Visit Rate* Equation



The Lagrange Multiplier test in Section 5 confirmed that, in the case of *Visit Rate*, the random effects specification fits the data better than a simple OLS model with one overall constant. A quick comparison with the one-constant OLS model (the results of which are not reported here) showed that the estimated parameters in a random effects model could differ considerably in magnitude. Hence, unless the sample size is very large (in which case the point about estimator efficiency is moot), the use of random effects is recommended in future empirical research related to blogs. Indeed, the use of random effects models could well be the only feasible solution if some other model variables are correlated with blog-specific dummy variables.

Also, the results in Table 5 show that failing to account for endogeneity could lead to serious underestimation of the effect of the number of new comments on visit rate. So it is important for future blog re-

search to develop a more in-depth understanding of the dynamics of the interactions between bloggers and visitors, and to use such knowledge to guide effective empirical estimation.

Table 5. Estimation Results

VARIABLE	VISIT RATE[++]		NUMBER OF NEW COMMENTS
	GLS	**G2SLS**	**GLS**
Constant	2.7824	1.8939	3.4090***
Cumulative Number of Entries	0.0181***	0.0132	0.0450
Number of New Entries	-0.0061	-0.1137	1.2445***
Number of New Images	0.0511	0.0836*	-0.1060
Number of New Links	-0.0340	-0.0503*	0.2521**
Number of Words	0.0004	-0.0000	0.0052***
Blog Age	0.0747***	0.0821***	
Recency of Update	-0.1133	-0.0490	
Number of Ads	-0.1411	-0.1654	
Group Blog	1.3845	1.1172	
Top Navigation Bar	0.6102	0.2860	
Bottom Navigation Bar	0.9643	0.8324**	
Side Bar	-1.3237	-1.7817*	
Profile	-0.8168*	-0.8505**	
Photo	0.4497	0.1294	
Email	0.1317	-0.1434	
Archive	1.2084	2.0434**	
Categorized Entries	0.4570*	0.8330***	
Tag Box	-0.7600**	-0.3685	
Trackback	0.1623	-0.1102	
External Links	-2.0768*	-1.6813*	
Link to Personal Photo Gallery	-0.5854	-0.7571**	
Search Engine	0.0058	0.0431	
Number of New Visits	---------	--------	0.0018***
Number of New Comments	0.0622***	0.1388***	
Adjusted R^2	0.4661	0.4093	0.3336

Notes. We do not report the period fixed effects and random blog-specific effects for brevity. The sample size of the data set was 700 (100 blogs 7 days). [++] We used *Average Visit Length* and *Total Page Views* as instruments for the *Number of New Comments* in the G2SLS estimation. Signif.: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

7. CONCLUSION

This chapter serves two purposes. First, we illustrate how econometric techniques can be applied in empirical information systems research to facilitate the analysis of massive amounts of data that comprise a large number of cross-sectional units and multiple time periods, and that cannot be systematically con-

trolled by the researchers.[13] Such data are becoming available because of increased information transparency on the Internet, faster network transmission, and advances in Web bot and intelligent agent technologies, all of which make large-scale data collection from external sources much easier and affordable to a researcher. Our second purpose is to explore the communication dynamics exhibited in blogs – one of the fastest growing phenomena on the Internet. We reiterate our specific contributions and discuss the limitations of this research below.

7.1. Contributions

Using a set of panel data collected from 100 blogs for seven days in February 2006, we studied the visit and interaction patterns between bloggers and visitors, as a means of illustrating the nature of empirical advances and discovery of new information that are possible with interesting new sources of data that have become available on the Internet. Our methodology accounted for unobservable blog-specific effects, even when these effects were perfectly correlated with some independent variables and hence could not be estimated directly in the traditional dummy-variable fashion. Our methodology also explicitly allows for mutual influences between blog visits and comments, which has not received much attention in previous blog studies.

We found that blog popularity, in terms of actual number of visits, is systematically influenced by social and technological factors. People submit more comments in response to new entries, and these comments enhance the popularity of the blogs. Our findings also provide support for the notion that some people may read blogs to seek specific types of information. Hence, blogs could potentially serve as information repositories or an interaction medium as well. Future research should explore how blogs could be better used for different managerial or social applications.

7.2. Limitations

There are several limitations in this study. First, our snapshot was very short at only seven days, and only a demonstration of our approach, instead of a fully-conclusive treatment of these issues. Hence, we were not able to assess the actual diffusion of the studied blogs. A longer time window, possibly in the magnitude of months or years, is necessary to gain further insights into the evolution of blogs, and to examine if blog visits are subject to other social influences, such as word of mouth, fads, or the timing of a particular discussion topic. The number of blogs that we sampled was small compared with the size of the overall blog population too, and this may threaten the external validity of our findings. Future research should survey more blogs and observe each blog for a longer period of time to augment the knowledge gained in this study.

[13] Examples of cross-sectional units on the Internet may include blogs, the book titles contained in an online bookstore, the items that are auctioned in eBay, etc.

We also focused exclusively on objective data that we could directly obtain from the blogs that we studied or from Site Meter. We did not include other dimensions such as content quality or blogger reputation, however, and these are important variables that obviously will affect people's inclination to visit, read, and respond to blogs. The incorporation of random effects and the use of the standardized *Visit Rate* variable in our model should have captured the effects due to some of these qualitative variables. These two measures, however, could only account for blog-specific heterogeneity persists over time. If the effects of the missing variables are temporal, then the use of random effects and standardized variables is insufficient, and the estimated parameters may be inconsistent. In this case, the only solution is to collect additional information on the missing variables. How to operationalize and measure these mostly qualitative variables is a key challenge for future research.

Finally, it is worth noting that an important assumption of the random effects model is that the random effects, in our case, the u_i and v_i in Equations 3 and 4, are uncorrelated with the independent variables. If this assumption does not hold (e.g., if a certain blog feature systematically increases the visit rate of some blogs more than the others), then the random effects specification will again lead to inconsistent estimates. In this case, the fixed effects model or instrumental variable regression should be used.

REFERENCES

[1] Adamic, L. and N. Glance. The political blogosphere and the 2004 U.S. election: Divided they blog. Presentation at the Second Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis, and Dynamics, Chiba, Japan, May 10, 2005.

[2] Advertising Age. Sony pays $25,000 a month for gawker blog, 76, 5, January 31, 2005. Referenced by www.marketingtom.com/2005/02/index.html. Accessed on January 15, 2007.

[3] Ajzen, I. and M. Fishbein. *Understand Attitudes and Predicting Social Behavior*. Prentice Hall, Englewood Cliffs, NJ, 1980.

[4] Anderson, C. The long tail. *Wired Magazine*, October 2004.

[5] Angrist, J.D. and A.B. Krueger. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, 15, 4, 2001, 69-85.

[6] Baltagi, B.H. *Econometric Analysis of Panel Data*, 2nd Ed. John Wiley and Sons, West Sussex, UK, 2001.

[7] Blood, R. *We've Got Blog: How Weblogs Are Changing Our Culture*. Perseus Publishing, Cambridge, UK, 2002.

[8] Brynjolfsson, E., Y.J. Hu and M.D. Smith. From niches to riches: Anatomy of the long tail. *Sloan Management Review*, 47, 4, 2006, 67-71.

[9] Bucy, E.P. and A. Lang. Formal features of cyberspace: Relationships between Web page complexity and site traffic. *Journal of the American Society for Information Science*, 50, 13, 1999, 1246-1256.

[10] Davis, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 3, 1989, 319-340.

[11] Delone, W.H. and E.R. McLean. Information systems success: The quest for the dependent variable. *Information Systems Research*, 13, 1, 1992, 60-95.

[12] Du, H.S. and C. Wagner Weblog success: Exploring the role of technology. *International Journal of Human Computer Studies*, 64, 9, 2006, 789-798.

[13] Gelder, P.V., G. Beijer and M. Berger. Statistical analysis of page views on Web sites. In A. Zanasi, C.A. Brebbia, N.F.F. Ebecken and P. Melli (Editors). *Proceedings of the 3rd International Conference on Data Mining Methods and Databases, Data Mining III*, Series: Management Information Systems, Volume 6, WIT Press, Southampton, UK, 2002, 979-988.

[14] Greene, W.H. *Econometric Analysis*. Prentice Hall, Englewood Cliffs, NJ, 2003.

[15] Gruhl, D. R. Guha, R. Kumar, J. Novak and A. Tomkins. The predictive power of online chatter. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, August 21-24, 2005, ACM Press, New York, NY, 2004, 78-87.

[16] Herring, S.C., L.A. Scheidt, E. Wright and S. Bonus. Weblogs as a bridging genre. *Information Technology and People*, 18, 2, 2005, 142-171.

[17] Hindman, M.S., K. Tsioutsiouliklis and J.A. Johnson. Measuring media diversity online and offline: Evidence from political Web sites. Presented at the 32nd Research Conference on Communication, Information and Internet Policy, National Center for Technology and Law, School of Law, George Mason University, October 1-3, 2004.

[18] Hoffman, D. L. and T.P. Novak Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing*, 60(3), July 1996, 50-68.

[19] Lim, K.H. and I. Benbasat. The effect of multimedia on perceived equivocality and perceived usefulness of information systems. *MIS Quarterly*, 42, 3, 2000, 153-179.

[20] McAfee, A.P. Enterprise 2.0: The dawn of emergent collaboration. *Sloan Management Review*, 47, 3, 2006, 20-28.

[21] McKinney, V., K. Yoon and F.M. Zahedi. The measurement of Web customer satisfaction: An expectation and disconfirmation approach. *Information Systems Research*, 13, 3, 2002, 296-315.

[22] Menchen, E. Blogger motivations: Power, pull and positive feedback. Presented at the 6th *International and Interdisciplinary Association of Internet Researchers*, Chicago, IL, October 5-9, 2005. blog.erickamenchen.net/ MenchenBlogMotivations.pdf. Accessed on January 5, 2007.

[23] Nardi, B.A., D.J. Schiano and M. Gumbrecht. Blogging as social activity, or would you let 900 million people read your diary? In J. D. Herbsleb and G. M. Olson (Eds.), *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, Chicago, IL, November 16-24, 2004, ACM Press, New York, NY, 2004a, 222-231.

[24] Nardi, B.A., D.J. Schiano, M. Gumbrecht and L. Swartz. Why we blog? *Communications of the ACM*, 47, 12, 2004b, 41-46.

[25] Nielsen, J. *Designing Web Usability*. New Riders Publishing, Indianapolis, IN, 2000.

[26] Novak, T.P., D.L. Hoffman and Y.F. Yung. Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science*, 19, 1, 2000, 22-42.

[27] Palmer, J.W. Web site usability, design and performance metrics. *Information Systems Research*, 13, 2, 2002, 151-167.

[28] Shirky, C. Power laws, weblogs, and inequality. Entry on author's Networks, Economics and Culture mailing list, February 8, 2003. shirky.com/writings/powerlaw_weblog. html. Accessed on January 18, 2007.

[29] Simmons, E. *The Impact of the Weblog: A Case Study of the United States and Iran*. Unpublished senior honor's thesis, Ohio State University, Columbus, OH, 2005. kb.osu.edu/dspace/bitstream/1811/371/1/BlogImpact.pdf. Accessed on January 18, 2006.

[30] Technorati.com. About us. www.technorati.com/about/. Accessed on December 25, 2006.

[31] Trammell, K.D. and A. Keshelashvili. Examining the new influencers: A self-presentation study of A-list blogs. *Journalism and Mass Communication Quarterly*, 82, 4, 2005, 968-982.

[32] Viegas, F.B. Bloggers' expectations of privacy and accountability: An initial survey. *Journal of Computer Mediated Communication*, 10, 3, 2005, 78-87.