THE EFFECTS OF TRADITIONAL AND SOCIAL EARNED MEDIA ON SALES:

A STUDY OF A MICROLENDING MARKETPLACE

Andrew T. Stephen and Jeff Galak*

Andrew T. Stephen
Assistant Professor of Business Administration & Katz Fellow in Marketing
University of Pittsburgh
Joseph M. Katz Graduate School of Business
318 Mervis Hall, Pittsburgh, PA 15260, USA
Email. astephen@katz.pitt.edu, Tel. (412) 648-1517

Jeff Galak
Assistant Professor of Marketing
Carnegie Mellon University
Tepper School of Business
5000 Forbes Avenue, Pittsburgh, PA 15213, USA
Email. jgalak@cmu.edu, Tel. (412) 268-5810

Forthcoming, *Journal of Marketing Research*

* Andrew T. Stephen (astephen@katz.pitt.edu) is an Assistant Professor of Business Administration & Katz Fellow in Marketing at the Joseph M. Katz Graduate School of Business, University of Pittsburgh, 318 Mervis, Pittsburgh, PA 15260, USA. Jeff Galak (jgalak@cmu.edu) is an Assistant Professor of Marketing at the Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA. The authors thank Peter Boatwright, Inyoung Chae, Isaac Dinner, Jacob Goldenberg, Don Lehmann, Olivier Toubia, Kaifu Zhang, and seminar participants at Carnegie Mellon University, INSEAD, Queen's University, University of Miami, and the 2011 Marketing Science conference for their helpful suggestions and comments, and Kiva.org for providing data. The authors are also grateful to the *JMR* editor, associate editor, and two anonymous reviewers for their many helpful suggestions.

THE EFFECTS OF TRADITIONAL AND SOCIAL EARNED MEDIA ON SALES: A STUDY OF A MICROLENDING MARKETPLACE

ABSTRACT

Marketers distinguish between three types of media: paid (e.g., advertising), owned (e.g., company website), and earned (e.g., publicity). The effects of paid media on sales have been extensively covered in the marketing literature. The effects of earned media, however, have received limited attention. This article examines how two types of earned media, traditional (e.g., publicity and press mentions) and social (e.g., blog and online community posts), affect sales and activity in each other. Fourteen months of daily sales and media activity data from a microlending marketplace website are analyzed using a multivariate autoregressive time series model. The authors find that (i) both traditional and social earned media affect sales, (ii) the per-event sales impact of traditional earned media activity is larger than for social earned media, (iii) however, because of the greater frequency of social earned media activity, after adjusting for event frequency social earned media's sales elasticity is significantly greater than traditional earned media's, and (iv) social earned media appears to play an important role in driving traditional earned media activity.

Over the past decade, the media landscape has dramatically changed with social media outlets (SMOs) like blogs, online discussion forums, and online communities now supplementing traditional media outlets (TMOs) like newspapers, magazines, and television programs. Furthermore, while social media was once the domain of younger, tech-savvy consumers who were faster to adopt new technologies, it is now generally considered to have entered the mainstream and covers a broad demographic spectrum with 75% of Internet-using adults in the United States using online social media regularly (Bernoff, Pflaum, and Bowen 2008). This large number of users makes it critical to understand how social media influences consumers and how it operates alongside traditional media.

This paper focuses on how traditional and social earned media activity affect sales of a microlending website, *Kiva*. Earned media refers to media activity that a company does not directly generate, such as press mentions in traditional media and online community posts in consumer-generated social media. Marketers distinguish earned media from paid media (advertising) and owned media (activity in channels a company owns, such as a company website), and it is common for firms to consider all three types of media when developing marketing communications strategies. While earned media is not new, in the past it has mostly been limited to TMOs. However, the popularity of social media and the increased emphasis placed by marketers on social approaches (e.g., word-of-mouth [WOM]) have made it more important for firms to understand how multi-channel earned media affects their sales.

The way in which earned media activity is generated and the way in which it impacts sales is not well understood, particularly in the current environment in which traditional and social media channels coexist. Past research has demonstrated that traditional media publicity can affect marketing outcomes (e.g., Agarwal and Kamakura 1995; Elberse 2007; Trusov, Bucklin, and Pauwels 2009), that online consumer-generated content such as online reviews can affect sales (e.g., Chevalier and Mayzlin 2006), and that, sometimes, even negative publicity can have a positive marketing effect (e.g., Ahluwalia, Burnkrant and Unnava 2000; Berger, Sorensen, and Rasmussen 2010). However, this same literature has tended to examine the isolated impact of either traditional or social media on sales (or other performance metrics), thus neglecting any effects that traditional and social media might have on each other and, indirectly, through one-another, on sales. Although some prior work has examined interrelations between media vehicles, this has typically been done within a single broad channel such as Internet advertising and not across a variety of media channels (Danaher 2007; Danaher, Lee, and Kerbache 2010). In order to properly understand the total impact of both traditional and social earned media sources on sales, an integrated perspective is needed that considers them jointly.

This paper addresses two research questions: (i) what are the relative impacts of traditional earned media and social earned media on sales? And (ii) in what ways do these earned media types influence each other? To answer these questions we analyze a novel dataset of 14 months of time series data covering the daily sales and earned media activity across multiple TMOs and SMOs for an online microlending marketplace, *Kiva* (www.kiva.org). A sale in this context is a small, low-risk loan made by a lender (customer) to a qualified borrower in a developing country. The number of loans made on the website each day is daily sales volume, separated into new versus repeat customer sales. Earned media activity data consists of daily counts of activity in TMOs (newspapers, magazines, television, radio) and SMOs (blogs, online communities) over the 14 months for which sales data were available. We use a multivariate time series model for count data to capture effects of past activity on present activity through autoregressive models and contemporaneous correlations among variables through a copula. We estimate long-run effects of earned media activity from multiple TMOs and SMOs on sales and each other. We also capture effects of new sales on repeat sales and vice versa, which reflect WOM effects on sales not detected by SMO activity.

BACKGROUND AND CONCEPTUALIZATION

Paid, Owned, and Earned Media

Marketers have adopted a typology for offline and online media activity that falls into three categories: paid media, owned media, and earned media (Corcoran 2009; Goodall 2009). *Paid media,* or advertising, refers to media activity that is generated by the company (or its agents). *Owned media* refers to media activity that is generated by the company (or its agents) in channels it controls. Common forms of owned media include press releases, brochures, and posts made by company representatives on an official company blog or website. *Earned media*—the central focus of this paper—refers to media activity that is not directly generated by the company but rather by other entities such as customers, in the case of WOM, or journalists, in the case of traditional media publicity. Marketing actions can help to generate earned media activity but marketers do not directly generate the activity. See Table 1 for examples.

[INSERT TABLE 1 ABOUT HERE]

A further distinction can be made for earned media according to whether the media source or outlet is “social” or “traditional.” The former refers to company-related media activity generated through consumers’ social online and offline interactions. This includes consumer-generated media such as blog posts, conversations in online discussion forums and communities, and newer behaviors such as tweets on Twitter and status updates on Facebook. The latter refers to traditional media activity generated by professional media outlets. This covers publicity and press coverage where professional media organizations and journalists generate earned media activity through professionally published or broadcast content.

Prior Research on Media Effects on Marketing Outcomes

Whereas advertising (paid media) effects on sales and related outcomes have received more attention, little research has examined the marketing- and financially-relevant consequences of *earned* media. Moreover, with few exceptions (e.g., Danaher et al. 2010), relationships between multiple media channels have not been extensively studied, particularly in the case of earned media (much of the extant multi-channel literature focuses on paid media).

Word-of-mouth research. Recently, there has been a growing interest in understanding how WOM, particularly online WOM, impacts sales, diffusion, and other marketing performance measures (e.g., Bruce, Foutz, and Kolsarici 2012; Chevalier and Mayzlin 2006; Godes and Mayzlin 2004; Trusov et al. 2009; Van den Bulte and Lilien 2001; Villanueva, Yoo and Hanssens 2008). For example, Godes and Mayzlin (2004) examine how online discussion forum activity affects television show ratings, Chevalier and Mayzlin (2006) show that user-generated online book reviews influence book sales, Trusov et al. (2009) examine how referrals to join an online social network affect a website's growth, and Onishi and Manchanda (2011) study how blog activity and TV advertising affect sales for a set of Japanese brands. Much of the extant literature examines one type of WOM and does not contrast it with traditional earned media, making it difficult to compare the relative sizes of the effects of traditional versus social media on key marketing outcomes such as sales. In line with a recent call for more comprehensive and multi-faceted research on WOM (Libai et al. 2010), we examine joint effects of social and traditional earned media on sales.

Publicity research. The impact of traditional forms of publicity on marketing outcomes has received disproportionately less attention in the marketing literature than other kinds of earned media such as WOM. Much of the extant research on effects of media publicity centers on the how *negative* publicity or *bad* news influence outcomes such as sales and consumer demand (e.g., Ahluwalia et al. 2000; Berger et al. 2010; Eliashberg and Shugan 1997). Consistent with the findings on how online WOM impacts performance, this literature generally finds publicity can affect a product's success in the marketplace. Like the online WOM literature, often one source of publicity is examined, precluding comparisons between different kinds of channels (e.g., online/offline).

Integrated marketing communications research. While past research has established that specific types of paid and earned media activity are important because they impact various performance metrics, more integrated study is warranted. The integrated marketing communications (IMC) literature assumes that media channels operate as a system. However, the IMC literature mostly focuses on paid media and optimal resource-allocations across coordinated media channels (e.g., Schultz, Tannenbaum, and Lauterborn 1993; Vakratsas and Ma 2005). IMC research has shown, for example, that there is a complex interplay among marketing communications efforts (Smith, Gopalakrishna, and Chatterjee 2006), and advertising in one media channel can increase promotion effectiveness in other channels (Leclerc and Little 1997).

Conceptualization and Potential Mechanisms

In this section we outline hypotheses and potential mechanisms through which these effects may operate. However, given our goal is to document these effects rather than to test specific mechanisms, these are only possible explanations for our empirical findings.

Earned media effects on sales. It may seem logical to expect that traditional earned media activity will have a larger impact on sales than social earned media activity because TMOs generally reach more people than SMOs. However, reach may not translate into increased sales. While we expect both traditional and social earned media activity to impact sales, we hypothesize that the impact will be greater for social earned media. Recent research on WOM suggests that social earned media could have a larger impact on consumer actions (e.g., website signups) than both paid and traditional earned media (Trusov et al. 2009). Thus, receiving information from a social source (e.g., through an online community) may be more influential in shifting consumers' opinions and, ultimately, triggering purchasing behavior.

Earned media effects on itself. Research on diffusion of information suggests that WOM (and, more generally, social influence) operates later in a consumer's decision process, coming after mass/traditional media (e.g., Katz and Lazarsfeld 1955; Rogers 2003). For instance, in Katz and Lazarsfeld's (1955) two-step flow of communication model, information flows from mass media to "opinion leader" individuals (step one), who then socially spread information to everyone else (step two). We may therefore expect traditional media to drive social media. However, the reverse is also plausible and, arguably, more likely in certain settings. For example, grassroots political campaigns and social movements usually start by building a following through WOM and social earned media, and then only once they have achieved a relatively high level of prominence do they begin to attract the attention of TMOs (e.g., the early days of Obama's first presidential campaign in 2006, the "Occupy Wall Street" movement in 2011, and the "Kony 2012" campaign in 2012). This process also occurs for brands and products like Facebook that, like many other startups, grew and gained prominence first through online and offline social interactions. Only after Facebook became well known and talked about did it begin to attract the attention of TMOs. In fact, more generally, journalists, editors, and producers at TMOs often look to social media for story ideas (Arrington 2009). Thus, we hypothesize that social media will influence traditional media, particularly when a brand is not well known.

Potential mechanisms. There are at least three plausible reasons why social earned media influences traditional earned media. First, the two forms of media may reach different types of consumers through a selectivity mechanism (Kornish and Li 2010; Schmitt, Skiera, and Van den Bulte 2011). That is, despite a typically larger reach, TMOs may not reach high-involvement consumers who are actively engaged in the focal brand or product. In contrast, because SMOs are often topic-specific (e.g., online brand communities) and socially interactive, social media more selectively appeals to high-involvement consumers who want to discuss topics of mutual interest with like-minded others. Thus, SMOs may more effectively reach high-involvement consumers who want to socially interact and who are more inclined to purchase. This mechanism is most likely to manifest in settings where high-involvement consumers have a substantially higher purchasing propensity (or value) and where their social interactions have implications for sales and generating buzz. This is the case in our research setting, and would likely apply to settings involving other kinds of social movements, activism, and brand evangelism.

Second, social earned media may drive traditional earned media because of “preference minorities” (Choi and Bell 2011). Specifically, brands and products vary based on how broadly or narrowly they appeal to consumers. Many brands and products do not have broad appeal and instead appeal to a preference minority with a niche interest. In this case, topic-specific media outlets are likely to be more effective in influencing behavior and generating buzz. SMOs are more likely to be specialized and better cater to preference minorities than large-reach mass-media TMOs. Thus, for a niche brand like *Kiva*, SMOs are more likely to be the channel where preference-minority consumers can find information that influences behavior. This could then explain how social media can drive traditional media: over time, as the preference minority grows in size through WOM and online social interactions, a niche brand becomes more likely to garner attention from TMOs.

Finally, social earned media's driving of traditional earned media may be caused by particular characteristics of social earned media itself. Specifically, compared to TMOs, SMOs generally have higher publication frequencies and less stringent guidelines (if any) dictating what users can publish. Accordingly, social earned media may be better than traditional earned media as an early-stage indicator of what is popular (for a related discussion about Internet search activity and sales see Goel et al. 2010 and Kulkarni, Kannan, and Moe 2012). If this is the case, an effect of social media on traditional media may simply reflect that social media is predictive of traditional earned media activity. Similarly, if sales are a function of how popular or prominent a product is, social earned media activity may simply predict (but not cause) sales activity. While we acknowledge this mechanism, we attempt to rule it out by controlling for the prominence of *Kiva* and microlending in our analyses.

DATA

Our data are provided by *Kiva*, a non-profit organization that operates an online marketplace for microloans. These small (i.e., "micro") loans are made to entrepreneurs in developing countries who do not have access to major financial institutions, have no collateral to post against the loan, and who do not require large sums of money to finance their businesses. Because the vast majority of these loans are directed to the poorest of the poor, microfinance has engendered widespread support as a way to fight global poverty.

Kiva has created a system that allows individuals in the developed world to join their website (for free) and make small loans (minimum $25; average $33) to pre-approved borrowers, all of whom are entrepreneurs in developing countries. For the purpose of this paper, we treat *Kiva* lenders as customers and the loans they make as sales. Customers select from a list of prospective borrowers and decide to whom to lend based on a variety of factors (e.g. risk, intended use of funds). Small loans are then bundled into larger loans and given to the borrowers (bundled loans are small; average bundle = $430). Between late 2005, when *Kiva* was founded, and the end of March 2012, approximately 744,000 customers (with the vast majority in the U.S.) have loaned $300 million to 759,000 borrowers in 218 countries, with a 98.94% repayment rate. Since our focus is on aggregate daily sales and earned media activity we do not go further into customers' lending decisions and factors that affect these individual-level actions. Galak, Small, and Stephen (2011) recently studied these decisions from a psychological perspective and readers are referred to their research for an individual-level account of lending on *Kiva*.

Variables

Kiva provided daily loan sales data for the period January 1, 2007 to March 2, 2008 (427 days), which we combined with media activity data from *Kiva* and other public sources.

Sales. For each day we know the number of loans made by new first-time lenders and repeat lenders. We denote the new and repeat sales volumes on day t as $SalesNew_t$ and $SalesRepeat_t$. Descriptive statistics are reported in Table 2. During our 427-day observation window, 152,439 ($5 million) new sales and 426,415 ($14 million) repeat sales were made. The two sales variables are plotted in Figure 1.

[INSERT FIGURE 1 & TABLE 2 ABOUT HERE]

Earned media activity. *Kiva* did not engage in any paid media during the time of our data and instead relied on earned media for acquiring new customers and retaining existing ones. This is a special case of a more general situation where a company would have a combination of paid, owned, and earned media activity. However, having no paid media activity is quite common, particularly for small businesses and startups operating on very tight budgets. In the discussion section we speculate on how the addition of paid media activity might affect our results. An earned media event occurred when *Kiva* was referred to in either a traditional or social media outlet. We know how many *Kiva* media events of each type occurred on each day in our observation window. For traditional earned media, our data cover newspapers, magazines, television, and radio, and for social earned media our data cover blogs and discussion forum posts, most of which were in a *Kiva*-focused online customer community/network.[1] Before describing the variables and how they were constructed, we first address three important issues related to our media variables.

First, traditional earned media activity for *Kiva* occurs infrequently, as would be expected for almost all companies, brands, or entities except for very high profile ones. The sparseness of traditional media activity within each type (newspapers, magazines, television, radio) required us to aggregate these variables into a single time series of traditional media activity.[2]

Second, we treat the two types of social media—blog posts and online community posts—separately because there is sufficient data on each type, and, conceptually, they may operate differently. For instance, online communities tend to have a higher degree of social

[1] Although not exhaustive, this set is consistent with the types of SMOs examined in previous literature. Data from other SMOs were not collected by *Kiva* and were unavailable from public sources for our observation window.
[2] Aggregating over days to create time series with weekly resolution (instead of daily) did not allow us to overcome this issue since traditional media activity was spaced out and rarely occurred on multiple days in the same week.

interactivity than blogs. Separating social earned media activity into these two types allows us to see whether social interactivity makes a difference in how social earned media affects sales.

Third, we acknowledge that the lines are increasingly blurred between traditional and social media (e.g., newspapers have blogs). Accordingly, to delineate between traditional and social media activity we considered the content's source. Content authored by a traditional professional media outlet was counted as traditional media, and consumer-generated content was counted as social media. Note that while our SMO data captures online social earned media only, our model allows us to infer social effects not tracked by the SMO variables (i.e., WOM).

Traditional media. *Kiva* supplied publicity-tracking data listing traditional media events by day and outlet. We supplemented these with Dow Jones Factiva data. $Traditional_t$ is the number of *Kiva* events on day t in the following outlets: newspapers (11 U.S. national[3], 41 U.S. local/regional, and 7 non-U.S mentions), magazines (13 mentions), television (5 national network program, 2 local program, and 4 cable program mentions), and radio (3 mentions). There were 86 traditional media events for *Kiva*.

Blogs. Google Blog Search was used to compile a daily count of numbers of non-*Kiva* initiated blog posts mentioning *Kiva*, denoted by $Blogs_t$. Over 427 days there were 2,485 blog posts about *Kiva*. These posts were not made by *Kiva* (see below for owned media).

Online communities. The number of posts on online communities/discussion forums that mentioned *Kiva* on day t is a measure of online community posting activity that we denote $CommPosts_t$. We gathered data from two sources: (i) daily posting logs from *Kivafriends* (www.kivafriends.org), an online customer community and social network dedicated to conversations about *Kiva* and microlending, and (ii) two discussion forum search engines: Omgili and Google Groups. Overall, there were 23,862 *Kiva*-related forum posts during our

[3] The *New York Times*, the *Wall Street Journal*, *USA Today*, and the *Washington Post*.

observation window. Of these, almost all (23,821) were made in the *Kivafriends* community. Search results from Omgili.com and Google either turned up posts within *Kivafriends* (which were not double counted) or unrelated discussions. Given that nearly all community discussion activity observed for *Kiva* came from the online community, we treat $CommPosts_t$ as a measure of online WOM conversation activity within the social network for *Kiva* customers and fans. Additionally, we collected data on the size of the *Kivafriends* community as a control variable. $CommMembers_t$ is the number of member registrations on *Kivafriends* made on day t.

Other variables. We used a number of other variables as controls in our model.

Owned media. *Kiva* operates a company blog (kivanews.blogspot.com) and posted 86 times during our observation window. We denote the number of posts made on the company blog on day t as $OwnedBlog_t$. In addition, *Kiva* issued two press releases, denoted $OwnedPR_t$.

Interest in *Kiva* and microfinance. We control for general interest in *Kiva* and the topic of microfinance since the earned media variables may simply be picking up on collective interest in *Kiva* and microfinance. If this is the case then earned media effects on sales may not be a result of, say, a mention of *Kiva* in a newspaper article, but rather the fact that *Kiva* is just more culturally prominent. While it is impossible to perfectly measure daily interest in a website or topic, data from Google Trends is a reasonable proxy. Google indexes and normalizes search volume data and therefore does not provide a count of searches on a given keyword. The data gives an indication of trends in search interest. We collected data for the keywords "kiva" and "microfinance." These variables are $SearchKiva_t$ and $SearchTopic_t$. They are indexed against a common baseline: searches on "microfinance" at the beginning of Google's data.

Christmas and weekend effects. The dummy variables $Christmas_t$ and $Weekend_t$ respectively control for Christmas/end-of-year and weekend effects. *Christmas* days were the last two weeks of December and the first two weeks of January. Results were robust to alternative definitions, including only the last two weeks of December, only the last week of December, and the last week of December and the first week of January.

Number of borrowers. We control for the number of loan requests on *Kiva* on day *t*, $Borrowers_t$. This should be unaffected by media and sales because the borrowers generally do not access TMOs and SMOs in the U.S. and directly deal with field agents, not *Kiva*. This allows us to control for the level of "supply" in the marketplace, which we expect to affect sales but is not of substantive interest given our research questions.

MODELING MULTIVARIATE TIME SERIES COUNT DATA

Data Challenges and Modeling Approach

Our multivariate time series data have three features that our modeling approach must accommodate. First, all sales and media variables are time series *counts* (i.e., non-negative integers). Second, for some key variables (e.g., *Traditional*) there are many zeros. Third, the variables may be contemporaneously correlated (e.g., same-day cross-media effects) and not just related through autoregressive lags that capture the effect of past values on current values. Accordingly, we need a time series model that makes appropriate distributional assumptions for counts, accommodates excess zeros, and estimates contemporaneous correlations. Although a common model for multivariate time series data is vector autoregression (VAR) (e.g., Dekimpe and Hanssens 2004; Stephen and Toubia 2010; Trusov et al. 2009; Villanueva et al. 2008), we do not use it here because it does not satisfy these requirements.

We turn to the literature on multivariate count data models in statistics, finance, and transport engineering (e.g., Karlis and Meligkotsidou 2005; Trivedi and Zimmer 2005; Wedel et al. 2003), as well as marketing (e.g., Ansari et al. 2011; Choi et al. 2012; Danaher 2007; Dong et al. 2012; Park and Fader 2004). Many multivariate count models are seemingly unrelated regressions with appropriate distributions such as Poisson and negative binomial. These models were not designed to handle time series count data, making them inappropriate for our use (we explain why below). Instead, we build on recent work in finance on modeling multiple stocks' trading volumes that led to a flexible multivariate autoregressive model for time series count data (Heinen and Rengifo 2007).

Zero-Inflated Multivariate Autoregressive Double Poisson Model

We extend the Heinen and Rengifo (2007) model to explicitly accommodate rare events (e.g., traditional media activity) that generate excess zero counts in time series. Our model has four components: (i) double Poisson (DP) distributions for the marginal distributions, i.e., for each endogenous time series count variable, (ii) autoregressive models for the conditional means of the DP distributions to capture lagged effects of variables on each other (cross-variable effects) and themselves (own effects), (iii) a multivariate normal copula to connect the marginal distributions by estimating contemporaneous correlations between time series not picked up in lagged cross-variable effects, and (iv) a zero-inflated specification to accommodate excess zeros for variables that track occurrences of infrequent events.

Double Poisson distribution. For count data we need a discrete probability distribution with non-negative support. Although Poisson or negative binomial distributions are typically used, the former assumes equidispersion (mean = variance) and the latter allows for overdispersion (mean < variance) but not underdispersion (mean > variance). Time series count data can be both over- and underdispersed (Efron 1986). If the conditional mean is modeled as an autoregressive process (consistent with traditional time series models) then overdispersion is generated (Heinen and Rengifo 2007). At the same time overdispersion in the data may in fact be less than overdispersion resulting from autocorrelation, which gives rise to (conditional) underdispersion in the underlying marginal distributions after controlling for autocorrelation (Efron 1986). Thus, each time series count variable in our data needs a marginal model based on a flexible discrete probability distribution with non-negative support that allows for over- *and* underdispersion. Following Efron (1986), we use the double Poisson (DP) distribution as it satisfies this requirement (see also Cameron and Trivedi [1998] and Winkelmann [2008]). The DP probability density function (pdf) for data y, mean μ, and dispersion $\phi > 0$ is:

$$f_{DP}(y|\mu,\phi) = k(\mu,\phi)\sqrt{\phi}\exp(-\phi\mu)\left(\frac{\exp(-y)y^{y}}{y!}\right)\left(\frac{\mu \mathrm{e}}{y}\right)^{\phi y} \tag{1}$$

Where e is Euler's number and $k(\mu,\phi)$ is a normalizing constant. Efron (1986) approximates the constant as:

$$\frac{1}{k(\mu,\phi)} \approx 1+\frac{1-\phi}{12\mu\phi}\left(1+\frac{1}{\mu\phi}\right) \tag{2}$$

Note the dispersion parameter ϕ allows for overdispersion ($\phi > 1$), underdispersion ($\phi < 1$), and the special case of equidispersion ($\phi = 1$, in which case Equation 1 reduces to the Poisson pdf).

Autoregressive conditional mean model. As with most parametric nonlinear models, we model the mean of the distribution using a linear model and a link function.[4] Let Y_{it} be a count of the number of events of type i that occurred in time period t, where each event type is represented by its own time series of counts, $i = 1, \ldots, M$, and $t = 1, \ldots, T$. Each DP-distributed data point is conditional on the past of all variables; i.e., $Y_{it} \mid \Gamma_{-t} \sim DP(\mu_{it},\phi_i)$ where Γ_{-t} is the past

[4] The identity link is appropriate since regressors are non-negative. We checked estimated means and none were negative. Other link functions, such as a logit or exponential, would also be appropriate for more general cases.

of all M series (and exogenous control variables) up to and including the previous period, μ_{it} is the mean for variable i at time t, and ϕ_i is the dispersion for variable i. We use M first-order linear autoregression with exogenous variables (ARX1) models to estimate lagged cross-variable and own effects and exogenous covariate effects through the DP conditional means:

$$E(Y_{it} \mid \Gamma_{t-1}) = \mu_{it} = \sum_{j=1}^{M} \beta_{ji} Y_{j,t-1} + \sum_{l=1}^{L} \gamma_{li} X_{lt} \tag{3}$$

Where β_{ji} is the effect of $Y_{j,t\text{-}1}$ on Y_{it} (i.e., own for $j = i$, cross-variable for $j \neq i$), and γ_{li} is the effect of an exogenous covariate X_{lt} on Y_{it} (for L exogenous covariates). Equation 3 describes the conditional mean model for a first-order autoregression (i.e., lag order = 1). It is trivial to extend this to allow lag order > 1. The appropriate lag order is empirically determined by testing different lag orders and selecting the one that gives the best-fitting model, consistent with standard time series modeling practice. Estimating higher-order systems is difficult for $M > 3$ since the number of parameters to estimate becomes large without imposing a reduced rank structure on **B** (the M-by-M matrix of β_{ji}) (cf. Ben Omrane and Heinen 2010).

From the set of eleven time series variables listed above we determined which ones were endogenous (**Y**) and which were exogenous (**X**) using Granger causality tests (Granger 1969), following standard multivariate time series modeling procedure (Dekimpe and Hanssens 2004). A variable is deemed to be exogenous (in Granger causality terms) if it is not caused by any other variables in the system (it may, however, be caused by lags of itself). Details are given in the results section.

Multivariate normal copula. The third part of the model ties the M marginal models together to capture conditional associations between the variables that are not captured through the autoregressive conditional means models; i.e., contemporaneous or "same day" correlations.

For the sake of brevity we do not provide a full technical description of copulas. We encourage readers to consult Danaher and Smith (2011) for a thorough explanation of copulas and how they can be used in marketing science research.

In simple terms, copulas flexibly permit the combining of multiple univariate marginal distributions to form a multivariate joint distribution without requiring that all marginal distributions are the same. A variety of copulas are described in the literature (see Trivedi and Zimmer 2005), but most can only accommodate $M = 2$. Following Heinen and Rengifo (2007) and Danaher and Smith (2011), we use the multivariate normal copula since it can accommodate $M > 2$ (see Web Appendix). The practical benefit of using a copula here is that it allows for contemporaneous correlations between the M endogenous time series count variables to be estimated. This contemporaneous interdependence is after controlling for cross-variable lagged effects in the autoregressive conditional mean models. We cannot, however, infer the presence or absence of any contemporaneous *causal* relationships with the copula.

Since we use the copula estimates for computing long-run effects (see below) and our research questions do not consider short-term same-period causal effects, the copula is sufficient for our application. If we sought to estimate direct contemporaneous *causal* effects an alternative approach would be the structural vector autoregression (SVAR) models used in macroeconometrics. Cross-variable contemporaneous effects would be directly entered into each conditional mean model and a copula would be unnecessary. Although possible (and attempted below as a robustness check), it can require assumptions (identification restrictions) on contemporaneous effects (Sims 1980). To avoid making such assumptions the preferred approach here is to allow for contemporaneous correlations through the copula. Further, in the economics literature the structural approach is usually employed only in cases where the time resolution of data is low, such as annual macroeconomic data (Sims 1980). In such cases same-period effects are expected and may not be captured through autoregressive lag effects given the longer period length. This problem is unlikely in our case since we use daily data.

Zero-inflation. While the DP distribution can accommodate zeros, explicit treatment of these zeros can improve estimates and model fit. In keeping with other count models with excess zeros, we introduce zero-inflation using a finite mixture in the same way that, for example, the zero-inflated Poisson regression model is constructed (Cameron and Trivedi 1998), with $Y_{it} = 0$ either from the DP distribution or a degenerate distribution with mass concentrated at zero. We describe this specification below, and use it only for the marginal models for *Traditional*, *CommPosts*, and *CommMembers* since these are the only variables with some zero values.

Model Estimation and Evaluation

Joint density. In general, the joint density at time t for all M endogenous variables (with Θ as a vector of all parameters in the marginal models) is simply the product of the products of the M marginal densities and the multivariate normal copula $c(q_t;\Sigma)$:

$$h(Y_{1t},\ldots,Y_{Mt},\Theta,\Sigma) = \prod_{i=1}^{M} f_{mixture}(Y_{it},\mu_{it},\phi_i,\pi_i)\cdot c(q_t;\Sigma) \tag{4}$$

Where $f_{mixture}(Y_{it},\mu_{it},\phi_i,\pi_i)$ is the density of a finite mixture of the DP distribution and a degenerate distribution of a mass at zero such that for variable i at time t:

$$\begin{aligned} \Pr(Y_{it} = 0) &= \pi_i + (1-\pi_i)\left[\sqrt{\phi_i}\exp(-\phi_i\mu_{it})\right] \\ \Pr(Y_{it} > 0) &= (1-\pi_i) f_{DP}(Y_{it},\mu_{it},\phi_i) \end{aligned} \tag{5}$$

$\pi_i \in [0,1]$ is the mixture parameter for the zero-inflated model (without zero-inflation $\pi_i = 0$). Consistent with standard practice in zero-inflated count models, we use a logit link for this parameter ($\mathbf{w}_i$ can be a vector of ones or covariates); i.e., $\pi_i = \exp(\mathbf{w}_i'\lambda)/[1+\exp(\mathbf{w}_i'\lambda)]$. For the copula $c(q_t;\Sigma)$, q_t are the M-dimensional vectors of normal quantiles of the probability integral transforms (PITs) of the count data under the marginal DP densities, and Σ is the covariance matrix of the multivariate normal copula representing contemporaneous interdependence. Note that since PIT theory applies to continuous random variables, the continuous extension argument in Denuit and Lambert (2005) is followed to allow us to apply copulas to discrete marginal distributions.

Likelihood function. Based on the joint density, the log-likelihood is:

$$LL = \sum_{t=1}^{T} \log h(Y_{1t},\ldots,Y_{Mt},\Theta,\Sigma) = \sum_{t=1}^{T}\sum_{i=1}^{M} \log f_{mixture}(Y_{it},\mu_{it},\phi_i,\pi_i) + \log c(q_t;\Sigma) \tag{6}$$

Since the log-likelihood is additive we can split it into two parts. For the marginal models, LL_1:

$$\begin{aligned} LL_1 = & \sum_{i=1}^{M}\sum_{t=2}^{T} \mathbf{1}_{Y_{it}=0} \log\left[\pi_i + (1-\pi_i)\sqrt{\phi_i}\exp(-\mu_{it}\phi_i)\right] \\ & + \sum_{i=1}^{M}\sum_{t=2}^{T} \mathbf{1}_{Y_{it}>0} \log\left[(1-\pi_i)k(\mu_{it},\phi_i)\sqrt{\phi_i}\exp(-\mu_{it}\phi_i)\frac{\exp(-Y_{it})Y_{it}^{Y_{it}}}{Y_{it}!}\left(\frac{\mu_{it}\mathrm{e}}{Y_{it}}\right)^{\phi_i Y_{it}}\right] \end{aligned} \tag{7}$$

Where $\mathbf{1}_{Y_{it}=0}$ is an indicator variable that equals 1 when $Y_{it} = 0$ and 0 otherwise, and $\mathbf{1}_{Y_{it}>0}$ equals 1 when $Y_{it} > 0$ and 0 otherwise. If $\boldsymbol{Y}_i$ does not contain any zeros we set $\pi_i = 0$. For the multivariate normal copula, $LL_2 = \sum_{t=1}^{T}\left[\tfrac{1}{2}q_t'(\mathbf{I}_M - \Sigma^{-1})q_t - \tfrac{1}{2}\log|\Sigma|\right]$, where q_t is the M-by-1 vector of the normal quantiles of the PITs of the continued extension of the data (see appendix) and $\mathbf{I}_M$ is an M-dimensional identity matrix.

Estimation. In theory, all parameters can be jointly estimated by maximizing the log-likelihood in Equation 6. However, as noted by Heinen and Rengifo (2007), this is infeasible given the number of parameters to be estimated, nonlinearities in the copula, and the nonlinear normalizing constant in the DP pdf. Instead, we follow a two-step procedure suggested by Patton (2006) and followed by Heinen and Rengifo (2007), which yields consistent estimates for all parameters (Patton 2006). The first step estimates parameters of the *M* marginal models (yielding estimates of parameter matrix Θ) by maximizing LL_1. The second step estimates the copula variance-covariance matrix Σ under the step 1 parameter estimates (which give the marginal DP distributions). The use of the multivariate normal copula makes this a computationally trivial step since the maximum likelihood estimate of a zero-mean multivariate normal distribution's variance-covariance matrix is its sample counterpart; i.e., $\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} q_t q_t'$. LL_2 is obtained by plugging the estimated variance-covariance matrix and q_t into the equation for LL_2. Note q_t depends on the data and the step 1 parameter estimates because the PITs from which the q_t are computed are conditional on DP distributions with estimated means and dispersion parameters for each data point from step 1. Further, while the parameter estimates in step 1 affect the copula estimates in step 2, the reverse is not the case; i.e., the copula does not affect the point estimates, statistical inference, or descriptive fit of the marginal models. The parameter and copula estimates are used together, however, to compute the long-run effects upon which we base our findings.

Specification tests and fit. Standard time series modeling practices are employed. First, stationarity tests check whether variables can be modeled in levels (stationary) or differences (evolving). Second, Granger causality tests determine which variables are endogenous in the sense that they are affected by lags of themselves and other variables, versus which variables are exogenous and only affected by lags of themselves. Third, fit is evaluated based on log-likelihood (*LL*). Fourth, standardized residuals, $\varepsilon_{it} = (Y_{it} - \hat{\mu}_{it}) / \sqrt{(\hat{\mu}_{it} / \hat{\phi}_i)}$, are checked to see that they have mean close to 0, variance close to 1, and minimal residual autocorrelation.

RESULTS

Specification Tests and Model Fit

Stationarity. All time series variables were subjected to an Augmented Dickey-Fuller (ADF) unit root test that used test statistics based on parameter estimates from our DP model (i.e., the DP distributional assumption was upheld). The ADF test's null hypothesis is that a series is evolving and therefore non-stationary. In all but two cases this null hypothesis was rejected (*p*s < .01). The two non-stationary time series were *SearchKiva* and *SearchTopic*. Following convention these variables were first-differenced (Dekimpe and Hanssens 2004). Both Δ*SearchKiva* and Δ*SearchTopic* were stationary. Another test of stationarity is that the eigenvalues of (**I** – **B**) lie within the unit circle, which was the case (note: **I** is an *M*-dimensional identity matrix and **B** is the *M*-by-*M* matrix of parameter estimates for β_{ji} in Equation 2). It is therefore appropriate to use the autoregressive model for the DP distributions' conditional means and except for the search interest variables the data can be used "as is" in their levels without differencing (Dekimpe and Hanssens 2004).

Endogenous variables. Our conceptualization suggests that the sales and earned media activity variables are endogenous; i.e., influenced by themselves and each other. We checked each of the eleven time series variables with Granger causality tests (Granger 1969). If a variable was Granger caused by its own past and not by any other variables then it was exogenous, otherwise it was endogenous. Six endogenous variables were found ($M = 6$): *SalesNew*, *SalesRepeat*, *Traditional*, *Blogs*, *CommPosts*, and *CommMembers*. The other variables were treated as exogenous covariates (i.e., **X** in Equation 2). As an additional check, we also modeled a simpler system of $M = 4$ variables with *SalesNew*, *SalesRepeat*, *Traditional*, and *Social* where *Social* = *Blogs* + *CommPosts*. Results from this model were similar to and consistent with those from the $M = 6$ system.

Lag order. The standard approach in time series modeling is to select an appropriate lag order by varying the lag order and seeing which variant model fits best (based on a likelihood-based information criterion). We estimated the full model using lags of one, two, and three periods. The Akaike Information Criterion (AIC) for the one-, two-, and three-lag models, respectively, was 18,174, 18,483, and 23,665 (this ordering of model fits was also found with the Bayesian Information Criterion [BIC]). Thus, the best-fitting model was the one-lag version.

Residuals and fit. Each endogenous variable's vector of standardized residuals (ε_i) had a mean close to 0, variance close 1, and residual autocorrelation functions with no evidence of strong residual autocorrelations. Based on a likelihood ratio test, this model fit better than a base model that included exogenous covariates and lagged own effects but no lagged cross-variable effects ($\chi^2 = 742.30$, $df = 30$, $p < .001$).

Main Findings

Table 3 reports marginal model estimates, the copula correlation matrix, and fit statistics. Long-run effects based on cumulative orthogonalized impulse response functions (COIRFs) are reported in Table 4 and long-run elasticities are reported in Tables 5 and 6.

[INSERT TABLES 3-6 ABOUT HERE]

In reporting the results we focus on addressing our research questions related to the effects of the earned media variables on sales and each other. Although some exogenous variables have significant parameter estimates in Table 3 we do not focus on these since they are not central to our research questions.[5] We base our findings on the long-run effects reported in Table 4 since it is infeasible to interpret the parameter estimates in Table 3 directly in multivariate time series models (Sims 1980). The COIRFs capture the total (direct and indirect) over-time impact of a unit increase (shock) in a variable over its baseline on all endogenous variables in the dynamic system (Dekimpe and Hanssens 2004). Importantly, they also account for contemporaneous correlations because a shock to any variable may be accompanied by shocks in other variables at the same time. The estimated parameters in **B** were used to compute unorthogonalized IRFs. Uncertainty in parameter estimates in **B** was accommodated following standard procedures described by, for example, Sims and Zha (1999) and Trusov et al. (2009): we computed standard errors (using the delta method) and t-values, and deemed an IRF not significantly different from zero if $|t| < 1$. Cumulative IRFs were then computed[6] and adjusted for contemporaneous correlations (orthogonalization) using the Cholesky decomposition of the estimated copula covariance matrix (Sims 1981).[7] We interpret and compute elasticities for only those effects that had significant COIRFs in Table 4.

Earned media effects on sales. All three earned media variables have long-run positive effects on new and repeat sales. The sizes of these effects—across media variables and across sales variables—differ greatly. Traditional earned media activity has the largest *per-event* long-run impact: an extra unit of media publicity from a TMO approximately generates an additional 894 new and 403 repeat sales. This is much greater than the per-event impacts of blog posts (90

---

[5] For example, long-run effects of *Borrowers* on *SalesNew* and *SalesRepeat* are 22.76 and 11.11, respectively (non-significant on other variables). These may reflect effects of increasing “assortment size” on sales.

[6] The cumulative unorthogonalized IRF matrix, $\mathbf{B}_{CIRF} = \sum_{s=1}^{\infty} \mathbf{B}^{s}$. In practice, for each element of **B** we summed from $s = 1$ to the point at which that element’s IRF became non-significant.

[7] $\Sigma = \mathbf{PP}'$, where **P** is the Cholesky decomposition of Σ and is a lower triangular M-by-M matrix with a positive diagonal. The COIRF matrix is $\mathbf{B}_{CIRF}\mathbf{P}$. Since **P** can change depending on the ordering of variables in Σ we followed Sims (1981) and checked sensitivity to different orders. **P** did not markedly change over different permutations of Σ.

new and 63 repeat sales) and community forum posts (99 new and 48 repeat sales). This is not surprising since TMOs have larger reach. While an interesting analysis would be to look at the sales responses after adjusting for the reach of the respective media types, audience size/reach data are not available for these data. We therefore leave further exploration of the returns to media with different-sized audiences to future research.

We cannot properly compare these media effects on sales, however, without acknowledging that the different types of media events occur at vastly different frequencies. While a single traditional media event may have a large impact on *SalesNew* and *SalesRepeat*, these events are very infrequent. Social media activity, however, occurs much more frequently. A way to look at these findings and to compare across earned media variables is to look at the long-run elasticities of sales to media activities in Table 5. Specifically, we compute arc elasticities using the COIRFs in Table 4 and the means in Table 1. The arc elasticity for the effect of impulse j on response i using COIRFs is η_{ji} = COIRF$_{ji}$ × E(Y_j) / E(Y_i). Of particular interest is the size of the elasticity for community posting activity on sales. The sales elasticities to community posting are over 30 times larger than the sales elasticities to traditional activity, and around 10 times larger than the sales elasticities to blog posting. The same pattern is seen in Table 6, panel B. After adjusting for the differences in the frequencies with which these earned media events occur, we see that social earned media is more effective in generating extra sales than traditional earned media. This is particularly the case for Kiva's online community.

Sales effects on itself. We next consider how new sales affect repeat sales and vice versa. For both sales variables, the largest sales elasticity comes not from community activity but from repeat sales (22.74 for new sales elasticity to repeat sales, and 5.65 for repeat sales elasticity to repeat sales). The endogenous sales effects suggest some WOM effects not captured in the social earned media variables. The effect of repeat sales on new sales likely picks up the effect of WOM from existing customers to non-customers (who become new customers). The effect of repeat sales on repeat sales likely picks up both WOM (from existing customers to other existing customers) and within-customer repeat sales (e.g., loyalty). The other significant endogenous sales effect is for new customer sales on itself, which is likely also a WOM effect; i.e., the ability of WOM from new customers to acquire additional new customers (elasticity = 2.37). The strength of these effects suggest that social influences such as WOM from existing customers play a critical role in driving both the acquisition of new customers as well as in promoting retention and growth. Importantly, these implied WOM (and repeat sales/loyalty) effects are in addition to, and after controlling for, online community activity.

Earned media effects on itself. The only long-run effect on traditional media activity comes from blogs (elasticity = 1.16), which itself is strongly influenced by community posting activity (elasticity = 4.42). At least in the case of *Kiva*, as hypothesized earlier, traditional earned media activity flows from social media but not the other way.

Additional Analyses and Robustness Checks

Direct contemporaneous effects. As a first robustness check we estimated an analog to an SVAR model (Bernanke 1986; Sims 1980, 1986) that entered contemporaneous cross-variable effects directly into the model instead of using a copula and is equivalent to assuming that all off-diagonal elements in the copula covariance matrix are zero. This assumption satisfies SVAR identification requirements. See the web appendix for details. We provide additional details and results in the web appendix. As expected, a number of direct contemporaneous effects were significant (see Table C1 in the Web Appendix). The substantive findings were conceptually consistent with those reported above for the main model.

Traditional media reach. As a second robustness check we explored the role of traditional media reach in driving our results. We did this by splitting *Traditional* into *TraditionalHighReach* and *TraditionalOther*. The high-reach variable measured daily activity for high-reach TMOs. We defined *TraditionalHighReach* (and, accordingly, *TraditionalOther*) in two ways. First, we defined high-reach TMOs as those with a national reach (e.g., CNN, *New York Times*, *Oprah*; 36 events in total). Second, we defined high-reach TMOs as those that had events coinciding (same or previous day) with the days of the very-high sales outliers in Figure 1 (specifically: *New York Times*, *Oprah*, network TV news; 11 events in total). We re-estimated our model replacing *Traditional* with *TraditionalHighReach* and *TraditionalOther* (both ways). Irrespective of how these variables were defined, the long-run results were substantively consistent with those reported above for the main model. The only differences were (i) *SalesNew* was positively affected by *TraditionalOther* but not *TraditionalHighReach*, (ii) *SalesRepeat* was positively affected by *TraditionalHighReach* but not *TraditionalOther*, (iii) *TraditionalHighReach* was positively affected by lagged *TraditionalOther*, and (iv) the positive blog post effect on *Traditional* in the previous model was found for *TraditionalOther* but not for *TraditionalHighReach*. This suggests that the influence of high-reach, national-audience TMOs on sales may be limited for Kiva, particularly with respect to acquiring new customers and generating new sales. High-reach national TMOs may instead serve a “memory jogging” role whereby they are most effective in activating existing customers and thus generating repeat sales. Further, the effect of social earned media on traditional earned media flows first through

TraditionalOther and then to the high-reach TMOs. In other words, news about *Kiva* spreads first into more local traditional media outlets and then into the mainstream national outlets.

Related to this analysis, as a third robustness check we tested the sensitivity of the substantive findings to traditional earned media events that coincided with the sales outlier days in Figure 1 (i.e., TMOs defined the second way for *TraditionalHighReach*). We re-estimated the main model using *Traditional* and added a lagged control dummy variable that equaled one on days that the high-reach TMOs had events. Although some parameter estimates changed slightly, none of the above-reported substantive findings changed.

Counterfactual analysis. To illustrate the importance of cross-media effects we performed simple counterfactual analyses. For each earned media variable we estimated a version of the model with the effects of that variable on the other earned media variables fixed to zero, computed the COIRFs, and compared the sizes of the long-run media effects on sales to those reported in Table 4. The comparisons are reported in Table 7 and indicate how much sales are affected when a variable's cross-media effects are removed. The largest impact on the system happens when online community posting cannot affect the other media variables, resulting in drops in long-run effects on sales of between 21% and 37%. Specifically, the impacts of traditional earned media events on sales drop by 21% (new) and 23% (repeat) when online community activity cannot affect the other media variables. This highlights the "behind the scenes" role played here by online community activity. Traditional media events have a large per-event effect on sales that is substantially lowered when removing indirect online community effects are removed. This was not the case when other media variables were restricted.

[INSERT TABLE 7 & FIGURE 2 ABOUT HERE]

DISCUSSION

Our objective was to study how traditional and social earned media affect sales and each other. The long-run elasticities are summarized in Figure 2. Traditional earned media has a per-event effect on sales that is much larger than the corresponding per-event effects of social earned media. This is not surprising given that TMOs typically have larger reach than SMOs. However, as is very often the case, traditional earned media events were quite infrequent when compared to social media events. After taking the disparity in event frequencies into account there are significant effects of social earned media on sales, and, particularly, for online community activity. These small per-event impacts of social earned media activity can accumulate to have a substantial long-run impact on sales. From a practical standpoint, our findings emphasize the importance of both types of earned media. However, since traditional earned media events are usually rare and difficult to instigate, marketers hoping to generate sales through earned media would likely be well served by focusing on trying to generate social media activity and WOM.

Our results appear to fit the selectivity mechanism described earlier (Kornish and Li 2010; Schmitt et al. 2011), where blogs and online communities that discuss Kiva selectively attract audiences who are more interested and involved in the topic than the audiences who are exposed *en masse* through TMOs. People who visit blogs or online communities and read about *Kiva* choose to expose themselves to the content, suggesting that they may be more involved and likely to act.[8] People exposed through TMOs, particularly large-reach outlets, may just stumble upon content and are thus not expected to have high involvement *a priori*.

Taking the above-mentioned mechanisms and our findings into account, the smaller influence of TMOs on sales may be due to their low targetability of high-involvement people and their incompatibility with niche topics. Their incompatibility with niche topics (and the superior

[8] This assumes positive content. Since *Kiva* is trying to alleviate poverty, virtually all content was positive.

compatibility of SMOs) appears consistent with the preference minorities mechanism discussed above (Choi and Bell 2011). TMOs reach many consumers but they are not necessarily the consumers who are most interested, involved, and likely to engage. Reaching a more engaged audience through more focused targeting efforts may improve the effectiveness of traditional earned media in driving sales. This may be particularly helpful for brands or products that, like *Kiva*, have a niche appeal. Further, focusing on smaller-reach TMOs that have more relevant audiences would likely help (e.g., specialist magazines that are read by people with a particular niche interest). However, the lack of social interactivity in TMOs may nevertheless hinder their ability to be equally or more effective in driving consumers to take action as interactive SMOs, particularly in settings where high-involvement consumers are necessary and consumer-to-consumer social interactions play an important role in sustaining interest and triggering behavior.

Finally, we acknowledge that our results may also be accounted for by the above-discussed mechanism whereby social earned media simply acts as a better early-stage indicator of what is popular, which would suggest that effects of social earned media activity on traditional earned media activity are predictive but not causal. Prior research on the effects of Internet search activity on sales (Goel et al. 2010; Kulkarni et al. 2012) showed that searches (e.g., on Google) can be predictive of sales and can indicate how popular something is. We controlled for the popularity of Kiva and microlending using Google search data (*SearchKiva*, *SearchTopic*). If social earned media activity was simply picking up the popularity of Kiva then we would expect direct and indirect effects of *Blogs* and *CommPosts* on *Traditional* to be mitigated when search activity was controlled for. This was not the case here. While this does not mean we cannot entirely rule out this potential mechanism, it does suggest this mechanism is unlikely the sole explanation for our results. Nevertheless, it does mean that at least part of the effects of SMOs on TMOs found in our analysis may be merely predictive rather than of a more causal nature.

Limitations and Future Research Directions

The conclusions that we draw from our analyses are limited in at least six ways. First, despite gathering data from a variety of reliable sources for multiple media types, our set of earned media variables is not exhaustive. Second, our results apply to a single organization in a specialized domain. We make no claims of generalizability, and encourage future research to examine similar research questions in different contexts. However, we expect that endogenous earned media effects are present in many domains. Long-run sales elasticities for different types of earned media will likely differ depending on factors such as whether high-involvement consumers are necessary, whether social interactions play an important role in influencing purchasing, and whether the product has broad or niche appeal. Third, our results are based on an environment without paid media activity. This is not uncommon, particularly for non-profits, startups, and small businesses. With paid media we would not expect dramatically different findings for earned media, since advertising would likely increase interest in *Kiva*, thus helping earned media be more effective in driving sales. An interesting future research direction is to examine what happens when paid media activity is present and whether paid media can be used to amplify the positive effects of earned media (traditional and/or social) on sales and each other. Fourth, we were unable to consider the valence of media events. Since *Kiva* is a non-profit trying to alleviate world poverty, almost all content we sampled was positive. Future research could look at how valence moderates the impact of earned media activity on sales. Fifth, while we drew a clear line between traditional and social earned media based on whether the source was a professional media organization, the line has become blurred and a hybrid media type that is professionally sourced but online and socially interactive has emerged. Future work could examine this type. Finally, as noted above, since we cannot fully rule out the "early-predictor" mechanism using available control variables it is possible that at least part of the effects of SMOs on TMOs found here are more predictive than they are causal.

An important direction for future research is to more deeply examine the theoretical mechanisms underpinning our findings. More work is needed to fully understand when, how, and why traditional and social earned media affect sales and each other. Two avenues for future research are particularly promising. First, research that helps disambiguate various mechanisms that potentially explain earned media effects on sales will be helpful. For example, to determine whether social earned media's impact on traditional earned media occurs because of factors inherent to SMOs (e.g., social interactivity) or because they have high targetability of high-involvement consumers, a setting where some TMOs have high targetability of high-involvement consumers could be studied (e.g., special-interest magazines). Second, research that considers lower-involvement contexts and broad-appeal products would be useful for testing if the strong role played by social earned media found here would hold or whether traditional earned media's role would become more important.

A final direction for future research involves "zooming in" on same-day interactions between social media and sales. The estimated copula correlations reported in Table 3 suggest very strong contemporaneous relations between social media and sales. Further investigation of these same-day relations, ideally using data with a finer (within-day) time resolution is likely to give rise to interesting new findings related to the speed at which social media can influence marketing outcomes. A related opportunity for future research is to attempt to better understand what observable factors such as characteristics of social media channels and their audiences explain the sizes of contemporaneous relations between social media activity and sales.

To conclude, our findings point to a fairly complex characterization of how earned media activity for a company can affect sales. We hope that this research encourages additional research to further improve our understanding of how different media channels work in concert to drive marketing outcomes.

REFERENCES

Agrawal, Jagdish and Wagner A. Kamakura (1995), "The Economic Worth of Celebrity Endorsers: An Event Study Analysis," *Journal of Marketing*, 59 (3), 56-62.

Ahluwalia, Rohini, Robert Burnkrant, and H. Unnava (2000), "Consumer Response to Negative Publicity: The Moderating Role of Commitment," *Journal of Marketing Research*, 37 (May), 203–214.

Ansari, Asim, Oded Koenigsberg, and Florian Stahl (2011), "Modeling Multiple Relationships in Social Networks," *Journal of Marketing Research*, 48 (4), 713-728.

Arrington, Michael (2009), "NYTimes Tech Editor Reads TechCrunch Every Morning for Story Ideas," *TechCrunch.com*, July 13.

Ben Omrane, Walid and Andréas Heinen (2010), "Public News Announcements and Quoting Activity in the Euro/Dollar Foreign Exchange Market," *Computation Statistics and Data Analysis*, 54, 2419-2431.

Berger, Jonah, Alan T. Sorensen, and Scott J. Rasmussen (2010), "Positive Effects of Negative Publicity: When Negative Reviews Increase Sales," *Marketing Science*, 29 (5), 815-827.

Bernanke, Ben S. (1986), "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25, 49-100.

Bernoff, Josh, Cynthia N. Pflaum, and Emily Bowen (2008), "The Growth of Social Technology Adoption," *Forrester Research Report*, Cambridge, MA.

Bruce, Norris I., Natasha Zhang Foutz, and Ceren Kolsarici (2012), "Dynamic Effectiveness of Advertising and Word-of-Mouth in the Sequential Distribution of Short Life Cycle Products," *Journal of Marketing Research*, forthcoming.

Cameron, A. Colin and Pravin K. Trivedi (1998), *Regression Analysis of Count Data*, Cambridge, U.K.: Cambridge University Press.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345-354.

Choi, Jeonghye and David R. Bell (2011), "Preference Minorities and the Internet," *Journal of Marketing Research*, 48 (August), 670-682.

Choi, Jeonghye, David R. Bell, and Leonard M. Lodish (2012), "Traditional and IS-Enabled Customer Acquisition on the Internet," *Management Science*, in press.

Corcoran, Sean (2009), *Defining Earned, Owned, and Paid Media,* http://blogs.forrester.com.

Danaher, Peter (2007), "Modeling Page Views Across Multiple Websites with an Application to Internet Reach and Frequency Prediction," *Marketing Science*, 26 (3), 422-437.

Danaher, Peter, Janghyuk Lee, and Laoucine Kerbache (2010), "Optimal Internet Media Selection," *Marketing Science*, 29 (2), 336-347.

Danaher, Peter and Michael S. Smith (2011), "Modeling Multivariate Distributions Using Copulas: Applications in Marketing," *Marketing Science*, 30 (1), 4-21.

Dekimpe, Marnik G. and Dominique M. Hanssens (2004), "Persistence Modeling for Assessing Marketing Strategy Performance." In Christine Moorman and Donald R. Lehmann (eds), *Assessing Marketing Strategy Performance*, Cambridge, MA: Marketing Science Institute.

Denuit, Michel and Philippe Lambert (2005), "Constraints on Concordance Measures in Bivariate Discrete Data," *Journal of Multivariate Analysis*, 93 (1), 40-57.

Dong, Xiaojing, Pradeep K. Chintagunta, and Puneet Manchanda (2011), "A New Multivariate Count Data Model to Study Multi-Category Physician Prescription Behavior," *Quantitative Marketing and Economics*, 9 (3), 301-337.

Efron, Bradley (1986), "Double Exponential Families and Their Use in Generalized Linear Regression," *Journal of the American Statistical Association*, 81 (395), 709-721.

Elberse, Anita (2007), "The Power of Stars: Do Start Actors Drive the Success of Movies?" *Journal of Marketing*, 71 (4), 102-120.

Eliashberg, Jehoshua and Steven M. Shugan (1997), "Film Critics: Influencers or Predictors?" *Journal of Marketing*, 61 (2), 68-79.

Galak, Jeff, Deborah Small, and Andrew T. Stephen (2011), "Micro-Finance Decision Making: A Field Study of Prosocial Lending," *Journal of Marketing Research*, 48 (special issue), S130-S137.

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

Godes, David and Dina Mayzlin (2009), "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science*, 28 (4), 721-739.

Goel, Sharad, Jake M. Hofman, Sébastien Lahaie, David M. Pennock, and Duncan J. Watts (2011), "Predicting Consumer Behavior with Web Search," *Proceedings of the National Academy of Sciences*, 107 (41), 17486-17490.

Goodall, Daniel (2009), *Owned, Bought, and Earned Media*, Nokia, http://danielgoodall.com.

Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, 37 (3), 424-438.

Heinen, Andréas and Erick Rengifo (2007), "Multivariate Autoregressive Modeling of Time Series Count Data Using Copulas," *Journal of Empirical Finance*, 14, 564-583.

Karlis, Dimitris and Loukia Meligkotsidou (2005), "Multivariate Poisson Regression with Covariance Structure," *Statistics and Computing*, 15 (4), 255-265.

Katz, Elihu and Paul F. Lazarsfeld (1955), *Personal Influence,* Glencoe, IL: Free Press.

Kornish, Laura J. and Qiuping Li (2010), "Optimal Referral Bonuses with Asymmetric Information: Firm-Offered and Interpersonal Incentives," *Marketing Science*, 29 (1), 108-121.

Kulkarni, Gauri, P.K. Kannan and Wendy Moe (2012), "Using Online Search Data to Forecast New Product Sales, *Decision Support Systems,* 52 (3), 604-611.

Leclerc, France and John D. C. Little (1997), "Can Advertising Copy Make FSI Coupons More Effective?" *Journal of Marketing Research*, 34 (4), 473-484.

Libai, Barak, R. Bolton, M. Bügel, K. deRuyter, O. Goetz, H. Risselada and Andrew T. Stephen (2010), "Customer-to-Customer Interactions," *Journal of Service Research*, 13 (3), 267-282.

Onishi, Hiroshi and Puneet Manchanda (2011), "Marketing Activity, Blogging, and Sales," *International Journal of Research in Marketing*, forthcoming.

Park, Young-Hoon and Peter S. Fader (2004), "Modeling Browsing Behavior at Multiple Websites," *Marketing Science*, 23 (3), 280-303.

Patton, Andrew J. (2006), "Estimation of Multivariate Models for Time Series of Possibly Different Lengths," *Journal of Applied Econometrics*, 21, 147-173.

Rogers, Everett (2003), *Diffusion of Innovations*, New York, NY: Free Press.

Schmitt, Philipp, Bernd Skiera, and Christophe Van den Bulte (2011), "Referral Programs and Customer Value," *Journal of Marketing*, 75 (January), 46-59.

Schultz, Don E., Stanely I. Tannenbaum, and Robert F. Lauterborn (1993), *Integrated Marketing Communications*, Chicago, IL: NTC Business Books.

Sklar, A. (1959), *Fonctions de Repartitions à n Dimensions et Leurs Marges*, Public Institute of Statistics of the University of Paris 8, p. 229-231.

Sims, Christopher A. (1980), “Macroeconomics and Reality,” *Econometrica*, 48 (1), 1-48.

Sims, Christopher A. (1981), “An Autoregressive Index Model for the U.S. 1948-1975.” In J. Kmenta and J. B. Ramsey (eds), *Large-Scale Macro-Econometric Models*, Amsterdam: North-Holland.

Sims, Christopher A. (1986), “Are Forecasting Models Usable for Policy Analysis?” *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 2-16.

Sims, Christopher A. and Tao Zha (1999), “Error Bands for Impulse Responses,” *Econometrica*, 67 (5), 1113-1155.

Smith, Timothy M., Srinath Gopalakrishna, and Rabikar Chatterjee (2006), “A Three-Stage Model of Integrated Marketing Communications at the Marketing—Sales Interface,” *Journal of Marketing Research*, 43 (4), 564-579.

Stephen, Andrew T. and Olivier Toubia (2010), “Deriving Value from Social Commerce Networks,” *Journal of Marketing Research*, 47 (2), 215-228.

Trivedi, Pravin and D. Zimmer (2005), “Copula Modeling: An Introduction for Practitioners,” *Foundations and Trends in Econometrics*, 1, 1-111.

Trusov, Michael, Randolph Bucklin, and Koen Pauwels (2009), “Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site,” *Journal of Marketing*, 73 (5), 90-102.

Vakratsas, Demetrios and Zhenfeng Ma (2005), “A Look at the Long-Run Effectiveness of Multimedia Advertising and Its Implications for Budget Allocation Decisions,” *Journal of Advertising Research*, 45 (2), 241-254.

Van den Bulte, Christope and Gary L. Lilien (2001), "Medical Innovation Revisited: Social Contagion Versus Marketing Effort," *American Journal of Sociology*, 106 (5), 1409–1435.

Villanueva, Julian, Shijin Yoo, and Dominique M. Hanssens (2008), "The Impact of Marketing-Induced Versus Word-of-Mouth Customer Acquisition on Customer Equity Growth," *Journal of Marketing Research*, 45 (1), 48-59.

Wedel, Michel, Ulf Böckenholt, and Wagner A. Kamakura (2003), "Factor Models for Multivariate Count Data," *Journal of Multivariate Analysis*, 87, 356-369.

Winkelmann, Rainer (2008), *Econometric Analysis of Count Data*, Berlin, Germany: Springer.

FIGURE 1
DAILY TIME SERIES PLOTS FOR NEW AND REPEAT SALES FROM KIVA LOANS

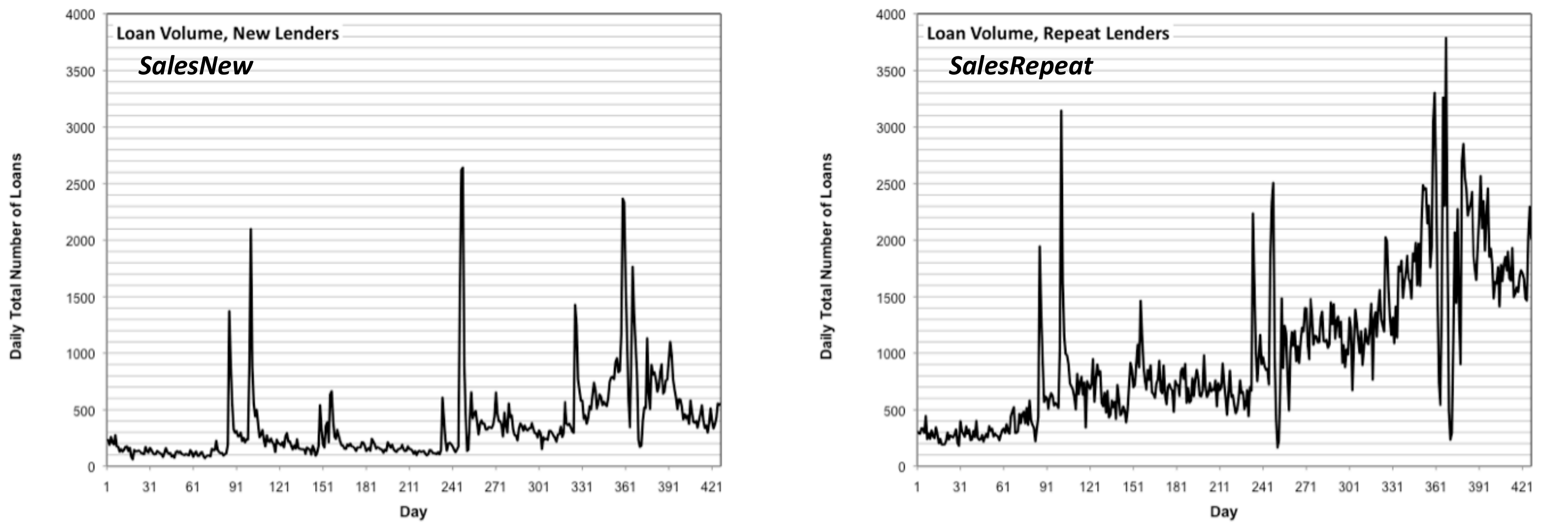


FIGURE 2
LONG-RUN ELASTICITIES OF TRADITIONAL AND SOCIAL EARNED MEDIA ACTIVITY ON SALES

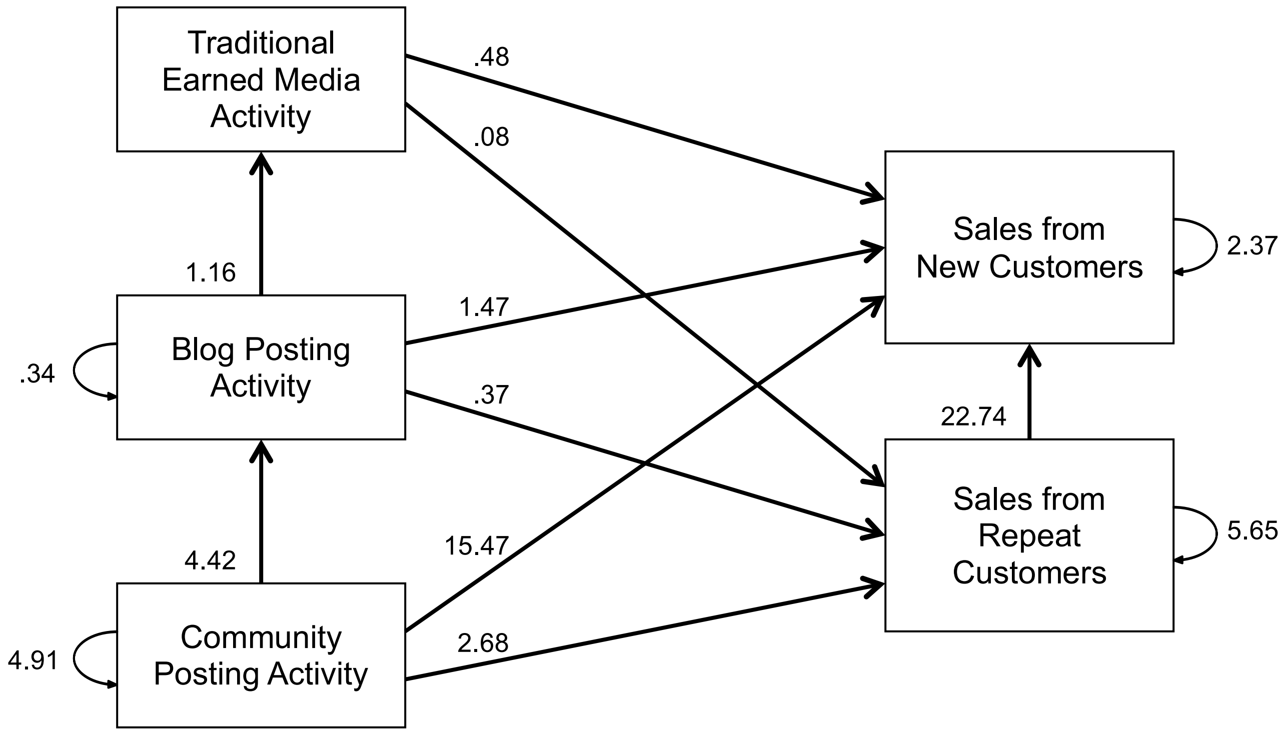


TABLE 1
PAID, OWNED, AND EARNED MEDIA

Type	Definition	Offline Examples	Online Examples
Paid	Media activity related to a company or brand that is generated by the company or its agents	• Traditional advertising (e.g., TV, radio, print, outdoor) • Sponsorships • Direct mail	• Display/banner advertising • Search advertising (e.g., Google AdWords) • Social network advertising (e.g., Facebook ads) • Electronic direct mail (e.g., email advertisements)
Owned	Media activity related to a company or brand that is generated by the company or its agents in channels it controls	• Retail in-store visual merchandising or displays • Brochures • Company press releases	• Company/brand website • Company/brand blog • Company-owned pages/accounts in online social networks (e.g., Twitter account, Facebook brand page)
Earned	Media activity related to a company or brand that is not directly generated by the company or its agents, but rather by other entities such as customers or journalists	• Traditional publicity mentions in professional media outlets • Ratings and reviews in traditional media outlets (e.g., movie reviews) • Consumer-to-consumer word-of-mouth conversations about products, including advice and referrals • Consumers showing or demonstrating products to each other	• Traditional publicity mentions in digital media outlets (e.g., professional blogs) • Online word-of-mouth referrals (e.g., invitations to join a website) • Posts in online communities or social networks (e.g., status updates, tweets) • Online ratings and reviews (e.g., Yelp.com for restaurants, Amazon.com for products)

TABLE 2
DESCRIPTIVE STATISTICS

Type	Variable	Definition	Mean (Standard Deviation)	Median	Minimum	Maximum
Sales	*SalesNew*	The total number of loans made by new (first-time) lenders on day *t*	356.99 (350.06)	237	60	2,641
	SalesRepeat	The total number of loans made by repeat lenders on day *t*	998.63 (654.59)	237	162	3,788
Traditional Earned Media	*Traditional*	The number of mentions in one of the four kinds of traditional media (newspapers, magazines, television, and radio) on day *t*	.20 (.48)	0	0	3
Social Earned Media	*Blogs*	The number of blog[a] posts made on day *t*	5.82 (2.94)	5	1	18
	CommPosts	The number of community forum[b] posts made on day *t*	55.88 (47.23)	50	0	213
	CommMembers	The number of member registrations for *Kiva* forum made on day *t*	5.15 (4.83)	4	0	28
Owned Media	*OwnedBlog*	The number of blog posts made on the *Kiva*-run blog on day *t*	.20 (.44)	0	0	2
	OwnedPR	The number of press releases issued by *Kiva* on day *t*	.01 (.07)	0	0	1
Other Variables	*Borrowers*	The number of loan requests on *Kiva* on day *t*	65.78 (38.74)	58	0	495
	SearchKiva	The Google Trends search interest index for *Kiva* on day *t*	1.34 (.51)	1.25	.75	4.20
	SearchTopic	The Google Trends search interest index for microfinance on day *t*	1.17 (.10)	1.16	.95	1.45

The time period for all variables is from January 1, 2007 to March 2, 2008 (427 days).
[a] Blogs includes all blog websites indexed by Google Blog Search and not produced/published by traditional/professional media outlets.
[b] Mostly covers the *Kivafriends* online community, as well as relevant posts indexed by Omgili.com and Google Groups forum search engines.

TABLE 3
PARAMETER ESTIMATES AND FIT STATISTICS FOR MAIN MODEL

		SalesNew **(*i* = 1)**	***SalesRepeat*** **(*i* = 2)**	***Traditional*** **(*i* = 3)**	***Blogs*** **(*i* = 4)**	***CommPosts*** **(*i* = 5)**	***CommMembers*** **(*i* = 6)**
Marginal model parameter estimates							
$SalesNew_{t\text{-}1}$	β_{1i}	.70***	-.01	.001**	-.001	<.001	.003***
$SalesRepeat_{t\text{-}1}$	β_{2i}	.04	.73***	<.001	.002***	.02***	<.001
$Traditional_{t\text{-}1}$	β_{3i}	31.72*	53.26*	.08	.12	-.01	.26
$Blogs_{t\text{-}1}$	β_{4i}	-.94	5.67	.04**	.29***	.001	-.08
$CommPosts_{t\text{-}1}$	β_{5i}	.58**	1.97***	<.001	.02***	.66***	.03***
$CommMembers_{t\text{-}1}$	β_{6i}	1.29	4.03	.02	.05	.13	.43***
$\Delta SearchKiva_{t\text{-}1}$	γ_{1i}	-46.15	-68.91	.003	.33	.11	.62
$\Delta SearchTopic_{t\text{-}1}$	γ_{2i}	229.30	-78.31***	.002	17.67**	.20	-3.71
$Borrowers_{t}$	γ_{3i}	.46***	1.68***	<.001	.03***	-.05***	.01
$Christmas_{t}$	γ_{4i}	40.76	24.17	.21	.61	-.02	<.001
$Weekend_{t}$	γ_{5i}	-3.19	-45.61*	.19	-.83***	.01	-.14
$OwnedBlog_{t\text{-}1}$	γ_{6i}	17.74	7.96	-.20	.02	.06	.05
$OwnedPR_{t\text{-}1}$	γ_{7i}	-104.67	-479.41	-17.23	3.44	.07	.01
Dispersion	ϕ_i	.02***	.01**	1.01***	.82**	.11***	.64***
Zero-inflation [a]	λ_i	—	—	1.54***	—	-.61***	-1.73***
Copula correlation matrix							
SalesNew		1.00					
SalesRepeat		.82	1.00				
Traditional		.13	.09	1.00			
Blogs		.42	.46	.05	1.00		
CommPosts		.62	.68	.00	.37	1.00	
CommMembers		.72	.70	.03	.39	.72	1.00
Model fit							
Pseudo-*R* [b]		.80	.85	.13	.48	.85	.77
Mean(ε_i)		-.006	-.001	-.003	.068	-.008	-.036
Variance(ε_i)		1.459	1.061	1.167	1.037	.999	1.174
LL, AIC, BIC [c]				-9000.01; 18174.02; 18682.64			

* $p < .10$, ** $p < .05$, *** $p < .001$. [a] Zero-inflation parameters were only estimated for i = 3, 5 and 6; other variables did not have any zero values. $\pi_i = \text{Logit}(\lambda_i)$. [b] Computed as the correlation between actual and predicted values. [c] Likelihood ratio test for the full model vs. a base model with own effects only: $\chi^2(30) = 742.30$, $p < .001$.

TABLE 4
LONG-RUN RETURNS FROM CUMULATIVE ORTHOGONALIZED IMPULSE RESPONSE FUNCTIONS

***A one-unit shock to this variable…* (Impulse)**	***…in the long-run generates this many extra units of this variable* (Response)**					
	***SalesNew* (*i* = 1)**	***SalesRepeat* (*i* = 2)**	***Traditional* (*i* = 3)**	***Blogs* (*i* = 4)**	***CommPosts* (*i* = 5)**	***CommMembers* (*i* = 6)**
SalesNew	2.37**	.00ns	.00ns	.00ns	.00ns	.00ns
SalesRepeat	8.13**	5.65**	.00ns	.00ns	.02ns	.00ns
Traditional	894.23**	402.77**	.00ns	.00ns	.00ns	.00ns
Blogs	90.15**	62.88**	.04**	.34**	.00ns	.00ns
CommPosts	98.83**	47.96**	-.47ns	.46**	4.91**	.02ns
CommMembers	.43ns	.12ns	-.03ns	.03ns	.18ns	.36**

Notes: (i) 95% (99%) of the long-run response is achieved 8 (12) weeks after an impulse. (ii) Orthogonalized IRFs are used to control for contemporaneous dependence between variables captured in the copula covariance matrix.
** Significant (95% confidence interval does not contain 0); ns Not significant (95% confidence interval contains 0)

TABLE 5
LONG-RUN ELASTICITIES

A 1% increase in this variable…	***…in the long-run generates the following percentage increase in activity in this variable***					
	***SalesNew* (*i* = 1)**	***SalesRepeat* (*i* = 2)**	***Traditional* (*i* = 3)**	***Blogs* (*i* = 4)**	***CommPosts* (*i* = 5)**	***CommMembers* (*i* = 6)**
SalesNew	2.37	—	—	—	—	—
SalesRepeat	22.74	5.65	—	—	—	—
Traditional	.48	.08	—	—	—	—
Blogs	1.47	.37	1.16	.34	—	—
CommPosts	15.47	2.68	—	4.42	4.91	—
CommMembers	—	—	—	—	—	.36

TABLE 6
LONG-RUN RETURNS WITH RESPECT TO AVERAGE DAILY ACTIVITY

A.

A one-unit increase in this variable…	***…in the long-run generates the following percentage increase in activity in this variable above average daily activity***					
	***SalesNew* (*i* = 1)**	***SalesRepeat* (*i* = 2)**	***Traditional* (*i* = 3)**	***Blogs* (*i* = 4)**	***CommPosts* (*i* = 5)**	***CommMembers* (*i* = 6)**
SalesNew	.66	—	—	—	—	—
SalesRepeat	2.28	.57	—	—	—	—
Traditional	250.49	40.33	—	—	—	—
Blogs	25.25	6.30	20.00	5.84	—	—
CommPosts	27.68	4.80	—	7.90	8.79	—
CommMembers	—	—	—	—	—	6.98

B.

A 1% increase in average daily activity in this variable…	***…in the long-run generates this many extra units of this variable***					
	***SalesNew* (*i* = 1)**	***SalesRepeat* (*i* = 2)**	***Traditional* (*i* = 3)**	***Blogs* (*i* = 4)**	***CommPosts* (*i* = 5)**	***CommMembers* (*i* = 6)**
SalesNew	8.46	—	—	—	—	—
SalesRepeat	81.19	56.42	—	—	—	—
Traditional	1.70	.77	—	—	—	—
Blogs	5.25	3.66	.002	.02	—	—
CommPosts	55.23	26.80	—	.26	2.74	—
CommMembers	—	—	—	—	—	.02

TABLE 7
COUNTERFACTUAL LONG-RUN EFFECTS ON SALES WHEN CROSS-MEDIA EFFECTS REMOVED FROM MODEL

	***Traditional* Removed**		***Blogs* Removed**		***CommPosts* Removed**		***CommMembers* Removed**	
Impulse Variable	***Sales New***	***Sales Repeat***	***Sales New***	***Sales Repeat***	***Sales New***	***Sales Repeat***	***Sales New***	***Sales Repeat***
SalesNew	2.37 (0%)	.00 (0%)	2.39 (+.8%)	.00 (0%)	2.38 (+.4%)	.00 (0%)	2.37 (0%)	.00 (0%)
SalesRepeat	8.03 (-1.2%)	5.58 (-1.2%)	7.54 (-7.3%)	5.23 (-7.3%)	5.95 (-26.8%)	4.13 (-26.9%)	7.62 (-6.3%)	5.29 (-6.4%)
Traditional	833.54 (-6.8%)	373.02 (-7.4%)	834.77 (-6.6%)	376.86 (-6.4%)	706.85 (-21.0%)	309.12 (-23.3%)	833.24 (-6.8%)	372.27 (-7.6%)
Blogs	87.52 (-2.9%)	61.05 (-2.9%)	64.46 (-28.5%)	44.96 (-28.5%)	68.62 (-23.9%)	47.86 (-23.9%)	86.26 (-4.3%)	60.17 (-4.3%)
CommPosts	97.53 (-1.3%)	47.31 (-1.4%)	90.10 (-8.8%)	44.16 (-7.9%)	63.30 (-36.0%)	30.25 (-36.9%)	90.90 (-8.0%)	44.00 (-8.3%)
CommMembers	.43 (0%)	.12 (0%)	.43 (0%)	.12 (0%)	.43 (0%)	.12 (0%)	.43 (0%)	.12 (0%)
Mean % change	-2.0%	-2.2%	-8.4%	-8.4%	-17.9%	-18.5%	-4.2%	-4.4%

Note: percent changes in long-run effects from original model (Table 4) to restricted model are shown in parentheses.

The Effects of Traditional and Social Earned Media on Sales:
A Study of a Microlending Marketplace

Andrew T. Stephen & Jeff Galak

WEB APPENDIX A

MULTIVARIATE NORMAL COPULA

This copula's distribution is given by $C(z_1,...,z_M;\Sigma) = \Phi^M\left(q_1,...,q_M;\Sigma\right)$, where Φ^M is the M-dimensional multivariate standard normal distribution function, and Σ is the variance-covariance matrix for the dependence between the M variables in the system. The q_i are the normal quantiles of the probability integral transforms of the data under the marginal densities; i.e., $q_i = \Phi^{-1}(z_i)$, where Φ^{-1} is the inverse standard normal distribution function, and z_i are the probability integral transforms (PITs) of the data. The copula's density is $c(q_t;\Sigma)$, which can be rewritten as $c(z_1,...,z_M;\Sigma) = |\Sigma|^{-1/2}\exp\left(\frac{1}{2}\mathbf{q}'(\mathbf{I}_M - \Sigma^{-1})\mathbf{q}\right)$, where $\mathbf{q}$ is the vector of q_i (for i = 1, …, M). Following Sklar's (1959) copula theorem, the joint density at time t of the M times series (with Θ as a vector of all parameters in the conditional mean equations) is simply the product of the products of the M univariate densities and the copula:

$$h(Y_{1t},...,Y_{Mt},\Theta,\Sigma) = \prod_{i=1}^{M} f_{mixture}(Y_{it},\mu_{it},\phi_i,\pi_i)\cdot c(z_{1t},...,z_{Mt};\Sigma) \tag{W1}$$

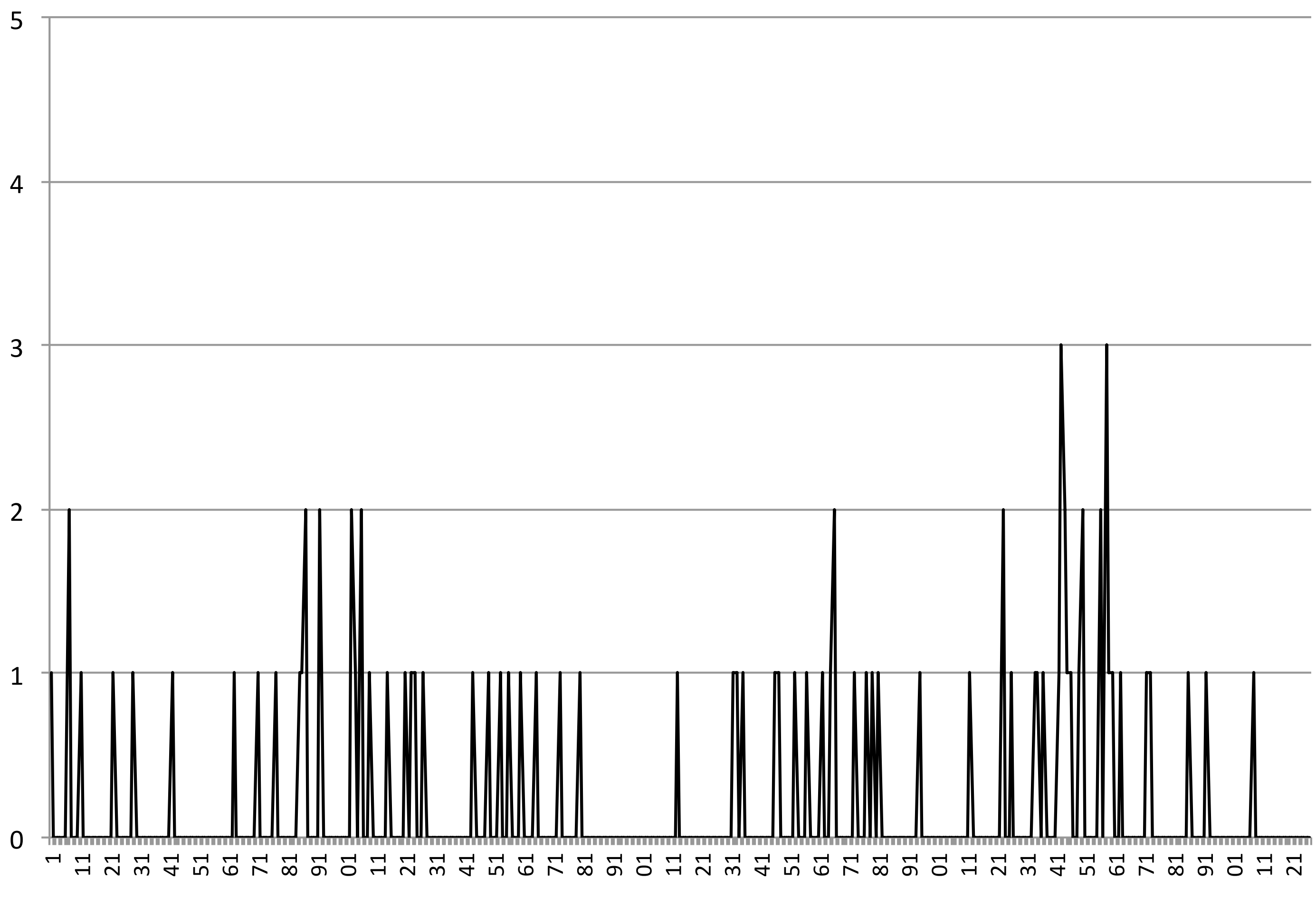
The above copula model works under the assumption of all marginal distributions being continuous, which is obviously not the case with count data. To overcome this we create a continued extension of each discrete variable that adds a continuous variable U to each discrete variable. U has a strictly increasing cdf, is valued in [0,1], is independent of the discrete variable, and has no parameters in common with the discrete variable's distribution (Denuit and Lambert

2005). We use $U \sim$ uniform(0,1), and for discrete variable Y create its continued extension $Y^* = Y + (U - 1)$. PITs are based on Y^*. Specifically, if z_{it} is the PIT of the continued extension of the DP-distributed variable i at time t it then follows, from Heinen and Rengifo (2007), that $z_{it} = F^*_{DP,it}(Y^*_{it}) = F_{DP,it}(Y_{it} - 1) + f_{DP,it}(Y_{it}) * U_{it}$, where U_{it} is a uniform(0,1) random variable.

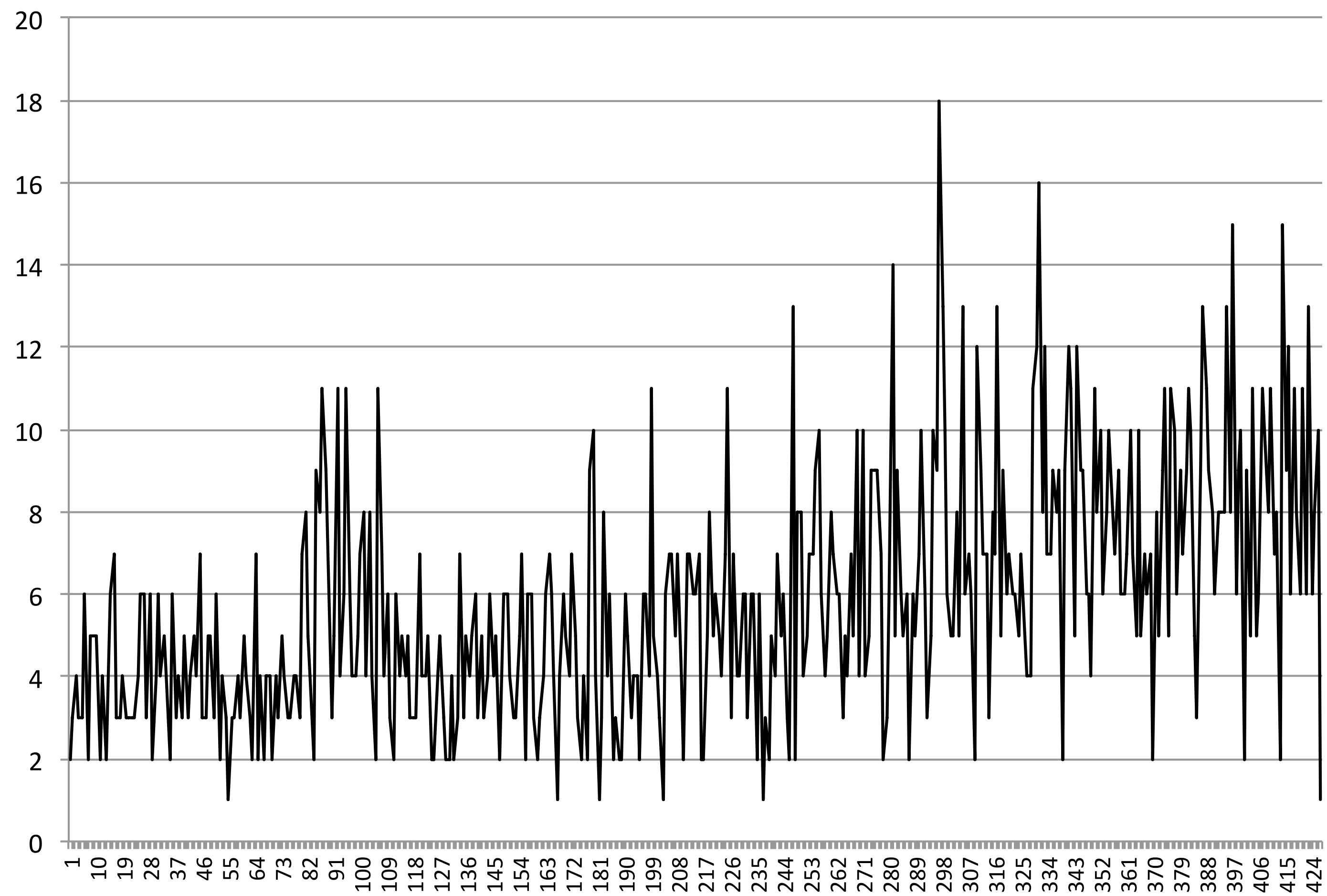
WEB APPENDIX B

ADDITIONAL TIME SERIES PLOTS AND CORRELATION MATRICES

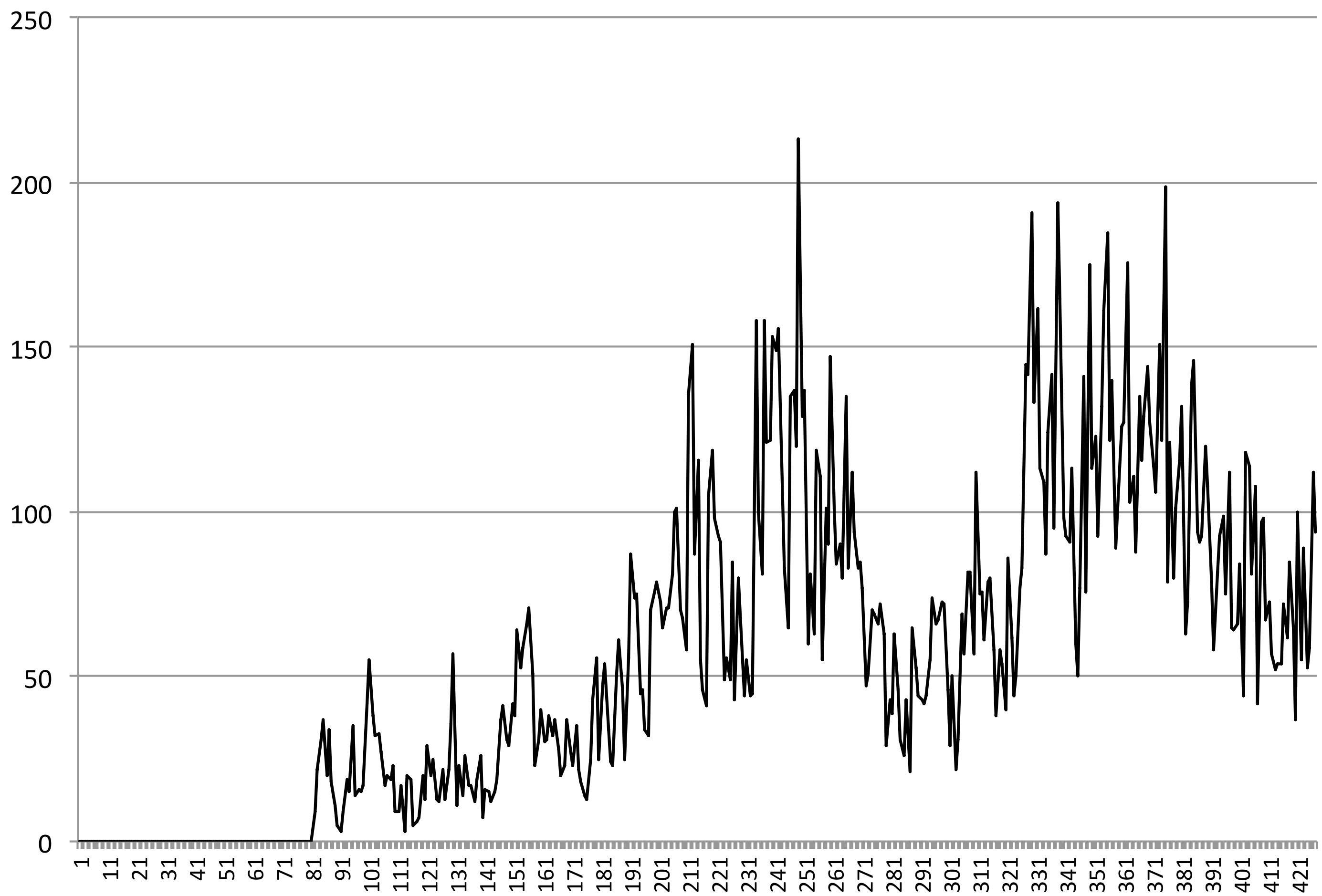
$Traditional_t$ – number of traditional earned media events on day *t*



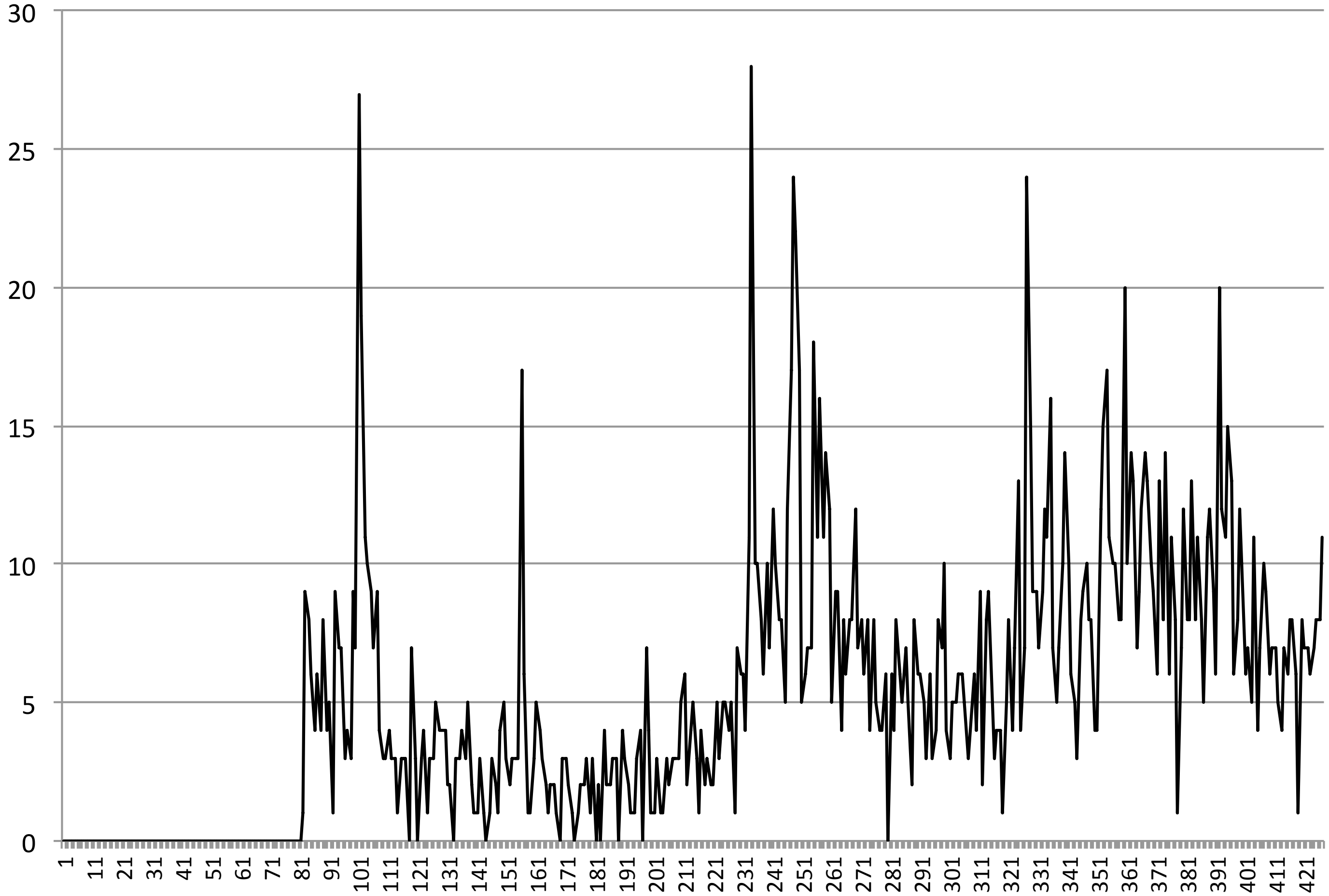
***Blogs**$_t$* – number of blog posts made on day *t*



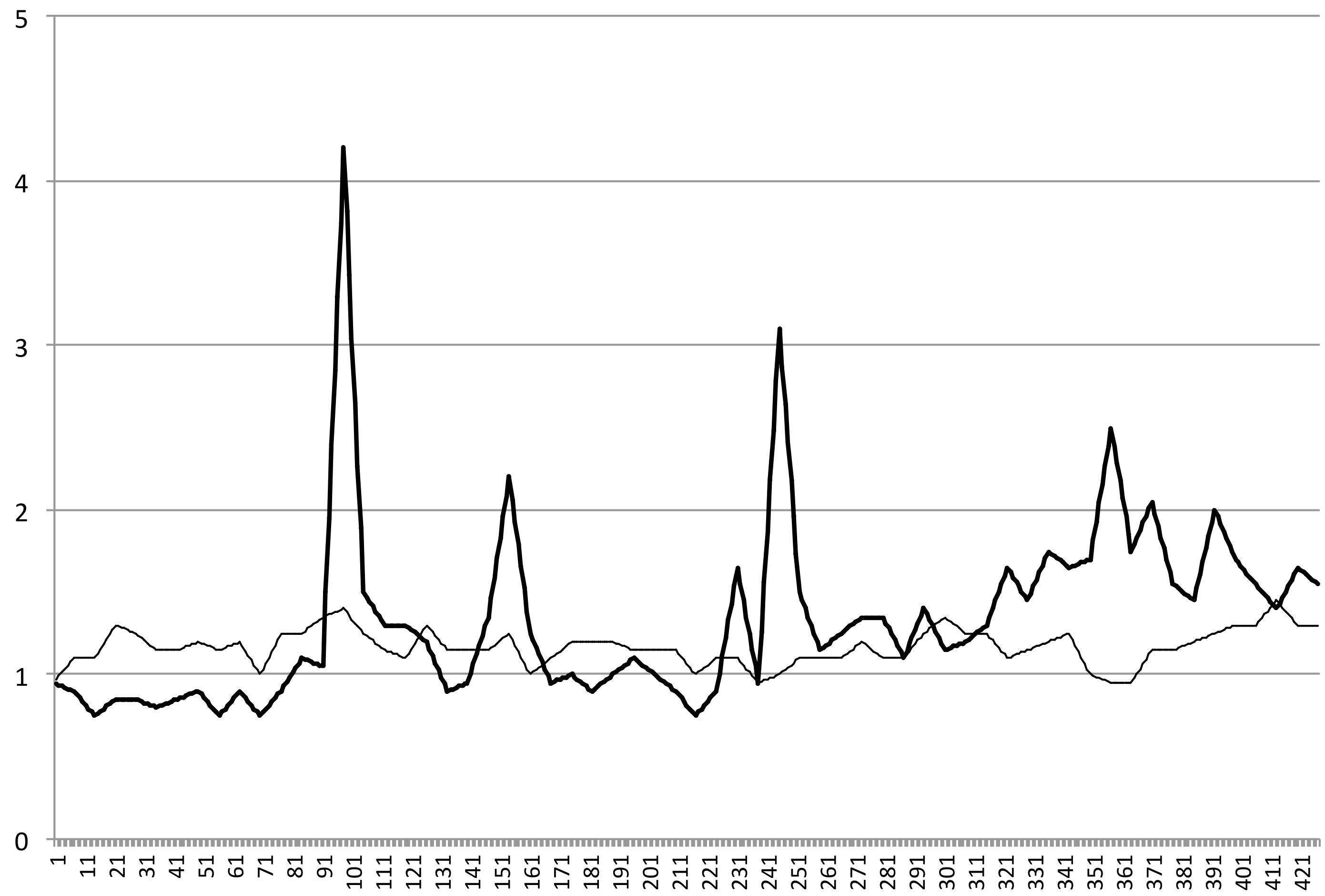
CommPosts$_t$ – number of posts made in online communities and discussion forums on day *t*



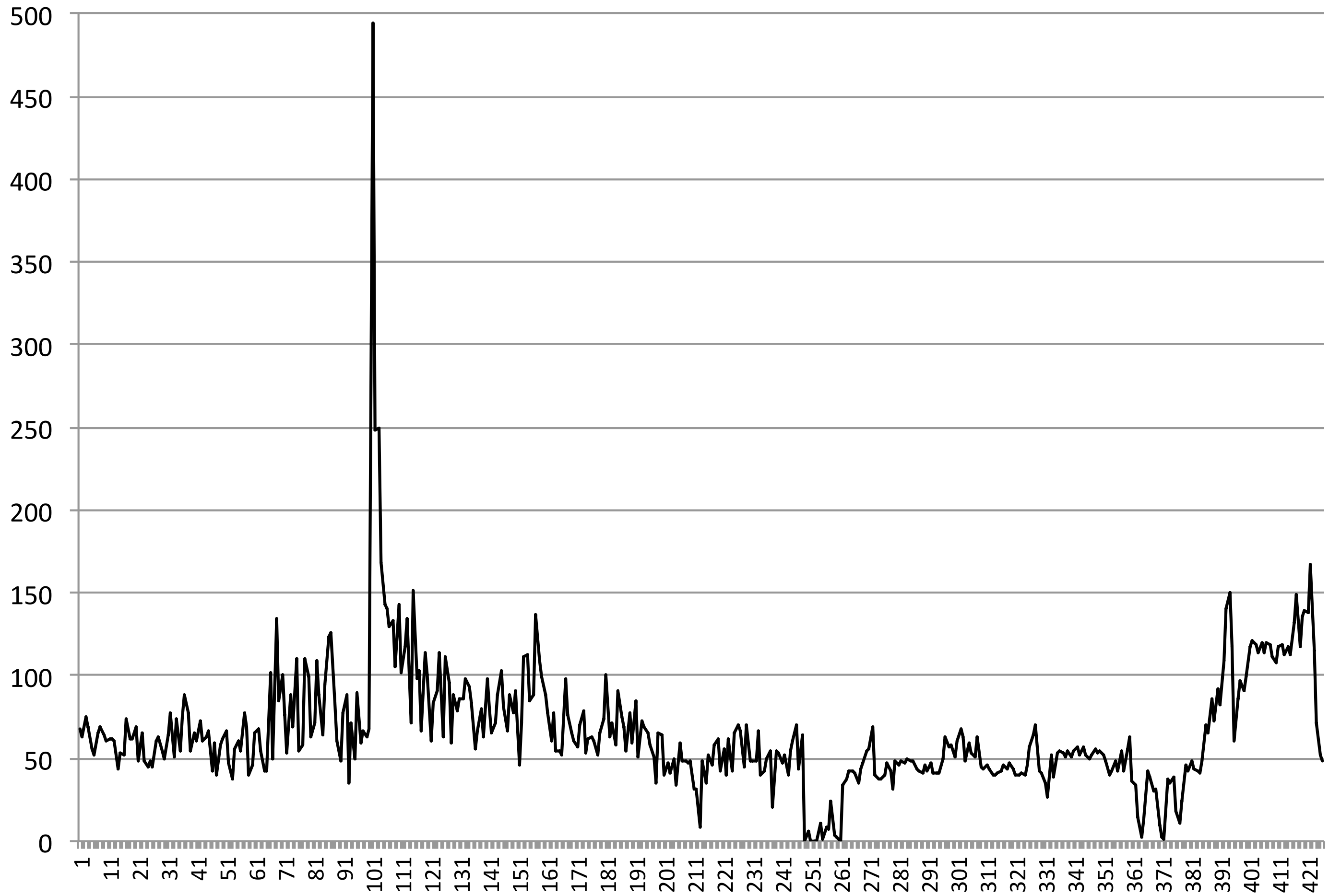
CommMembers$_t$ – number of new registrations for online community on day *t*



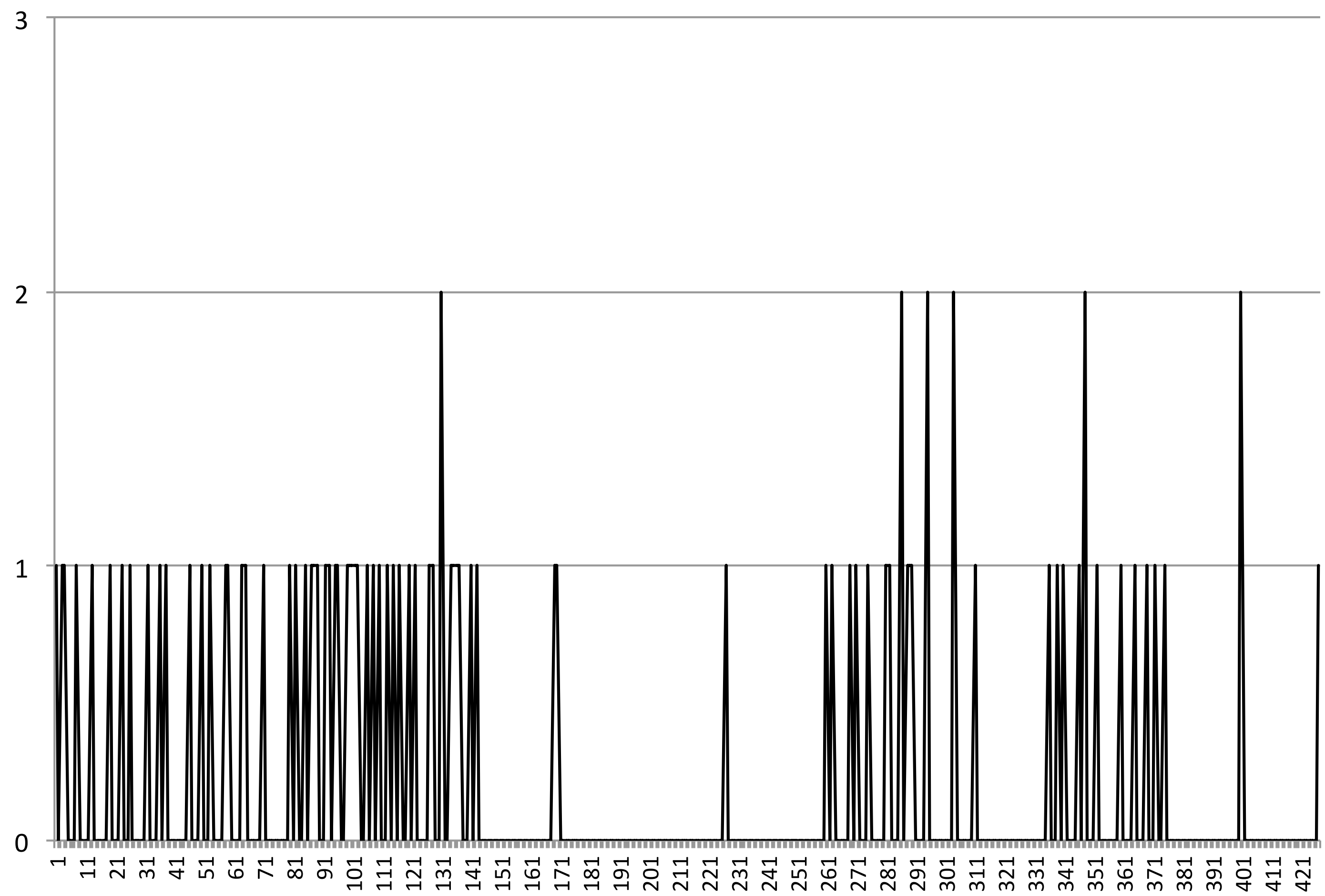
***SearchKiva**$_t$* (thicker line) and ***SearchTopic**$_t$* (thinner line) – indexed and normalized Google Trends scale



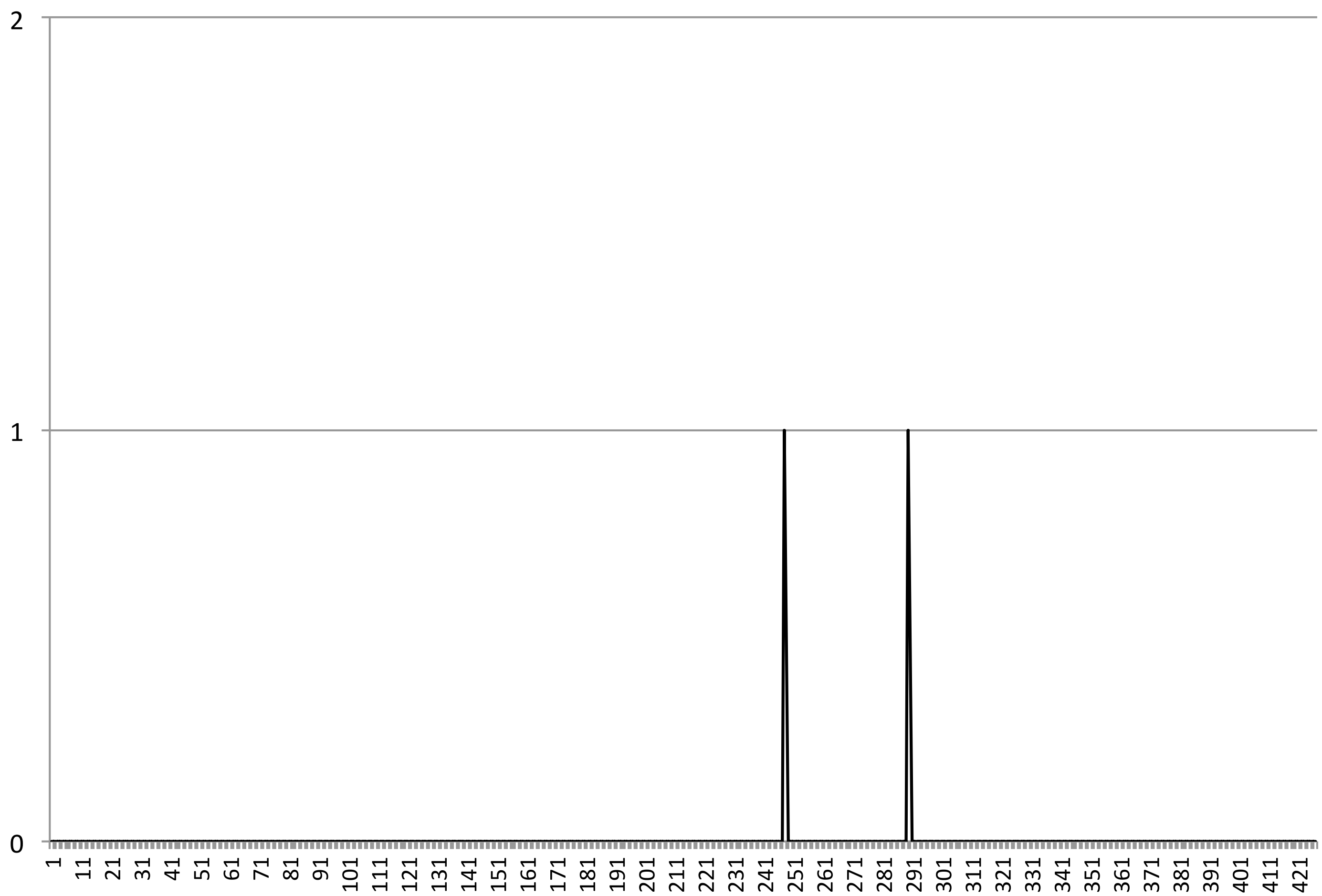
Borrowers$_t$ – number of loan requests made on *Kiva* on day *t*



OwnedBlog$_t$ – number of blog posts made on the official *Kiva* blog on day *t*



$OwnedPR_t$ – number of press releases issued by *Kiva* on day *t*



Correlations: Same day with same day

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. $SalesNew_t$	1.00												
2. $SalesRepeat_t$	.76	1.00											
3. $Traditional_t$	.20	.13	1.00										
4. $Blogs_t$	.30	.47	.09	1.00									
5. $CommPosts_t$	.50	.63	.04	.39	1.00								
6. $CommMembers_t$	.66	.68	.09	.36	.67	1.00							
7. $\Delta SearchKiva_t$	-.25	-.10	-.09	.03	-.03	-.19	1.00						
8. $\Delta SearchTopic_t$	.04	-.01	-.05	.02	.00	-.02	.22	1.00					
9. $Borrowers_t$	.10	.13	.04	.00	-.26	.04	-.20	-.15	1.00				
10. $Christmas_t$	.29	.18	.10	.04	.19	.13	-.03	.13	-.20	1.00			
11. $Weekend_t$	-.08	-.11	.01	-.30	-.03	-.08	.00	.00	-.05	.00	1.00		
12. $OwnedBlog_t$	.01	-.07	.02	.01	-.16	-.01	-.02	-.03	.11	.04	-.24	1.00	
13. $OwnedPR_t$	.21	.10	-.03	-.03	.04	.14	-.06	.08	-.02	-.02	-.04	.05	1.00

Correlations: Same day with previous day

	Lag(1)	Lag(2)	Lag(3)	Lag(4)	Lag(5)	Lag(6)	Lag(7)	Lag(8)	Lag(9)	Lag(10)	Lag(11)	Lag(12)	Lag(13)
1. $SalesNew_t$	.78	.67	.25	.31	.47	.51	-.21	.03	-.01	.29	-.03	.00	.05
2. $SalesRepeat_t$	.67	.83	.16	.44	.59	.58	-.09	-.03	.03	.18	-.06	-.05	.01
3. $Traditional_t$	.13	.10	.14	.04	.11	.08	-.09	-.03	.02	.08	.16	-.02	-.03
4. $Blogs_t$	.33	.45	.05	.32	.39	.35	-.02	.05	-.04	.05	-.03	-.07	.07
5. $CommPosts_t$	.53	.64	.08	.36	.84	.68	-.03	-.01	-.26	.19	-.06	-.16	.10
6. $CommMembers_t$	.68	.65	.13	.35	.61	.70	-.15	-.03	-.01	.13	-.09	-.04	.13
7. $\Delta SearchKiva_t$	-.25	-.10	-.09	.02	-.02	-.18	.79	.17	-.21	-.04	.00	-.02	-.06
8. $\Delta SearchTopic_t$	.05	-.01	-.10	.01	.00	-.02	.17	.82	-.14	.15	-.01	-.02	.08
9. $Borrowers_t$	.05	.06	.01	-.01	-.27	-.03	-.21	-.15	.67	-.20	-.03	.09	-.08
10. $Christmas_t$	.31	.19	.13	.06	.21	.14	-.02	.11	-.19	.96	-.01	.05	-.02
11. $Weekend_t$	-.09	-.06	-.06	-.15	-.01	-.01	.00	.00	-.01	.00	.30	-.08	-.04
12. $OwnedBlog_t$	.01	-.03	.00	-.02	-.14	-.02	-.02	-.04	.10	.03	-.02	.02	-.03
13. $OwnedPR_t$	.22	.08	.04	.07	.06	.10	-.06	.08	-.04	-.02	-.04	.05	.00

Correlations: Previous day with previous day

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. $SalesNew_{t-1}$	1.00												
2. $SalesRepeat_{t-1}$	.80	1.00											
3. $Traditional_{t-1}$	.21	.13	1.00										
4. $Blogs_{t-1}$	.33	.47	.09	1.00									
5. $CommPosts_{t-1}$	.50	.63	.04	.40	1.00								
6. $CommMembers_{t-1}$	.67	.69	.09	.37	.67	1.00							
7. $\Delta SearchKiva_{t-1}$	-.25	-.10	-.09	.03	-.03	-.19	1.00						
8. $\Delta SearchTopic_{t-1}$	.05	-.01	-.05	.02	.00	-.02	.22	1.00					
9. $Borrowers_{t-1}$	.12	.13	.04	-.01	-.26	.04	-.20	-.15	1.00				
10. $Christmas_{t-1}$	.30	.18	.10	.04	.19	.13	-.03	.13	-.20	1.00			
11. $Weekend_{t-1}$	-.10	-.10	.02	-.29	-.04	-.08	.00	.00	-.05	.00	1.00		
12. $OwnedBlog_{t-1}$	-.01	-.06	.03	.01	-.17	-.02	-.02	-.03	.12	.05	-.24	1.00	
13. $OwnedPR_{t-1}$	.22	.10	-.03	-.03	.04	.14	-.06	.08	-.02	-.02	-.04	.05	1.00

Correlations: With dummy variables for days that had an Oprah or feature New York Times traditional earned media event

	Oprah	NYT
1. $SalesNew_t$	.32	.14
2. $SalesRepeat_t$	.12	.07
3. $Traditional_t$	.12	.08
4. $Blogs_t$	.07	.04
5. $CommPosts_t$	.08	-.04
6. $CommMembers_t$	.10	-.01
7. $\Delta SearchKiva_t$	-.09	-.01
8. $\Delta SearchTopic_t$	.04	.04
9. $Borrowers_t$	-.03	.07
10. $Christmas_t$	-.02	-.02
11. $Weekend_t$	-.04	-.03
12. $OwnedBlog_t$	-.03	-.02
13. $OwnedPR_t$	-.01	-.01

	Oprah	NYT
1. $SalesNew_{t-1}$	.11	-.02
2. $SalesRepeat_{t-1}$	.06	-.04
3. $Traditional_{t-1}$	.04	.08
4. $Blogs_{t-1}$	-.03	.05
5. $CommPosts_{t-1}$	.07	-.02
6. $CommMembers_{t-1}$	.04	.01
7. $\Delta SearchKiva_{t-1}$	-.09	-.01
8. $\Delta SearchTopic_{t-1}$	.05	.04
9. $Borrowers_{t-1}$	-.01	.04
10. $Christmas_{t-1}$	-.02	-.02
11. $Weekend_{t-1}$	-.04	-.03
12. $OwnedBlog_{t-1}$	-.03	.09
13. $OwnedPR_{t-1}$	-.01	-.01

	Lag (Oprah)	Lag (NYT)
1. $SalesNew_t$	.30	.07
2. $SalesRepeat_t$	.13	.03
3. $Traditional_t$	-.03	.18
4. $Blogs_t$	-.07	.09
5. $CommPosts_t$	.11	-.02
6. $CommMembers_t$	.27	.01
7. $\Delta SearchKiva_t$	-.09	-.01
8. $\Delta SearchTopic_t$	.05	.04
9. $Borrowers_t$	-.01	.08
10. $Christmas_t$	-.02	-.02
11. $Weekend_t$	-.04	-.03
12. $OwnedBlog_t$	-.03	.09
13. $OwnedPR_t$	.50	-.01

WEB APPENDIX C

DIRECT CONTEMPORANEOUS EFFECTS ROBUSTNESS CHECK

Two approaches were considered for capturing contemporaneous relations between endogenous variables in our M =6 system of equations. The first approach was the one used in the paper and based on standard reduced-form multivariate time series models: estimate the model with lagged effects, use those estimates to estimate copula correlations to capture contemporaneous relations between endogenous variables, and then use the estimated lagged effects and copula correlations to compute cumulative orthogonalized impulse-response functions that provide estimates of long-run effects accounting for contemporaneous effects.

The second approach was related to the structural vector autoregression (SVAR) models developed in the macroeconomics literature (e.g., Bernanke 1986; Sims 1980, 1986) and was used as a robustness check. This involved estimating a model where, in each marginal model, direct contemporaneous other-variable effects (e.g., effects of $SalesRepeat_t$, $Traditional_t$, $Blogs_t$, and $CommPosts_t$ on $SalesNew_t$) and autoregressive lag effects were estimated together and then used to compute cumulative impulse-response functions for estimating long-run effects. We did not estimate copula correlations between the M = 6 equations because this model specification assumes that contemporaneous relations between endogenous variables are picked up by the direct contemporaneous effects. This assumption also helps with model identification. Specifically, for M = 6 equations, $(M^2 - M)/2 = 15$ identification restrictions are needed (Bernanke 1986). Thus, we assumed that the marginal models were orthogonal (i.e., uncorrelated), thus making a copula model unnecessary. An alternative would be to restrict the model parameters themselves, although that would require stronger theory-based assumptions.

For the SVAR-type robustness check model we report results in Figure C1 and Tables C1, C2, C3, and C4. In Table C1 we report parameter estimates (similar to Table 3 in the paper). In Figure C1 we summarize the long-run elasticities (similar to Figure 2 in the paper). Tables C2, C3, and C4 report long-run effects based on computed cumulative IRFs (similar to Tables 4, 5, and 6 in the paper). A number of direct contemporaneous effects reported in Table C1 are significant and, overall, the long-run results reported in Figure C1 and Tables C2, C3, and C4 are similar to and conceptually consistent with those derived from the main model in the paper.

FIGURE C1
LONG-RUN ELASTICITIES OF TRADITIONAL AND SOCIAL EARNED MEDIA ACTIVITY ON SALES FOR DIRECT CONTEMPORANEOUS EFFECTS MODEL

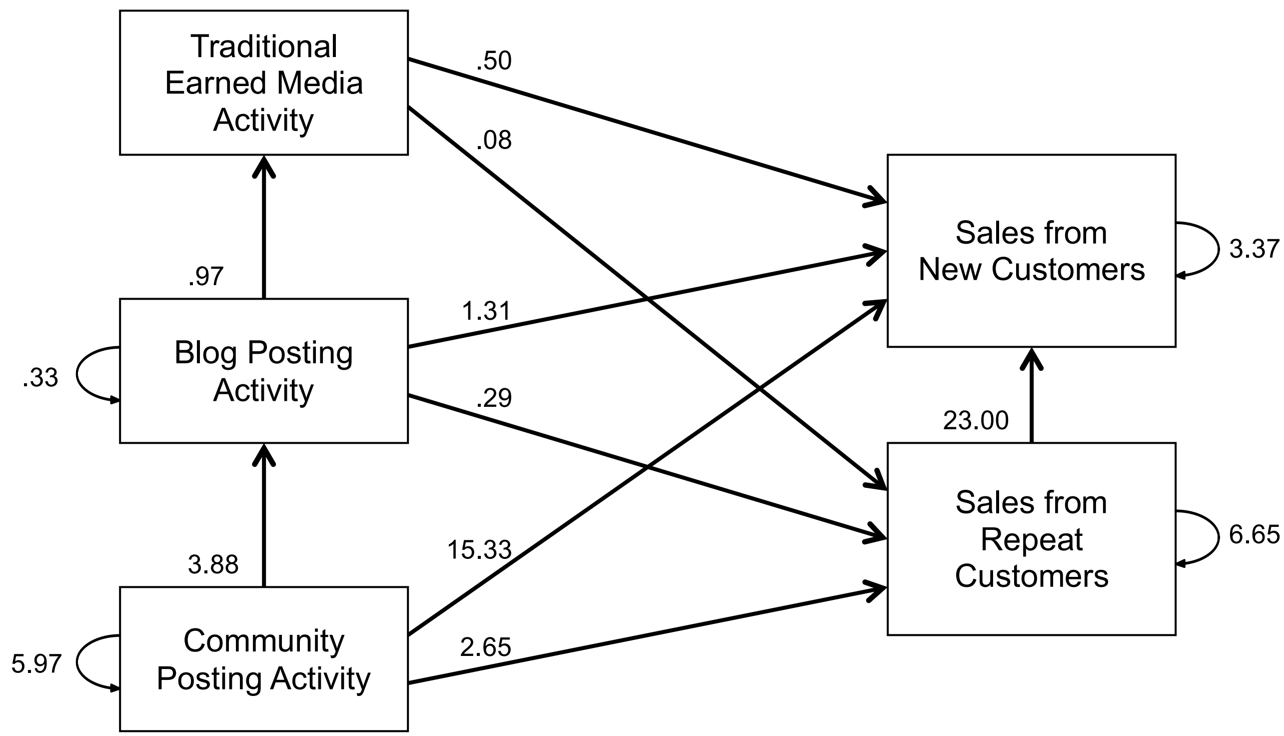


TABLE C1
PARAMETER ESTIMATES AND FIT STATISTICS FOR DIRECT CONTEMPORANEOUS EFFECTS MODEL

		***SalesNew* (i = 1)**	***SalesRepeat* (i = 2)**	***Traditional* (i = 3)**	***Blogs* (i = 4)**	***CommPosts* (i = 5)**	***CommMembers* (i = 6)**
Estimates for lagged and exogenous effects							
$SalesNew_{t\text{-}1}$	β_{1i}	$.58^{***}$	$-.84^{**}$	<.001	<.001	<.001	$.001^{***}$
$SalesRepeat_{t\text{-}1}$	β_{2i}	$-.05^{**}$	$.70^{**}$	<.001	.001	$.02^{***}$	$<.001^{***}$
$Traditional_{t\text{-}1}$	β_{3i}	7.22^{**}	2.88	.08	.06	.01	.16
$Blogs_{t\text{-}1}$	β_{4i}	1.28	3.89	$.04^{**}$	$.27^{***}$	.001	$.05^{**}$
$CommPosts_{t\text{-}1}$	β_{5i}	-.15	-.35	<.001	$.01^{**}$	$.43^{***}$	.001
$CommMembers_{t\text{-}1}$	β_{6i}	5.80^{**}	4.98	.01	.01	.03	$.70^{***}$
$\Delta SearchKiva_{t\text{-}1}$	γ_{1i}	-59.70	7.87	.001	.31	.11	.57
$\Delta SearchTopic_{t\text{-}1}$	γ_{2i}	155.59	-275.13	.002	18.65^{**}	.20	-3.70
$Borrowers_t$	γ_{3i}	$.19^{*}$	$.77^{**}$	<.001	$.03^{***}$	$-.11^{***}$	.01
$Christmas_t$	γ_{4i}	40.50^{*}	-29.91	.20	.52	-.01	.01
$Weekend_t$	γ_{5i}	-16.38	-22.63	.17	$-.77^{***}$	.02	-.03
$OwnedBlog_{t\text{-}1}$	γ_{6i}	17.70	7.94	-.21	.02	.05	.04
$OwnedPR_{t\text{-}1}$	γ_{7i}	-74.13	-479.36	-17.17	3.42	.07	.01
Dispersion	ϕ_i	$.02^{***}$	$.02^{***}$	1.02^{***}	$.84^{***}$	$.30^{***}$	$.64^{***}$
Zero-inflation [a]	λ_i	—	—	1.55^{***}	—	$-.10^{***}$	-1.50^{***}
Estimates for contemporaneous effects							
SalesNew		—	1.11^{***}	-.01	$.01^{*}$	.01	$.02^{***}$
SalesRepeat		$.21^{***}$	—	.09	$.01^{**}$	$.02^{***}$	$.01^{***}$
Traditional		41.85^{***}	-2.61	—	.00	.00	.00
Blogs		-2.58	6.26^{*}	.01	—	.11	.11
CommPosts		-.15	1.42^{***}	.00	$.01^{**}$	—	$.06^{***}$
CommMembers		9.27^{***}	11.92^{***}	.01	.01	.33	—
Model fit							
Pseudo-R [b]		.87	.90	.12	.50	.84	.77
Mean(ε_i)		-.007	-.002	-1.289	.070	-.008	-.668
Variance(ε_i)		1.23	.95	.76	1.04	.99	.83
LL, AIC, BIC		-10373; 20981; 21665					

* $p < .10$, ** $p < .05$, *** $p < .001$. [a] Zero-inflation parameters were only estimated for i = 3, 5 and 6; other variables did not have any zero values. π_i = Logit(λ_i). [b] Computed as the correlation between actual and predicted values.

TABLE C2
LONG-RUN RETURNS FROM CUMULATIVE IMPULSE RESPONSE FUNCTIONS FOR DIRECT CONTEMPORANEOUS EFFECTS MODEL

***A one-unit shock to this variable…* (Impulse)**	***…in the long-run generates this many extra units of this variable* (Response)**					
	***SalesNew* (*i* = 1)**	***SalesRepeat* (*i* = 2)**	***Traditional* (*i* = 3)**	***Blogs* (*i* = 4)**	***CommPosts* (*i* = 5)**	***CommMembers* (*i* = 6)**
SalesNew	3.37**	.42ns	.00ns	.00ns	.00ns	.01ns
SalesRepeat	8.22**	6.65**	.00ns	.00ns	.06ns	.00ns
Traditional	939.33**	403.02**	.00ns	.00ns	.00ns	.00ns
Blogs	80.29**	49.08**	.04**	.33**	.00ns	.30ns
CommPosts	97.93**	47.28**	-.18ns	.40**	5.97**	.04ns
CommMembers	.43ns	.13ns	-.00ns	.02ns	3.72ns	1.35**

Notes: 95% (99%) of the long-run response is achieved 8 (12) weeks after an impulse. ** Significant (95% confidence interval does not contain 0); ns Not significant (95% confidence interval contains 0)

TABLE C3
LONG-RUN ELASTICITIES FOR DIRECT CONTEMPORANEOUS EFFECTS MODEL

A 1% increase in this variable…	***…in the long-run generates the following percentage increase in activity in this variable***					
	***SalesNew* (*i* = 1)**	***SalesRepeat* (*i* = 2)**	***Traditional* (*i* = 3)**	***Blogs* (*i* = 4)**	***CommPosts* (*i* = 5)**	***CommMembers* (*i* = 6)**
SalesNew	3.37	—	—	—	—	—
SalesRepeat	23.00	6.65	—	—	—	—
Traditional	.50	.08	—	—	—	—
Blogs	1.31	.29	.97	.33	—	—
CommPosts	15.33	2.65	—	3.88	5.97	—
CommMembers	—	—	—	—	—	1.35

TABLE C4
LONG-RUN RETURNS WITH RESPECT TO AVERAGE DAILY ACTIVITY FOR DIRECT CONTEMPORANEOUS EFFECTS MODEL

A.

A one-unit increase in this variable…	***…in the long-run generates the following percentage increase in activity in this variable above average daily activity***					
	***SalesNew* (*i* = 1)**	***SalesRepeat* (*i* = 2)**	***Traditional* (*i* = 3)**	***Blogs* (*i* = 4)**	***CommPosts* (*i* = 5)**	***CommMembers* (*i* = 6)**
SalesNew	.94	—	—	—	—	—
SalesRepeat	2.30	.67	—	—	—	—
Traditional	263.13	40.36	—	—	—	—
Blogs	22.49	4.91	16.75	5.72	—	—
CommPosts	27.43	4.73	—	6.94	10.68	—
CommMembers	—	—	—	—	—	26.18

B.

A 1% increase in average daily activity in this variable…	***…in the long-run generates this many extra units of this variable***					
	***SalesNew* (*i* = 1)**	***SalesRepeat* (*i* = 2)**	***Traditional* (*i* = 3)**	***Blogs* (*i* = 4)**	***CommPosts* (*i* = 5)**	***CommMembers* (*i* = 6)**
SalesNew	12.04	—	—	—	—	—
SalesRepeat	82.10	66.42	—	—	—	—
Traditional	1.78	.77	—	—	—	—
Blogs	4.67	2.86	.002	.02	—	—
CommPosts	54.73	26.42	—	.23	3.33	—
CommMembers	—	—	—	—	—	.07