The Complementary Roles of Traditional and Social Media Publicity in Driving Marketing Performance

Andrew T. STEPHEN
Jeff GALAK
2010/97/MKT
(Revised version of 2009/52/MKT)
The Complementary Roles of Traditional and Social Media Publicity in Driving Marketing Performance

Andrew T. Stephen*

Jeff Galak**

Revised version of 2009/52/MKT

October 5, 2010

The authors thank Peter Boatwright, Jacob Goldenberg, Don Lehmann, Olivier Toubia, Kaifu Zhang, Puja Agarwal, Inyoung Chae, and seminar participants at Carnegie Mellon University and INSEAD for their helpful suggestions and comments, and Kiva.org for providing data.

* Assistant Professor of Marketing at INSEAD, Boulevard de Constance, 77305 Fontainebleau, France. Ph: +33 (0)1 60 71 26 47 Email: andrew.stephen@insead.edu

** Assistant Professor of Marketing at Tepper School of Business, Carnegie Mellon University 5000 Forbes Avenue, Pittsburgh, PA 15213, USA. Ph: +1 (412) 268 5810 Email: jgalak@cmu.edu

A Working Paper is the author’s intellectual property. It is intended as a means to promote research to interested readers. Its content should not be copied or hosted on any server without written permission from publications.fb@insead.edu

Click here to access the INSEAD Working Paper collection
THE COMPLEMENTARY ROLES OF TRADITIONAL AND SOCIAL MEDIA
PUBLICITY IN DRIVING MARKETING PERFORMANCE

ABSTRACT

The media landscape has dramatically changed, with traditional media (e.g., newspapers, television) now supplemented by social media (e.g., blogs, online communities). This situation is not well understood with respect to the relative impacts of these media types on marketing performance (e.g., sales), and how they influence each other. These issues are examined using 14 months of daily count data for sales and media activity for a micro-lending website. Multivariate time series count data pose a number of statistical challenges, which are overcome using a copula-based multivariate autoregressive count model. The authors find that both traditional and social media affect sales, directly and indirectly through effects on each other. While the unit sales impact for traditional media is larger than for social media, the greater frequency of social media activity results in it having a comparable effect to traditional media in the case of blogs, and a larger effect in the case of online communities. Overall, the results emphasize the critical role that interactive, conversational online social media plays in driving sales.

Keywords: traditional media, social media, marketing performance, sales, publicity, media modeling, multivariate time series, count data, copula model.
Over the past decade, the media landscape has dramatically changed with social media outlets (SMOs; e.g., blogs, online discussion forums, and online communities) now supplementing traditional media outlets (TMOs; e.g., newspapers, magazines, and television programs). To the extent that these media outlets each have an effect on marketing performance (e.g., sales), it is critical to understand their relative importance and their interrelatedness. 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 such social media (Bernoff, Pflaum, and Bowen 2008; Miller 2009a; Owyang, Bernoff, Van Boshirk, Pflaum, and Polanco 2009). This large number of users makes it critical to understand not only how social media influences consumers, but also how it operates alongside traditional media. This is the primary focus of this paper. Specifically, we examine how media publicity—sometimes called “earned” media—generated by both traditional and social media outlets affects marketing performance.

Despite the importance of this issue, the generation of such publicity and how it impacts firms’ marketing performance is not yet well understood. Past research, for example, has demonstrated that traditional media publicity can affect marketing outcomes (e.g., the literature on celebrity endorsement and star power; Agarwal and Kamakura 1995; Elberse 2007), that online user-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 impact of traditional or social media on marketing performance in isolation, thus neglecting potential effects that traditional and social
media can have on each other. Although some prior work has examined interrelations between media vehicles, this has typically been within a single broad channel such as Internet advertising (e.g., ad placements across many websites) and not across a variety of media channels (e.g., Danaher 2007; Danaher, Lee, and Kerbache 2010). In order to properly understand the total impact of both traditional and social media sources, an integrated perspective is needed that jointly considers how they affect each other and sales.

We address the following questions in this paper: (i) what are the relative impacts of traditional media and online social media on marketing performance? (ii) In what ways do these media types influence each other? And (iii) how similar or different are the roles of traditional media and social media in driving marketing performance? As we later demonstrate, answering these questions leads to novel and interesting insights on how both traditional and social media operate, and how different types of media outlets fulfill different (but potentially complementary) roles in influencing customer behavior.

We use a novel dataset that features 14 months of time series data covering the daily sales and “earned” media publicity activity\(^1\) across multiple TMOs and SMOs for a microfinance website, Kiva.org. A sale on this website is in the form of a small (minimum $25, average $33), low-risk loan made by an individual to a qualified entrepreneur in a developing country. We use the number of loans made on the website each day as our measure of daily sales volume, and are able to break it down into whether the transaction came from a new customer (first-time lender) or an existing one. The media activity data consists of daily counts of activity from a variety of TMOs (newspapers, magazines, television, radio) and SMOs (blogs, online community/forums) over the 14 months for which sales data were available.

\(^1\) Media activity is any reference to a company, brand, product or service by a media outlet in an article, program, post, or story.
We address our research questions by comparing the effects of media activity from multiple TMOs and SMOs on sales (i.e., the sales response to media activity) and themselves (i.e., endogenous media response to media activity). Through endogenous sales response to sales activity we are also able to infer implicit word-of-mouth (WOM) effects on sales that are not explicitly captured through social media activity. We model this endogenous system using a multivariate time series model for count data that captures the effects of past activity on present activity through a system of autoregressive models, and captures the contemporaneous dependence between variables through a copula model.

The paper is organized as follows. First, we provide an overview of extant literature on publicity, advertising, and media effects on marketing outcomes and outline our conceptual framework. Second, we describe our data. Third, we present the statistical model used to analyze these data. Fourth, we report the results of our empirical analyses. Finally, we conclude with a discussion of our findings and their implications.

**BACKGROUND AND THEORY**

**Past Research on Media Effects on Marketing Outcomes**

Little research has examined marketing- and financially-relevant consequences of publicity. In contrast, advertising (“paid media”) effects on marketing outcomes (typically sales) have received much attention. Also, with few exceptions (e.g., Danaher et al. 2010), the relationships between media types and outlets have not been extensively studied—particularly from a publicity-generation perspective (much of the extant literature focuses on advertising).

*Word-of-mouth research.* Recently, there has been a growing interest in understanding how WOM, particularly online WOM (which is a form of social media), impacts sales, diffusion,
and other marketing performance measures (e.g., Chevalier and Mayzlin 2006; Godes and Mayzlin 2004; Trusov, Bucklin and Pauwels 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, and Trusov et al. (2009) examine how referrals (invitations) to join an online social network affect website membership growth.

This growing body of literature on the impacts of various forms of WOM suggests that WOM—or, more generally, social media—indeed can affect relevant marketing outcomes. Often, however, only one type of WOM is examined in this literature, and is not contrasted with traditional media, making it difficult to compare the relative sizes of the effects of traditional versus social media on key marketing outcomes. In line with a recent call for more comprehensive and multi-faceted research on WOM (Libai, Bolton, Bügel, deRuyter, Goetz, Risselada, and Stephen 2010), this paper examines joint effects of traditional and social media.

Publicity research. Social media is a relatively new form of publicity, and yet the impact of more traditional forms of publicity on marketing outcomes has received disproportionately less attention than that of WOM and, of course, that of firm-sponsored activities (e.g., advertising, promotions). Much of the extant research on effects of media publicity centers on the how negative publicity, bad news, or unfavorable reviews influence outcomes such as sales and consumer demand (e.g., Ahluwalia, Burnkrant and Unnava 2000; Berger et al. 2010; Eliashberg and Shugan 1997). Consistent with the findings on how 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 or media attention is examined, precluding comparisons across media types.
An Integrated System Perspective

A general limitation of the above-described research on the effects of WOM and publicity on marketing outcomes, is that the interplay between various types of media is often not considered. Multiple types of media outlets need to be examined together in order to draw comparisons between them and, critically, to understand how they influence each other and operate jointly to affect, for example, sales. Thus, while past research has established that various types of traditional and social media (or WOM) are important because they impact a range of performance metrics, more integrated study is warranted—which is the central aim of the current research. The dataset we use here covers a number of media types and outlets, and, though not perfectly comprehensive, is sufficiently rich to enable use to address our research questions related to how traditional and social media affect sales and each other.

The question of how traditional and social media compare with respect to the nature and size of their respective impacts on marketing performance cannot be answered without first realizing that media outlets, traditional and social, are unlikely to exert any publicity influence on marketing outcomes in isolation. Rather, all media types are likely related in the sense that they have an influence on each other. For example, the technology editor of the New York Times reportedly uses the technology blog TechCrunch.com as a daily source of new story ideas (Arrington 2009). Thus, in order to understand how traditional and social media affect marketing performance we must also understand how TMOs and SMOs influence each other’s content generation and treat various outlets as parts of a broader system.

The notion that media operates as an “integrated system” shares conceptual similarities with the literature on integrated marketing communications (IMC) where a firm’s “paid media” (advertising and promotions) are coordinated across channels (cf. Schultz, Tannenbaum, and
Lauterborn 1993; Vakratsas and Ma 2005). IMC research has shown, for example, there is a complex interplay among marketing communications efforts (Smith, Gopalakrishna, and Chatterjee 2006), and advertising in one media channel can increase the effectiveness of promotions in other channels (Leclerc and Little 1997).

We argue that media outlets, traditional and social, are part of an informal interconnected system of media outlets that endogenously influence each other over time. The effect of traditional media on social media is expected because TMOs tend to be mainstream and have relatively large audience (reach). The reverse effect is also predicted, though the mechanism is less obvious. Since online social media tends to reflect collective public interest (Miller 2009b) and TMOs tend to report on topics that people are aware of, interested in, and talking about, social media activity, over time, can plausibly trigger corresponding traditional media activity.

The following example illustrates these concepts. Suppose that an unknown indie band releases a new album on virtually no advertising budget. Being unknown makes it highly unlikely that news of this new release will be picked up by any TMOs. They do, however, have an active fanbase who talks about them in social media. The new album release spurs fans to blog, tweet, post, and chat more online. This activity leads to more activity over time, with social media buzz continuing until eventually a sufficient amount of awareness and buzz has been generated. In other words, the band’s prominence in the public’s collective consciousness has risen. At some point this draws the attention of TMOs and generates some traditional media activity (e.g., an article in a newspaper or magazine). Since TMOs have greater reach than the SMOs that have carried past conversations about the band, this activity likely triggers a “sales bump.” If we only considered the direct sales response we would attribute this to traditional media. However, an information diffusion process unfolded over time prior to the traditional
media activity, and originated in and was carried through social media (which kept buzz alive and growing). Thus, although traditional media may have ultimately been responsible for affecting marketing performance, social media indirectly played a crucial role.

DATA

Our data are from Kiva, a not-for-profit organization which operates a website (www.kiva.org) that serves as an intermediary for microfinance loans. Microfinancing is the provision of small (i.e., “micro”) financial services (e.g., loans) 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. Microfinance has engendered widespread support and is a way that the U.N., the U.S., and other wealthy countries help fight global poverty.

Kiva is a “microlending marketplace” that individual consumers join (for free) so that they can make small loans (minimum $25; average $33) to pre-approved borrowers, all of whom are entrepreneurs in developing countries. Small loans are bundled into larger loans and given to the borrowers (bundled loans are typically still small; average bundle = $430). Between late 2005 (when Kiva was founded) and September 2010, approximately 754,000 “customers” (with the vast majority in the U.S.) have loaned $159 million to 412,000 borrowers in 181 countries. The loan repayment rate is 99%.

Variables

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

---

2 During the time of our data borrowers came only from developing countries. More recently (but after our data ended) Kiva has extended its platform to allow borrowers to also come from the U.S.
Throughout this paper, consistent with standard marketing terminology, we refer to Kiva lenders (i.e., the people who loan money) as “customers” and the loans they make as “sales.” We now describe the variables used in our analysis.

**Sales.** For each day we know the number of loans made by (i) new first-time lenders and (ii) repeat lenders. We denote the new and repeat sales volumes on day $t$ as $\text{Sales}_{\text{New}}^t$ and $\text{Sales}_{\text{Repeat}}^t$. Descriptive statistics are reported in Table 1. During our 427-day observation window, 152,439 new sales and 425,292 repeat sales were made (total sales = 577,731, or approximately $19$ million). The two sales variables are plotted in Figure 1.

[INSERT FIGURE 1 & TABLE 1 ABOUT HERE]

**Media activity.** Kiva relies entirely on media publicity and WOM for acquiring new customers and retaining existing ones. Kiva did not engage in any advertising during the time of our data. We focus on media events across various TMOs and SMOs. A media event occurred when Kiva was referred to in a media outlet, either traditional or social. We know how many Kiva media events of each type occurred on each day in our observation window.

The media activity variables and descriptive statistics are listed in Table 1 and described below. For traditional media our data covers newspapers, magazines, television, and radio, and for social media our data covers blogs and discussion forum posts in a Kiva-focused online customer community/network. Before describing the variables and how they were compiled, we first address three important issues related to our media variables.

First, traditional media activity for Kiva occurs infrequently, as would be expected for almost all companies, brands, or entities except for exceptionally high profile ones (e.g., Apple or Google). The sparseness of traditional media activity within each type (newspapers,

---

3 Although not an exhaustive set of social media sources, this set is consistent with the types of SMOs examined in previous literature. Data from other SMOs, particularly major social networks such as Facebook were not collected by Kiva and not available from public sources for our observation window.
magazines, television, radio) required us to aggregate these variables into a single time series of traditional media activity. While it would have been ideal to treat each one separately in our analysis, this was simply not possible on technical grounds and we would not be able to identify all parameters in our model without enough data (see the next section for model details). This, however, does not preclude us from answering our research questions about how traditional and social media affect sales and each other because all TMOs here are professional mass media outlets based on the same “one-to-many broadcasting” model for distributing content. Thus, we do not expect different types of traditional media to have drastically different effects on sales or social media activity.

Second, we treat the two types of social media—blogs and discussions—separately because (i) there is sufficient data on each type to allow this, and (ii) at a conceptual level they could potentially work in different ways. On the one hand, blogs are quite similar to the one-to-many broadcasting model seen in traditional media. While blogs are consumer-generated media and not produced by professional media outlets and can be interactive (e.g., through readers’ comments on posts), they are more of a one-way form of social media. On the other hand, discussions in Internet forums or online communities are more two-way, socially interactive, conversational, and thus resemble WOM. They cover consumer-to-consumer social interactions around a topic making them simultaneously “one-to-one” and “many-to-many” and inherently more conversational. Accordingly, blogs and discussions should not be combined into a single social media activity variable since these distinctions make it interesting to examine how they impact each other as well as traditional media and sales.

---

4 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.
Third, we make a clear distinction between traditional and social media given that the lines between these broad classes of media are increasingly blurred (e.g., newspapers have blogs). To delineate between traditional and social media activity we considered the source of the content. Content authored by a traditional, professional media outlet (e.g., the *New York Times*) was counted as traditional media, and consumer-generated online content (e.g., a blog post) was counted as online social media.

We now describe the data collection process for each type of media in our data.

**Traditional media.** *Kiva* supplied their own media tracking data that listed media events by day and outlet, which we supplemented with data from the Dow Jones Factiva database in case any events had been missed by *Kiva*’s staff. *Traditional* is the number of *Kiva* media events that appeared on day *t* in any one of the following outlets: newspapers (comprising of 11 U.S. national\(^5\) mentions, 41 U.S. local/regional mentions, and 7 non-U.S mentions), magazines (comprising of 13 mentions), television (comprising of 5 national network program mentions, 2 local program mentions, and 4 cable program mentions), and radio (comprising of 3 mentions). In total there were 86 separate traditional media events for *Kiva* during our observation window.

**Blogs.** Google Blog Search was used to compile a daily count of numbers of blog posts mentioning *Kiva*, denoted by *Blogs*. Over 427 days there were 2,485 blog posts about *Kiva*.\(^6\)

**Discussion forums/online community.** The number of posts on Internet community forums that mentioned *Kiva* on day *t* is a measure of online community conversation activity that we denote *Community*. 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:

---

\(^{5}\) The *New York Times*, the *Wall Street Journal*, *USA Today*, and the *Washington Post*.

\(^{6}\) This excludes posts authored by traditional media outlets and published on their blog websites (e.g., blogs on nytimes.com). Posts by traditional media outlets were counted as traditional media (but were not double counted).
Omgili.com, a forum-specific search engine that indexes millions of discussion threads from over 100,000 separate forums, and Google Groups search. Overall, there were 23,862 Kiva-related forum posts during our 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 Community as a measure of online WOM conversation activity within the dedicated social network for Kiva customers and fans.

MODELING MULTIVARIATE TIME SERIES COUNT DATA

The data described above and summarized in Table 1 provides a 427 day-long time series for five variables: (i) sales volume from new customers (SalesNew), (ii) sales volume from repeat customers (SalesRepeat), (iii) traditional media activity (Traditional), (iv) blog activity (Blogs), and (v) online community forum activity (Community). We treat these as endogenous in the sense that they likely influence each other (as suggested by significant correlations between them), consistent with our theory of media operating as a complex system to jointly influence sales. We thus take a multivariate approach to estimating relationships between these variables.

Data Challenges and Modeling Approach

Certain features of the data preclude the use of multivariate time series models commonly used in the marketing literature, such as vector autoregression (VAR) models (e.g., Stephen and Toubia 2010; Trusov et al. 2009; Villnueva et al. 2008). First, all variables are time series counts (i.e., non-negative integers). Second, for some variables (particularly 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 captures contemporaneous correlations.

As a starting point we turn to the literature on multivariate count data models. This literature comes mostly from statistics, finance, and transport engineering (e.g., Karlis and Meligkotsidou 2005; King 1989), but not marketing (Danaher’s [2007] multivariate negative binomial model is a notable exception). Most of the multivariate count models are seemingly unrelated regressions with appropriate distributions for count data 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 jointly modeling multiple stocks’ trading activities that led to the development of a multivariate autoregressive model for time series count data (Heinen and Rengifo 2007).

**Zero-Inflated Multivariate Autoregressive Double Poisson Model**

We extend the multivariate autoregressive conditional double Poisson model introduced by Heinen and Rengifo (2007) in the finance literature for multivariate estimation of relationships between a series of time series count variables 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 univariate 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

---

7 Standard VAR models have been used for such data in prior work in marketing and other fields, though are less appropriate given these data characteristics. E.g., the specification tests for VAR models make distributional assumptions that are incompatible with count data; and in VAR models contemporaneous effects are picked up in the error covariance matrix, which does not disentangle them from random disturbances.
connect the univariate marginal distributions to capture contemporaneous dependence 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 very infrequent events.

**Double Poisson distribution.** The first step is to make appropriate distributional assumptions for the marginal models. Since all variables are counts we need a discrete probability distribution with non-negative support. The options are usually Poisson or negative binomial distributions (Winkelmann 2008). An issue that must be considered in choosing a count distribution is dispersion (ratio of mean to variance), and a common challenge is handling overdispersion (mean < variance). Since the oft-used Poisson distribution assumes equidispersion (mean = variance) it is not appropriate for overdispersed data; the negative binomial distribution, however, is (but does not allow for underdispersion: mean > variance).

It is possible for time series count data to 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 will be generated (Heinen and Rengifo 2007). At the same time, as pointed out by Efron (1986), overdispersion in the data may in fact be less than overdispersion resulting from autocorrelations, which could give rise to (conditional) underdispersion in the underlying marginal distributions after controlling for autocorrelation. 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. The probability density function (pdf) for data \( y \), mean \( \mu \), and dispersion \( \theta \) (mean/variance) is:

\[
f_{dp}(y|\mu, \phi) = k(\mu, \phi) \sqrt{\phi} \exp(-\phi \mu) \left( \frac{\exp(-y)}{y!} \right) \left( \frac{\mu e}{y} \right)^{\phi y}
\]

(1)

---

Where \( e \) is Euler’s number, and \( k(\mu, \phi) \) is a normalizing constant. Efron (1986) approximates the constant as 
\[
\frac{1}{k(\mu, \phi)} = 1 + \frac{1-\phi}{12\mu} (1 + \frac{1}{\mu\phi}).
\]
Note when \( \phi = 1 \) the pdf 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.\(^9\) 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 \( t = 1, \ldots, T \). For our data \( M = 5 \) and \( T = 426 \) (we drop the observations at \( t = 1 \) because of lags). Events can be sales (\( \text{SalesNew} \) and \( \text{SalesRepeat} \), \( i = 1 \) and 2, respectively) or media activity (\( \text{Traditional}, \text{Blogs}, \) and \( \text{Community} \), \( i = 3 \) to 5, respectively). Each DP-distributed data point is conditional on the past of all variables; i.e., \( Y_{it} | \Gamma_{t-1} \sim DP(\mu_{it}, \phi_i) \) where \( \Gamma_{t-1} \) is the past of all \( M \) series up to and including the previous period, \( \mu_{it} \) is the mean for variable \( i \) at time \( t \), and \( \phi_i \) is the dispersion for variable \( i \).

We use a system of \( M \) 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:

\[
E(Y_{it} | \Gamma_{t-1}) = \mu_{it} = \sum_{j=1}^{M} \beta_{ji} Y_{jt-1} + \sum_{l=1}^{L} \gamma_{li} X_{lt}
\]

Where \( \beta_{ji} \) is the effect of \( Y_{jt-1} \) on \( Y_{it} \) (i.e., own for \( j = i \), cross-variable for \( j \neq i \)), and \( \gamma_{li} \) is the effect of an exogenous covariate \( X_{lt} \) on \( Y_{it} \) (for \( L \) exogenous covariates). We use a single covariate, \( \text{Borrowers}_t \), which is the number of loans requested by borrowers on \( \text{Kiva} \) on day \( t \) (i.e., a measure of loan demand).\(^{10}\)

---

\(^9\) The identity link is appropriate since all 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.

\(^{10}\) Loan demand—the number of borrowers/entrepreneurs in developing countries who want to borrow money from \( \text{Kiva} \) lenders—is exogenous to media and sales activity since the entrepreneurs generally do not have access to TMOs and SMOs in the U.S., and do not deal directly with \( \text{Kiva} \) (they deal with field agents).
We tested some other exogenous covariates typical in time series modeling—a linear
time trend and a seasonal effect (for Christmas, since *Kiva* gift certificates were popular holiday
gifts)—but these controls were unnecessary (i.e., fit did not significantly improve when
included). Also, we do not go beyond first-order lags since the number of parameters to estimate
in higher-order systems with $M$ greater than 2 or 3 becomes large without imposing a reduced
rank structure on $B$ (the $M$-by-$M$ matrix of $\beta_0$) (cf. Ben Omrane and Heinen 2010). Nevertheless,
second- and third-order AR lags were tested, but they did not significantly improve model fit
(likelihood ratio tests for ARX1 vs. ARX2/ARX3 each favored ARX1; $p < .001$).

*Multivariate normal copula model.* The third part of the model ties the $M$ univariate
marginal models together to capture the dependence of the variables on each other that is not
captured through the lagged effects in the ARX1 conditional means models; i.e.,
contemporaneous or “same day” effects. Copulas have received very little attention in the
marketing literature, although their popularity is growing in finance and econometrics (e.g., Ben
Omrane and Heinen 2010; Heinen and Rengifo 2007; Patton 2006; Trivedi and Zimmer 2005).
For the sake of brevity we do not provide a general technical description of copula models here.
Instead, we encourage readers to consult Danaher and Smith (2010) for a thorough explanation
of copula models and how they can be applied 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 the marginals all
come from the same distributional family. The practical benefit of using a copula model here is
that it allows for the $M$ endogenous time series count variables to be contemporaneously
correlated and for these correlations to be estimated. We can therefore capture dependence
between these variables beyond that captured through cross-variable lagged effects in the
autoregressive conditional mean models. This is critical because of the increasingly “real-time”
nature of media, particularly online. For instance, sales responses to media activity may occur
very quickly (e.g., a customer reads a blog about *Kiva* and is prompted to immediately visit *Kiva*
to lend some money), as could cross-media effects (e.g., when a TV show host proclaims *Kiva* to
be the solution to poverty this immediately triggers a slew of community discussion posts and
blog entries). Without data at very fine time resolution (e.g., hourly), contemporaneous
dependence would be missed if we only looked at the lagged cross-variable effects.

Although a variety of copula models are described in the literature (see Trivedi and
Zimmer 2005), most can only handle bivariate systems (i.e., $M = 2$). Following Heinen and
Rengifo (2007) and Danaher and Smith (2010), we use the multivariate normal (or Gaussian)
copula since it is a general and robust copula that can accommodate $M > 2$ (see appendix).

**Zero-inflation.** *Traditional* has a large number of days with zero activity counts. While
the DP distribution can accommodate many 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 (cf. 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* and *Community*
since these are the only variables with any zero data points.

**Model Estimation and Evaluation**

**Joint density.** In general, the joint density at time $t$ for all $M$ endogenous variables (with
$\Theta$ 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 $c(q_i; \Sigma)$:
where \( f_{\text{mixture}}(Y_{it}, \mu_{it}, \phi_{i}, \pi_{i}) \) 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 \( t \):

\[
\begin{align*}
\Pr(Y_{it} = 0) &= \pi_{i} + (1 - \pi_{i}) \sqrt{\phi_{i}} \exp(-\phi_{i} \mu_{it}) \\
\Pr(Y_{it} > 0) &= (1 - \pi_{i}) f_{\text{DP}}(Y_{it}, \mu_{it}, \phi_{i})
\end{align*}
\]

\( \pi_{i} \in [0,1] \) 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 (\( w_{i} \) can be a vector of ones or covariates); i.e., \( \pi_{i} = \exp(w_{i}' \lambda) / [1 + \exp(w_{i}' \lambda)] \). For the copula \( c(q; \Sigma), q_{i} \) are the \( M \)-dimensional vectors of normal quantiles of the probability integral transforms (PITs) of the count data\(^{11}\) under the marginal DP densities, and \( \Sigma \) is the covariance matrix of the multivariate normal copula representing contemporaneous interdependence.

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

\[
LL = \sum_{t=1}^{T} \log h(Y_{1t}, \ldots, Y_{Mt}, \Theta, \Sigma) = \sum_{t=1}^{T} \sum_{i=1}^{M} \log f_{\text{mixture}}(Y_{it}, \mu_{it}, \phi_{i}, \pi_{i}) + \log c(q_{i}; \Sigma)
\]

Since the log-likelihood is additive we can split it into two parts. For the marginal models, \( LL_{1} \):

\[
LL_{1} = \sum_{t=1}^{T} \sum_{i=1}^{M} 1_{Y_{it} = 0} \log \left[ \pi_{i} + (1 - \pi_{i}) \sqrt{\phi_{i}} \exp(-\mu_{it} \phi_{i}) \right] + \sum_{t=1}^{T} \sum_{i=1}^{M} 1_{Y_{it} > 0} \log \left[ (1 - \pi_{i}) k(\mu_{it}, \phi_{i}) \sqrt{\phi_{i}} \exp(-\mu_{it} \phi_{i}) \frac{\exp(-Y_{it}) Y_{it}^{\gamma_{i}}}{Y_{it}^{\gamma_{i}}} \left( \frac{\mu_{it} e}{Y_{it}} \right)^{\phi_{i}} \right]
\]

Where \( 1_{Y_{it} = 0} \) is an indicator variable that equals 1 when \( Y_{it} = 0 \) and 0 otherwise, and \( 1_{Y_{it} > 0} \) equals 1 when \( Y_{it} > 0 \) and 0 otherwise. And for the multivariate normal copula, \( LL_{2} \):

\(^{11}\) 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.
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 \( I_M \) is an \( M \)-dimensional identity matrix.

**Two-step estimation procedure.** In theory, all parameters can be jointly estimated by maximizing the log-likelihood in equation 5. 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 Patton (2006) shows yields consistent estimates for all parameters. The first step estimates parameters of the \( M \) univariate marginal models (yielding estimates of parameter matrix \( \Theta \)) by maximizing \( LL_1 \). The second step estimates the copula (yielding an estimate of variance-covariance matrix \( \Sigma \)) 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., \( \hat{\Sigma} = \frac{1}{T} \sum_{i=1}^{T} q_iq_i' \). \( LL_2 \) is then obtained by plugging the estimated variance-covariance matrix and \( q_t \) into equation 8. Note that \( q_t \) 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.

**Tests and fit.** First, following standard practice in time series econometrics, stationarity tests are performed (stationarity is assumed for the marginal models), along with Granger causality tests to check that each of the \( M \) variables in the system is endogenous. Second, Heinen and Rengifo (2007) suggest that model fit be evaluated based on the log-likelihood \( (LL) \). Third,
to check the model is correctly specified the standardized residuals, \( \epsilon_i = \frac{Y_i - \hat{\mu}_i}{\hat{\phi}_i} \), are checked to see that they have mean close to 0, variance close to 1, and minimal autocorrelation.

**RESULTS**

In this section we present the estimation results for the \( M = 5 \) system of variables to show the interdependence between *Traditional*, *Blogs*, and *Community* and how each of these variables directly and/or indirectly affects *SalesNew* and *SalesRepeat*.\(^{12}\)

**Specification Tests and Model Fit**

*Stationarity*. Each variable was 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 tests confirmed that all time series used here are stationary; i.e., the ADF test’s null hypothesis of an evolving series was rejected in all cases; \( ps < .01 \). Another test of stationarity is that the eigenvalues of \((I - B)\) lie within the unit circle, which was the case. It is therefore appropriate to use the autoregressive model for the DP distributions’ conditional means and the data can be used “as is” in their levels without differencing (cf. Dekimpe and Hanssens 2004).

*Endogeneity*. Our conceptualization suggests that the sales and media activity variables are endogenous. This was confirmed with a Granger causality test for each variable (Granger 1969). In all five tests we rejected the null hypothesis that the test variable was exogenous and caused by itself and not by any of the other variables in the system (\( ps < .05 \)).

*Residuals*. Each variable’s vector of standardized residuals (\( \epsilon_i \)) was checked to see if it had mean close to 0, variance close to 1, and minimal autocorrelation. The residuals’ means and

\(^{12}\) We also modeled a simpler system of \( M = 4 \) variables with *SalesNew*, *SalesRepeat*, *Traditional*, and *Social*. *Social* = *Blogs* + *Community*. Results from this model were similar to and consistent with those from the \( M = 5 \) system.
variances are reported at the bottom of Table 2 for each variable. All residual means are close to 0, all variances are reasonably close 1, and a check of each variable’s residual autocorrelation function showed no evidence of strong autocorrelation in any variable’s residuals.

**Model fit.** We report fit statistics for the model and where appropriate each univariate marginal model in Table 2. Based on a likelihood ratio test, this model fits better than a base model that includes lagged own effects but no lagged cross-variable effects in ($\chi^2 = 664.66$, $df = 20$, $p < .001$). Capturing cross-variable effects is therefore appropriate. Also, all univariate fit statistics (pseudo-$R^2$ and median absolute deviation [MAD]) indicate reasonable fit, are better than the base model.

**Main Findings**

Table 2 reports the estimates for the marginal models and the copula correlation matrix. Long-run effects based on cumulative orthogonalized impulse response functions (COIRFs) are reported in Table 3 and long-run elasticities are reported in Table 4.

[INSERT TABLES 2, 3 & 4 ABOUT HERE]

As a general first result, as hypothesized, this system of variables exhibits significant interdependence. This is evidenced by the copula correlation matrix for contemporaneous effects being significantly different from an identity matrix ($p < .001$) and many significant cross-variable lagged effects in the off-diagonals of matrix $B$.

We base our findings on the long-run effects (COIRFs) reported in Table 3 (since it is infeasible to interpret parameter estimates directly in multivariate time series models; cf. Sims 1980). The COIRFs capture the total (direct + 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 effects
(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, after first setting all non-significant \((p > .05)\) elements of B to zero.\(^{13}\) These were then adjusted for contemporaneous effects (orthogonalization) that are in the copula correlation matrix but not in B using the Cholesky decomposition of the estimated copula covariance matrix (Sims 1981).\(^{14}\)

**Media effects on sales.** In Table 2 we see that traditional media and community forums are the only media variables to *directly* affect new and repeat sales in terms of an immediate sales response to media activity. In the long run, blog activity also has an effect on sales, indirectly through the direct effect of Blog on Traditional. Thus, all media variables affect both sales variables. The sizes of these effects—across media variables and across sales variables—differ greatly, however. Based on the COIRFs in Table 3, in the long run traditional media activity has the largest *per event* impact: an extra unit of media publicity from a TMO approximately generates an extra 951 new and 367 repeat sales. This is far greater than the unit per-event impacts of blog posts (22 new and 9 repeat sales) and community forum posts (130 new and 55 repeat sales). This is not surprising since TMOs typically have larger reach.\(^{15}\)

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 piece of traditional media publicity may have a large impact on *SalesNew* and *SalesRepeat*, new traditional media content arises only very infrequently. Social media activity, 

---

\(^{13}\) The cumulative unorthogonalized IRF matrix, \(B_{CIRF} = \sum_{t=1}^{\infty} B^t\).

\(^{14}\) \(\Sigma = PP',\) where P is the Cholesky decomposition of \(\Sigma\) and is a lower triangular \(M\)-by-\(M\) matrix with a positive diagonal. The COIRF matrix is \(B_{CIRF}P\). Since P can change depending on the ordering of variables in \(\Sigma\) we followed Sims (1981) and checked sensitivity to different orders. P did not markedly change over different permutations of \(\Sigma\). For the results in Tables 3 and 4 we used the order in which the variables are listed in Table 1.

\(^{15}\) An interesting analysis would be to look at the sales responses after adjusting for the reach of the respective media types. However, audience size/reach data are not available for these data. We leave further exploration of the returns to media with different-sized audiences as a direction for future research.
in *Kiva*’s case particularly from their online community, occurs much more frequently. We can take frequency into account by approximating the extra sales expected from an average day’s worth of media activity of each type. The expected number of extra sales (new + repeat) from an average day of media activity are about 264 for traditional media, 179 for blogs, and 10,328 for online community forums.

Interestingly, the frequency-adjusted impacts of blogs and traditional media are similarly sized. This is because blogs tend to act more like TMOs than their more interactive and conversational community social media counterpart. Of particular note is the massive frequency-adjusted effect of community activity on sales. This is likely because of what the community allows customers to do, and what those actions subsequently induce. The community provides a forum for current and potential customers to socialize around the *Kiva* brand and concept by discussing about it with like-minded people. This social interactivity (which is something that blogs largely lack) spurs greater ongoing interest in *Kiva* and, as we see here, more sales. Theoretically, this finding is consistent with the notion that frequent social interactions are a way to keep consumers actively engaged with and enthusiastic about a product (or website) to spur persistent usage (Stephen and Berger 2010).

Another way to look at these findings and to compare across media variables is to look at the long-run elasticities of sales to media activities, which we report in Table 4.¹⁶ It is interesting to see that all media variables have a larger long run effect on new customer sales than repeat customer sales. Media activity here is better at generating new rather than repeat sales. We also once again see that community activity is a larger sales driver than blogs and traditional media.

¹⁶ We compute the arc elasticities using the COIRFs in Table 3 and the means in Table 1. The arc elasticity for the effect of impulse \( j \) on response \( i \) using COIRFs is \( \eta_{ji} = \text{COIRF}_{ji} \times \frac{E(Y_j)}{E(Y_i)} \).
**Endogenous sales effects.** The other sales elasticities in Table 4 are for the endogenous sales effects; i.e., the effects of the sales variables on themselves. In fact, for both sales variables, the largest sales elasticity comes not from community activity but rather from repeat sales (31.67 for new sales elasticity to repeat sales, and 5.24 for repeat sales elasticity to repeat sales). These effects detect underlying social contagion (or WOM) that is not picked up through social media and therefore may occur in private conversations (on- or offline; or perhaps passively simply by observation; cf. Libai et al. 2010). The effect of repeat sales on new and repeat sales can be thought of as WOM from existing customers to non-customers (who become new customers) and to other existing customers, respectively. (The effect of repeat sales on itself also could be due to loyalty, which cannot be disentangled from implicit WOM.) The other large endogenous sales effect is for new customer sales on itself; i.e., the ability of WOM from new customers to acquire additional new customers (elasticity = 5.15). The strength of these effects show that social influences (like 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, note that these implied WOM effects are in addition to and after controlling for community activity, which of course itself is a form of online WOM.

**Endogenous media effects.** The sales response results cannot be fully understood without consideration of how the media types influence activity in each other. Traditional media activity not only can increase the amount of online social media buzz, which is not surprising, but also can be influenced by social media. This is consistent with our above conceptualization of media outlets operating as system or informal network, as well as anecdotal evidence of, for example, *New York Times* editors getting story ideas from blogs. The largest elasticity for effects on traditional media activity is from repeat sales (9.99)—which is itself strongly driven by
community forum activity. The elasticity of traditional media activity to community activity is also substantial (8.38). In fact, interestingly, at least in the case of Kiva, we find that traditional media activity is more strongly influenced in the long run by social media activity in communities than it is by itself. Put simply, here traditional media follows social media, not the other way around.

More generally, since online social media activity for a brand, company or product is a reflection of its prominence in the collective consciousness or zeitgeist (Miller 2009b), more social media activity should translate into more traditional media activity because TMOs are more likely to pick up on topics that people are aware of and are talking about. To the extent that social media activity is a proxy for general buzz about a topic (in this case Kiva), this shows that frequent, ongoing buzz helps to generate publicity events in broader-reaching TMOs. Infrequent but high-impact traditional media publicity events are undoubtedly valuable, and social media appears to play a key role in generating such events. Moreover, in between these events, social media, particularly online communities, helps to keep customers engaged and the buzz alive.

Finally, we find that traditional media has a very small impact on social media activity, community forum activity has a big impact on itself (indicating that more conversations spur even more chatter) and blogs, though blogs do not have much of an effect on community forum activity. In fact, a general finding here is that blogs are not highly impactful at all, at least in the case of Kiva during the time of our data. This may be due to blogs being more one-to-many than truly socially interactive, as discussed above. Across all dependent variables, looking at the elasticities the effect of blog activity is always approximately of the same order of magnitude as the effect of traditional media activity, which is always of a lower order of magnitude than the effect of community activity.
DISCUSSION AND CONCLUSION

Given the fast-changing media landscape, the ubiquity of different types of online social media (e.g., consumer-generated content broadcast to others through blogs, and more social and “conversational” interactions through online communities and social networks), and the increasingly dire predictions about the future of traditional media (particularly newspapers and magazines), our overall objective was to study the interrelatedness between traditional and social media to see not only how these different types of media influence an important marketing outcome (sales), but also how they influence each other.

Traditional media has a strong effect on sales (after controlling for the effects of social media). On a per-event basis, this effect is much larger than the comparable effect of social media. However, because the frequency of social media activity (particularly in online communities) for Kiva is much greater than the frequency of traditional media activity, the impact of social media is by no means negligible (i.e., small unit impacts quickly accumulate). Although the sales spikes seen in Figure 1 correspond to traditional media events (e.g., Kiva being written about by a high-profile New York Times columnist or being mentioned on Oprah), social media contributes to long-run sales growth both directly and indirectly through its positive influence on traditional media. In fact, a good characterization of social media here is as an ongoing, persistent driver of media activity and sales activity. Traditional media publicity can pack a sizeable punch in lifting sales, but these effects are too infrequent to rely on. We find here that social media, on the other hand, is more a reliable engine for driving sales.

In these data we saw that forum activity—conversations—in an online community had the largest total impact on sales out of all three media variables. Combined with the implicit social WOM-type effects detected through the endogenous sales response effects, this
emphasizes the importance of consumer-to-consumer social interactions in driving sales. Being able to discuss a topic of mutual interest, whether through direct personal contact or an online community or social network, appears to strongly help drive customers’ actions—in this case, sales in the form of micro-loans. Interestingly, for Kiva this helps more than traditional media events (even big, high profile ones). Further, as noted above and in recent work (Libai et al. 2010; Stephen and Berger 2010), ongoing conversations and buzz may be what is needed to keep a topic (e.g., product, website, brand) alive and prominent in many consumers’ minds.

Another important finding of the current work is the high degree of interdependence between traditional and social media outlets. This interdependence, or influence of one media outlet’s activity on another outlet’s activity, occurred among all three media variables, and most notably with traditional media being influenced by social media (more than the other way around). This result supports our claim that media outlets that generate publicity for products, brands and companies should be seen as part of a single system. Taken further, if we see media outlets of various types as part of an integrated system, a key finding is that the SMOs in this system—e.g., bloggers, fans, customers, and other regular people who generate content by posting messages on blogs, online communities, and networks—are powerful. Social media may in fact serve a critical “transfer” function, similar to that described earlier in our indie band example, where it takes information from lesser-known, lower-reach outlets and communities, builds awareness and buzz over time, and eventually draws the attention of mainstream traditional media. To the extent that social media taps the public consciousness or “collective brain” (Miller 2009b), this implies that the individuals who participate in social media collectively play an important information dissemination role in the broader media landscape.
In the case of *Kiva*, a critical function of *truly social* social media (i.e., WOM in an online community) appears to be that it spurs additional media activity, which in turn helps to drive further sales and media activity. Online community activity therefore plays a valuable “behind the scenes” role that should not be undervalued (but cannot be seen simply by looking at the COIRFs in Table 3). As a simple illustration of this, we estimated a version of the model with the effects of *Community* on *Traditional* and *Blogs* fixed to zero, computed the COIRFs, and compared the sizes of the long-run media effects on sales to those reported in Table 3. Simply by not allowing *Community* to affect other media variables results in 54%, 53%, and 90% smaller sales responses to traditional media, blog, and community activity, respectively.

To summarize, our findings contribute to the existing literatures on media and publicity effects on marketing outcomes, and the literature on online WOM by deepening our understanding of how traditional and social media work together to impact sales and, critically, how they affect each other and the different roles that they play. To the best of our knowledge, this research is the first to examine offline traditional and online social media’s joint effects on a marketing outcome and each other. From a practical standpoint, our results emphasize the importance of both types of media. Further, seeing that social media (particularly *truly social* media based around communities and networks) plays such a central role, marketing practitioners hoping to generate sales through publicity efforts would be well served by focusing on building consumer buzz and online WOM (either organically or through techniques such as viral marketing campaigns). Given the importance of social interactions that we uncovered here, these results add to the nascent literature showing the value and benefits of viral marketing programs that encourage and foster consumer-to-consumer social interactions (e.g., Godes and
Mayzlin 2009; Libai et al. 2010). Finally, as a methodological contribution, we introduced a new type of multivariate time series model designed to handle time series counts with excess zeros.

The current research is an important step forward in the growing streams of research on multi-media effects on marketing outcomes, social media, and online marketing. It is, of course, not without limitations. First, despite gathering data from a variety of reliable sources for multiple media types and outlets, our set of media variables is not exhaustive. We do, however, cover the major types and major outlets within each type. In particular, given that since the time of our data (2007 and early 2008) more social media platforms have emerged (e.g., Facebook and Twitter are now major sources of social media, and are searchable), future research should consider incorporating additional social media types. Second, we only examined how media publicity affects marketing performance for a single organization. We of course make no claims of generalizability, and we encourage future research to examine the same set (or an expanded set) of media types for other products, brands, and companies, and to consider other marketing performance variables in addition to sales. Third, we could not consider the valence of the media events. Since Kiva is a non-profit trying to alleviate world poverty, almost all media content we sampled was positive. It would be interesting for future work to look at how content valence moderates the impact of media activity on sales. Finally, while we drew a clear line between traditional and social media based on whether the content source was a professional organization or not, in reality the lines are blurred and a “hybrid” form of media has emerged that is professionally sourced but online and social. Future work could examine how this hybrid form of media fits into the system we studied. We hope that researchers consider these and related issues in future research in this interesting and complex area.
REFERENCES


FIGURE 1
DAILY TIME SERIES PLOTS FOR NEW AND REPEAT SALES FROM KIVA LOANS
<table>
<thead>
<tr>
<th>Type</th>
<th>Variable</th>
<th>Definition</th>
<th>Mean (Standard Deviation)</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales</td>
<td>SalesNew</td>
<td>The total number of loans made by new (first-time) lenders on day t</td>
<td>356.99 (350.06)</td>
<td>237</td>
<td>60</td>
<td>2,641</td>
</tr>
<tr>
<td></td>
<td>SalesRepeat</td>
<td>The total number of loans made by repeat lenders on day t</td>
<td>998.63 (654.59)</td>
<td>237</td>
<td>162</td>
<td>3,788</td>
</tr>
<tr>
<td>Traditional Media</td>
<td>Traditional</td>
<td>The number of mentions in one of the four kinds of traditional media (newspapers, magazines, television, and radio) on day t</td>
<td>.20 (.48)</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Social Media</td>
<td>Blogs</td>
<td>The number of blog(^a) posts made on day t</td>
<td>5.82 (2.94)</td>
<td>5</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Community</td>
<td>The number of community forum(^b) posts made on day t</td>
<td>55.88 (47.23)</td>
<td>50</td>
<td>0</td>
<td>213</td>
</tr>
</tbody>
</table>

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 2

**PARAMETER ESTIMATES AND FIT STATISTICS**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>SalesNew (i = 1)</th>
<th>SalesRepeat (i = 2)</th>
<th>Traditional (i = 3)</th>
<th>Blogs (i = 4)</th>
<th>Community (i = 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Univariate marginal model parameter estimates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SalesNew_{t-1}</td>
<td>$\beta_{1i}$</td>
<td>.78**</td>
<td>.01</td>
<td>.001*</td>
<td>-.01</td>
</tr>
<tr>
<td>SalesRepeat_{t-1}</td>
<td>$\beta_{2i}$</td>
<td>-.01</td>
<td>.73**</td>
<td>&lt;.001</td>
<td>.002**</td>
</tr>
<tr>
<td>Traditional_{t-1}</td>
<td>$\beta_{3i}$</td>
<td>37.31*</td>
<td>59.37*</td>
<td>.08</td>
<td>.14</td>
</tr>
<tr>
<td>Blogs_{t-1}</td>
<td>$\beta_{4i}$</td>
<td>-.51</td>
<td>6.64</td>
<td>.02*</td>
<td>.31**</td>
</tr>
<tr>
<td>Community_{t-1}</td>
<td>$\beta_{5i}$</td>
<td>.72**</td>
<td>2.12**</td>
<td>.01</td>
<td>.02**</td>
</tr>
<tr>
<td>Borrowers_{t}</td>
<td>$\gamma_{1i}$</td>
<td>.61**</td>
<td>1.45**</td>
<td>&lt;.001</td>
<td>.03**</td>
</tr>
<tr>
<td>Dispersion</td>
<td>$\phi_{1i}$</td>
<td>.02**</td>
<td>.01**</td>
<td>1.01**</td>
<td>.80**</td>
</tr>
<tr>
<td>Zero-inflation a</td>
<td>$\lambda_{i}$</td>
<td>—</td>
<td>—</td>
<td>1.50*</td>
<td>—</td>
</tr>
<tr>
<td><strong>Copula correlation estimates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SalesNew</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SalesRepeat</td>
<td></td>
<td>.85</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional</td>
<td></td>
<td>.13</td>
<td>.06</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Blogs</td>
<td></td>
<td>.45</td>
<td>.45</td>
<td>.01</td>
<td>1.00</td>
</tr>
<tr>
<td>Community</td>
<td></td>
<td>.62</td>
<td>.69</td>
<td>.05</td>
<td>.38</td>
</tr>
<tr>
<td><strong>Model fit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudo-$R^2$</td>
<td>.81</td>
<td>.86</td>
<td>.15</td>
<td>.45</td>
<td>.85</td>
</tr>
<tr>
<td>MAD</td>
<td>45.53</td>
<td>129.74</td>
<td>.18</td>
<td>1.66</td>
<td>9.60</td>
</tr>
<tr>
<td>Mean($\epsilon_{i}$)</td>
<td>-.003</td>
<td>-.006</td>
<td>.024</td>
<td>.066</td>
<td>.024</td>
</tr>
<tr>
<td>Variance($\epsilon_{i}$)</td>
<td>1.42</td>
<td>1.11</td>
<td>1.23</td>
<td>1.04</td>
<td>.99</td>
</tr>
<tr>
<td>LL, AIC, BIC (full model)</td>
<td></td>
<td></td>
<td>-8,076; 16,226; 16,435</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LL, AIC, BIC (own effect) b</td>
<td></td>
<td></td>
<td>-8,410; 16,854; 16,950</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $p < .05$, ** $p < .001$.

a Zero-inflation parameters were only estimated for $i = 3$ and $i = 5$; other variables did not have any zero values. $\pi_i = \text{Logit}(\lambda_i)$.

b Likelihood ratio test for the full model vs. the own-effects-only base model: $\chi^2(20) = 667.78, p < .001$. 

---

37
### TABLE 3
LONG-RUN EFFECTS 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)

<table>
<thead>
<tr>
<th></th>
<th>SalesNew ((i = 1))</th>
<th>SalesRepeat ((i = 2))</th>
<th>Traditional ((i = 3))</th>
<th>Blogs ((i = 4))</th>
<th>Community ((i = 5))</th>
</tr>
</thead>
<tbody>
<tr>
<td>SalesNew</td>
<td>5.15</td>
<td>.66</td>
<td>.002</td>
<td>.01</td>
<td>.07</td>
</tr>
<tr>
<td>SalesRepeat</td>
<td>11.32</td>
<td>5.24</td>
<td>.002</td>
<td>.10</td>
<td>.57</td>
</tr>
<tr>
<td>Traditional</td>
<td>950.83</td>
<td>366.91</td>
<td>.19</td>
<td>6.34</td>
<td>35.96</td>
</tr>
<tr>
<td>Blogs</td>
<td>22.22</td>
<td>8.56</td>
<td>.03</td>
<td>.54</td>
<td>.83</td>
</tr>
<tr>
<td>Community</td>
<td>130.09</td>
<td>54.74</td>
<td>.03</td>
<td>1.25</td>
<td>7.55</td>
</tr>
</tbody>
</table>

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.
TABLE 4
LONG-RUN ELASTICITIES

A 1% increase in this variable… in the long-run generates the following percentage increase in activity in this variable

<table>
<thead>
<tr>
<th></th>
<th>SalesNew (i = 1)</th>
<th>SalesRepeat (i = 2)</th>
<th>Traditional (i = 3)</th>
<th>Blogs (i = 4)</th>
<th>Community (i = 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SalesNew</td>
<td>5.15</td>
<td>.24</td>
<td>3.57</td>
<td>.61</td>
<td>.45</td>
</tr>
<tr>
<td>SalesRepeat</td>
<td>31.67</td>
<td>5.24</td>
<td>9.99</td>
<td>17.16</td>
<td>10.19</td>
</tr>
<tr>
<td>Traditional</td>
<td>.53</td>
<td>.07</td>
<td>.19</td>
<td>.22</td>
<td>.13</td>
</tr>
<tr>
<td>Blogs</td>
<td>.36</td>
<td>.05</td>
<td>.87</td>
<td>.54</td>
<td>.09</td>
</tr>
<tr>
<td>Community</td>
<td>20.36</td>
<td>3.06</td>
<td>8.38</td>
<td>12.00</td>
<td>7.55</td>
</tr>
</tbody>
</table>
WEB APPENDIX
MULTIVARIATE NORMAL COPULA

This copula’s distribution is given by $C(z_1, ..., z_M; \Sigma) = \Phi^M(q_1, ..., q_M; \Sigma)$, where $\Phi^M$ is the $M$-dimensional multivariate standard normal distribution function, and $\Sigma$ 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 $\Phi^{-1}$ is the inverse standard normal distribution function, and $z_i$ are the probability integral transforms (PITs) of the data. The copula’s density is $c(q; \Sigma)$, which can be rewritten as

$$c(z_1, ..., z_M; \Sigma) = |\Sigma|^{-1/2} \exp\left(\frac{1}{2} q(I_M - \Sigma^{-1})q^T\right),$$

where $q$ 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 $\Theta$ 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:

$$h(Y_{it}, ..., Y_{Mt}, \Theta, \Sigma) = \prod_{i=1}^{M} f_{mixure}(Y_{it}, \mu_i, \phi_i, \pi_i) \cdot c(z_{i1}, ..., z_{iM}; \Sigma) \tag{A1}$$

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 $U \sim \text{uniform}(0,1)$, 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 $z_{it} = F_{DP,i}^*(Y_{it}^*) = F_{DP,i}(Y_{it} - 1) + f_{DP,i}(Y_{it}) \cdot U_{it}$, where $U_{it}$ is a uniform(0,1) random variable.
Europe Campus
Boulevard de Constance
77305 Fontainebleau Cedex, France
Tel: +33 (0)1 60 72 40 00
Fax: +33 (0)1 60 74 55 00/01

Asia Campus
1 Ayer Rajah Avenue, Singapore 138676
Tel: +65 67 99 53 88
Fax: +65 67 99 53 99

Abu Dhabi Campus
Muroor Road - Street No 4
P.O. Box 48049
Abu Dhabi, United Arab Emirates
Tel: +971 2 651 5200
Fax: +971 2 443 9461

www.insead.edu