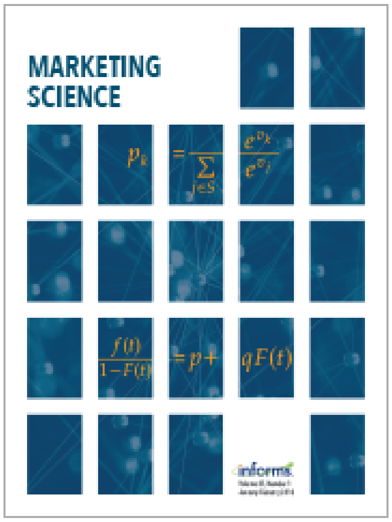


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance

Seshadri Tirunillai, Gerard J. Tellis,

MARKETING SCIENCE
Vol. 31, No. 2, March–April 2012, pp. 198–215
ISSN 0732-2399 (print) | ISSN 1526-548X (online)
http://dx.doi.org/10.1287/mksc.1110.0682

Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance

Seshadri Tirunillai
C.T. Bauer College of Business, University of Houston, Houston, Texas 77204, seshadri@bauer.uh.edu

Gerard J. Tellis
Marshall School of Business, University of Southern California, Los Angeles, California 90089,
tellis@usc.edu

This study examines whether user-generated content (UGC) is related to stock market performance, which metric of UGC has the strongest relationship, and what the dynamics of the relationship are. We aggregate UGC from multiple websites over a four-year period across 6 markets and 15 firms. We derive multiple metrics of UGC and use multivariate time-series models to assess the relationship between UGC and stock market performance.

Volume of chatter significantly leads abnormal returns by a few days (supported by Granger causality tests). Of all the metrics of UGC, volume of chatter has the strongest positive effect on abnormal returns and trading volume. The effect of negative and positive metrics of UGC on abnormal returns is asymmetric. Whereas negative UGC has a significant negative effect on abnormal returns with a short "wear-in" and long "wear-out," positive UGC has no significant effect on these metrics. The volume of chatter and negative chatter have a significant positive effect on trading volume. Idiosyncratic risk increases significantly with negative information in UGC. Positive information does not have much influence on the risk of the firm. An increase in off-line advertising significantly increases the volume of chatter and decreases negative chatter. These results have important implications for managers and investors.

Key words: user-generated content (UGC); stock returns; online word of mouth; vector autoregression (VAR); computational text processing
History: Received: June 17, 2009; accepted: July 10, 2011; Eric Bradlow and then Preyas Desai served as the editor-in-chief and Alan Montgomery served as associate editor for this article. Published online in *Articles in Advance* January 24, 2012.

Introduction

With the popularity of online media, consumers now go beyond their role as passive seekers of information to actively and regularly share their experience with others on online platforms such as communities, blogs, product reviews, and wikis. The body of information that consumers so generate is popularly known as user-generated content (UGC). UGC is one aspect of a broad phenomenon of interconsumer communication called word of mouth. Whereas consumer input is available to a small extent in traditional media (e.g., letters to editors and consumer complaints), its popularity in new online media is relatively massive (thousands per day) and growing rapidly. Estimates for 2010 are that 131.4 million users consume and 95.3 million users create UGC (Verna 2009). UGC may be more important and impactful than traditional word of mouth because of its instant availability, low cost, ease of use, wide subscription, wide access, and wide reach. It is also growing more rapidly and is easier for firms to measure and monitor than traditional forms of word of mouth. Studies suggest that UGC has higher impact on subsequent consumer choice than traditional marketing activities of the firm (Trusov et al. 2009). Most importantly, UGC is available at a highly disaggregate temporal level (days, hours, minutes). Thus, when sales or customer satisfaction are not available at such a disaggregate level, UGC could provide some indication of how the brand is doing.

This study focuses on the relationship between UGC and stock market performance of the firm. Prior research indicates that the stock market responds to the assessment of quality by experts' published reviews of products (Tellis and Johnson 2007). Studies in finance investigate the impact on stock market performance of print news (e.g., Tetlock 2007, Tetlock et al. 2008) or traders' online financial recommendations in Internet message boards (e.g., Das and Chen 2007). Unlike the print media or third-party reviews, UGC reflects the opinions and experiences of a large number of active users of the product, thus benefitting from the "wisdom of crowds" (Surowiecki 2005). It reflects the accumulation of consumers' ongoing experiences over an extended period of time in real environments. It also evolves with time and can be tapped continuously over time.

We restrict our study to two forms of UGC—product reviews and product ratings. These two forms of UGC are rich in product-related information. Other forms of UGC such as videos and blogs have a great deal of information with little relevance to specific products and brands. Thus, the signal-to-noise ratio in product ratings and reviews is much higher than that in videos, blogs, and networking sites.

Table 1 classifies the literature on consumer generated data on five dimensions. The current study differs from prior research in many respects (see Table 1). First, whereas Luo (2007, 2009) conducts an insightful study on the impact of negative word of mouth on stock returns, his study draws data from traditional media and uses one form of negative content (complaints) in one industry (airlines). UGC, unlike other forms of word-of-mouth communication, leaves a digital trail of consumer opinion on all aspects of a product, including both positive and negative word of mouth. Second, several studies examine the interaction among consumers or the impact of UGC on demand/sales (e.g., Dellarocas et al. 2007, Dhar and Chang 2009, Liu 2006, Chintagunta et al. 2010, Chevalier and Mayzlin 2006, Trusov et al. 2009, Moe and Trusov 2011) but not the impact of UGC on stock market performance. Third, most studies on UGC use a narrow set of metrics such as numerical ratings or volume, ignoring the information content of text in these reviews, which is rich in consumer expressions (see Table 1). Fourth, all prior studies focus on a single market such as movies, books, or airlines. In contrast, the current study is unique in that it focuses on the impact of UGC on stock market performance using four metrics of UGC and three metrics of stock performance across six different markets (see Table 1).

In particular, this study seeks answers to the following questions:

- Is there a relationship between UGC and the stock market performance of the firm?
- If a relationship exists, what is the direction of causality?
- Among the various metrics of UGC, which metric best relates to stock performance of the firm?
- What are the dynamics of the relationship in terms of wear-in, wear-out, and duration?

Our choice of stock market performance as the dependent variable has three benefits. First, as a measure of shareholder value, market performance is the ultimate concern of the firm and increasingly used in marketing studies (Srinivasan and Hanssens 2009). Second, stock market performance is available at the daily level, allowing for granular analysis, whereas sales, profits, and earnings are not as easily available at this level for most firms. Third, by indicating what metrics of UGC impact stock performance, this study suggests the metrics of UGC that should concern managers.

The rest of this paper is organized as follows. The second section presents the theory, the third section explains the method, the fourth section describes the measures, the fifth section describes the models, and the last two sections present the results and discussion.

Table 1 Classification of Relevant Literature on UGC

Study	Online medium open to consumers	Online user (consumer) generated content	Contrast of positive vs. negative content	Focus on stock market performance	Coverage of multiple markets
This study	Yes	Yes	Yes	Yes	Yes, six markets
Luo (2009)	No	No (complaints filed with Dept. of Transportation)	No	Yes	No, just airline service
Dellarocas et al. (2007)	Yes	Yes	No	No	No, only movies
Forman et al. (2008)	Yes	Yes	No	No	No, only books
Dhar and Chang (2009)	Yes	Yes	No	No	No, only music
Chevalier and Mayzlin (2006)	Yes	Yes	No	No	No, only books
Trusov et al. (2009)	Yes	No (social networking site)	No	No	No
Moe and Trusov (2011)	Yes	Yes	No	No	No
Liu (2006)	Yes	Yes	No	No	No, only movies
Chintagunta et al. (2009)	Yes	Yes	No	No	No, only movies
Godes and Mayzlin (2004)	Yes	Yes	No	No	No, only TV
Zhu and Zhang (2010)	Yes	Yes	No	No	No, only video games

Theory: Relationship Between UGC and Stock Market Performance

This section builds the theory for the relationship between UGC and stock market performance. It answers four questions: Why might UGC be value relevant to investors? Why would UGC be a leading indicator of stock market performance? Why would its effect be delayed? Why would this effect be asymmetric across measures?

Value Relevance of UGC to Investors

Although firms provide information to investors through financial statements and other disclosures, numerous studies have shown the existence of asymmetric information between firms (e.g., managers) and investors, resulting in capital market imperfections (Healy and Palepu 2001). Hence, investors may seek information on firms' performance from alternate sources such as investigative reports or experts' reviews in various news media. Investigative reports have a direct impact on stock market performance (e.g., Mitchell and Mulherin 1994); the effect of an expert's reviews on stock prices is also known and is fairly instantaneous (Tellis and Johnson 2007). However, UGC is a direct expression of consumers' personal experience and uncovers feedback on products that may not be evident in investigative reports or experts' reviews in the media. Investors may consult UGC for such unanticipated information about product performance that is not already available in established media reports or experts' reviews. In a study by the Brunswick group of over 448 investment personnel (equity analyst or institutional investors), approximately 43% of them suggested that UGC has become an important determinant in their investment decision in recent years.[1] Also, our discussions with investment houses suggested that many of them regularly monitor new media for various firm-related information.

UGC as a Leading Indicator of Stock Market Performance

UGC could predict stock market performance for two reasons. First, in a perfect market with transparent information systems, all information on the firm would be available immediately to all investors. Unfortunately, information on the firm is usually available sporadically (e.g., disclosures, company press releases) or at a low frequency, usually monthly or quarterly (e.g., corporate reports, earning statements, sales reports). Most investors probably depend on firm-specific information in media stories, site visits, or reports from security or industry analysts, which are also available at a low frequency. Unlike these sources of firm performance, UGC could be observed at a relatively high frequency (daily). Thus, UGC could present new information about the current performance of the firm at a greater temporal frequency than otherwise available to investors. Subject to Granger causality tests (Granger 1969), it also has the potential to act as a leading indicator of firm performance (for more details, refer to Srinivasan et al. 2010).

Second, consumers who have bought the product might chat about the product on the Web. Other consumers who are uninformed about products or who are undecided about which brand to buy may consult UGC to finalize their decision. The online reviews and discussion could subsequently affect their decisions. At the aggregate level, these decisions would translate into future sales, cash flows, and stock market performance. Thus, daily UGC could predict future performance ahead of firms' reports on quarterly sales and cash flows.

Delay in Response to UGC

The information in UGC might take from a few days to weeks to be fully reflected in stock market performance, for three reasons. First, for traders to benefit from the information, they have to systematically monitor and extract information from UGC, as has been done in this study. Unfortunately, most investors may as yet not have the awareness and sophistication to monitor UGC continuously at a daily level. Investors might as yet incorporate the information slowly over a couple of days. Second, transaction costs for the investors may be high enough that they cannot make profitable trades by immediately acting on the information gained. Third, these dynamics can also be attributed to the slow diffusion of information about the firms (e.g., Lo and MacKinlay 1988, Hong and Stein 1999). This pattern holds even more strongly in the case of UGC because the information has to diffuse systematically and evenly between the consumer markets and stock markets. Thus, the information in UGC might take a few days or weeks to be fully reflected in stock market performance.

Asymmetric Response Across UGC Metrics

The content of UGC, as either overall positive or overall negative, could affect the decisions of investors asymmetrically for three reasons. First, investors discount or overlook the positive information because they suspect it is unreliable or because they find it less diagnostic than negative information. This effect may be due to negativity bias, according to which negative information elicits a stronger response than positive information (Baumeister et al. 2001, Rozin and Royzman 2001). Second, for investors, negative information may be more important than positive information because losses loom larger than

[1] See Duckworth et al. (2009).

gains (Tversky and Kahneman 1981). Third, positive information about products, especially new products, is usually well known and actively propagated by firms even before their launch through actions such as announcements and advertisements, whereas negative information is usually not anticipated as it is an uncontrolled outcome of experience of consumers. For any of these reasons, negative metrics may have a stronger impact on returns than positive metrics of UGC.

In sum, because investors do not currently have perfect information about the firm's performance from firm and media reports, UGC could provide an additional source of information and may affect investors' decisions and stock market performance.

Method

This section describes the research design and data collection.

Research Design

The following subsections explain the sampling of firms and markets, time, and media.

Sampling of Firms and Markets. We select firms, products, and markets on several criteria to ensure the feasibility, validity, and reliability of the study. First, the product categories must have rich data on UGC across the time period of investigation. Digital products, high-tech products, and popular consumer durables fall in this class. Second, the products reviewed have to constitute a major fraction of the firm's sales, so that Web chatter about the product would provide a strong signal about the firm's performance. For example, mobile phones constitute the bulk of Nokia's sales. However, for highly diversified firms such as General Electric, Procter & Gamble, or Johnson & Johnson, the chatter about individual products may not always be congruent and may not provide a single clear signal. Third, firms have to be listed on one of the U.S. stock exchanges (NASDAQ/NYSE/AMEX) because stock market performance is the dependent variable. Fourth, firms included in the sample should not have undergone identity changes (such as by mergers or acquisition) during the given time period. Fifth, the sampled markets should represent a cross section of markets.

The use of these criteria leads to our selection of the following six markets: (firms sampled are in parentheses)—personal computing (Hewlett-Packard and Dell), cellular phones (Motorola and Nokia), personal digital assistants or smartphones (Research In Motion and Palm), footwear (Skechers USA, Timberland, and Nike), toys (Mattel, Hasbro, and Leap Frog), and data storage (Seagate Technology, Western Digital, and SanDisk). In each of these markets, we sample as many firms as fit the above criteria. To the best of our knowledge, this is the broadest sample among all published studies on UGC in marketing.

Time Sampling. We choose four and a half years, from June 2005 to January 2010, for our analysis. UGC on selected websites is rather sparse before June 2005. Because both UGC and stock prices are available at very low granularity, we chose the daily level of analysis. Higher levels of aggregation (weekly or monthly) may lead to biased estimates (Tellis and Franses 2006), whereas lower levels of aggregation (hourly) have sparse data on UGC.

Sampling of Media and Sources. We use four criteria to select the format and sources of the media for this study.

First, consumers generate content in a variety of online media formats such as text (e.g., blogs, product reviews) and videos (e.g., video blogs, postings on video-sharing sites such as YouTube). However, during the period under consideration, consumers did not use media such as video much to voice their opinions on products. They used UGC such as product reviews extensively for this purpose.

Second, we choose consumer reviews of products because consumer reviews focus tightly on product evaluations, unlike blogs, forums, and bulletin boards, where the conversations digress greatly from product-related issues, greatly diluting the signal-to-noise ratio.

Third, we focused on consumer reviews rather than expert reviews (e.g., cnet.com, zdnet.com) because consumer reviews reveal experiences of a large number of consumers of actual product performance over extended periods in natural environments (Chen and Xie 2008). As such, they provide strength in numbers or the "wisdom of crowds" (Surowiecki 2005). Moreover, consumer reviews have a big and growing influence on their purchase decisions (A. C. Nielsen 2007).

Fourth, for sources of consumer reviews, we use three popular websites—Amazon.com, Epinions.com (a service of Shopping.com, Inc., an eBay company), and Yahoo! Shopping (a service of Yahoo.com that allows us to access reviews from affiliate sites). These sources are among the most popular websites for consumer reviews and have a very high reach and acceptance among consumers, reflected in the number of unique visitors and the number of years of existence. The parent websites of all the three services fall among the top 10 visited sites (Yahoo!—2, eBay—8, and Amazon—10) as ranked by the unique visitors during data collection (Nielsen 2007). This ranking is also consistent with the comScore ranking (Yahoo! sites—2, eBay—6, and Amazon—8) (comScore 2008).

UGC Data Collection

Because the UGC data are not efficient to collect or process manually, we resort to automated techniques for data collection and analysis. Here, we briefly outline the procedure adopted for data collection and preprocessing. Details are in Online Appendix A (available as part of the online version that can be found at http://mktsci.journal.informs.org/).

The data collection for UGC varies for each site, because no one standard method can get data from these disparate sites. In the cases where the sites allow access to data through some application programming interface (e.g., Amazon Web Services or Yahoo! Shopping), we use that approach. In the case of Epinions.com, where no Web service is available, we develop scripts to collect data from the site periodically. We store the collected reviews at a disaggregate (individual-review) level so that we can parse the individual reviews and aggregate them for our analysis. We extract and store numerical data, such as ratings, in numeric format. In total, the sample contains 347,628 reviews between June 2005 and Jan 2010.

To arrive at the valence of the reviews, we use two algorithms proven to be reliable for text classification applications in a specific domain (such as product evaluation): the naïve Bayesian classifier and the support vector machine classifier. The details can be found in Online Appendix A. We use voting between these algorithms to arrive at the valence of the review, as has been done in the prior literature (Das and Chen 2007). Almost 94% of classifications of the reviews are in agreement between the two algorithms. Among the rest, we resort to manual classification if the total number of words in the reviews is more than 10. We discard those reviews that have 10 words or fewer because of a lack of information content in these texts. We then aggregate the individual consumer reviews of a particular firm in any given day. Doing so, we derive the daily time-series data for each metric of UGC for every firm in our sample.

Measures

This section describes the measures of UGC, stock market performance, and the control variables.

Measures of UGC

UGC can be characterized by several metrics (e.g., Godes and Mayzlin 2004, Liu 2006). We restrict our analysis to four important metrics: ratings, chatter, positive valence, and negative valence. We explain each of these metrics below.

Ratings. The simplest measure that we use is consumer ratings. Ratings are the numerical assessment of the product by consumers based on a numeric scale designed by each website (on a scale of 1 (bad) to 5 (good)). We measure the aggregate rating of a firm by taking the arithmetic mean of all the individual ratings of a firm across websites in a day.

Volume of Chatter. Volume of chatter refers to the total number of reviews posted by consumers about the products of a firm in a day. This measure reflects the magnitude of coverage received by the firm in UGC.

Valence. Valence of UGC refers to whether the overall review is positive or negative. We derive the valence of a review from an analysis of the text in the reviews using computational procedures as explained in Online Appendix A. The statistical algorithms used for the binary text classification of the reviews are proven to be robust (e.g., Domingos and Pazzani 1997, Joachims 1998). We classify the review as positive or negative and then count the number of classified reviews for a given firm in each time period, which constitutes the positive or negative chatter, respectively. In the tests of robustness, we test alternate measures of valence by counting the number of positively (or negatively) laden words for each firm in a given time period.

Measures of Stock Market Performance

This subsection explains the method for the measurement of the three measures of stock market performance: abnormal returns, idiosyncratic risk, and trading volume. We first specify the firm's daily expected return using the Fama–French (1993) three-factor (plus the Carhart 1997 momentum factor) model, where the factors account for marketwide factors influencing a firm's returns. Prior studies in marketing have used this technique to measure for firm's abnormal returns and idiosyncratic risk (e.g., Sood and Tellis 2009, Luo 2009, Wiles et al. 2010, Tuli and Bharadwaj 2009). We derive two financial measures from this model: the estimated abnormal returns and the estimated conditional volatility (risk).

We model the conditional variance of the error in this model as an EGARCH process (exponential generalized autoregressive conditional heteroskedasticity; see Nelson 1991) to account for the time-varying nature of the firm's risk. The EGARCH specification is proven to have a better fit than any of the other conditional heteroskedasticity specifications and helps capture the asymmetry of conditional volatilities (e.g., Fu 2009, Huang et al. 2010, Engle and Ng 1993). We specify the returns of a firm as

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{i,MKT}(R_{MKT,t} - R_{f,t}) + \beta_{i,SMB}SMB_t + \beta_{i,HML}HML_t + \beta_{i,MOM}MOM_t + \varepsilon_{i,t},$$

where

$$\varepsilon_{i,t} \sim \mathrm{N}(0, \sigma_{i,t}),$$

$$\ln(\sigma_{i,t}^2) = a_i + \sum_{j=1}^{p} b_{i,j} \ln(\sigma_{i,t-j}^2) + \sum_{k=1}^{q} c_{i,k} \left\{ \Theta\left(\frac{\varepsilon_{i,t-k}}{\sigma_{i,t-k}}\right) + \Gamma\left(\left|\frac{\varepsilon_{i,t-k}}{\sigma_{i,t-k}}\right| - \left(\frac{2}{\pi}\right)^{1/2} \right) \right\}, \tag{1}$$

where t is the subscript for the time period, i is the firm-specific index, $R_{i,t}$ is the return[2] of the stock i in time t, $R_{MKT,t}$ is the market portfolio return, $R_{f,t}$ is the risk-free rate of return (30-day treasury bill), SMB_t is the small-minus-big capitalization factor, HML_t is the high-minus-low book-to-market equity factor, and MOM_t is the momentum factor in the given time period. The βs are the factor loading coefficients, and a, b, and c are the coefficients of the risk process.

Prior studies across marketing and finance use similar measures of idiosyncratic abnormal returns and risk (e.g., Tuli and Bharadwaj 2009, Luo 2009, Fu 2009, Huang et al. 2010). We run this regression for a rolling window of 250 trading days prior to the target day to get the estimated factor coefficients of the Fama–French model (Equation (1)). The risk-adjusted returns (abnormal returns) on time period $t+1$ is taken as the estimated abnormal returns of a firm for that period (e.g., Avramov et al. 2009). Specifically, the abnormal return for a firm in time period $t+1$ is

$$[R_{i,t+1} - R_{f,t+1}] - \{\hat{\beta}_{i,MKT}(R_{MKT,t+1} - R_{f,t+1}) + \hat{\beta}_{i,SMB} \cdot SMB_{t+1} + \hat{\beta}_{i,HML} HML_{t+1} + \hat{\beta}_{i,MOM} MOM_{t+1}\}. \tag{2}$$

Henceforth, in the interests of brevity, we use the term "returns" for "abnormal returns." The idiosyncratic risk is the estimated conditional variance $\sigma_{i,t}$ of the residuals. Prior studies (e.g., Fu 2009) show that the estimated conditional variance provides an accurate estimate of the idiosyncratic risk.[3] We repeat this procedure on a rolling basis for all the trading days in the sample to get the returns and idiosyncratic risk for each day. The statistical properties of the abnormal returns are summarized in Online Appendix C5.

We measure trading volume as the daily turnover that is measured as the volume of trade of a firm in a given day adjusted for the number of shares outstanding at the end of the day (e.g., Campbell et al. 1993, Chordia and Swaminathan 2000), thus controlling for any potential firm size effects. All the financial data for calculating stock returns, trading volume, and the factors used in the model are available from CRSP.

[2] Stock returns are calculated on the daily closing price adjusted for corporate actions (e.g., dividends).

[3] We find different permutations of EGARCH (p, q) specifications (EGARCH $(1, 1)$, EGARCH $(1, 2)$, and so on till EGARCH $(3, 3)$), for day $t+1$ for each of the individual stocks. We then choose the specification with the lowest Akaike information criterion (AIC) (Fu 2009).

Measures of Control Variables

We describe the measures of four control variables: *analysts' forecasts*, *advertising*, *media citations*, and *new product announcements*.

Analysts' Forecasts. The analysts' forecasts are measured as the median estimate of the analysts' consensus forecast of the earnings obtained from the I/B/E/S database.

Advertising. A firm's advertising spending is the daily dollars spent on television advertising by a firm obtained from TNS Media Intelligence.

Media Citations. Media citations are the number of articles in the print media on the firm in any given day from LexisNexis and Factiva. We search for relevant firms or firms in these databases as follows. First, we do an elaborate search for all articles that mention the name of the firm or brands (or subbrands) in our sample. We search across major newspapers, dailies, and news wire services (including the *Wall Street Journal*, which is covered by Factiva) in each of these databases for each day within the time horizon of our study. LexisNexis assigns a relevancy score for each article in its index and assigns it in the corresponding tags in the results. We use this score to ensure that the articles do indeed discuss the brand or the firm of interest and are not incidental mentions in the article. Specifically, we identify the articles that are relevant to the firm if an article is identified as having a relevancy score of 60% or more in the LexisNexis indexing with respect to the given firm name. Similarly, for Factiva, we use the company tag, which shows if the search result is indeed relevant to the firm under investigation.

New Product Announcements. We also use the above databases and method to obtain announcements of new products for each firm in each category. In addition to the rules above, we identify firm-specific new product announcements using the procedure outlined in Sood and Tellis (2009).

We then merge the UGC data with the financial data (returns, idiosyncratic risk, trading volume), advertising data, new products data, and analysts' forecasts based on firm and date to get the 1,112 trading days of data.

Models

This section describes the models for estimating the relationships between measures of UGC and measures of stock market performance. It covers the rationale for the model, tests for stationarity and units roots, the test for Granger causality, model specification, and modeling dynamics.

Rationale for VAR

We adopt the persistence-modeling framework using the vector autoregression (VAR) for our empirical investigation. VAR is suitable for examining the dynamics of the relationship between UGC and stock performance for several reasons. First, VAR is appropriate for this study as opposed to event studies, because UGC is generated continuously over time and is not a discrete event (Campbell et al. 1997, Srinivasan and Hanssens 2009). The VAR model allows us to examine the immediate and lagged-term effects of different UGC metrics on stock market performance (Dekimpe and Hanssens 2004). Second, it allows us to account for direct and indirect feedback effects among the endogenous variables (here, the three stock market performance metrics and four UGC metrics) through the system of equations. Third, it captures the dynamics of carryover effects over time through the generalized impulse response functions, which helps in and also assesses the relative contribution of the different metrics of UGC through the generalized forecast error variance decomposition (Pesaran and Shin 1998)—both of which are robust to the assumptions of causal ordering of the variables. Fourth, it helps control for trends, seasonality, nonstationarity, serial correlation, and reverse causality (Luo 2009), and it is robust to such deviations. Prior studies in marketing and finance have successfully used vector autoregressive models for similar purposes (e.g., Joshi and Hanssens 2010, Luo 2009, Pauwels et al. 2004, Tetlock 2007). Thus, our system of equations in the VAR model assesses the dynamics of the relation between the UGC metrics and the stock performance metrics accounting for exogenous factors and is robust to deviations from the assumptions of stationarity, heteroskedasticity, and serial autocorrelation.

We carry out the analysis in the following steps: (1) Estimate the stationarity properties of various metrics of stock market performance and UGC using the unit root and cointegration tests. (2) Test for causal relationship among the variables through the Granger causality test. (3) Estimate dynamics of carryover effects through analysis of the impulse response functions. (4) Estimate the contribution of the metrics using variance decomposition. Finally, to check the generalizability of the results across categories, in the robustness section we run a panel VAR model.

Test for Stationarity, Unit Roots, and Cointegration

We conduct stationarity and unit root tests to examine the stability of the statistical properties of the metrics of stock market performance and UGC. These unit root tests investigate whether the variables entering the system evolve continually or are stationary. Following the literature (Dekimpe and Hanssens 2004), we used both the Augmented Dickey–Fuller (ADF) test for the evolution of variables and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test (Kwiatkowski et al. 1992) so as to ensure that our results are robust to alternate formulations of the null hypothesis. If the variables are integrated in their levels, we run Johansen's procedure to test for cointegration (Johansen 1995) to examine whether any linear combination of these variables exhibit lower order of integration. The results would indicate whether the variables share an equilibrium relationship over a long time, i.e., exhibit a permanent shift in the equilibrium rather than reverting to their mean levels after a few time periods. Depending on the results of the stationarity tests and the cointegration tests, we choose the appropriate specification for the variables entering the VAR model. If the variables are stationary, we choose it in the levels; if the variables evolve, we use the first difference of the series in the model; and if the series are cointegrated, we have to adopt error correction models with an error term for adjustment. The metrics of stock market performance are stationary in the levels, and the metrics of UGC are evolving but not cointegrated; hence, we use the first differences of the UGC metrics. The details of implementation of these steps are discussed along with the results.

Test for Granger Causality

The details of the tests for Granger causality are in Online Appendix C2.

Model Specification

We specify the relationship among the metrics of UGC and stock market performance through the following VAR model (with exogenous variables):

$$Y_t = \sum_{n=1}^{p} \Gamma_n Y_{t-n} + \Phi X_t + e_t, \tag{3}$$

where, $t \in \{T_0, T_1, T_2, \ldots, T\}$ is the time period index, Y is the vector of the endogenous variables in the system, Γ_n are the coefficients matrices of the lags of endogenous variables, X is the vector of control variables and Φ is its coefficients, and e is the error term. The optimal lag order ("n") for the VAR model is chosen by the (Schwarz) Bayesian information criterion (BIC) (Lütkepohl 2005). We specify the variables in levels or first differences, depending on the order of integration of the variables, which is determined through the unit root and cointegration tests. For example, the metrics of UGC and the trading volume enter the model in first differences because they are stationary in their first difference but not in the levels, and they also do not exhibit cointegration (discussed in detail in the Results section). Taking these factors into account, Equation (4) is specified as

$$\begin{bmatrix} AbRet_t \\ IdioRisk_t \\ TrdVol_t \\ \Delta VolChtr_t \\ \Delta Ratings_t \\ \Delta NegChtr_t \\ \Delta PosChtr_t \\ \Delta CompChtr_t \\ \Delta CompNegChtr_t \end{bmatrix} = \sum_{n=1}^{p} \begin{pmatrix} \gamma_{11}^{n} & \cdots & \gamma_{19}^{n} \\ \vdots & \ddots & \vdots \\ \gamma_{91} & \cdots & \gamma_{99} \end{pmatrix} \cdot \begin{bmatrix} AbRet_{t-n} \\ IdioRisk_{t-n} \\ TrdVol_{t-n} \\ \Delta VolChtr_{t-n} \\ \Delta Ratings_{t-n} \\ \Delta NegChtr_{t-n} \\ \Delta PosChtr_{t-n} \\ \Delta CompChtr_{t-n} \\ \Delta CompNegChtr_{t-n} \end{bmatrix} + \begin{pmatrix} \phi_{11} & \cdots & \phi_{1p} \\ \vdots & \ddots & \vdots \\ \phi_{91} & \cdots & \phi_{9p} \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} + \begin{bmatrix} e_{AbRet,\,t} \\ e_{IdioRisk,\,t} \\ e_{TrdVol,\,t} \\ e_{VolChtr,\,t} \\ e_{Ratings,\,t} \\ e_{NegChtr,\,t} \\ e_{PosChtr,\,t} \\ e_{CompChtr,\,t} \\ e_{CompNegChtr,\,t} \end{bmatrix}, \tag{4}$$

where *AbRet* is the abnormal returns, *IdioRisk* is the idiosyncratic risk, and *TrdVol* is the trading volume (daily turnover). These three are the measures of stock performance. The UGC metrics in the model are the *VolChtr*, the volume of chatter; *Ratings*, the UGC ratings; *NegChtr*, the negative chatter; *PosChtr*, the positive chatter; *CompChtr*, the competitor chatter; and *CompNegChtr*, the negative chatter of the competitor. The off-diagonal terms of the matrix $\Gamma - \gamma_{kl}^{n}$ (k and l being the rows and columns of the matrix, respectively) estimate the indirect or the cross-carryover effects among the endogenous variables, and the diagonal elements (γ_{kk}^{n}) estimate the direct effect. The vector X comprises the p control variables—analysts' forecasts, advertising, media citations, new product announcements, and the seasonal dummies (month of the year for November, December, and January; holidays such as Labor Day, Memorial Day). Each of the equations in the system explains the influence of UGC after controlling for the above control variables.

We estimate the model using generalized method of moments (GMM) (Hansen 1982), which does not make distributional assumptions on the data and controls for heteroskedasticity and temporal autocorrelation in measures of stock performance and UGC. Campbell et al. (1997, pp. 173–175) suggest that estimating models explaining the returns using firm (or marketing) characteristics by GMM are consistent and unbiased. Furthermore, they suggest using heteroskedasticity-consistent standard errors to overcome any temporal heteroskedasticity. To this end, we use the heteroskedasticity- and autocorrelation-consistent covariance estimator (also referred as the Newey–West standard error; see Newey and West 1987), calculated for up to five lags. More details of the estimation procedure and its econometrics properties can be found in Hayashi (2000, Chapter 6) and Hamilton (1994, pp. 409–416).

Modeling Dynamics and the Relative Contribution of Metrics

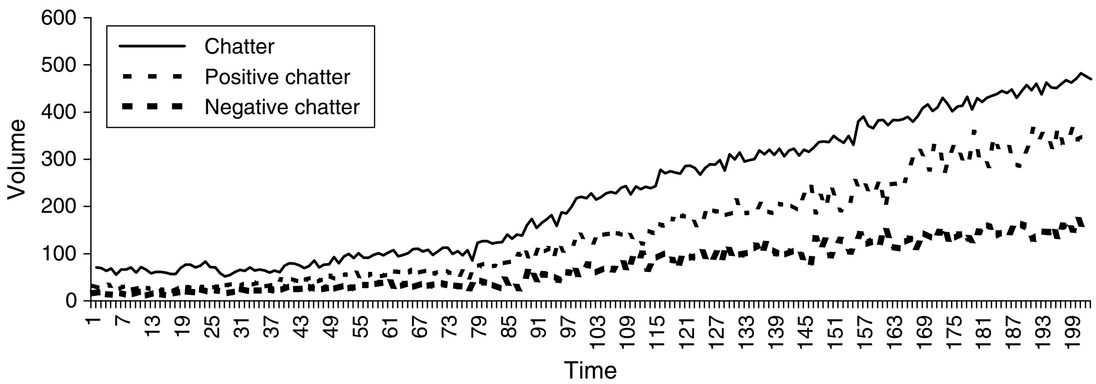
We model the dynamics in the system with the help of the generalized impulse response function (Pesaran and Shin 1998) based on the above VAR estimation. We use the generalized impulse response function for these because it is not sensitive to the causal ordering of the variable entering the system. We compute the confidence intervals of the impulse response functions using Monte Carlo simulations. We further assess the relative impact of the metrics of UGC on the stock performance using the generalized forecast error variance decomposition techniques (Koop et al. 1996, Pesaran and Shin 1998).

Table 2 **Distribution of Metrics of UGC Across the Markets**

Markets	% Positive chatter	% Negative chatter
Personal computer	75.25	22.17
Cell phone	70.25	25.36
PDA/smartphone	75.24	22.89
Data storage	65.42	30.75
Toys	78.25	18.74
Footwear	87.56	10.87
Average	75.33	21.79

Note. The percentage of positive and negative might not add to 100 because some of the observations could not be classified either as positive or negative chatter and are not included for further analysis of valence of chatter.

Figure 1 Time-Series Plot of Selected Metrics of UGC



Results

Table 2 shows the distribution of positive and negative metrics of UGC across the various markets for the sampled time. Note that positive comments dominate the reviews across markets. Figure 1 depicts the trend of metrics of UGC over time aggregated to the weekly level. Note how all metrics show an upward trend as UGC becomes a more popular medium in the population. Figure 2 shows the time-series plots of chatter and stock prices (adjusted for corporate actions such as splits and bonuses). The sampled time covers a period when the overall stock market went up substantially and fell steeply. Moreover, each stock shows a varying and deferring pattern in upward and downward movement. Thus, there is substantial variation in the raw data. The figure shows a moderate relationship between stock prices and chatter for all firms. The subsequent subsections examine this relationship systematically using the models outlined above.

We first present the tests for stationarity, unit root, cointegration, and Granger causality. We next estimate the model and derive the impulse response functions to examine the short-term and accumulated (long-term) relationship between UGC and stock performance and the duration of the impact. We next estimate the relative importance of the metrics of UGC by generalized forecast error variance decomposition. Finally, we describe a portfolio analysis and various tests for robustness.

Tests for Stationarity and Unit Root and Cointegration

We use the ADF test to determine whether the variables are evolving or stationary. We also examine the variables with the KPSS test (Kwiatkowski et al. 1992) to ensure that our results are robust to assumptions of null hypothesis of unit root. For the returns, the presence of a unit root in the ADF test is safely rejected (see Table 3), suggesting that the returns are stable and stationary in the given time period. Alternatively, the KPSS test also confirms that the absence of unit root for stock returns as the null hypothesis for this test cannot be rejected at 95% confidence level (here, the null hypothesis assumes stationarity or absence of unit root). We run these tests on the various metrics of UGC with and without the trend to account for growth of the online media during the time period under consideration. The results of the ADF tests on the levels of various metrics of UGC have values ranging from −0.25 to −3.26 (much less than the range of critical values at 0.05 level of significance), suggesting that the metrics of UGC evolve in levels. We then ran the cointegration tests (Johansen 1995) to see whether any combination of the UGC metrics and stock performance variables (returns, trading volume, or risk) show cointegration over the given time period (the details of the cointegration tests are in Online Appendix C1). Because no cointegration is observed among the variables, we take the first differences of the metrics of UGC and run the ADF and the KPSS tests to assess their suitability as possible endogenous variables in the VAR model.[4] The first differences of the metrics of UGC are stationary, as shown in Table 3.

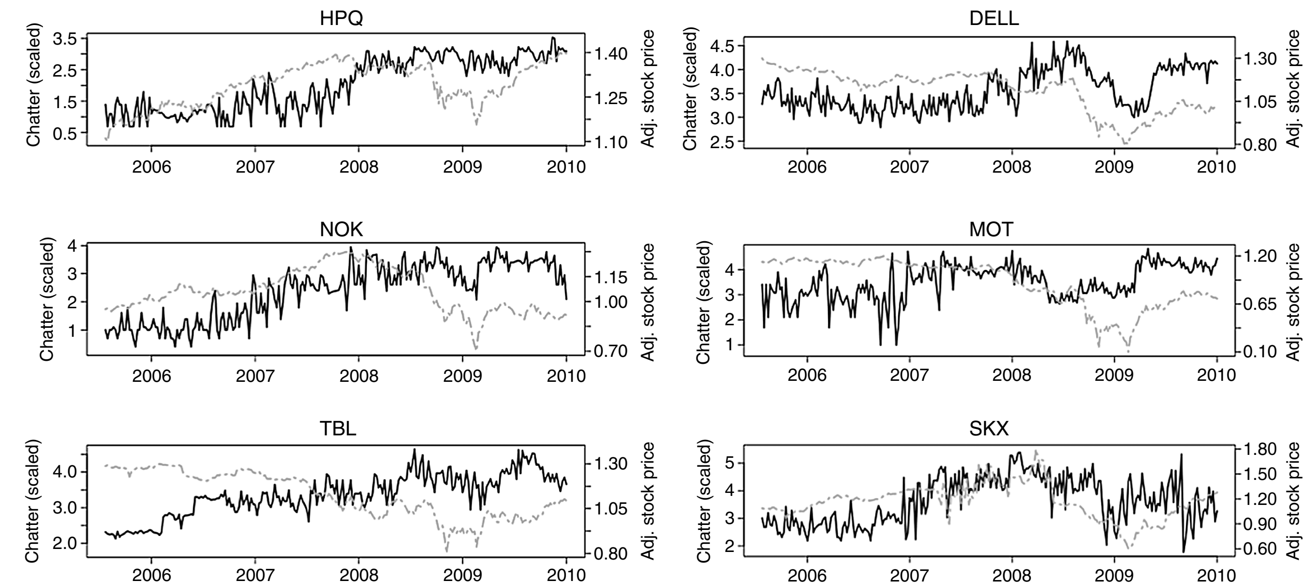
Granger Causality Tests

We test for the presence and direction of a causal relationship between each metric of UGC and returns or risk by conducting the Granger causality tests. In the first set of tests, we assume the null hypothesis that the metric of UGC (e.g., *Volume of chatter*) does not "Granger cause" returns. We use five[5] lags considering the Akaike information criterion and the Schwarz Bayesian information criterion. Table 4 shows the median p-values of the Wald test (chi-square) statistic for Granger causality. The results suggest that some metrics of UGC significantly Granger cause returns

[4] More details on the steps of VAR framework can be found in Dekimpe and Hanssens (2004).

[5] We repeat these tests for lags up to 25 days on each of the metrics. There was no significant temporal causation seen in these higher-order lags.

Figure 2 Time-Series Plot of Chatter and Stock Prices of Selected Markets (January 2005 to December 2009)



Note. HPQ, Hewlett-Packard; MOT, Motorola; NOK, Nokia; SKX, Skechers USA; TBL, Timberland.

while others do not. Volume of chatter shows a strong impact on returns (median $p = 0.04$, ranging from 0.02 to 0.07 across firms), followed by negative chatter (median $p = 0.05$ ranging from 0.03 to 0.07) and then positive chatter. The average ratings do not Granger cause stock market performance. The reverse feedback (causality) from returns to metrics of UGC is not significant (median p-value ranging being 0.23, ranging from 0.13 to 0.49). The instantaneous causality also emerges to be insignificant across the metrics (median p-value of 0.29). These results confirm the temporal causal relationship between the metrics of UGC and stock returns.

We do similar tests for risk and trading volume. The details of the result are in Table 4. The results suggests that negative chatter Granger causes risk (median $p = 0.05$; range, 0.02 to 0.08). Other metrics (volume of chatter, positive chatter, or average ratings) do not exhibit statistical significance. The reverse feedback from risk to any of the metrics of UGC is not significant (median $p = 0.28$). We do not find any significant change in the results of returns or risk because of the inclusion of higher-order lags.

Table 3 Summary of Unit Root/Stationarity Tests of the Endogenous Variables

	ADF test	KPSS test
Returns	−4.62	0.11
Idiosyncratic risk	−4.48	0.26
Trading volume	−4.42	0.28
Δ*Volume of chatter*	−4.79	0.23
Δ*Negative chatter*	−4.76	0.25
Δ*Positive chatter*	−4.60	0.20
Δ*Ratings*	−4.63	0.26
Δ*Competitor chatter*	−4.58	0.20
Δ*Competitor negative chatter*	−4.62	0.25

Notes. All the metrics of UGC are measured in first differences. The critical value at 5% is −3.44 for the ADF test and 0.463 for the KPSS test. The details can found in Tables B1 and B2 (in Online Appendix B).

Table 4 Summary of the Results of Granger Causality Tests

	Δ*Chatter*	Δ *Negative chatter*	Δ*Positive chatter*	Δ*Ratings*
Returns	**0.04**	**0.05**	0.13	0.15
Risk	0.11	**0.05**	0.18	0.19
Trading volume	**0.03**	0.06	0.08	0.13

Notes. The estimates of the Granger causality are the median of the significance of the joint chi-square statistics. Significant values are represented in bold. Results by individual firms are in Table B3 in Online Appendix B.

Short- and Long-Term Relationship Between UGC and Stock Performance

We estimate the VAR model with the returns, risk, trading volume, and the metrics of UGC (volume of chatter and negative chatter) as the endogenous variables. We test the residuals of the models for some of the assumptions of the VAR model using standard diagnostic tests.[6] The optimal lag length for the VAR models is three, as determined by Schwarz's Bayesian information criterion.

[6] We test the assumption of autocorrelation using the Durbin–Watson and Lagrange multiplier tests; we test the assumption of multivariate normality using the Jarque–Bera test. Because we do not get significant positive results in any of these tests, we safely conclude that the models in our analysis are robust to any of these misspecifications.

We show the short-term and long-term dynamics between the metrics of UGC, returns, and risk using the estimated VAR model through simulations of the generalized impulse response function (Pesaran and Shin 1998, Dekimpe and Hanssens 1999). The generalized impulse response function uses the VAR estimates to trace the effect of a unit shock (one standard deviation) in any one of the variables (a measure of UGC) on all other variables (returns, risk, or trading volume) in the system over subsequent periods. We define the short-term impact as the effect derived from estimates of the VAR model for the first three time periods, the average time taken for the effect of UGC metrics to reach their peak effect. We define the long-term or cumulative impact as the accumulated impact of the impulse response function (of the metrics on UGC on metric of stock performance) to reach its asymptote. Most of the accumulated effect on stock returns reaches the long-run (asymptotic) levels within 10 time periods. Hence, in effect, we take the long-term (accumulated) duration as a period of 15 days, ensuring that we have captured the total effect of UGC metrics on stock performance. To avoid ambiguities in the representation of the effect of returns, we follow the convention in the finance and accounting literatures, and we express the short-term and cumulative impact on stock returns in basis points (one basis point is one-hundredth of a percentage).

Table 5 presents the results, averaged across the firms, of the short-term and accumulated impact of the metrics of UGC on returns and risk.[7] Among the various metrics, volume of chatter has a high positive impact on returns in the short term (6.3 basis points, $p < 0.01$) and a similar accumulated impact of 14.8 basis points ($p < 0.01$). Volume of chatter also has a significant positive influence on the trading volume with the short term and an accumulated impact of 0.08% and 0.10% of daily turnover, respectively ($p < 0.01$). Although volume of chatter has a significant impact on returns, it has no significant impact on firm's risk. Positive chatter does not have a significant impact on returns, risk, or trading volume. Negative chatter, on the other hand, influences returns negatively both in the short term (−4.2 basis points, $p < 0.05$) and in the long term (−8.4 basis points, $p < 0.05$). This figure is comparable with that in the study of Tellis and Johnson (2007). In contrast to the effect on stock returns, negative chatter increases the risk and trading volume of the firm. These results are over and above the effects of advertising, which is separately accounted for in our model. This result shows that the word-of-mouth effects are significantly larger than advertising.

[7] The detailed results for each firm are in Online Appendix B in Tables B4, B5, and B6 for stock returns, volatility, and trading volume, respectively. The values presented here are the median values.

Table 5 **Short-Term and Accumulated Impact of UGC Metrics on Stock Performance**

	Returns	*Risk*	*ΔTrading volume*
ΔChatter			
Immediate	**6.31**	0.001	**0.075**
Accumulated	**14.82**	0.021	**0.110**
ΔRatings			
Immediate	1.81	0.002	0.045
Accumulated	2.34	0.009	0.079
ΔNegative chatter			
Immediate	**−4.21**	**0.005**	**0.025**
Accumulated	**−8.37**	**0.011**	**0.058**
ΔPositive chatter			
Immediate	3.21	0.002	0.038
Accumulated	6.27	0.006	0.086
ΔCompetitor chatter			
Immediate	**−2.18**	**0.002**	**0.023**
Accumulated	**−5.13**	**0.010**	**0.069**
ΔCompetitor negative chatter			
Immediate	**3.16**	0.003	0.051
Accumulated	**7.36**	0.008	0.076

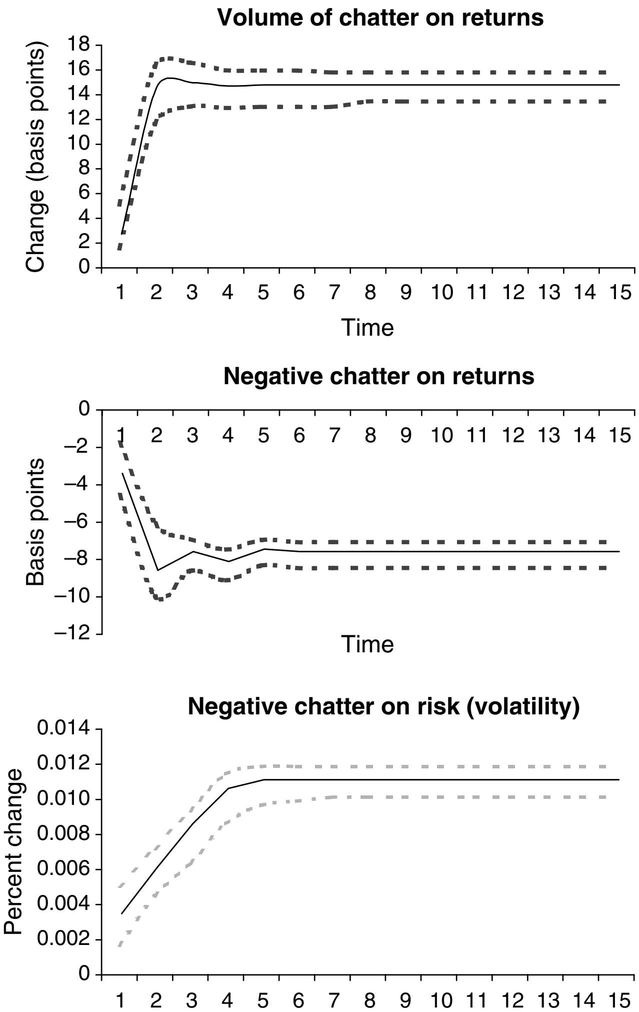
Notes. Estimates from the VAR model and the impulse response functions. Only the relevant endogenous variables are shown in this table. The values for returns are in basis points (1 basis point = one hundredth of a percentage). The values for risk are measured as change in stock prices, and the trading volume is measured as percentage daily turnover. Significant values are represented in bold. The details for individual firms can be found in Tables B4–B6 (see Online Appendix B).

Although these effects seem to be small in basis points, they have a substantial impact in terms of the dollar value. In monetary terms, the negative relationship between chatter and returns could translate into a substantial impact on the market capitalization of the firms. Other factors remaining equal, a unit shock to negative UGC could erode about $1.4 million from the average market capitalization in the short term and an accumulated value of $3.3 million over the 15 days following the chatter. Surprisingly, the influence of the ratings assigned by consumers to the products does not seem to have any significant effect on returns or trading volume. We also test the model by separating the ratings as positive (greater than 3) and negative (less than 3). The inclusion of the positive and negative ratings in the Granger causality test or the VAR model does not change the results. This result suggests that textual content of the reviews has more information content than a summary measure of rating.

Effects of Competitor UGC

Similar to the effect of own UGC on the returns, we also assess the effect of the competitor UGC on a firm's UGC metrics and financial metrics (returns, risk, and trading volume). Increase in competitor negative chatter UGC increases the volume of chatter of the firm. A percentage increase in competitive chatter increases the volume of chatter of the target firm by 7% on an average (range 2%–13%). The last two rows of Table 5 show that an increase in competitor chatter has an adverse effect on the returns of the target firm. An increase in competitive chatter decreases the returns of a target firm by 2.2 basis points in the short term and by 5.1 basis points cumulatively (both significant at 0.01 levels). Competitor chatter also has an influence on the risk of the target firm (0.2% change in stock price in the short term and 1.0% cumulatively, $p < 0.01$). Similar results hold for the effect of competitor chatter on the target firm's trading volume. On the other hand, an increase in competitor's negative chatter has a positive influence on the firm's returns. The firm stands to gain 3.2 basis points from changes in competitor's negative chatter with an accumulated gain of 7.4 basis points. However, the negative chatter of competitors does not seem to influence the risk or the trading volume.

Figure 3 **Accumulated Impulse Response Functions of Key UGC Metrics**



Notes. The vertical axis shows the returns (in basis points) in panels 1 and 2. In panel 3, the vertical axis is the idiosyncratic risk. In all the panels, the horizontal axis depicts the time in days. The graph shows the confidence band (light) on either side of the response function (bold).

Duration of Impact

Figure 3 illustrates the results of the impulse response function for Hewlett-Packard in the personal computer market. The first panel of Figure 3 shows that a one-unit shock in volume of chatter has an increasing impact on returns, reaching a peak (wear-in) in the second day. Returns then wear out over the next few days to long-term equilibrium, resulting in an accumulated value of approximately 14.8 basis points. The impulse response of returns to negative metrics of UGC contrasts sharply with that to volume of chatter. The second and third panels of Figure 3 depict the impact of a unit shock of negative chatter on stock returns and volatility, respectively. Negative chatter has a negative impact on returns immediately with an accumulated loss of 8.4 basis points over the next few days, whereas the negative chatter increases the risk associated with the stock, which peaks in about four days. The negative metrics of UGC show a strong immediate impact on returns on the subsequent day and decay to the equilibrium levels over the next four days. Table 6 summarizes duration of these effects.

Relative Importance of Metrics of UGC

We use the generalized forecast error variance decomposition (Pesaran and Shin 1998) of returns and risk to assess the relative importance of the various metrics of UGC. The decomposition ascertains the extent to which change in various metrics of UGC contributes to the deviation of returns, risk, and trading volume from baseline expectations. We use the different metrics of UGC (ratings, volume of chatter, negative chatter of those of the firms, and the volume of chatter and negative chatter of the competitors) endogenously with the stock performance variables in the VAR specification described in Equation (4).

Table 7 shows the relative importance of these metrics in determining the stock returns, risk, and trading volume. The results show volume of chatter to be the most important metric that influences the stock performance, followed by negative chatter. Whereas chatter explains 6.5% of variance of returns and 5.6% of the variance of trading volume, negative chatter explains 4.7% of variance of returns and 4.6% of variance of change in trading volume.[8] Competitor's UGC explains 0.83%, with competitor's chatter

[8] The total variation has been normalized and the remaining variation in returns, risk, or trading volume is contributed by the historical values of the respective stock performance variable.

Table 6 Duration of the Short and Accumulated Impact

	Δ*Chatter*		Δ*Negative chatter*		Δ*Positive chatter*		Δ*Competitor*		Δ*Competitor negative chatter*	
	Wear-in	Wear-out	Wear-in	Wear-out	Wear-in	Wear-out	Wear-in	Wear-out	Wear-in	Wear-out
Returns	2	3	2	5	1	2	2	1	1	2
Risk	1	5	4	7	1	3	1	2	1	2
Trading volume	3	5	1	4	2	2	2	3	1	3

Notes. From the generalized impulse response function. All values are depicted in mean time (in days).

Table 7 Variance in Stock Performance Explained by the Metrics of UGC

	Return	*Risk*	*Trading volume*
Δ*Chatter*	6.54	3.65	5.64
Δ*Ratings*	0.3	0.24	1.13
Δ*Negative chatter*	4.65	4.56	2.53
Δ*Positive chatter*	0.56	0.79	0.74
Δ*Competitor chatter*	0.45	0.27	0.35
Δ*Competitor negative chatter*	0.38	0.64	0.21

Notes. From generalized forecast error variance decomposition. All figures are in percentages. The table depicts the normalized percentage contributions by metrics of UGC for the first period only. The other contributions by the variables own past values or by other stock performance variables are not shown.

contributing 0.45% and competitor's negative chatter contributing 0.38% of variance. Similar results hold for the influence of chatter on risk, where negative chatter accounts for the largest variation (4.56%) in risk among all the UGC metrics. We could safely conclude that the firm's risk increases because of the increase in negative information and uncertainty of future stock market performance.

Consistent with prior results, the effect of chatter persists for about six days (see Figure 4). After an immediate influence on returns, the contribution of volume of chatter and negative chatter to the variance of the stock returns diminishes rapidly after four days. Similar results hold good for the influence of chatter on trading volume. Negative chatter shows a similar influence on returns over time, although the percentage of variance explained is not as much as that of volume of chatter. However, the influence of negative chatter on risk is strong and persists for about six days.

Impact of Control Variables

The impact of chatter on stock market performance holds after controlling for several exogenous variables: advertising, new product announcements, media citations, and analysts' forecasts. We next describe the impact of these control variables.

Advertising has a significant positive influence on chatter. A 1% increase in advertising expenditure increases volume of chatter by 0.09 ($p = 0.05$), averaged across the firms. These figures translate into an advertising elasticity of 0.09. These estimates of advertising elasticity are slightly lower than the average advertising elasticity of 0.1 in the literature (Sethuraman et al. 2011, Tellis 2004). Although advertising has a strong influence on the volume of chatter, it does not have a statistically significant effect on the positive chatter or ratings of the firms. The elasticity is higher for firms in market categories such as data storage products, footwear, and toys than for firms in other categories. This difference could occur because of the greater brand differentiation in the latter categories, especially footwear and toys. Advertising has a negative impact on the negative metrics. For every percent increase in advertising, the negative chatter decreases by 0.064.

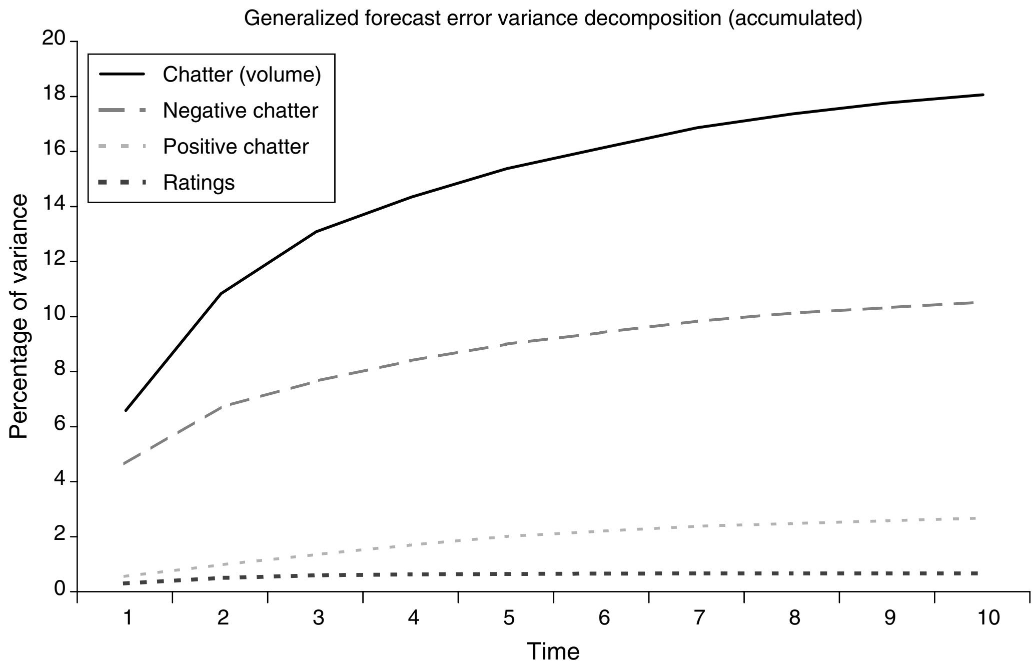
New product announcements have a positive impact on chatter in certain product categories such as cell phones and computers. However, this relationship is not statistically significant for product categories such as toys and footwear. This result may be because online consumers are more responsive to digital products than to other products. The elasticities of all the metrics are shown in Table 8.

Media citations and analyst forecasts do not influence any of the UGC metrics. The results of the effect of metrics of UGC on the stock performance hold good even after these factors are accounted for in our model.

Analysis Over Time

We repeat the analysis over rolling windows of time to investigate whether the phenomenon changes over time. We use a rolling window of two years starting from June 2005 till December 2009, repeatedly reestimating Equation (4). Table B8 (see Online Appendix B) summarizes the results for the three UGC metrics, which were significant in the prior analysis. We find that, on average, the magnitude of the effect seems to be decreasing to a small extent over time for all the metrics (both in the short term and in long term). The size of the effect is much more pronounced for the accumulated response than for the short-term response. The immediate response to volume of chatter decreases over time, from 6.85 basis points (in the period 2005–2007) to 6.4 basis points (in the period 2007–2009). This amounts to an

Figure 4 Explained Variance of Returns by Metrics of UGC



Notes. From the accumulated generalized forecast error variance decomposition. The figure depicts the normalized percentage contributions by metrics of UGC for each time period. The contributions from all the variables at each time point sums to 100%. The rest of the contributions to the variance are by the past returns, risk, and trading volume (not shown here to retain the readability of the graph).

average decrease of 3.1% for the short-term response and an accumulated decrease of 8.8% over time to the change in volume of chatter. Similarly, we find that the immediate response to negative chatter averages a 2.5% decrease over the years and the accumulated response decreases by an average of 12.1%. These results may imply increasing market efficiency as investors increasingly become aware and process UGC. However, we must be cautious in interpretation because this analysis over time is done over a period of only three years.

Robustness Tests

We carry out several tests to ascertain the robustness of the above results.

Analysis Using Panel Time Series. Alternatively, we also model the relation between the metrics of UGC and stock performance as a dynamic system of simultaneous equations using the panel vector auto regressive model (e.g., Horváth et al. 2005, Pauwels and Srinivasan 2004) to account for possible heterogeneity among firms and find out the joint significance of the effects. To run a pooled model, we have to first ensure that the different firms could be pooled (the overview of the test for pooling is Online Appendix C3, and the complete details are in Baltagi 2008, pp. 57–63). After ensuring that the model can be pooled across firms, we specify the relationship among the metrics of UGC and stock market performance as

$$Y_{i,t} = \sum_{n=1}^{p} \Gamma_n Y_{i,t-n} + \Phi X_{i,t} + V_{it} + e_{it}, \tag{5}$$

where $i = \{1, 2, \ldots, 15\}$ is the firm index, $t \in \{T_0, T_1, T_2, \ldots, T\}$ is the time period index, and v accounts for firm-specific effects. The rest of the definitions of variables remain the same as in Equation (3). We further assume that the errors are orthogonal to the lagged values of the endogenous variables as well as the control variables and that $E(e_{it} \mid Y_{it-1}, \ldots, Y_{it}) = 0$. As in the model in Equation (3), the variables enter the panel VAR model in first differences because of the stationarity properties of the time series (details are in Online Appendix C2). Note that the time dimension is much larger than the number of cross sections. Hence, because of the large

Table 8 Influence of Other Marketing Variables on UGC Metrics

	Advertising	New product announcements	Media citations
Δ*Chatter*	**0.09**	0.03	0.04
Δ*Ratings*	0.01	0.01	0.02
Δ*Negative chatter*	**−0.06**	0.04	0.03
Δ*Positive chatter*	0.02	0.02	0.01
Δ*Competitor chatter*	0.01	0.03	0.07
Δ*Competitor negative chatter*	0.01	0.05	0.05

Notes. Estimates of elasticities of the marketing mix variables from the VAR model. Significant values are represented in bold.

time-series asymptotic properties, we could circumvent the estimation problems typical in analysis of dynamic panels (e.g., Holtz-Eakin et al. 1988, Pesaran and Shin 1995). We estimate the model simultaneously in Equation (5) using feasible generalized least squares (FGLS) with the Newey–West standard error (Newey and West 1997), controlling for any temporal heteroskedasticity and autocorrelation in the errors. As in the unpooled model, we estimate the impulse response functions and derive the short-term and cumulative (long-term) effects. We estimate the latter model up to five lags. The results of the pooled model are summarized in Table B7 (in Online Appendix B). The results of the pooled model are consistent with the original results. This confirms that our original findings are robust to market- or firm-specific heterogeneity and are not sensitive to the specification of the model in Equation (3). We also estimate the system using only the abnormal returns, the results of which are summarized in Online Appendix C4.

Alternative Metrics of UGC. We use alternative means of measuring the valence of the reviews to check the sensitivity of the model to the measurement of valence using the text classification algorithms. Instead of classifying the review as positive or negative (which was done earlier based on the naïve Bayesian and support vector machine algorithms), we now count the total number of positive or negative words and phrases across all the reviews of a given brand in a given day, which we label positive expressions or negative expressions, respectively. We use a dictionary-based frequency count of positive and negative terms for a given product. The effect of positive expressions is not statistically significant, whereas the negative expressions have a significant effect on the stock performance metrics (details in Online Appendix Tables B3 and B4). These results are consistent with the earlier analysis of the valence analyzed using the text classification method, and hence we could safely conclude that the results are not sensitive to alternate measures of the valence.

Portfolio Analysis. We carry out a portfolio analysis to check the robustness of the VAR analysis and ascertain investment opportunities in the results. Specifically, we try to address the following question: Do the asymmetric returns between negative and positive chatter hold when we try to use these metrics for investments? We use the calendar portfolio method (Jaffe 1974, Mandelker 1974) to answer this question. Calendar portfolio estimates the risk-adjusted returns across *all* the firms experiencing a similar event. In this approach, we form positive or negative portfolios by including all the firms experiencing positive or negative chatter, respectively, in a given day and track the returns to these portfolios. Specifically, we include a firm in the positive (or negative) portfolio if the volume of positive (or negative) chatter for the firm in a given day is greater than 50% of the overall volume of chatter. The composition of each of the portfolios changes daily as firms are added or deleted, depending on the valence of the chatter in a given day. We estimate the returns on the portfolio using the Fama–French three-factor model with Carhart momentum:

Table 9 **Results of the Calendar-Time Portfolio Analysis**

Average number of firms	α	*MKT*	*SMB*	*HML*	*MOM*	R^2
Positive portfolio 5	0.06	**0.48**	**0.31**	**0.43**	0.15	0.82
Negative portfolio 6	**−0.17**	2.36	**0.02**	−0.21	0.72	0.79

Notes. Estimated as the abnormal returns to portfolio from the four-factor model. Bold statistics imply statistical significance. α is the value of the abnormal returns for the portfolio, and *MKT*, *SMB*, *HML*, and *MOM* refer to the market, size, book-to-market, and momentum factors, respectively.

$$\begin{aligned} R_{i,t}-R_{f,t} &= \alpha_i+\beta_{i,MKT}(R_{MKT,t}-R_{f,t})+\beta_{i,SMB}SMB_t \\ &\quad +\beta_{i,HML}HML_t+\beta_{i,MOM}MOM_t+e_{i,t} \qquad (6)\\ e &\sim \mathrm{N}(0,\sigma^2). \end{aligned}$$

The estimates of α in Equation (6) give us the returns of the portfolio over and above that predicted by the four factors included in the model. The results (in Table 9) corroborate the asymmetry of the results reported earlier in the estimates of the VAR models and the impulse response function. The returns of the negative portfolio emerge as negative and are statistically significant. This result contrasts with that for the positive portfolio, which has an α that is not statistically significant. The overall gains using this strategy with an initial investment of a hundred million dollars yields an average annual profit of $7.9 million over the four years in our sample.

Discussion

UGC has been rapidly growing in the last few years. This study sought to ascertain whether UGC is related to stock market returns; if so, which metric of UGC is the most important; and what are the dynamics of this relationship. This section summarizes the main findings from the study, lists the contributions, discusses key issues, draws implications, and lists limitations.

Summary

The main findings of this study are the following:

• Volume of chatter has a significant positive lead effect on returns of a few days. Granger causality tests suggest that UGC predicts returns and trading volume, but not vice versa. The impact of UGC on returns prevails even after controlling for analysts' forecasts, media citations, advertising, and new product announcements.

• Of all metrics of UGC, volume of chatter has the strongest relationship with returns in the short term and long term (cumulatively). Numerical ratings do not have any significant impact on returns.

• Positive and negative metrics of UGC have an asymmetric impact on returns. Negative metrics have a stronger influence on returns than do positive metrics. In absolute terms, the erosion of value by negative UGC is greater than the accrual of value because of positive UGC. Whereas negative UGC has a significant negative effect on returns with a short wear-in and long wear-out, positive UGC has a nonsignificant effect on returns.

• The volume of chatter and negative chatter is found to positively influence the trading volume in both the short term and the long term.

• Competitor chatter is found to negatively influence a firm's stock returns while having a positive influence in the risk and trading volume.

• Off-line television advertising increases the volume of chatter while decreasing negative chatter.

• Idiosyncratic risk increases significantly with negative information in UGC. Idiosyncratic risk is not influenced by other metrics of UGC.

Contributions

This study makes several contributions. First, it is the first study to show that UGC is related to the stock market performance of the firm. Second, it is the first study to show which of four metrics—volume of chatter, ratings, positive valence, or negative valence—are important. Third, this study makes a methodological contribution by demonstrating techniques to aggregate and derive these dimensions from UGC. Fourth, it is the first study to examine the dynamics of the relationship between the various metrics of UGC and stock market performance (returns, risk, and trading volume) using VAR. Fifth, although studies suggest the adverse impact of negative word of mouth on returns (e.g., Luo 2009), no study has shown conclusively a comparison between positive and negative metrics.

Issues

The results raise the following three issues: Why does UGC predict returns? What do the dynamics of the relationship mean? Why is the effect asymmetric?

Why Does UGC Predict Returns? Because of the lack of perfect information, investors depend on site visits, reports from industry analysts, company press releases, experts' reviews in media, and regular sales and earnings announcements. The first four of these channels are sporadic and infrequent, whereas the latter two are presented in monthly or quarterly intervals. Thus, UGC represents new information, otherwise unavailable to investors, who may rely on this information. So, UGC may predict returns.

What Do the Dynamics Mean? We find that UGC leads returns by a few days. This leading effect could be attributed to a variety of economic and behavioral reasons. First, the extraction of information from UGC requires sophisticated methods, as demonstrated in this paper. Traders may not yet have easy recourse to these methods. Second, the gains from same-day trades may currently fall below the costs of processing the UGC immediately. Third, diffusion of knowledge about firms is known to be relatively slow, more so when it occurs through the distributed databases of consumer reviews in disparate websites (e.g., Lo and MacKinlay 1988). For these reasons, UGC may lead returns.

Why Is the Effect Asymmetric? We find that negative metrics are more strongly related to UGC than positive metrics. This could occur for three reasons. First, negative information is less diagnostic than positive information. Therefore, investors find the negative information more useful than positive information. Also, they may want to know the worst about the brand rather than the best. Second, investors may be loss averse. That is, the damage from negative information weighs more heavily than the gain from positive information. Third, positive information may be well known to investors from advertising and press releases. For these reasons, negative metrics may have a stronger impact on returns than positive metrics do.

Implications

This study has three implications for investing, strategy, and policy.

First, because UGC predicts returns, it contains valuable information about a firm's performance over and above that contained in standard sources of information. Thus, marketing managers should monitor UGC as a part of their marketing research.

Second, our results suggest that marketing managers should focus on negative chatter more than positive chatter or average ratings. Textual analysis of negative chatter could signal potential problems or discontent among consumers that deserve serious and immediate attention. Taking corrective action could avert any long-term damage to shareholder value. Our simulation indicates that unmitigated negative chatter can lead to an overall loss in brand value of $3.3 million over two weeks.

Third, because off-line television advertising increases the volume of chatter while decreasing negative chatter, firms can use it to favorably influence chatter. Such a use would require managing the content of advertising to appropriately influence the issues raised in negative chatter.

Fourth, this study highlights the importance of informational content of UGC to investors and other stakeholders. Portfolio analysis indicates that buying and selling stocks based on the valence of chatter can lead to an average gain of $27 million over one year, even for only 6 markets and 15 brands.

Limitations and Future Research

This study suffers from several limitations that could be the focus of future research. First, some of the steps in analyzing UGC are time consuming and computationally intensive. For implementation in managerial settings, practitioners would have to scale up and implement efficient computational procedures, especially in real-time monitoring of UGC. Second, to keep the research manageable, we had to restrict our sample to 1,112 days and 5 metrics over the 15 firms in 6 markets. It would be worthwhile to assess the generalizability of the results. We restrict our study to only two forms of UGC—product reviews and product ratings. We do this because of the high signal-to-noise ratio in ratings and reviews relative to these other sources of UGC. Assessing the impact of UGC on stock market performance at a daily level could have its limitations, such as nonsynchronous trading (Lo and MacKinlay 1988). That could help in assessing the impact on a finer level (e.g., at an hourly level or even a higher frequency). Third, we are unable to get sales data at the daily level. It would be insightful to get such data and analyze the direct path through which UGC affects stock market performance of firms.

Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://mktsci.journal.informs.org/.

Acknowledgments

This study benefited from a grant by the Marketing Science Institute and a generous gift from Don Murray to the USC Marshall Center for Global Innovation.

References

Avramov, D., T. Chordia, G. Jostova, A. Philipov. 2009. Credit ratings and the cross-section of stock returns. *J. Financial Markets* **12**(3) 469–499.

Baltagi, B. 2008. *Econometric Analysis of Panel Data*. John Wiley & Sons, Chichester, UK.

Baumeister, R. F., E. Bratslavsky, C. Finkenauer, K. D. Vohs. 2001. Bad is stronger than good. *Rev. General Psych.* **5**(4) 323–370.

Campbell, J. Y., S. J. Grossman, J. Wang. 1993. Trading volume and serial correlation in stock returns. *Quart. J. Econom.* **108**(4) 905–939.

Campbell, J. Y., A. W. Lo, A. C. McKinley. 1997. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ.

Carhart, M. M. 1997. On persistence in mutual fund performance. *J. Finance* **52**(1) 57–82.

Chen, Y., J. Xie. 2008. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Sci.* **54**(3) 477–491.

Chevalier, J. A., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* **43**(3) 345–354.

Chintagunta, P. K., S. Gopinath, S. Venkataraman. 2010. The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Sci.* **29**(5) 944–957.

Chordia, T., B. Swaminathan. 2000. Trading volume and cross-autocorrelations in stock returns. *J. Finance* **55**(2) 913–935.

comScore. 2008. comScore Media Metrix ranks top 50 U.S. Web properties for November 2008. Press release (December 16), comScore, Reston, VA. http://www.comscore.com/Press_Events/Press_Releases/2008/12/Top_50_US_Internet_Sites.

Das, S. R., M. Y. Chen. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the Web. *Management Sci.* **53**(9) 1375–1388.

Dekimpe, M. G., D. M. Hanssens. 1999. Sustained spending and persistent response: A new look at long-term marketing profitability. *J. Marketing Res.* **36**(4) 397–412.

Dekimpe, M., D. Hanssens. 2004. Persistence modeling for assessing marketing strategy performance. D. Lehmann, C. Moorman, eds. *Cool Tools in Marketing Strategy Research*, Chapter 4. Marketing Science Institute, Cambridge, MA.

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

Dhar, V., E. A. Chang. 2009. Does chatter matter? The impact of user-generated content on music sales. *J. Interactive Marketing* **23**(4) 300–307.

Domingos, P., M. Pazzani. 1997. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learn.* **29**(2) 103–130.

Duckworth, A., J. Golz, G. Trayner. 2009. Are analysts and investors engaging with new media? *Brunswick Rev.* (2) http://www.brunswickgroup.com/insights-analysis/brunswick-review/brunswick-review-issue-2/research/engaging-with-new-media.aspx.

Engle, R., V. Ng. 1993. Measuring and testing the impact of news on volatility. *J. Finance* **48**(5) 1749–1778.

Fama, E. F., K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *J. Financial Econom.* **33**(1) 3–56.

Forman, C., A. Ghose, B. Wiesenfeld. 2008. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Inform. Systems Res.* **19**(3) 291–313.

Fu, F. 2009. Idiosyncratic risk and the cross-section of expected stock returns. *J. Financial Econom.* **91**(1) 24–37.

Godes, D., D. Mayzlin. 2004. Using online conversations to study word-of-mouth communication. *Marketing Sci.* **23**(4) 545–560.

Granger, C. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**(3) 424–438.

Hamilton, J. D. 1994. *Time Series Analysis*, 1st ed. Princeton University Press, Princeton, NJ.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* **50**(4) 1029–1054.

Hayashi, F. 2000. *Econometrics.* Princeton University Press, Princeton, NJ.

Healy, P. M., K. G. Palepu. 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *J. Accounting Econom.* **31**(1–3) 405–440.

Holtz-Eakin, D., W. Newey, H. S. Rosen. 1988. Estimating vector autoregressions with panel data. *Econometrica* **56**(6) 1371–1395.

Hong, H., J. C. Stein. 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *J. Finance* **54**(6) 2143–2184.

Horvath, C., P. S. H. Leeflang, J. E. Wieringa, D. R. Wittink. 2005. Competitive reaction-and feedback effects based on VARX models of pooled store data. *Internat. J. Res. Marketing* **22**(4) 415–426.

Huang, W., Q. Liu, S. G. Rhee, L. Zhang. 2010. Return reversals, idiosyncratic risk, ad expected returns. *Rev. Financial Stud.* **23**(1) 147–168.

Jaffe, J. F. 1974. Special information and insider trading. *J. Bus.* **47**(3) 410–428.

Joachims, T. 1998. Text categorization with support vector machines: Learning with many relevant features. C. Nédellec, C. Rouveirol, eds. *Machine Learning: ECML-98*. Lecture Notes in Artificial Intelligence, Vol. 1398. Springer, Berlin, 137–142.

Johansen, S. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, New York.

Joshi, A., D. M. Hanssens. 2010. The direct and indirect effects of advertising spending on firm value. *J. Marketing* **74**(1) 20–33.

Koop, G., M. H. Pesaran, S. M. Potter. 1996. Impulse response analysis in nonlinear multivariate models. *J. Econometrics* **74**(1) 119–147.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, Y. Shin. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econometrics* **54**(1–3) 159–178.

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

Lo, A. W., A. C. MacKinlay. 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. *Rev. Financial Stud.* **1**(1) 41–66.

Luo, X. 2007. Consumer negative voice and firm-idiosyncratic stock returns. *J. Marketing* **71**(3) 75–88.

Luo, X. 2009. Quantifying the long-term impact of negative word of mouth on cash flows and stock prices. *Marketing Sci.* **28**(1) 148–165.

Lütkepohl, H. 2005. *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.

Mandelker, G. 1974. Risk and return: The case of merging firms. *J. Financial Econom.* **1**(4) 303–335.

Mitchell, M. L., J. H. Mulherin. 1994. The impact of public information on the stock market. *J. Finance* **49**(3) 923–950.

Moe, W., M. Trusov. 2011. Measuring the value of social dynamics in online product ratings forums. *J. Marketing Res.* **48**(3) 444–456.

Nelson, D. B. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59**(2) 347–370.

Newey, W. K., K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**(3) 703–708.

Nielsen Company. 2007. A. C. Nielsen global online consumer survey. Report, Nielsen Company, New York.

Pauwels, K., S. Srinivasan. 2004. Who benefits from store brand entry? *Marketing Sci.* **23**(3) 364–390.

Pauwels, K., J. Silva-Risso, S. Srinivasan, D. M. Hanssens. 2004. New products, sales promotions, and firm value: The case of the automobile industry. *J. Marketing* **68**(4) 142–156.

Pesaran, M. H., Y. Shin. 1998. Generalized impulse response analysis in linear multivariate models. *Econom. Lett.* **58**(1) 17–29.

Pesaran, M. H., R. P. Smith. 1995. Estimating long-run relationships from dynamic heterogeneous panels. *J. Econometrics* **68**(1) 79–113.

Rozin, P., E. B. Royzman. 2001. Negativity bias, negativity dominance, and contagion. *Personality Soc. Psych. Rev.* **5**(4) 296–320.

Sethuraman, R., G. J. Tellis, R. A. Briesch. 2011. How well does advertising work? Generalizations from a meta-analysis of brand advertising elasticity. *J. Marketing Res.* **48**(3) 457–471.

Sood, A., G. J. Tellis. 2009. Do innovations really pay off? Total stock market returns to innovation. *Marketing Sci.* **28**(3) 442–456.

Srinivasan, S., D. M. Hanssens. 2009. Marketing and firm value: Metrics, methods, findings, and future directions. *J. Marketing Res.* **46**(3) 293–312.

Srinivasan, S., M. Vanhuele, K. Pauwels. 2010. Mind-set metrics in market response models: An integrative approach. *J. Marketing Res.* **47**(4) 672–684.

Surowiecki, J. 2005. *The Wisdom of Crowds*. Anchor Books, New York.

Tellis, G. J. 2004. *Effective Advertising: Understanding How, When, and Why Advertising Works.* Sage, Thousand Oaks, CA.

Tellis, G. J., P. H. Franses. 2006. Optimal data interval for estimating advertising response. *Marketing Sci.* **25**(3) 217–229.

Tellis, G. J., J. Johnson. 2007. The value of quality. *Marketing Sci.* **26**(6) 758–773.

Tetlock, P. C. 2007. Giving content to investor sentiment: The role of media in the stock market. *J. Finance* **62**(3) 1139–1168.

Tetlock, P. C., M. Saar-Tsechansky, S. Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *J. Finance* **63**(3) 1437–1467.

Trusov, M., R. E. Bucklin, K. Pauwels. 2009. Effects of word-of-mouth versus traditional marketing: Findings from an Internet social networking site. *J. Marketing* **73**(5) 90–102.

Tuli, K. R., S. G. Bharadwaj. 2009. Customer satisfaction and stock returns risk. *J. Marketing* **73**(6) 184–197.

Tversky, A., D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* **211**(4481) 453–458.

Verna, P. 2009. User-generated content: More popular than profitable. Report, eMarketer, New York.

Wiles, M. A., S. P. Jain, S. Mishra, C. Lindsey. 2010. Stock market response to regulatory reports of deceptive advertising: The moderating effect of omission bias and firm reputation. *Marketing Sci.* **29**(5) 828–845.

Zhu, F., X. Zhang. 2010. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *J. Marketing* **74**(2) 133–148.