

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

Seshadri Tirunillai

Gerard J. Tellis

Seshadri Tirunillai is a doctoral candidate and **Gerard J. Tellis** is Professor of Marketing, Management and Organization, Director of the Center for Global Innovation, and Neely Chair in American Enterprise at Marshall School of Business, University of Southern California. This study benefitted from a grant by the Marketing Science Institute and a generous gift of Don Murray to the USC Marshall Center for Global Innovation. Address: HOH 814, Los Angeles, California, 90089-0443. USA. Tel: +1.213.740.5031, fax: +1.213.740.7828, E-mail: tellis@usc.edu.

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

Abstract

User-Generated Content (UGC) in online platforms provides a valuable source of consumer feedback on market performance of firms. This study examines whether UGC is related to stock market performance, which metric of UGC has the strongest relationship, and what the dynamics of the relationship are. The authors aggregate UGC (in the form of product reviews) from multiple websites over a four year period across six markets and fifteen firms. They derive multiple metrics of UGC and use multivariate time series models to assess the short and long term relationship between the metrics of UGC and stock market performance (abnormal returns, risk, and trading volume).

The findings reveal that UGC is moderately correlated with abnormal returns, risk, and trading volume. Of all the metrics of UGC, chatter (volume) shows the strongest relationship with abnormal returns and trading volume. The effect of negative and positive metrics of UGC on abnormal returns is asymmetric. Whereas negative UGC has a strong effect on abnormal returns and trading volume with a short “wearin” and long “wearout,” positive UGC has no effect on these metrics. Idiosyncratic risk increases with negative information in UGC. These results have important implications for managers and investors.

Keywords: User-Generated Content, stock returns, online word-of-mouth, Vector Auto-Regression (VAR), computational text processing.

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 referred to as User-Generated Content (UGC). UGC is one aspect of a broad phenomenon of inter-consumer communication called word-of-mouth. While 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 (*eMarketer*, January 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 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 greater 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, while Luo (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 2009; 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 6 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 bests relate to stock performance of the firm?
- What are the dynamics of the relationship in terms of wearin, wearout, 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 the 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.

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 which 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 Brunswick group of over 448 investment personnel (equity analyst or institutional investors) about 43% of them suggested that UGC has become an important determinant in

their investment decision in the recent years¹. 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, or 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, it also has the potential to act as a leading indicator of firm performance (for more details refer 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. Firstly, for traders to benefit from the information, they have to

¹ <http://www.brunswickgroup.com/insights-analysis/brunswick-review/brunswick-review-issue-2/research/engaging-with-new-media/page2.aspx>

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. Secondly, transaction costs for the investors may be high enough that they cannot make profitable trades by immediately acting on the information gained. Thirdly, these dynamics can also be attributed to the slow diffusion of information about the firms (e.g., Lo and MacKinlay 1990; Hong and Stein 1999). This pattern holds even more strongly in the case of UGC as the information has to diffuse systematically and evenly between the consumer markets and the 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 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 because losses loom larger than gains (Tversky and Kahneman 1981). Third, positive information about products, especially on new products, are usually well known and actively propagated by firms even before the 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 GE, P&G, or J&J, 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: (firm sampled are in parentheses) - personal computing (Hewlett Packard Company and Dell, Inc.), cellular phones (Motorola, Inc. and Nokia Corporation), personal digital assistants or smartphones (Research In Motion Limited and Palm, Inc.), footwear (Skechers USA , Inc., Timberland Company and Nike, Inc.), toys (Mattel, Inc., Hasbro, Inc., and Leap Frog, Inc.) and data storage (Seagate Technology, Western Digital Corporation and SanDisk, Inc.). 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) while 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 (blogs, product reviews) and videos (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 report 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 (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 ten visited sites (Yahoo!– 2, eBay– 8 and Amazon– 10) as ranked by

the unique visitors during data collection (Nielsen report 2007). This ranking is also consistent with the comScore ranking (Yahoo! sites – 2, eBay – 6 and Amazon – 7) (comScore Media Metrix, 2008).

UGC Data Collection

Since the UGC data is 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 pre-processing. Details are in the attached Online Appendix A.

The data collection for UGC varies for each site, as 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 level (individual review) 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): Naïve Bayesian classifier and Support Vector Machine classifier. *The details are in the attached 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 ten. We discard those reviews that have ten words or less due to lack of information content in these text. 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 one (bad) to five (good) scale). 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.

Chatter

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 the 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. Joachim 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 market wide 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 estimated conditional volatility (risk) this model.

We model the conditional variance of the error in this model as an EGARCH process (Exponential Generalized Autoregressive Conditional Heteroskedasticity, 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:

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

$$\varepsilon_{i,t} \sim N(0, \sigma_{it}^2) \quad (1)$$

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

where, t is the subscript for the time period, i the firm specific index, $R_{i,t}$ is the returns² of the stock i in time t, $R_{MKT,t}$ is the market portfolio return (CRSP value weighted index), $R_{f,t}$ is the risk-free rate of return (thirty 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 β 's are the factor loading 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

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

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:

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

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 risk³. 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.

We measure trading volume as the daily turnover which 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, Grossman, Wang 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 from the Center for Research in Security Prices (CRSP).

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.

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).

Advertising

A firm's advertising spending is the daily dollars spending 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 Lexis Nexis 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 sub-brands) in our sample. We search across major newspapers, dailies, and news wire services (including Wall Street Journal which is covered by Factiva) in each of these databases for each day within the time horizon of our study. Lexis Nexis 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 Lexis Nexis 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 1112 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 rationale for the model, tests for stationarity and units roots, test for Granger Causality, model specification, and modeling dynamics.

Rationale for VAR

We adopt the persistence modeling framework using the Vector Auto-Regression (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, as UGC is generated continuously over time and not a discrete event (Campbell, Lo and McKinley, 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 (GIRFs) which helps in and also assesses the relative contribution of the different metrics of UGC through the Generalized Forecast Error Variance Decomposition (GFEVD) (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, non-stationarity, serial correlation, reverse causality (Luo 2009) and 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 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 if the variables entering the system evolve continually or are stationarity. Following the literature (Dekimpe and Hanssens 2004), we used both Augmented Dickey-Fuller (ADF) test for the evolution of variables and 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 the Johansen's procedure to test for cointegration (Johansen 1995) to examine if 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):

$$Y_t = \sum_{n=1}^p \Gamma_n Y_{t-n} + \Phi X_t + e_t \quad 4$$

where $t \in \{T_0, T_1, T_2, \dots, T\}$ 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 (Schwartz's) Bayesian Information Criteria (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 co-integration tests. For example, the metrics of UGC and the trading volume enter the model in first differences as 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). Taking these factors into account, Equation (5) is specified as:

$$(5) \quad \begin{bmatrix} AbRet_t \\ IdioRisk_t \\ TrdVol_t \\ \Delta Chtr_t \\ \Delta Ratings_t \\ \Delta NegChtr_t \\ \Delta PosChtr_t \\ \Delta CompChtr_t \\ \Delta CompNegChtr_t \end{bmatrix} = \sum_{n=1}^p \begin{bmatrix} \gamma_{11}^n & \dots & \gamma_{19}^n \\ \vdots & \ddots & \vdots \\ \gamma_{91} & \dots & \gamma_{99} \end{bmatrix} \begin{bmatrix} AbRet_{t-n} \\ IdioRisk_{t-n} \\ TrdVol_{t-n} \\ \Delta Chtr_{t-n} \\ \Delta Ratings_{t-n} \\ \Delta NegChtr_{t-n} \\ \Delta PosChtr_{t-n} \\ \Delta CompChtr_{t-n} \\ \Delta CompNegChtr_{t-n} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \dots & \phi_{1p} \\ \vdots & \ddots & \vdots \\ \phi_{p1} & \dots & \phi_{pp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} + \begin{bmatrix} e_{AbRet,t} \\ e_{IdioRisk,t} \\ e_{TrdVol,t} \\ e_{Chtr,t} \\ e_{Ratings,t} \\ e_{NegChtr,t} \\ e_{PosChtr,t} \\ e_{CompChtr,t} \\ e_{CompNegChtr,t} \end{bmatrix}$$

where $AbRet$ is the abnormal returns, $IdioRisk$ the idiosyncratic risk, $TrdVol$ the trading volume (daily turnover). These three are the measures of stock performance. The UGC metrics in the model are the $Chtr$ the UGC 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) 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, Lo, and MacKinklay (1997, pp. 173-175) suggest that estimating models explaining the returns using firm (or marketing) characteristics by GMM are consistent and unbiased (pp. 173-175). Further, they suggest using Heteroskedasticity consistent standard errors to overcome any temporal heteroskedasticity. To this end, we use the Heteroskedasticity and Autocorrelation Consistent covariance estimator⁴ (also referred as the Newey-West standard error, 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 Chapter 14 (p 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 as 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).

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 sub-sections 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 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 & Unit Root and Cointegration

We use the Augmented Dickey-Fuller (ADF) test to determine if the variables are evolving or stationary. We also examine the variables with the Kwiatkowski-Phillips-Schmidt-Shin (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 ADF test is safely rejected (Table 3), suggesting that the returns are stable and stationary in the given time period. Alternately, the KPSS test also confirms 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 1988; Johansen and Juselius 1990) to see if 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). Since 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 model⁵. 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., chatter) does not Granger cause returns. We use five⁶ 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 while others do not. 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

⁵ More details on the steps of VAR framework can be found in Dekimpe and Hanssens 2004.

⁶We repeat these tests for lags up to twenty five days on each of the metrics. There was no significant temporal causation seen in these higher order lags.

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, (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 due to the inclusion of higher order lags.

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 (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 tests⁷. The optimal lag length for the VAR models is three as determined by the Schwartz's Bayesian Information Criterion.

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

⁷ 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. Since 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.

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 express the short and cumulative impact on stock returns in basis points (one basis point is one hundredth of a percentage).

Table 5a presents the results, averaged across the firms, of the short term and accumulated impact of the metrics of UGC on returns and risk⁸. Among the various metrics, chatter shows a high positive impact on returns in the short term (6.3 basis points $p < 0.01$) and similar accumulated impact of 14.8 basis points ($p < 0.01$). Chatter also has a significant positive influence on the trading volume with the short term and an accumulated impact of 0.08 percent and 0.10 percent of daily turnover respectively ($p < 0.01$). While 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.

Though these effects seem to be small in basis points, yet 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 fifteen days following the chatter.

⁸ The detailed results for each firm are in the online appendix B in Tables B3, B4, and B5 for stock returns, volatility and trading volume, respectively. The values presented here are the median values.

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 three) and negative (less than three). 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 overall chatter of the firm. A percentage increase in competitive chatter increases the overall chatter of the target firm by seven percent on an average (range 2%-13%). The last two rows of Table 5a 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 have an influence on the risk of the target firm (0.2 percent change in stock price in the short term and 1.0 percent cumulatively, $p < 0.01$). Similar result holds 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. But the negative chatter of competitors does not seem to influence the risk or the trading volume.

Duration of Impact

Figures 3 illustrate the results of the impulse response function for the Hewlett Packard in the personal computer market. The first panel on Figure 3 shows that a one unit shock in chatter has an increasing impact on returns, reaching a peak (wearin) in the second day. Returns then wear out over the next few days to long term equilibrium resulting in an accumulated value of around 14.8 basis points. The impulse response of returns to negative metrics of UGC contrasts sharply with that to chatter. The second

and third panel of Figure 3 depicts 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; while the negative chatter increases the risk associated with the stock which peaks in about 4 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, chatter, negative chatter of those of the firms and the 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 chatter to be the most important metric that influences the stock performance followed by negative chatter. While chatter explains 6.5% of variance of returns and 5.6% of the variance of trading volume; negative chatter explain 4.7% of variance of returns and 4.6% of variance of change in trading volume⁹. Competitor's UGC explains 0.83% with competitor's chatter contributing 0.45% and competitor's negative chatter contributing 0.38% of variance). Similar results holds 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 due to 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 (Figure 4). After an immediate influence on returns, the contribution of chatter and negative chatter to the variance of the

⁹ The total variation have been normalized and the remaining variation in returns, risk or trading volume is contributed by the historical values of the respective stock performance variable.

stock returns diminishes rapidly after four days. Similar results holds good for the influence of chatter on trading volume. Negative chatter shows a similar influence on returns over time though the percent of variance explained is not as much as that of chatter. But 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 one percent increase in advertising expenditure increases chatter by 9.5% and decreases negative chatter by 18.9% ($p = 0.05$), averaged across the firms. This effect is significant at 0.05 levels in all 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 0.12 in the literature (Sethuraman, Tellis and Briesch 2011). Though advertising has a strong influence on the 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 6.4%.

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 5b.

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 holds good even after these factors are accounted in our model.

Analysis Over Time

We repeat the analysis over rolling windows of time to investigate if the phenomenon changes over time. We use a rolling window of two years starting from June 2005 till December 2009, repeatedly re-estimating Equation 4. Table B7 (Online Appendix B) summarizes the results for the three UGC metrics, which were significant in the prior analysis. We find that on an average, the magnitude of the effect seems to be decreasing to a small extent over time for all the metrics (both in short and 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 chatter decreases over time from 6.85 basis points (in the period 2005-07) to 6.4 basis points, (in the period 2007-09). This amounts to an average decrease of 3.1% for the short term response and an accumulated decrease of 8.8% over time to the change in 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 as 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

Alternately, 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. Horvath 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 are Online Appendix C3 and the complete details are in Baltagi 2008 p. 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:

$$Y_{i,t} = \sum_{n=1}^p \Gamma_n Y_{i,t-n} + \Phi X_{i,t} + \nu_{it} + e_{it} \quad 6$$

where $i \in \{1, 2, \dots, 15\}$ is the firm index, $t \in \{T_0, T_1, T_2, \dots, T\}$ is the time period index and ν accounts for firm specific effects. The rest of definitions of variables remain same as in Equation 4. We further assume that the errors are orthogonal to the lagged values of the endogenous variables as well as the control variables and that $E(e_{it} | Y_{it-1}, \dots, Y_{iT}) = 0$. As in the model in Equation 4, the variables enter the Panel VAR model in first differences due to the stationarity properties of the time series (details are in Online Appendix C2). It is to be noted that the time dimension is much larger than the number of cross section. Hence due to the large time series asymptotic properties, we could circumvent the estimation problems typical in analysis of dynamic panels (e.g., Holtz-Eakin, Newey, and Rosen 1988; Pesaran and Shin 1995). We estimate the model simultaneously in Equation 6 using Feasible Generalized Least Square (FGLS) with Newey-West standard error (Newey and West 1997), controlling for any temporal heteroskedasticity and autocorrelation in the errors. As in the un-pooled model, we estimate the Impulse Response Functions and derive the short and cumulative (long term) effects. We estimate the latter model up to five lags. The results of the pooled model are summarized in Table B6 (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 4.

Alternative Metrics of UGC

We use alternate measures 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 the classifying the review as positive or negative (which was done earlier based on the Naïve Bayes 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 while the negative expressions have a significant effect on the stock performance metrics (details in online appendix table 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 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 3 Factor model with Carhart momentum.

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{i,MKT} (R_{MKT,t} - R_{f,t}) + \beta_{i,SMB} SMB_t + \beta_{i,HML} HML_t + \beta_{i,MOM} MOM_t + e_{i,t} \quad 7$$

$$e \sim N(0, \sigma^2)$$

The estimates of alpha in Equation 7 give us the returns of the portfolio over and above that predicted by the four factors included in the model. The results (in Table 8) 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 negative and are statistically significant. This result contrasts with that for the positive portfolio, which has an alpha 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 if 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:

- UGC and returns exhibit a significant positive correlation. Granger causality tests suggest that UGC predicts returns and trading volume. The impact of chatter on UGC prevails even after controlling for analysts' forecasts, media citations, advertising, and new product announcements.
- Of all metrics of UGC, 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 due to positive UGC. Whereas negative UGC has a strong effect with a short wearin and long wearout, positive UGC has a non-significant effect on returns.
- Offline 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.

- Granger causality tests suggest that neither returns nor risk have any significant effect on any of the metrics of UGC over time.

Contributions

This study makes several contributions. Firstly, 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—chatter, ratings, positive 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, though 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?

Due to 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 are sporadic and infrequent while the latter two are at 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 stock returns by a few days. This leading effect could be attributed to a variety of economic and behavioral reasons. Firstly, the extraction of information from UGC requires sophisticated methods as demonstrated in this paper. Traders may not yet have easy recourse to these

methods. Secondly, the gains from same day trades may currently fall below the costs of process the UGC immediately. Thirdly, diffusion about knowledge on firms is known to be relatively slow, more so when it occurs to through the distributed data bases of consumer reviews in disparate websites (e.g. Lo and MacKinlay 1990).

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 than positive information (typically 1 to 4). So, investors may find the negative information may useful. 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 weighs more heavily than the gain from positives. 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.

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 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 the two weeks. .

Third, because offline 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 appropriate 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. Firstly, 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 1112 days and five metrics over the fifteen firms in six 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 non-synchronous trading (Lo and MacKinlay 1990). That could help in assessing the impact on a finer level (e.g. at hourly level or even 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.

Figure 1: Time Series Plot of Selected Metrics of UGC

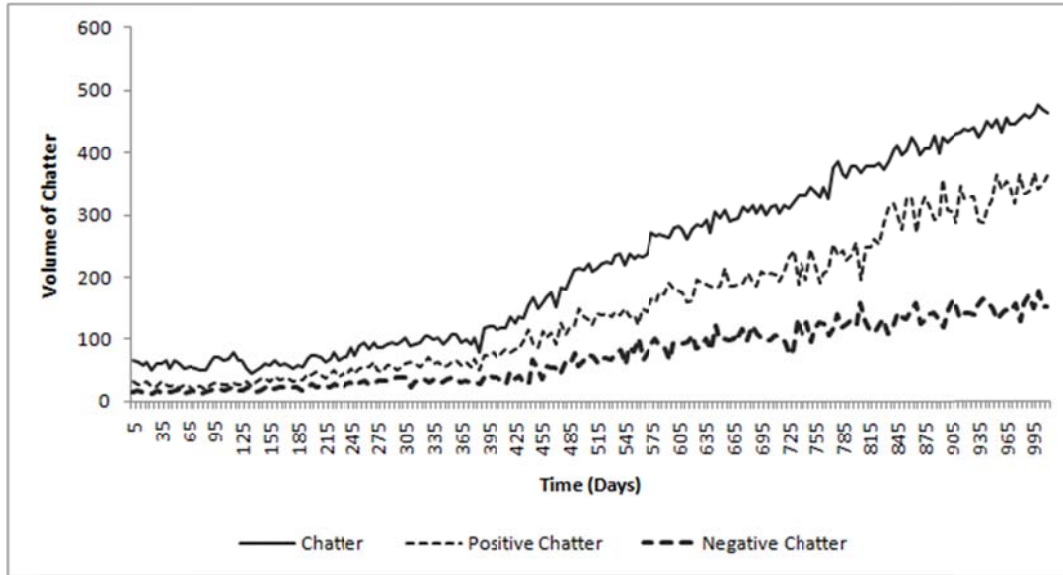


Figure 2: Time Series Plot of Chatter and Stock Prices of Selected Markets (Jan 2005 to Dec 2009)

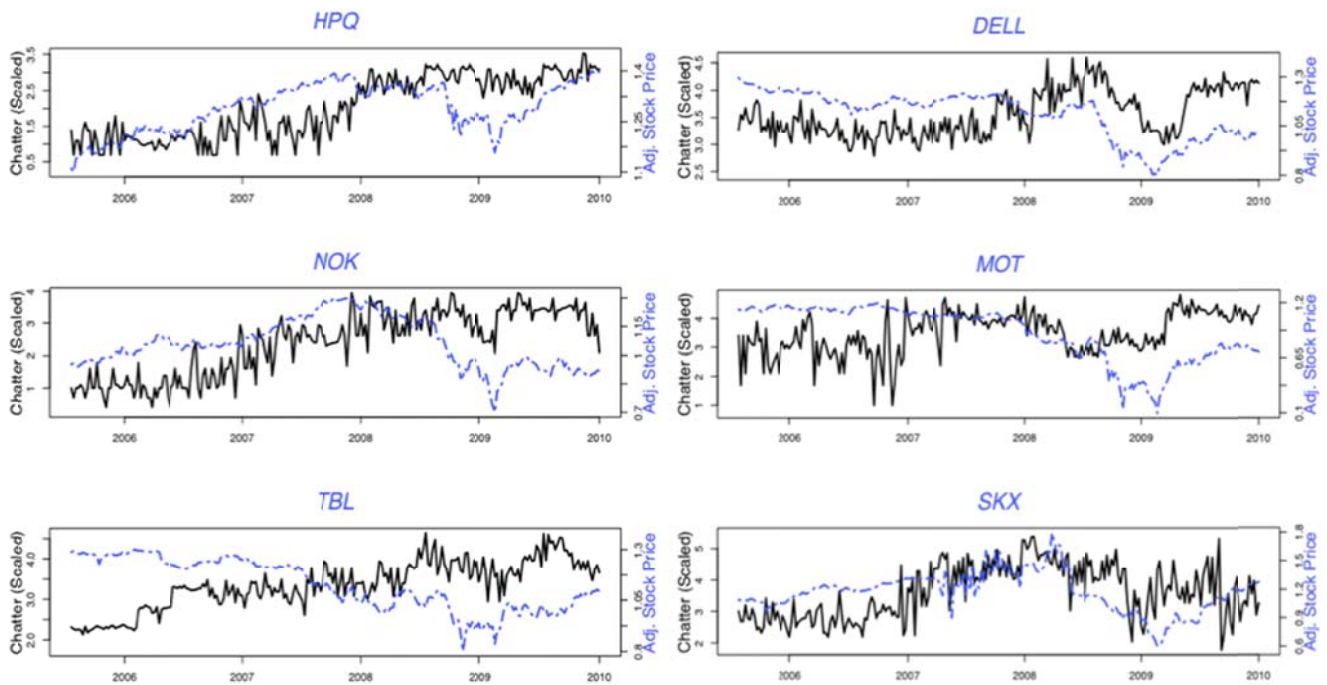
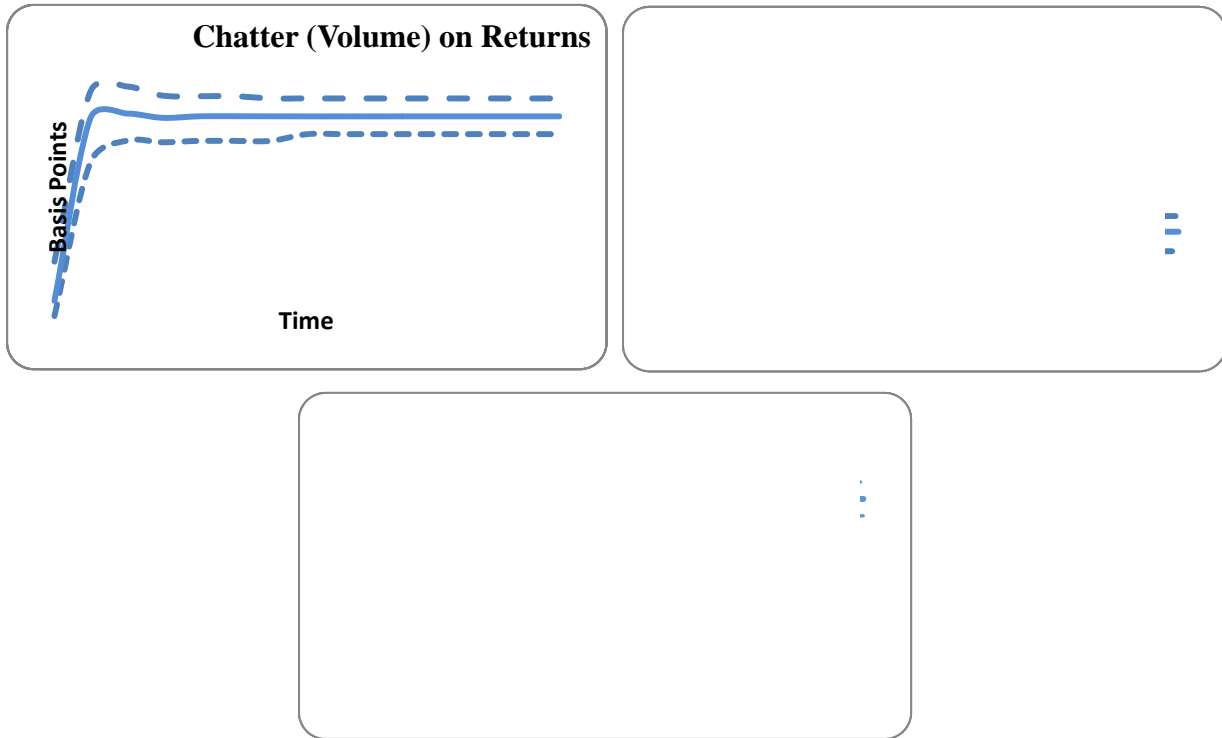
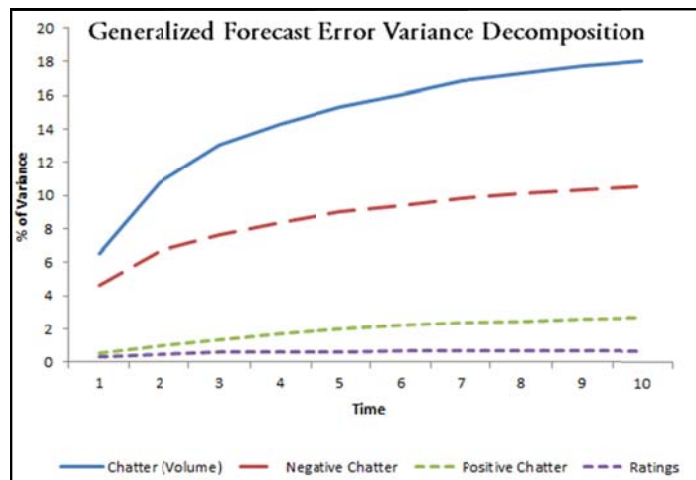


Figure 3: Accumulated Impulse Response Functions of Key UGC Metrics



The vertical axis shows the is the returns (in basis points) in panel 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).

**Figure 4: Explained Variance of Returns by Metrics of UGC
(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 hundred percent, Rest of the contributions to the variance is by the past returns, risk and trading volume, (not shown here to retain the readability of the graph).

Table 1: Classification of Relevant Literature on UGC

Article	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, 6 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 & Chang (2009)	Yes	Yes	No	No	No, only music
Chevalier & 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 & Mayzlin (2004)	Yes	Yes	No	No	No (only TV
Zhu & Zhang (2010)	Yes	Yes	No	No	No (only Video Games)

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
<i>Average</i>	75.33	21.79

The percentage of positive and negative might not add to 100 as 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.

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
ΔChatter	-4.79	0.23
ΔNegative chatter	-4.76	0.25
ΔPositive chatter	-4.60	0.20
ΔRatings	-4.63	0.26
Δ Competitor chatter	-4.58	0.20
Δ Competitor negative chatter	-4.62	0.25

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 be found in Table B1 and B2 (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

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 in Online Appendix B2.

Table 5a: Short-term and Accumulated Impact of UGC Metrics On Stock Performance (Estimates from the VAR model and the Impulse Response Functions)

		Returns	Risk	Δ Trading Volume
Δ Chatter	Immediate	6.31*	0.001	0.075*
	Accumulated	14.82*	0.021	0.110*
Δ Ratings	Immediate	1.81	0.002	0.045
	Accumulated	2.34	0.009	0.079
Δ Negative Chatter	Immediate	-4.21*	0.005*	0.025*
	Accumulated	-8.37*	0.011*	0.058*
Δ Positive Chatter	Immediate	3.21	0.002	0.038
	Accumulated	6.27	0.006	0.086
Δ Competitor Chatter	Immediate	-2.18*	0.002*	0.023*
	Accumulated	-5.13*	0.010*	0.069*
Δ Competitor Negative Chatter	Immediate	3.16*	0.003	0.051
	Accumulated	7.36*	0.008	0.076

Only the relevant endogenous variables are shown in this table. The values for returns are in Basis points (1 basis point = 1 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 B3, B4 and B5 (Online Appendix B)

Table 5b: Influence of other Marketing Variables on UGC Metrics
(Estimates of Elasticities of the Marketing Mix Variables from VAR model)

	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

Significant values are represented in bold.

Table 6: Duration of the Short and Accumulated Impact
(From the Generalized Impulse Response Function)

	Δ Chatter		Δ Negative Chatter		Δ Positive Chatter		Δ Competitor Chatter		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

All values are depicted in mean time (in days)

Table 7: Variance in Stock Performance Explained by the Metrics of UGC
(From Generalized Forecast Error Variance Decomposition)

	Return	Risk	Trading Volume
ΔChatter	6.54	3.65	5.64
ΔRatings	0.3	0.24	1.13
ΔNegative Chatter	4.65	4.56	2.53
ΔPositive Chatter	0.56	0.79	0.74
ΔCompetitor chatter	0.45	0.27	0.35
Δ Competitor Negative chatter	0.38	0.64	0.21

All figures are in percentages. The table depicts the normalized percentage contributions by metrics of UGC for the first period only. The other contribution by the variables own past values or by other stock performance variables and is not shown in the table.

Table 8: Results of the Calendar Time Portfolio Analysis
(Estimated as the abnormal returns to portfolio from the Four Factor Model)

	Average Number of firms	Alpha	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

Bold statistics implies statistical significance. Alpha 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.

References

- AC, N. (2007). A.C. Nielsen Report on Online Global Consumer Study.
- Avramov, D., Chordia, T., Jostova, G., & Philipov, A. (2009). Credit ratings and the cross-section of stock returns. *Journal of Financial Markets* , 12 (3), 469-499.
- Baltagi, B. (2008). *Econometrics of panel data*. Chichester: John Wiley & Sons.
- Baumeister, R., Bratslavsky, E., Finkenauer, C., & Vohs, K. (2001). Bad is stronger than good. *Review of general psychology* , 5 (4), 323-370.
- Campbell, J., Grossman, S., & Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics* , 108 (4), 905-939.
- Campbell, J., Lo, A., & McKinley, A. (1997). *The Econometrics of Financial Markets*. Princeton University Press, Princeton.
- Carhart, M. (1997). On persistence in mutual fund performance. *The Journal of Finance* , 52, 57-82.
- Chen, Y., & Xie, J. (2008). Online Consumer Review: Word-of-mouth as a New Element of Marketing Communication Mix. *Management Science* .
- Chevalier, J., & Mayzlin, D. (2006). The effect of word of mouth online: Online book reviews. *Journal of Marketing Research* , 43, 345-354.
- Chintagunta, P., Gopinath, S., & Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science* .
- Chordia, T., & Swaminathan, B. (2000). Trading volume and cross-autocorrelations in stock returns. *Journal of Finance* , 55 (2), 913-935.
- Chow, G. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica* , 28 (3), 591-605.
- comScore. (2008). comScore Media Metrix ranking report.
- Das, S., & Chen, M. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science* , 53, 1375.
- Dekimpe, M., & Hanssens, D. (2004). Persistence modeling for assessing marketing strategy performance. *Cool Tools in Marketing Strategy Research*, Donald Lehmann and Christine Moorman eds.
- Dekimpe, M., & Hanssens, D. (1999). Sustained spending and persistent response: A new look at long-term marketing profitability. *Journal of Marketing Research* , 397-412.
- Dellarocas, C., Zhang, X., & Awad, N. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing* , 21.
- Dhar, V., & Chang, E. A. (2009). Does Chatter Matter? The Impact of User-Generated Content on Music Sales. *Journal of Interactive Marketing* , 23, 300-307.
- Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine learning* , 29 (2), 103-130.
- eMarketer. (2009). User-Generated Content: More Popular than Profitable. January.
- Engle, R., & Ng, V. (1993). Measuring and testing the impact of news on volatility. *Journal of finance* , 48 (5), 1749-1778.
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research* , 19 (3), 291-313.
- Fu, F. (2009). Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics* , 91 (1), 24-37.
- Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society* , 37 (3), 424-438.
- Hamilton, J. D. (1994). *Time Series Analysis* (1 ed.). Princeton University Press.
- Hayashi, F. (2000). *Econometrics*. Princeton University Press.
- Healy, P., & Palepu, K. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics* , 31, 405-440.
- Holtz-Eakin, D., Newey, W., & Rosen, H. (1988). Estimating vector autoregressions with panel data. *Econometrica: Journal of the Econometric Society* , 56 (6), 1371-1395.
- Hong, H., & Stein, J. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance* , 54, 2143-2184.

- Horvath, C., Leeflang, P., Wieringa, J., & Wittink, D. (2005). Competitive reaction-and feedback effects based on VARX models of pooled store data. *International Journal of Research in Marketing* , 22 (4), 415-426.
- Im, K., Pesaran, M., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels* 1. *Journal of econometrics* , 115 (1), 53-74.
- Jaffe, J. (1974). Special information and insider trading. *The Journal of Business* , 47, 410-428.
- Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Machine Learning: ECML-98* , 137-142.
- Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press, USA.
- Joshi, A., & Hanssens, D. (2010). The Direct and Indirect Effects of Advertising Spending on Firm Value. *Journal of Marketing* , 74 (1), 20-33.
- Koop, G., Pesaran, M., & Potter, S. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* , 74, 119-147.
- Kwiatkowski, D., Phillips, P., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of econometrics* , 54, 159-178.
- Levin, A., Lin, C., & Chu, C. (2002). Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties. *Journal of Econometrics* , 108 (1), 1-24.
- Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing* , 70 (3), 74-89.
- Lo, A., & Andrew, W. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics* , 45 (1-2), 181-211.
- Lo, A., & MacKinlay, A. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies* , 1, 41-66.
- Luo, X. (2007). Consumer Negative Voice and Firm-Idiosyncratic Stock Returns. *Journal of Marketing* , 71, 75-88.
- Luo, X. (2009). Quantifying the Long-Term Impact of Negative Word of Mouth on Cash Flows and Stock Prices. *Marketing Science* .
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.
- Mandelker, G. (1974). Risk and return: the case of merging firms. *Journal of Financial Economics* , 1, 303-335.
- Mitchell, M. L., & Mulherin, J. H. (1994). The Impact of Public Information on the Stock Market. *The Journal of Finance* , 49, 923-950.
- Moe, W., & Trusov, M. (2011). Measuring the Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research* .
- Nelson, D. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica: Journal of the Econometric Society* , 59 (2), 347-370.
- Newey, W., & West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society* , 703-708.
- Pauwels, K., Silva-Risso, J., Srinivasan, S., & Hanssens, D. (2004). New products, sales promotions, and firm value: The case of the automobile industry. *Journal of Marketing* , 68, 142-156.
- Pesaran, H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters* , 58, 17-29.
- Pesaran, M., & Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels* 1. *Journal of econometrics* , 68 (1), 79-113.
- Rozin, P., & Royzman, E. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review* , 5 (4), 296.
- Sethuraman, R., Tellis, G. J., & Briesch, R. (2011). How Well Does Advertising Work? Generalizations from a Meta-Analysis of Brand Advertising Elasticity. *Journal of Marketing Research* .
- Sood, A., & Tellis, G. (2009). Do Innovations Really Pay Off? Total Stock Market Returns to Innovation. *Marketing Science* , 28 (3), 442-456.
- Srinivasan, S., & Hanssens, D. (2009). Marketing and Firm Value. *Journal of Marketing Research* .
- Srinivasan, S., Vanhuele, M., & Pauwels, K. (2010). Mind-Set Metrics in Market Response Models: An Integrative Approach. *Journal of Marketing Research* , 47 (4), 672-684.
- Surowiecki, J. (2005). *The wisdom of crowds: why the many are smarter than the few*. Abacus.
- Tellis, G., & Franses, P. (2006). Optimal data interval for estimating advertising response. *Marketing Science* , 25 (3), 217.
- Tellis, G., & Johnson, J. (2007). The Value of Quality. *Marketing Science* , 26, 758.

- Tetlock, P. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance* , 62, 1139-1168.
- Tetlock, P., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals, Forthcoming. *Journal of Finance* .
- Trusov, M., Bucklin, R., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. *Journal of Marketing* , 73 (5), 90-102.
- Tuli, K., & Bharadwaj, S. (2009). Customer satisfaction and stock returns risk. *Journal of marketing* , 73 (6), 184-193.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* , 211, 453-458.
- Wiles, M., Jain, S., Mishra, S., & Lindsey, C. (2010). Stock Market Response to Regulatory Reports of Deceptive Advertising: The Moderating Effect of Omission Bias and Firm Reputation. *Marketing Science* .
- Zhu, F., & Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing* , 74 (2), 133-148.

Online Appendix A: Algorithms Used in Text Analysis

We collect the data from the various websites as outlined in the data collection section. The table below outlines the steps in data collection.

Table A1: Steps in Data Collection

Step of Text processing	Procedure
Collection	Get data either by crawling or using the API as needed
Conversion and storage in machine readable format	Unify data from multiple sites and categories using UPC or model codes. Convert and store data in machine readable Extensible Markup Language (XML) format
Extraction	Parsing relevant fields to extract data, both numerical and textual
Preprocessing	Clean text for “noise” (e.g. unwanted HTML tags) and ensure data is analyzable
Analysis	Apply text classification algorithm to extract relevant measures (valence)

In this section, we focus on the analysis step (extraction of valence) and explain it in detail. For deriving the valence of the textual manifest we used two text algorithms – Bayesian classifier, a widely used text classifier, and Support Vector Machine, popular for its accuracy and reliability in supervised classification. First, the individual reviews are elaborately pre-processed sequentially to derive the corpus to run the classifiers. Data pre-processing used steps very similar to the ones adopted in the prior studies (e.g. Lee and Bradlow 2009; Feldman, Goldenberg and Netzer 2009). Unwanted words (such as HTML tags, URLs and telephone numbers) and stop words (e.g. I, and, or, etc.) are removed. The documents are then tokenized (converted to individual words or phrases) and stemmed (i.e. convert to the base form, for example like, liked, liking, etc. to “like”). This became the corpus for implementing the algorithms.

For example a review that is originally,

“This is a fast, powerful laptop with zippy wireless Internet performance. Acceptable battery life. Bluetooth wireless mouse is great accessory to add. Only nit is slightly washed out (low contrast) LCD monitor performance on web pages. Found notes on web on how to manually adjust video driver settings to optimize for crisper whites. But for the price, we are very happy with this purchase.”

After the processing of the text would become

“Fast powerful laptop zippy wireless Internet performance. Acceptable battery life. Bluetooth wireless mouse great accessory. Slightly washed out low contrast LCD monitor performance web. Found notes web manually adjust video driver settings optimize crisper whites. But price very happy purchase”

The pre-classified training vocabulary is assembled from multiple sources. We manually build a training vocabulary by labeling the words as positive or negative. The training data is market specific. We also used lexical sources such as the WordNet (Fellbaum 1998) and Roget’s thesaurus to augment the training data.

Naïve Bayesian Algorithm: This algorithm has been very popular in text classification applications ranging from spam filters in emails to large scale web-page classification. Our approach here is similar in spirit to some of the earlier studies using the algorithm (e.g. Das and Chen 2006). The main aim is to classify the documents (here reviews) into the valence categories (positive or negative) based on the input training set of the lexical words. We used the binary naïve Bayes model where the probability that the review (r) belongs to valence (j) is $P(j|r) \propto P(j) \prod_{i \leq k \leq n_r} P(t_k | j)$. The best fit for valence is given by the

likelihood $j = \arg \max_{j \in \{p,n\}} P(c) \prod_{i \leq k \leq n_r} P(t_k | j)$. The conditional probability $P(t_k | j)$ is estimated as the relative frequency of the term t in documents belonging to class “positive” (or “negative”).

$$P(t_k | j) = \frac{T_{jt}}{\sum_{i \in V} T_{ci}} \quad A1$$

(with Laplace smoothing applied to eliminate the zeros as $P(t_k | j) \neq 0$ (Manning, Raghavan and Schütze 2007 or Das and Chen 2006 for more details)).

Support Vector Machine Classification: Support vector machine, a semi-parametric classification technique, has been widely used in classification problems where the predictive validity is very important (Cui and Curry 2005). It has been shown to have a high reliability in text classification (Joachims 2002). We adopt this technique to classify the text based on their valence. The implementation is similar to the

one adopted previously - we use the same lexical resources and training sample in the Naïve Bayesian classification. Here, we adopt the text classification approach using this algorithm as follows. We have the training data of the form $(x_1, y_1), \dots, (x_n, y_n)$ where $x_i \in R^n$ and $y_i \in \{-1, +1\}$ where the negative words are classified as “-1” and positive ones as “+1”. The optimization function is given as

$$\text{Max } f(x) = \sum_{i=1}^n y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \quad \text{A4}$$

Where, K is the kernel function, b is the threshold (intercept) and α_i are weights. We maximize this subject to the constraint:

$$\begin{aligned} 0 \leq \alpha_i \leq C \text{ and} \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \quad \text{A5}$$

where C is the “misclassification cost” (“slack” variable). (For more details on this please refer Manning, Raghavan and Schutze 2007 or Joachims 2002). The implementation of the algorithm is done on a Linux machine using the libsvm package ¹⁰.

Calculation of the Accuracy of the Algorithms

We evaluate the quality of the two classification algorithms using the precision and sensitivity (recall) *Precision* is measured as the fraction of classified documents (true and false positives) that are correctly classified (true positives). While sensitivity (or recall) is the ratio of the number of documents correctly classified (true positives) divided by the total number of results that should have been returned (i.e. the sum of true positive and false negatives).

$$\text{Precision} = \text{true positives} / (\text{true positives} + \text{false positives})$$

$$\text{Sensitivity} = \text{true positives} / (\text{true positives} + \text{false negatives})$$

¹⁰ Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

The table below gives the average precision, recall and accuracy statistics for the markets in our sample.

	Average Precision	Average Sensitivity	Average Precision	Average Sensitivity	Inter-Algorithm Agreement
Personal Computer	75.31	81.29	81.28	72.65	97.25
Cell Phone	70.5	65.21	75.6	62.4	93.57
PDA/Smartphone	85.35	62.14	89.5	71.5	95.17
Data Storage	72.5	51.68	73.1	65.4	94.58
Toys	68.6	71.25	71.8	65.2	93.47
Footwear	82.64	64.25	77.6	68.4	92.37

References for Appendix A

Cui, D., & Curry, D. (2005). Prediction in marketing using the support vector machine. *Marketing Science* , 24, 595.

Das, S., & Chen, M. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science* , 53, 1375.

Feldman, Goldenberg, & Netzer. (2009). Mine Your Own Business - Market Structure Surveillance Through Text Mining. *Working Paper* .

Fellbaum, C., & others. (1998). *WordNet: An electronic lexical database*. MIT press Cambridge, MA.

Joachims, T. Learning to classify text using support vector machines: Methods, theory, and algorithms. *Computational Linguistics* , 29.

Lee, T., & Bradlow, E. (2006). *Automatic Construction of Conjoint Attributes and Levels from Online Customer Reviews*. Working Paper, Wharton School.

Manning, C., Raghavan, P., & Schütze, H. (2007). An Introduction to Information Retrieval. *An Introduction to Information Retrieval* . Cambridge University Press, Cambridge, England.

Online Appendix B

B1: Detailed Augmented Dickey Fuller Test for Unit Root

	Δ Stock Returns	Δ Chatter	Δ Negative Chatter	Δ Positive Chatter	Δ Competitor Chatter	Δ Negative Expressions	Δ Competitor Negative Chatter	Δ Ratings
HP (HPQ)	-6.25	-4.27	-4.83	-4.21	-4.38	-4.73	-3.97	-5.32
Dell (DELL)	-7.56	-5.89	-4.58	-4.29	-5.28	-4.08	-4.28	-4.28
Nokia (NOK)	-5.36	-4.28	-4.28	-3.10	-3.95	-5.47	-5.21	-4.58
Motorola (MOT)	-5.27	-3.99	-4.97	-4.27	-5.61	-4.97	-4.28	-4.29
Research in Motion (RIMM)	-5.68	-4.89	-4.92	-3.97	-4.38	-4.28	-5.61	-4.83
Palm (PALM)	-4.52	-5.24	-3.87	-6.23	-4.91	-5.24	-5.34	-4.68
Seagate (STX)	-3.57	-3.84	-3.97	-5.26	-4.28	-4.68	-4.95	-4.27
Western Digital (WDC)	-3.98	-4.28	-4.76	-4.95	-4.71	-5.61	-4.28	-4.61
Hasbro (HAS)	-4.01	-4.81	-3.94	-5.36	-4.79	-5.12	-3.97	-4.83
Mattel (MAT)	-4.85	-4.29	-4.38	-5.67	-5.04	-3.73	-3.75	-4.13
Skechers (SKX)	-5.27	-3.72	-4.28	-5.94	-4.58	-3.80	-5.04	-4.08
Timberland (TBL)	-6.27	-4.26	-4.21	-4.23	-5.17	-3.47	-4.92	-5.06

Note: The critical value at 5% level for the test is -3.44

B2: Detailed KPSS Test for stationarity

	Δ Stock Returns	Δ Chatter	Δ Negative Chatter	Δ Positive Chatter	Δ Competitor Chatter	Δ Negative Expressions	Δ Competitor Negative Chatter	Δ Ratings
HP (HPQ)	0.033	0.245	0.257	0.158	0.278	0.177	0.333	0.274
Dell (DELL)	0.128	0.310	0.240	0.220	0.158	0.270	0.258	0.350
Nokia (NOK)	0.058	0.254	0.124	0.187	0.166	0.197	0.063	0.125
Motorola (MOT)	0.089	0.025	0.012	0.248	0.147	0.257	0.306	0.324
Research in Motion (RIMM)	0.241	0.086	0.291	0.159	0.260	0.084	0.203	0.247
Palm (PALM)	0.087	0.235	0.280	0.157	0.268	0.086	0.348	0.214
Seagate (STX)	0.057	0.356	0.240	0.179	0.214	0.215	0.247	0.245
Western Digital (WDC)	0.088	0.234	0.097	0.230	0.188	0.223	0.215	0.284
Hasbro (HAS)	0.085	0.248	0.325	0.241	0.087	0.234	0.228	0.189
Mattel (MAT)	0.068	0.257	0.312	0.312	0.189	0.219	0.250	0.246
Skechers (SKX)	0.178	0.249	0.278	0.344	0.257	0.298	0.356	0.279
Timberland (TBL)	0.157	0.201	0.132	0.158	0.245	0.113	0.308	0.289

Note: All the metrics of UGC are measured in first difference. The critical value at 5% is 0.463.

Table B2: Granger Causality Tests By Firm Between UGC Metrics and Stock market Performance

	Returns				Idiosyncratic Risk			
	Chatter	Negative Chatter	Positive Chatter	Ratings	Chatter	Negative Chatter	Positive Chatter	Ratings
HP (HPQ)	0.02*	0.05*	0.09	0.14	0.11	0.05*	0.24	0.12
Dell (DELL)	0.03*	0.04*	0.11	0.23	0.09	0.06	0.13	0.09
Nokia (NOK)	0.05*	0.05*	0.08	0.17	0.14	0.08	0.35	0.11
Motorola (MOT)	0.03*	0.04*	0.12	0.14	0.09	0.03*	0.2	0.13
Research in Motion (RIMM)	0.04*	0.03*	0.08	0.1	0.18	0.05*	0.14	0.28
Palm (PALM)	0.02*	0.05*	0.17	0.09	0.07	0.02*	0.16	0.23
Seagate (STX)	0.04*	0.04*	0.08	0.14	0.09	0.04*	0.21	0.14
Western Digital (WDC)	0.03*	0.03*	0.25	0.19	0.07	0.06	0.22	0.15
Hasbro (HAS)	0.05*	0.06	0.13	0.22	0.15	0.03*	0.15	0.12
Mattel (MAT)	0.07	0.05*	0.19	0.13	0.13	0.02*	0.08	0.07
Skechers (SKX)	0.05*	0.06	0.07	0.14	0.09	0.07	0.4	0.06
Timberland (TBL)	0.02*	0.07	0.21	0.2	0.14	0.06	0.23	0.14
Leap Frog (LF)	0.04*	0.05*	0.05	0.09	0.11	0.04*	0.15	0.18
SanDisk (SNDK)	0.04*	0.04*	0.13	0.14	0.12	0.04*	0.24	0.2
Nike (NKE)	0.03*	0.03*	0.14	0.16	0.08	0.06	0.27	0.18

* The values shown here are the p values of the chi square test.

B3: Short and Accumulated Impact of Metrics of UGC on Returns By Firm

	Chatter		Negative Chatter		Negative Expressions	
	Short Term	Long Term	Short Term	Long Term	Short Term	Long Term
HP (HPQ)	6.32	14.81	-4.21	-8.63	-5.84	-4.10
Dell (DELL)	4.88	12.29	-3.71	-5.05	-1.66	-7.65
Nokia (NOK)	9.18	16.52	-5.39	-9.46	-1.28	-4.72
Motorola (MOT)	7.91	18.81	-2.48	-5.94	-2.45	-8.50
Research in Motion (RIMM)	5.48	16.51	-3.54	-5.58	-3.25	-8.76
Palm (PALM)	3.93	14.06	-4.82	-6.29	-1.24	-6.90
Seagate (STX)	4.58	11.32	-3.89	-5.20	-4.80	-7.69
Western Digital (WDC)	3.91	19.58	-4.63	-9.62	-1.31	-7.86
Hasbro (HAS)	6.98	11.67	-4.23	-8.42	-4.83	-5.84
Mattel (MAT)	3.03	18.98	-5.47	-9.11	-5.05	-6.92
Skechers (SKX)	9.27	15.31	-4.73	-9.28	-5.14	-7.13
Timberland (TBL)	6.94	12.80	-3.20	-6.61	-3.11	-6.82
Leap Frog (LF)	9.95	10.51	-2.27	-7.78	-2.14	-7.58
SanDisk (SNDK)	8.52	13.75	-4.77	-9.08	-3.81	-5.29
Nike (NKE)	2.62	16.98	-4.60	-9.35	-2.23	-7.96

Note: All values are in Basis points (1 basis point = 1 hundredth of a percentage)

Table B4: Short and Accumulated Impact of Metrics of UGC on Risk By Firm

	Negative Chatter		Competitor Chatter		Negative Expressions	
	Short Term	Long Term	Short Term	Long Term	Short Term	Long Term
HP (HPQ)	0.0055	0.0117	0.0033	0.0082	0.0043	0.0119
Dell (DELL)	0.0052	0.0090	0.0030	0.0068	0.0038	0.0108
Nokia (NOK)	0.0036	0.0089	0.0024	0.0120	0.0032	0.0127
Motorola (MOT)	0.0042	0.0073	0.0048	0.0084	0.0033	0.0139
Research in Motion (RIMM)	0.0057	0.0126	0.0032	0.0093	0.0044	0.0100
Palm (PALM)	0.0036	0.0071	0.0048	0.0069	0.0041	0.0140
Seagate (STX)	0.0051	0.0125	0.0040	0.0119	0.0043	0.0100
Western Digital (WDC)	0.0056	0.0107	0.0052	0.0144	0.0034	0.0143
Hasbro (HAS)	0.0040	0.0121	0.0029	0.0105	0.0042	0.0130
Mattel (MAT)	0.0042	0.0112	0.0027	0.0094	0.0030	0.0119
Skechers (SKX)	0.0048	0.0107	0.0018	0.0110	0.0040	0.0117
Timberland (TBL)	0.0048	0.0121	0.0028	0.0069	0.0040	0.0082
Leap Frog (LF)	0.0037	0.0133	0.0031	0.0102	0.0029	0.0080
SanDisk (SNDK)	0.0047	0.0094	0.0062	0.0094	0.0045	0.0118
Nike (NKE)	0.0052	0.0089	0.0060	0.0110	0.0031	0.0136

**Table B5: Short and Accumulated Impact of Metrics of UGC on Trading Volume
By Firm**

	Chatter		Negative Chatter		Competitor Chatter	
	Short Term	Long Term	Short Term	Long Term	Short Term	Long Term
HP (HPQ)	0.088	0.128	0.024	0.038	0.027	0.64
Dell (DELL)	0.075	0.138	0.025	0.065	0.035	0.085
Nokia (NOK)	0.072	0.104	0.031	0.061	0.017	0.047
Motorola (MOT)	0.081	0.109	0.015	0.056	0.037	0.038
Research in Motion (RIMM)	0.051	0.093	0.041	0.061	0.026	0.064
Palm (PALM)	0.076	0.133	0.023	0.045	0.032	0.041
Seagate (STX)	0.086	0.111	0.051	0.068	0.033	0.072
Western Digital (WDC)	0.076	0.101	0.026	0.045	0.037	0.058
Hasbro (HAS)	0.04	0.115	0.054	0.069	0.015	0.079
Mattel (MAT)	0.068	0.152	0.051	0.063	0.031	0.082
Skechers (SKX)	0.082	0.083	0.039	0.056	0.009	0.085
Timberland (TBL)	0.054	0.097	0.025	0.034	0.032	0.059
Leap Frog (LF)	0.056	0.109	0.038	0.049	0.011	0.053
SanDisk (SNDK)	0.081	0.117	0.048	0.061	0.028	0.037
Nike (NKE)	0.084	0.154	0.015	0.042	0.009	0.083

Table B6: Results of pooled -VAR model

		Returns	Risk	Trading Volume
Δ Chatter	Immediate	7.25	0.001	0.037
	Accumulated	14.35	0.005	0.086
Δ Ratings	Immediate	7.32	0.003	0.057
	Accumulated	8.54	0.008	0.096
Δ Negative chatter	Immediate	-3.27	0.004	0.027
	Accumulated	-6.57	0.008	0.067
Δ Positive Chatter	Immediate	2.31	0.002	0.034
	Accumulated	3.89	0.004	0.075
Δ Competitor Chatter	Immediate	-4.86	0.003	0.025
	Accumulated	-5.27	0.006	0.036
Δ Competitor Negative Chatter	Immediate	1.28	0.002	0.028
	Accumulated	4.97	0.006	0.039

Note: Significant values are represented in bold.

Table B7: Results of Analysis Over Time

	Δ Chatter		Δ Negative Chatter		Δ Competitor Chatter	
	Immediate	Accumulated	Immediate	Accumulated	Immediate	Accumulated
2005-07	6.85	15.62	-5.37	-8.64	-2.10	-5.62
2006-08	6.68	13.17	-5.29	-7.27	-2.56	-5.84
2007-09	6.43	12.91	-5.10	-6.67	-2.81	-5.19

Online Appendix C1 – Cointegration Test

We test the system of nine variables (three stock performance variables – abnormal returns of the stock, its risk and its trading volume- and the six metrics of UGC –chatter, positive and negative chatter, ratings, competitor chatter and negative chatter) to determine the number of cointegrating equations using Johansen’s multiple-trace test method (Johansen 1995). The null hypothesis of the trace statistic assumes that there are no more than “n” cointegrating relations between the endogenous variables in the system (details of the statistic can be found in Johansen 1995 Chapters 11 and 12). This method tests the presence of cointegrating relation between the variables sequentially starting with rank zero and accepts the first estimated value which fails to reject the null hypothesis as the cointegrating rank. The trace statistic estimate and the critical value (at 5% and 1% level) are given in the table below. As the table suggests, we do not find the presence of any cointegrating equation among the variables in the system.

Maximum Rank	Trace Statistic	Critical Value (5%)	Critical Value (1%)
0	351.2	192.89	204.95
1	196.85	156	168.36
2	143.87	124.24	133.57
3	135.24	94.15	103.18
4	113.76	68.52	76.07
5	78.37	47.21	54.46
6	58.34	29.68	35.65
7	37.83	15.41	20.04
8	21.91	3.76	6.65

Testing with the maximum eigenvector statistic also yields similar results for up to five lags, which strongly suggests the absence of cointegration among the variables.

Online Appendix C2: Tests for Granger Causality

We test for the existence of temporal causality between each measure of UGC and stock performance (returns, risk and trading volume) for individual firms using the Granger Causality test (Granger 1969). UGC is said to “*Granger cause*” returns if the lags of UGC predict returns even after controlling for lags of returns. We run the Granger causality test separately for each of the metrics of UGC with the metrics of stock performance entering endogenously in the models of Granger causality. We implement the Wald test for Granger causality using the chi-square test statistic from the Wald Test (refer Lütkepohl 2005 for more details). Specifically, if the data generating processes are represented by y_{1t} and y_{2t} {say, y_{1t} represents the metric of UGC and y_{2t} represents measures of stock performance (returns, risk and trading volume)}, then we specify the model as:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \sum_{i=1}^n \begin{pmatrix} \phi_{1,i} & \phi_{12,i} \\ \phi_{21,i} & \phi_{22,i} \end{pmatrix} \begin{pmatrix} y_{1,t-i} \\ y_{2,t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad 3$$

Where ϕ are the coefficients which test for the direction of causality. We test for the significance of ϕ by placing restrictions on the coefficients. More formally, we use the asymptotic results of these constraints under the null hypothesis $H_0 : C\beta = c$, where C is an $N \times (K^2N + K)$ matrix of restrictions (ones and zeros) and c is an $N \times 1$ vector (here assumed to be a zero vector for the null hypothesis of no causality). Using the same equation, we further test for instantaneous causality as outlined in Lütkepohl 2005 to confirm our results of Granger causality.

Online Appendix C3 : Tests for Pooling

We test the poolability of the data across the different firms in the sample using the extended form of Chow F-test (Chow 1960) applied to the time series cross sectional data (Joshi and Hanssens 2010). To execute this test, we first find the sum of square of residuals of the restricted pooled model ($SSE_{restricted}$) and the unrestricted model ($SSE_{unrestricted}$). We find the F-statistic as the ratio of the difference in the sum of squares of these residuals to that of the unrestricted sum of squares of residuals (adjusted for their appropriate degrees of freedom). Details of the test can be found in Baltagi 2008 (p. 57-63).

$$F = \frac{(SSE_{restricted} - SSE_{unrestricted})/n}{(SSE_{unrestricted}/m_{unrestricted})}$$

We also determine the order of integration of the variables entering the model using the Im-Pesaran-Shin panel unit root tests (Im, Pesaran and Shin 1997). Similar to the individual firm analysis, the metrics of UGC are found to be non-stationary in their levels, but at the same time do not exhibit cointegration to warrant an error correction specification.

Online Appendix C4 Tests of Robustness: Estimate of Returns in a Reduced Form Single Equation

We ran the VAR model using only the excess returns ($R_{i,t} - R_{f,t}$) as the focal financial variable in the system. The Fama French factors ($R_m - R_f$, SMB, HML) were all taken as exogenous to the system. The resulting equation is:

$$\begin{bmatrix} R_{i,t} - R_{f,t} \\ \Delta Chtr_t \\ \Delta Ratings_t \\ \Delta NegChtr_t \\ \Delta PosChtr_t \\ \Delta CompChtr_t \end{bmatrix} = \sum_{n=1}^p \begin{pmatrix} \gamma_{11}^n & \dots & \gamma_{19}^n \\ \vdots & \ddots & \vdots \\ \gamma_{91} & \dots & \gamma_{99} \end{pmatrix} \begin{bmatrix} R_{i,t} - R_{f,t} \\ \Delta Chtr_{t-n} \\ \Delta Ratings_{t-n} \\ \Delta NegChtr_{t-n} \\ \Delta PosChtr_{t-n} \\ \Delta CompChtr_{t-n} \\ \Delta CompNegChtr_{t-n} \end{bmatrix} + \begin{pmatrix} \phi_{11} & \dots & \phi_{1p} \\ \vdots & \ddots & \vdots \\ \phi_{p1} & \dots & \phi_{pp} \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} + \begin{bmatrix} e_{ex_returns} \\ e_{Chtr,i,t} \\ e_{Ratings,i,t} \\ e_{NegChtr,i,t} \\ e_{PosChtr,i,t} \\ e_{CompChtr,i,t} \\ e_{CompNegChtr,i,t} \end{bmatrix}$$

The results are in the table below:

		Excess Returns
ΔChatter	Immediate	5.87
	Accumulated	10.85
ΔRatings	Immediate	2.38
	Accumulated	6.76
ΔNegative Chatter	Immediate	-3.71
	Accumulated	-6.45
ΔPositive Chatter	Immediate	7.56
	Accumulated	12.37
ΔCompetitor Chatter	Immediate	-3.45
	Accumulated	-5.47
ΔCompetitor Negative Chatter	Immediate	3.87
	Accumulated	6.87

As can be seen from this table, the change in negative chatter has a negative influence of about 3.7 basis points on the returns while the changes in overall chatter has a positive influence

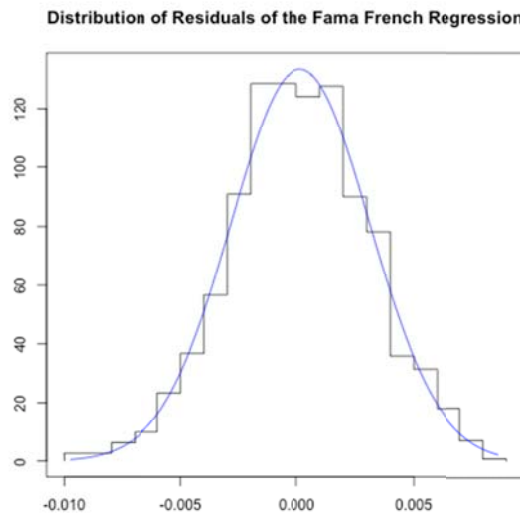
of about 5.9 basis points on the returns. Other variables do not have a significant influence on the excess returns of the firm.

Online Appendix C5 : Tests for Properties of Estimated Abnormal Returns

We test the normality and autocorrelation assumption of residuals of the Fama French Regression using an estimate window of 250 days.

Tests for Normality:

The density plot of the residuals (given below) indicates that the residuals are normally distributed with 0 mean (using the t test¹¹) and constant variance.



To formally test the assumptions of normality, we use the Jarque-Bera (1987) test. The Jarque–Bera test statistics for the regression equation assumes a null hypothesis of normality of the abnormal returns (residuals). The median JB χ^2 statistic across all the firms in the sample is 2.34 ($p = 0.28$) (range = [5.247 , 6.74]). Hence, we cannot reject the null of normality of residuals

¹¹ We use t tests to test if the average of residual of the Fama French regression is different from zero in the sample time period. The median t value of 1.25 (range 0.79 – 2.01) across the firms suggests that we cannot accept the alternate that the average of the residuals over the sample time period is different from zero.

and we can safely conclude that the distribution of the residual of the model satisfies the normality assumption.

Test for Autocorrelation:

We calculate the autocorrelation for multiple lags of the residuals to test for serial correlation. For this purpose, we use the Ljung-Box (Ljung and Box 1978) test which is a portmanteau test of white noise. The result indicates that the null hypotheses of absence of serial correlation cannot be rejected. The Ljung box Q (χ^2) statistic has a median p value of 0.35 across all the firms (range = 0.23 to 0.58). This confirms that abnormal returns do not suffer from serial correlation. Moreover, this model is derived from the estimation assuming an EGARCH process which takes into consideration the temporal dependence among the observations. Further, the second stage VAR system is estimated with Heteroskedasticity and Autocorrelation Consistent estimator in the VAR model, accounting for any potential serial correlation.