Marketing Science Institute Working Paper Series 2020
Report No. 20-101

Satisfaction Surveys or Online Sentiment: Which Best Predicts Firm Performance?

Evert de Haan

Satisfaction Surveys or Online Sentiment: Which Best Predicts Firm Performance?

Evert de Haan[a]

October 11, 2019

[a] Department of Marketing, Faculty of Economics and Business, University of Groningen, Nettelbosje 2, 9747AE Groningen, phone: +31 50 36 32802, email: evert.de.haan@rug.nl. (corresponding author)

The author thanks Evgeny Nosenko for collecting a large part of the data and conducting initial analyses for his master thesis.

In the past couple of decades, marketing research has proved that customer satisfaction is an important driver and predictor of future customer behavior and firm performance. For example, studies have shown that customer satisfaction is positively related to customer-level performance indicators, such as revenue per customer (Ittner and Larcker 1998), customer retention (De Haan, Verhoef, and Wiesel 2015), share of wallet (Cooil et al. 2007), and cross-buying (Verhoef, Franses, and Hoekstra 2001). The positive association between customer satisfaction and firm-level performance is also well established; for example, customer satisfaction is positively related to firm revenue (Morgan and Rego 2006), market share (Rego, Morgan, and Fornell 2013), and (abnormal) stock return (Aksoy et al. 2008; Fornell et al. 2006). Having data on customer satisfaction and other customer feedback or mindset metrics is thus valuable for monitoring the customer base, which, according to Gupta, Lehmann and Stuart (2004) is one of the most valuable assets of a firm.

Many studies investigating customer satisfaction at the firm level have used data from the American Customer Satisfaction Index (ACSI), which annually measures customer satisfaction at about 400 firms in 46 industries and ten economic sectors by conducting surveys with roughly 300,000 customers (https://www.theacsi.org/about-acsi). A more recent development is to use electronic word of mouth (eWOM) instead of survey-based measures to assess the level of engagement with a firm and the opinions about, perceptions of, and attitudes toward the firm and then use the findings to predict future firm performance. Some studies use quantitative measures to evaluate eWOM. For example, Srinivasan, Rutz, and Pauwels (2016) find a positive relationship between the number of likes on Facebook and a brand's sales. Other studies trying to link eWOM with firm performance have also taken the (qualitative) content of eWOM into account. Tirunillai and Tellis (2012) find that the volume and valance of eWOM are differently related to abnormal stock return and a stock's idiosyncratic risk and trading volume.

To date, research has largely examined these two types of measures in isolation, with customer satisfaction and other survey-based measures on the one side and eWOM and related measures on the other side. Furthermore, for eWOM, most studies have focused only on one or a small number of firms and industries in a relative short time horizon. Given the vastly different setups, levels of aggregation, and types of analyses, these studies are difficult to compare with one another. What is known, however, is that both customer satisfaction and eWOM are important predictors and drivers of firm performance, as indicated by the empirical generalizations of Hanssens (2015); Morgan and Rego's (2005) and Van Doorn, Leeflang, and Tijs's (2013) findings of a strong relationship between customer satisfaction and firm performance; and the high elasticities of eWOM volume and valance found in You, Vadakkepatt, and Joshi's (2015) meta-analysis.

In this study, I first investigate the literature on the impact of both customer satisfaction and eWOM on firm performance and then explicate the potential strengths and weaknesses of both data sources. Thereafter, I empirically investigate how customer satisfaction data compares with eWOM data in terms of explaining and predicting firm performance and test whether these two data sources also complement each other. This study contributes to uncovering future research avenues (e.g., Are extracted online customer opinions a good alternative to survey-based research?) and managerial decision making, in terms of how best to track and monitor the attitudes and opinions of the customer base. Thus, I tackle one of the Marketing Science Institute's (2018) research priorities for 2018–2020 (i.e., "Capturing Information to Fuel Growth: What Key Performance Indices (KPIs)/Metrics Should Be Measured and How?").

For the empirical study, I collected data from 46 firms (344 firm-year observations) for the period 2010–2017. These firms are located in 11 different industries, and the data set includes annual ACSI scores, eWOM volume, and eWOM sentiment by scrapping 8,436,261 tweets and 13 different firm performance indicators (e.g., revenue, market share, gross margins, stock return). I find that ACSI and eWOM sentiment are significantly correlated with each other, but only to a small degree. The different predictors thus contain unique information, both when I compare them cross-sectionally as well as over time. Especially changes in eWOM sentiment are difficult to predict, while past ACSI scores are good predictors of future ACSI scores (i.e., updates of ACSI contain less "new" information). By estimating a series of regression models, I find that the changes in eWOM sentiment are good predictors of future firm performance, especially the changes (growth) in negative eWOM. In contrast with my expectations, however, predictions of eWOM sentiment do not seem to improve over time, despite the increase in social media use over the years and the user base becoming (somewhat) more representative.

LITERATURE REVIEW

A major part of this literature review focuses on comparing how customer satisfaction and eWOM sentiment and volume can predict future firm performance. Although other studies have tried to predict future customer behavior (e.g., Cooil et al. 2007; De Haan, Verhoef, and Wiesel 2015; Verhoef, Franses, and Hoekstra 2001) and its impact on firm performance (Anderson and Mittal 2000; Gupta and Zeithaml 2006), I do not focus on this topic. Instead, I concentrate on the studies that directly link customer attitudes to firm performance—that is, the link between unobservable metrics and financial performance, as Gupta and Zeithaml (2006) define it in their overview article. In this section, I focus first on the link between customer satisfaction and firm performance and then on the link between eWOM and firm performance. Next, I provide an overview and comparison of the literature.

Customer Satisfaction and Firm Performance

Research has extensively examined the link between customer satisfaction and firm performance since the early 1990s, especially after the establishment of the Swedish Customer Satisfaction Barometer (SCSB) in 1989 (Fornell 1992) and the ACSI in 1994 (Fornell et al. 1996; see also https://www.theacsi.org/about-acsi/history). Many studies in this research stream have made use of data from either the SCSB or the ACSI. For example, Anderson, Fornell, and Lehmann (1994) use data from the SCSB to investigate the antecedents of customer satisfaction. They find a positive relationship between customer satisfaction and firm profitability (in terms of return on assets [ROA]), making this the first study to empirically show the link between customer satisfaction and economic returns. Anderson, Fornell, and Rust (1997) also use data from the SCSB and similarly find a positive relationship between customer satisfaction and profitability. In addition, they show that this relationship is positively moderated by a firm’s productivity for goods and negatively moderated by a firm’s productivity for services. The impact of customer satisfaction on firm performance is thus positive but heterogeneous.

Ittner and Larcker (1998) use data from the ACSI to show that customer satisfaction provides new and incremental value for predicting stock performance. They also find that customer satisfaction is a leading indicator of firm performance in terms of revenue and growth in return on sales and that the growth in customer satisfaction is a leading indicator of the growth in the number of customers. Anderson, Fornell, and Mazvancheryl (2004) conceptualize the link between customer satisfaction and shareholder value and, with data from the ACSI, empirically find a positive link between customer satisfaction and Tobin's q, equity price, and price-to-book ratio, though there is significant heterogeneity between industries and firms within an industry. Gruca and Rego (2005) deepen the understanding of the link between customer satisfaction and firm performance by empirically showing, again using ACSI data, that customer satisfaction increases future cash flow growth and reduces cash flow variability. Subsequent studies show that with the use of ACSI, stock portfolios that have a higher (abnormal) stock return can be assembled (e.g., Aksoy et al. 2008; Fornell et al. 2006).

In their research, Morgan and Rego (2006) compare many measures extracted from the ACSI questionnaire in their ability to predict firm performance. To do so, they draw on annual data from 80 firms across seven years. Their main finding is that customer satisfaction is a better predictor of Tobin's q, cash flow, shareholder return, sales growth, gross margins, and market share than other customer feedback metrics, such as the number of complaints, number of promoters, and average repurchase likelihood. Van Doorn, Leeflang, and Tijs (2013) replicate those findings using Dutch data on customer satisfaction and related measures and come to the same conclusion.

The impact of customer satisfaction on firm performance is not always as straightforward as it seems, however, as Anderson, Fornell, and Lehmann (1994) show a negative correlation between customer satisfaction and a firm's market share. Using data from the ACSI, Rego, Morgan, and Fornell (2013) find that compared with direct competition, relative customer satisfaction positively drives future market share but that market share has an even more negative impact on future customer satisfaction. This is because firms with higher market shares often have a more heterogeneous customer base, which tends to be more difficult to satisfy.

eWOM and Firm Performance

Research has examined the link between eWOM and firm performance to a lesser degree than that between customer satisfaction and firm performance. One reason is that the former stream of research only began around the early 2000s. Another reason is that there is no standardized publicly available data set that covers eWOM volume and sentiment, as is the case for customer satisfaction with the SCSB and the ACSI. Therefore, researchers have collected eWOM data from different sources with varying levels of aggregation, using different measures for eWOM, usually limiting the scope in the number of firms and industries, and limiting the time horizon. These factors make the studies more difficult to compare with one another, to arrive at strong generalizations, and to compare eWOM measures with survey-based measures, such as customer satisfaction. Despite this, some studies have investigated the link between eWOM and firm performance, and these studies show that having data on eWOM can help predict future firm performance.

One of the earliest studies that have investigated eWOM on social media is that of De Vries, Gensler, and Leeflang (2012). These authors examine how the characteristics of a brand post on social media are related to the number of likes and comments the post receives. Srinivasan, Rutz, and Pauwels (2016) explore how eWOM in turn affects firm performance and find a positive relationship between the number of likes on Facebook and a brand's sales.

Other studies that try to link eWOM with firm performance have, next to the volume of eWOM, also examined the content of eWOM. For example, Tirunillai and Tellis (2012) investigate how product reviews affect a firm's stock performance by using text mining on the volume of reviews, the star rating of the review, and the valance of the textual content of the review. They find that review volume is positively related to abnormal stock returns and trading volume. In addition, negative reviews have a negative effect on abnormal stock returns and increase both the trading volume and the idiosyncratic risk of stocks, while positive reviews do not have a significant effect. As their study highlights, in addition to quantitative measures, the actual content of eWOM is important to examine.

Comparison of the Literature Streams

Table 1 provides a selective overview of articles that have investigated customer satisfaction in relation to (survey-based) measures and/or eWOM measures on the one hand and a firm's or brand's financial performance on the other hand. As the table and the preceding discussion indicate, both streams of research have some overlap. However, to my knowledge, no study to date has carried out a comparative review of these two data sources. Furthermore, in many cases the studies that do measure eWOM have examined only one or a small number of firms over a relatively short period. In this study, I use both customer satisfaction data and eWOM data for a large group of firms, across multiple industries, and spanning a longer period, which enables us to better compare the data sources and obtain more generalizable results.

STRENGTHS AND WEAKNESSES OF THE DATA SOURCES

Table 2 provides an overview of the advantages and disadvantages of collecting and using customer satisfaction (and other survey-based measures) and eWOM to investigate the customer base's and the brand's performance and to make predictions about future firm performance. In general, with survey-based measures a researcher has more control over data collection in terms of who, what, and when to ask. This helps make the sample representative of the entire population. By contrast, eWOM is provided only by customers who are actively engaging with the firm, making it nonrepresentative with only the motivated (e.g., very satisfied or dissatisfied) customers engaging in eWOM. Furthermore, the number of comments is not under the control of the researcher, and the content can be about any topic, making it impossible to have predefined constructs (e.g., it is impossible to measure customer satisfaction with service employees if the eWOM comments do not address this topic). Furthermore, information from surveys is structured, with the predefined constructs usually measured on a fixed (e.g., Likert) scale. Conversely, an advantage of eWOM is that it is actual outspoken opinions that can reach and influence other customers. The content of eWOM is not restricted a priori in terms of what is being measured, and as a result a new issue that arises can be immediately observed in the eWOM comments, while with surveys, this cannot be detected directly. Furthermore, while survey research can be expensive to conduct on a large scale, eWOM is “freely” available, and the amount of eWOM can offer further information in itself (Tirunillai and Tellis 2012).

HYPOTHESES

Given the strengths and weaknesses of both data sources, I do not have a clear a priori expectation of which data source better predicts firm performance, though I do have expectations about the relative differences across industries, between firms, and over time. As Table 2 indicates, one weakness of eWOM, compared with the ACSI, is that it is typically not representative of the entire population of (potential) customers. The reason for this is that people who talk about firms online are typically quite involved, they want their opinions to be heard by the firm and/or by other consumers, and they are more likely to be either very satisfied or very dissatisfied. Moreover, users of social media in general are not representative of the entire population, as they tend to be younger; for example, according to the Pew Research Center (2019), 90% of the people aged 18–29 years were on social media in 2019, while this was the case for only 40% of people aged 65 and over. Over the years though, the social media user base has become more representative of the entire population (Pew Research Center 2019). Furthermore, given the growth of social media (e.g., Twitter), the potential reach of a message has also increased over the years. Furthermore, some company actions, such as Nike's 2018 campaign featuring Colin Kaepernick and Gillette's 2019 toxic masculinity campaign, have received a great deal of (highly polarized) attention on social media, in which the eWOM spilled over to traditional media in terms of news coverage (*The Guardian* 2019). Therefore, although the people most likely to engage in eWOM are still far from representative of the entire customer base, over time eWOM has become more representative and influential. Thus, I propose the following:

H_1: Over time, (a) the sentiment of eWOM and (b) the volume of eWOM have become a better predictor of future firm performance.

When the volume of eWOM is low, the sentiment is less stable and less reliable than when the volume is high because, in the former caser, a few additional comments can have a greater impact on the sentiment. Furthermore, a low volume of eWOM indicates that people are not talking much about the firm and/or its products, indicating that eWOM is less important in this situation. Finally, when the volume is low, eWOM is reaching fewer people, meaning that the number of people potentially being influenced is lower, thereby reducing the impact of the sentiment of eWOM. Thus:

H_2: For firms that have a higher (vs. lower) eWOM volume, sentiment is a better predictor of future firm performance.

DATA

I separately discuss the three data sources I use: the customer satisfaction data from the ACSI, the eWOM data scrapped from Twitter, and the firm performance data collected from Morningstar and Yahoo Finance. After this, I provide descriptive statistics of the final data set.

Customer Satisfaction Data

The study uses a firm's overall annual ACSI score as the firm-level customer satisfaction score. The ACSI is a national cross-industry measure of customer satisfaction in the United States obtained through customer surveys. The index measures the satisfaction of consumers with the quality of products and services. As noted previously, the survey's sample is designed to be representative of the total U.S. population. The questionnaire contains 26 questions, with the answers serving as the basis for the six measures of customer satisfaction. Overall customer satisfaction (ACSI) is operationalized through three survey measures: (1) an overall rating of satisfaction, (2) the degree to which performance falls short of or exceeds expectations, and (3) a rating of performance relative to the customer's ideal good or service in the category. Fornell et al. (1996) offer more details on the ACSI.

I chose to use data from the ACSI because (1) the data are publicly available, (2) they cover many U.S. industries and firms, (3) they are collected yearly on a long time horizon (since 1994) with a consistent methodology, (4) the household sample is representative of the U.S. population, and (5) many previous studies have uncovered the ACSI's value in predicting firm performance (see Table 1). I restricted the study to firms that meet the following criteria:

- The firm needed to be included in the ACSI.

- The firm needed to be qualified as a monobrand firm, which Mizik and Jacobson (2008, p. 20-21) define as a firm “in which a single brand represents the bulk of [its] business.” Tamrakar, Pyo, and Gruca (2018) note that restricting the sample to monobrand firms is beneficial given the degree to which these firms focus on their corporate branding; for monobrand firms, any sentiment expressed about the brand online could potentially affect the entire firm’s financial market performance.
- All firms in the sample needed to be publicly traded because, for these firms, financial performance data are readily available.
- All firms needed to actively use Twitter as a social media platform for communicating with customers because I use these data to capture eWOM in this study.

The final sample includes 46 firms that met all four criteria. For the vast majority of the firms, the data are available for the 2010–2017 period. The firms cover 11 different industries, for 344 firm-year observations in total. Most firms in the data set have leading positions in their industries. The full list of firms included in the data set is available in the Appendix.

eWOM Data and Sentiment Analyses

After collecting the ACSI data, I needed to match these with eWOM data. I chose Twitter as the source of eWOM because (1) tweets are forever available unless they are deleted by the author or the platform; (2) many firms are active on Twitter and communicate with, receive messages from, and are discussed by their (current, potential, and past) customers; and (3) tweets can be easily scrapped, which aids data collection. These three points enable us to cover a long time horizon of tweets about the firms in the initial ACSI data set.

The tweets scrapped for this study were tweets in English and in which accounts of the firms in question were mentioned, as in this case, it is clear that the tweet is indeed about the firm in question and that the user wants his or her content to be heard by the firm (e.g., questions addressed to or feedback for the firm). I used Twitterscraper, a package for Python developed by Taspinar (https://github.com/taspinar/twitterscraper), to scrape all tweets in which the accounts of the firms are mentioned from 2010 to 2017. In total, I scrapped 8,436,261 tweets, which gives us a mean number of 183,397 tweets per firm.

I examine two dimensions of eWOM in this study: the volume and valence per firm per year. The number of tweets posted serves as the volume of eWOM, and the sentiment polarity score of a message serves as a measure of the eWOM valence. I obtain sentiment polarity scores using the R software package "qdap" (for details, see Rinker 2018). Polarity scores are measures of how positive or negative a text is; scores are assigned to each tweet that mentions a firm's profile. For each tweet, the algorithm first tags positive and negative (polarized) words using the sentiment dictionary of Hu and Liu (2004). Then, the algorithm evaluates the context in which the polarized words are used by taking a cluster of four words before and two words after each polarized word and labeling this as the "context cluster." The context cluster can contain four types of words: (1) neutral words, which do not affect the meaning of the polarized word; (2) negating words (e.g., "not," "don't"), which reverse the valance of the polarized word (e.g., "I'm not happy"); (3) amplifying words (e.g., "very," "seriously"), which strengthen the valance of the polarized word (e.g., "I am very happy"); and (4) deamplifying words (e.g., "barely," "mildly"), which weaken the valance of the polarized word (e.g., "I am mildly happy"). The polarity scores of the context clusters of a tweet are summed up and divided by the square root of the number of words, which gives the sentiment score per tweet (for technical details, see Rinker 2018).

After calculating the sentiment score of each of the 8,436,261 tweets, I calculated for each firm-year combination (1) the share of positive tweets, (2) the share of negative tweets, (3) the "net tweet sentiment" or the share of positive tweets less the share of negative tweets (i.e., similar in construction to Reichheld's [2003] Net Promoter Score), and (4) the standard deviation of the sentiment to determine how much heterogeneity is in the sentiment. I then matched these data to the ACSI data.

Firm Performance Data

I obtained yearly data on financial performance for 2010–2017 through a secondary data provider, Morningstar. I used 13 measures of financial performance. First, I assessed (1) revenues (in millions of dollars), (2) revenue growth (in percentage), (3) gross margin (in percentage), (4), operating margin (in percentage), (5) earnings before tax (EBT) margin (in percentage), (6) cash flow (in millions of dollars), (7) ROA (in percentage), and (8) return on equity (ROE) (in percentage). Second, using revenue I calculated (9) the market share, or the revenue of the firm in year t divided by the sum of the revenues from all firms in the same industry in year t. Third, I obtained stock performance data from Yahoo Finance, including (10) market value (in millions of dollars), (11) trading volume (in millions of shares traded), (12) number of times traded (shares traded divided by total shares outstanding), and (13) stock return (in percentage, including [reinvested] dividends). I matched these 13 firm performance indicators to the ACSI and eWOM data.

Descriptive Statistics

Table 3 provides an overview of the 344 firm-year observations. I have these data for all predictors of firm performance—namely, the ACSI, the net tweet sentiment, the share of positive and negative tweets, the standard deviation of the tweet sentiment, and the number of tweets per firm per year. I also have these data for most of the dependent (firm performance) variables. In the cases in which I have missing data (e.g., the gross and operating margins are not available for all firm-year observations), I estimate the models on this limited data set.

Table 4 shows the correlations between the predictors and the firm performance variables. Somewhat surprisingly, the bottom half of the table shows that in many cases, the sign of the correlation is in the opposite direction of what I expected (e.g., a negative and significant correlation between ACSI and revenue). This negative correlation can be explained by reverse causality; indeed, Anderson, Fornell, and Lehmann (1994) find a negative correlation between customer satisfaction and a firm's market share, which, as Rego, Morgan, and Fornell (2013) later explain, is because larger firms have more difficulty in satisfying their (more heterogeneous) customer base, resulting in a drop in satisfaction, even though satisfaction in itself increases market share as a result of positive word of mouth (WOM) and higher customer retention. This more complicated chain of effects may also affect the other correlations, which are in the opposite direction as expected and something I take into account in the methodology by only predicting future firm performance and controlling for current firm performance.

Table 4 also shows that the share of negative sentiment is, in most cases, more strongly correlated with firm performance than the share of positive sentiment (in terms of the absolute values of the correlation statistics); this is in line with Tirunillai and Tellis (2012), who show that negative reviews have a much stronger impact than positive reviews on stock performance. Earlier studies have also found that for traditional WOM, negative WOM has a stronger impact on purchase decisions than positive WOM. For example, Arndt (1967) finds that negative WOM decreases purchase likelihood as much as twice the degree to which positive WOM increases purchase likelihood. Skowronski and Carlston 1989) demonstrate that, in general, negative information has a stronger impact on judgment than positive information.

The top half of Table 4 shows the correlation between the different predictors. As expected, most of the correlations are in the expected direction. I find that the ACSI is positively correlated with the net tweet sentiment and the share of positive sentiment and negatively correlated with the share of negative sentiment. Regarding the absolute value of the correlations, the share of negative sentiment has the strongest correlation with the ACSI, but the correlations between the eWOM predictors and ACSI are limited in terms of size, indicating that, to a large extent, online sentiment and the ACSI are quite different (i.e., they are not good substitutes for each other).

I also find that a higher standard deviation in tweet sentiment is negatively correlated with the ACSI. This indicates that more mixed opinions, and thus potentially higher heterogeneity in the customer base, are related to lower levels of satisfaction, which is in line with the arguments of Rego, Morgan, and Fornell (2013). A higher number of tweets is also negatively correlated with the ACSI. Similarly, Rego, Morgan, and Fornell (2013) find a negative correlation between market share and the ACSI.

The high correlations of the net tweet sentiment with the other eWOM sentiment measures are mainly by design, as the net tweet sentiment is the share of positive tweets less the share of negative tweets. The negative correlation between the share of positive sentiment and the share of negative sentiment is as expected, but this negative correlation is relatively small in magnitude. This low correlation may be due to the large number of tweets with no sentiment (i.e., neutral tweets about the firm). In addition, an increase in the share of positive tweets does not automatically mean a decrease in the share of negative tweets, and vice versa. The relatively low correlations between most of the predictors allow us to test how well the metrics differ in predicting future firm performance and whether it makes sense to combine different predictors, as the level of multicollinearity will be limited.

METHODOLOGY

Before investigating the relationship between the predictors and future firm performance, I want to determine the degree to which information from the ACSI and eWOM is actually related to each other. In Table 4, I already show that the correlations between the ACSI and the eWOM variables are significant but relatively low in magnitude. A question that remains is whether past information of one predictor can help explain future values of the same and other predictors. If the past values of a predictor are highly related to future values of the same predictor, the annual update of this predictor will not provide much new information. Similarly, if one predictor can predict future values of another predictor well (i.e., one predictor is a leading indicator of the other predictor), the other predictor will also not provide much new information. I explore this issue to discover the degree to which the different predictors contain new and unique information and if they are potentially valuable for firms to collect. To do so, I estimate a series of regression models shown in Equation 1:

$$X_{mijt} = \alpha_0 + \sum_{m=1}^{M} \alpha_m \cdot X_{mijt-1} + e_{mijt}, \tag{1}$$

where X_{mijt} is predictor m for firm i in industry j in year t. I estimate the regression models of Equation 1 both excluding and including the lagged dependent variable. By excluding the lagged dependent variable, I can determine the extent to which the other predictors are able to predict the predictor of interest (i.e., the extent to which they can substitute the predictor). By including the lagged dependent variable, I can assess the extent to which the predictor can predict future values of itself and, in turn, the extent to which the annually updated figures (e.g., the newly released ACSI numbers) provide new information or are similar to those of the previous year.

In the second step, I investigate the extent to which I can predict future firm performance with the different predictors. For this, I use Equation 2:

$$Y_{nijt} = \beta_{0nmi} + \beta_{1nm} \cdot X_{mijt-1} + \beta_{2nm} \cdot Y_{njit-1} + \sum_{t=2011}^{2017} \beta_{tnm} \cdot year_t + u_{nijt}, \qquad (2)$$

where Y_{nijt} is firm performance outcome n for firm i in industry j in year t and $year_t$ is a dummy variable for each year. In Equation 2, I include a random intercept that captures the baseline differences between the different firms in the data set to incorporate firm-specific effects. I also include the lag of firm performance as a predictor, to determine the extent to which each predictor provides incremental value in its predictions. This setup is in line with the idea of Granger (1969) causality; that is, it provides an answer to the question whether one time series (in our case, the ACSI- or eWOM-based predictors) is useful to predict another time series (in our case, firm performance) and thus has an incremental predictive value.

Given that some of the variables are highly skewed, I use the log transformation of the revenue, the market value, the trading volume, the number of times traded, and the eWOM volume. When estimating Equation 2, I use three transformations for the predictors:

1. Absolute levels of the predictors, as presented in Equation 2;
2. The growth levels (i.e., yearly changes in or first differences) of the predictors, that is, ([Xmijt – Xmijt-1]/Xmijt); and
3. The relative levels of the predictors compared with similar firms in the same industry, that is, (Xmijt – $\overline{\text{X}}$mjt), where $\overline{\text{X}}$mjt is industry j's mean score on predictor m in year t.

To determine which model (i.e., which predictor) performs best, I calculate the Akaike weights in line with Wagenmakers and Farrell (2004) and De Haan, Verhoef, and Wiesel (2015). The Akaike weight can be interpreted as the likelihood that a model (in our case, predictor m) is the best performing of all estimated models (in our case, all predictors included in this study) to predict a certain outcome variable (in our case, firm performance variable n). I calculate the Akaike weights with Equations 3 and 4:

$$\Delta AIC_{m,n} = AIC_{m,n} - \min(AIC_n), \quad (3)$$

$$w\left(AIC_{m,n}\right) = \frac{\exp(-\frac{1}{2} \cdot \Delta AIC_{m,n})}{\sum_{k=1}^{K} \exp(-\frac{1}{2} \cdot \Delta AIC_{m,n})}, \quad (4)$$

where $AIC_{m,n}$ is the value of the Akaike information criterion (AIC) of the model with predictor m as the independent variable and firm performance variable n as the dependent variable, $\min(AIC_n)$ is the minimum value of the AIC of all the predictors trying to predict firm performance variable n, and $w\left(AIC_{m,n}\right)$ is the Akaike weight of the model with predictor m as the independent variable and firm performance variable n as the dependent variable.

Next, I extend the model in Equation 2 by estimating a stepwise regression, to assess whether a combination of predictors predicts firm performance even better than a single predictor. For this, I begin by including all the predictors at once in Equation 2, excluding the net tweet sentiment to avoid perfect multicollinearity. I then order the predictors by their contribution to the model and remove the predictors one by one (starting with the one with the lowest contribution), until removing a predictor significantly worsens the model based on the likelihood-ratio test.

Finally, I extend Equation 2 by including moderators for the eWOM predictors to test the hypotheses. For the moderators, I use a time-trend variable (to test H_{1a} and H_{1b}) and the volume of eWOM (to test H_2). For sentiment, I use the net tweet sentiment in this model, as this combines the shares of both positive and negative sentiment, without confronting problems with multicollinearity.

RESULTS

Table 5 shows the results of Equation 1, in terms of how the predictors help explain one another. Note that I do not include the net tweet sentiment in the equations, because doing so results in perfect multicollinearity, as the net tweet sentiment is a linear combination of the shares of positive and negative tweets. Table 5 shows that current negative sentiment is a significant predictor of next year's ACSI; when the share of negative sentiment increases by 1 percentage point (i.e., .01 in value), next year's ACSI is .851 points lower. When I control for current ACSI in predicting next year's ACSI, however, only the current ACSI is significant and the eWOM-based predictors turn nonsignificant. This indicates that the eWOM metrics have no incremental value in predicting future ACSI, meaning that ACSI is not driven (or Granger caused) by the eWOM predictors. I do however find that when including lagged ACSI, the R-square is rather high (.729), meaning that especially current ACSI does well in predicting next year's ACSI. This high autocorrelation indicates that the current ACSI does not provide much new information over the previous year's ACSI.

For the eWOM sentiment predictors, I find a somewhat similar pattern; the current year's ACSI is able to predict next year's online sentiment well, but most of this effect disappears when I control for current sentiment. The R-squares of the eWOM sentiment models are quite a bit lower though than those of the ACSI models, even when I include lagged eWOM sentiment. This means that the eWOM sentiment varies more over the years than the ACSI and thus is more difficult to predict from one year to the next. This could indicate that eWOM sentiment contains a great deal of new information, but it could also mean that this metric is less stable and thus may contain more noise.

Finally, I find that neither the eWOM sentiment nor the ACSI influences the eWOM volume. The volume of eWOM in the year before can however be well predicted the current eWOM volume (R^2 = .731). This high autocorrelation makes sense because when a firm or brand is popular in one year, it is also likely to be popular in the next year. Still, there remains unexplained variance, which means that every year, this predictor is bringing in new information. Given the findings in Table 5, I can thus conclude that all predictors contain unique information, in comparison with one another and when examining the changes over time.

As the metrics contain unique information, I test whether this is valuable for predicting future firm performance. Table 6 reports the results of Equation 2. Each parameter in the table is based on a separate regression model (i.e., this table presents all the β_{1nm}s from Equation 2). Regarding the prediction of revenue in the first column, I show that only the untransformed positive sentiment, untransformed negative sentiment, untransformed standard deviation in sentiment, and untransformed number of tweets are significantly related to future revenue. The parameters are mostly in line with the expectations that more negative tweets and more mixed opinions (i.e., a higher standard deviation in the sentiment) lead to lower revenue while a higher number of tweets is related to higher revenue. However, more positive tweets are also related to lower revenue; thus, it seems that emotional tweets, whether positive or negative, lead to lower revenue. Regarding the Akaike weights for the models explaining future revenue, with 60.31% certainty I find that the untransformed negative sentiment is the best-predicting predictor; however, it is not the best predictor for all the ACSI measures.

For all firm performance outcome variables, I show that in most cases, the growth in negative sentiment is the best predictor, as indicated by the Akaike weights. The growth in the standard deviation of the sentiment (i.e., having more mixed sentiment) and the growth in positive sentiment also have relatively high Akaike weights. So, for most predictions, the growth (or change) in one of the sentiment predictors is the best predictor of future firm performance, as the mean Akaike weights presented in the last column of Table 6 also make clear.

In terms of significance, I find somewhat different results. Here, the untransformed predictors are statistically significant in more cases than the growth and relative transformations. The ACSI and the share of positive sentiment are statistically significant for predicting 5 of the 13 firm performance outcomes and thus are the two best-performing predictors. The reason for the difference in conclusion from that of the Akaike weights, in which the growth in negative sentiment performed the best followed by the other growth transformed predictors, is that the Akaike weights consider the models as a whole, while the significance indicates something about the particular parameter. Similarly, Morgan and Rego (2006) find that some customer feedback metrics are nonsignificant in explaining firm performance, with the underlying model still having the highest R-square of all customer feedback metrics. Which one is more important, the significance or the Akaike weight, thus depends on whether the researcher is interested in the parameter estimate or the predictions of the model as a whole.

As indicated in the correlation matrix of Table 4 and the findings in Table 5, the different predictors are not strongly related to one another and provide unique information. Another question is whether adding the different metrics together provides better predictions than using a single metric, as I did in Table 6. To determine this, I estimated the stepwise regression models explained in the "Methodology" section and report the results in Table 7. As the table shows, for

4 of the 13 financial performance variables, a combination of at least two predictors works better than just one predictor. For example, for the revenue model, I show that four predictors (i.e., the positive sentiment, the relative positive sentiment, the negative sentiment, and the volume of tweets) jointly best predict the outcome variable. Thus, similar to De Haan, Verhoef, and Wiesel (2015), I can conclude that combining metrics is sometimes better than monitoring just one metric.

Finally, in Table 8 I test the hypotheses. In most models, the lag of the dependent variable is the most significant, often having a parameter with a value of around 1.000. This means that the previous year's performance is similar to the next year's performance with some noise around it. The lag dependent variable explains much of the variance in the dependent variable, leaving little room for the other variable to add any. This also explains why I find the many nonsignificant predictors in Table 6. There are a few exceptions in which the parameter of the lag of the dependent variable is close to zero; for example, the parameter for revenue growth is .153. In general, growth (in terms of first differences) is less easy to predict than the absolute value of performance, because it is less stable. The only nonsignificant lag dependent variable is stock return. Thus, past stock return cannot help predict future stock return of the firm, which is in line with the efficient-market hypothesis (Malkiel and Fama 1970).

I do not find strong support for the hypotheses. The growing importance of sentiment of eWOM over time only holds for cash flows and the number of times a stock is traded. Thus, I find only weak support for H_{1a}. Similarly, the tweet volume only increases in importance for gross margin, which gives little support to H_{1b}. Finally, the interaction between sentiment and volume of eWOM is nonsignificant for all 13 firm performance models, so I must reject H_2.

DISCUSSION

In this study, I collected data from 46 firms across 11 different industries for the period 2010–2017. I found that ACSI and eWOM sentiment are significantly correlated with each other, but only to a small degree. By estimating a series of regression models, I found that the changes in eWOM sentiment are good predictors of future firm performance and that, in general, the change (growth) in negative eWOM is the best predictor of future firm performance. Combining different predictors improves the prediction of future firm performance in a few cases. In contrast with the expectations, the predictive ability of eWOM-based predictors does not improve over time, even though social media use increases over the years and the user base is becoming (somewhat) more representative of the entire population.

Practical Implications

Given the research findings, I recommend that firms monitor (changes in) eWOM sentiment and volume, as these are good indicators of future firm performance. Firms could use eWOM data in combination with traditional customer feedback metrics to create a dashboard to better monitor their customer base. Moreover, using eWOM sentiment and volume as outcome variables (e.g., of marketing actions) can be useful for measuring the impact of marketing. Firms could use these variables, for example, when information on (disaggregate) sales or firm performance is not available or as an in-between stage for when the effect of marketing on performance is more long term. Here, eWOM sentiment and volume can be used in a similar way to the mindset metrics from Srinivasan, Vanhuele, and Pauwels (2010), i.e. as something which can be related to drivers such as marketing actions (e.g. changes in pricing, advertising, etc.) and to outcomes such as sales or other firm or customer level performance variables.

Firms could furthermore combine this with other information, to create a more elaborate dashboard of customer feedback. This can be similar eWOM information from other social media, other online communities, and review sites, and other information derived from these eWOM (e.g. based on topic analyses). Since combining metrics is beneficial, as shown in the results from Table 7 and demonstrated by De Haan, Verhoef, and Wiesel (2015), also including survey based measures could be included here.

Furthermore, firms can look at other internal data, e.g. by analyzing the messages they receive from their customers, such as the content of emails and of open response fields in surveys. As De Haan and Menichelli (2019) have shown, the sentiment and topic of written comments in a survey help to better predict customer churn, even when controlling for survey based metrics. Adding this information can thus improve a dashboard even further. All data sources might have their own advantages and disadvantages, as I also have shown in Table 2, but combining them and carefully crafting a customer feedback dashboard can provide managers the tools and insights to better monitor their customer base and, as shown in the current study, help predict (and with that potentially improve) future firm performance.

Limitations and Further Research

As with any research, this study has several limitations. First, I only make use of one source of eWOM—namely, tweets aimed at firms—but other social media (e.g., Facebook, Instagram) and online information sources (e.g., review sites) also contain rich information. Although incorporating these other data sources would bring new challenges, in terms of collecting the data (which can be difficult given website restrictions and deletion of past information) and combining the different sources into one or a convenient number of predictors, it would also be advantageous, as the different social media sites have different target groups, and thus combining them could provide better (more representative) scores of the eWOM-based metrics. Further research could focus on how to combine the data and metrics from different (online) sources to potentially develop an eWOM dashboard and on how to create standardized information on eWOM, as is done with customer satisfaction in the ACSI and the SCSB.

A second limitation is that I aggregated the rich eWOM data (i.e., information from 8,436,261 tweets) to 344 firm-year observations. I did this to make the eWOM metrics comparable to the ACSI and also because most firm performance variables are only available at the yearly or quarterly level. An exception is the stock performance data, which are available for every trading day. Although the eWOM predictors can vary within the year and, when having sufficient volume of eWOM, can even vary within a day, I do not have this rich information for customer satisfaction, making comparison based on this within-year variation unfair. Further research could use richer data on customer satisfaction (e.g., YouGov data) to compare changes per month, week, or even day with that of eWOM metrics and connect this with firm (e.g., stock, sales) performance.

Third, because of data restrictions, this study focuses only on large firms operating in the United States. The ACSI is also focused on the United States, while for eWOM metrics, I scrapped all (worldwide) English-language tweets. Given that firm performance is also measured at the global level, this study generates a bias in favor of the eWOM metrics, as eWOM metrics are somewhat more representative of the rest of the (English-language-tweeting) world than the U.S.-focused ACSI. Thus, further research could try to make the data sources more comparable by, for example, having customer satisfaction data from multiple countries, including smaller firms, and also incorporating non-English-language tweets. Such a study would provide more generalizable findings and, given the different data sources, would match geographically better with firm performance, thereby improving predictions.

APPENDIX

Overview of All 46 Firms Included in Data Set

Firm Name	Twitter		Industry (ACSI)	Tweets
	Account Name	Start Date		
1. Aetna	Aetna	01.09.2008	Health insurance	12,743
2. Allstate	Allstate	01.04.2008	Property insurance	26,487
3. Amazon.com	Amazon	01.02.2009	Internet retail	412,744
4. American Airlines	Americanair	01.03.2009	Airlines	636,266
5. Anthem	Askanthem	01.05.2010	Health insurance	5,875
6. Bank of America	Bankofamerica	01.10.2010	Banks	47,302
7. Barnes & Noble	Bnbuzz	01.04.2009	Specialty stores	53,812
8. Best Buy	Bestbuy	01.11.2008	Specialty stores	202,227
9. Chase	Chase	01.03.2011	Banks	54,040
10. Citibank	Citibank	01.10.2009	Banks	26,424
11. Costco	Costco	01.01.2013	Department stores	23,766
12. Delta	Delta	01.05.2007	Airlines	554,572
13. Dollar General	Dollargeneral	01.10.2009	Department stores	10,655
14. Domino's Pizza	Dominos	01.04.2009	Restaurants	273,053
15. Dunkin' Donuts	Dunkindonuts	01.09.2007	Restaurants	257,532
16. eBay	Ebay	01.01.2009	Internet retail	172,679
17. FedEx	Fedex	01.04.2010	Consumer shipping	103,135
18. Google	Google	01.02.2009	Internet services	390,856
19. The Home Depot	Homedepot	01.05.2008	Specialty stores	71,280
20. J.C. Penney	Jcpenney	01.11.2008	Department stores	87,168
21. JetBlue	Jetblue	01.05.2007	Airlines	262,943
22. Kohl's	Kohls	01.04.2009	Department stores	170,160
23. Levi Strauss	Levis	01.10.2010	Apparel	30,871
24. Lowe's	Lowes	01.01.2009	Specialty stores	70,829
25. Macy's	Macys	01.06.2009	Department stores	90,539
26. McDonald's	McDonalds	01.09.2009	Restaurants	437,445
27. MetLife	Metlife	01.09.2009	Life insurance	7,851
28. Nike	Nike	01.11.2011	Athletic shoes	179,500
29. Nordstrom	Nordstrom	01.06.2008	Department stores	122,315
30. Office Depot	Officedepot	01.01.2009	Specialty stores	20,809
31. Overstock	Overstock	01.11.2008	Internet retail	14,535
32. Papa John's	Papajohns	01.12.2008	Restaurants	160,345
33. Progressive	Progressive	01.04.2007	Property insurance	24,666
34. Prudential	Prudential	01.02.2013	Life insurance	7,320
35. Rite Aid	Riteaid	01.04.2010	Department stores	18,001
36. Sears	Sears	01.01.2009	Department stores	133,449

37. Southwest Airlines	Southwestair	01.07.2007	Airlines	492,101
38. Staples	Staples	01.09.2009	Specialty stores	77,713
39. Starbucks	Starbucks	01.11.2006	Restaurants	673,670
40. Target	Target	01.11.2009	Department stores	316,557
41. United	United	01.03.2011	Airlines	746,432
42. United Health	Unitedhealthgrp	01.10.2012	Health insurance	1,953
43. UPS	Ups	01.06.2010	Consumer shipping	151,885
44. Walgreens	Walgreens	01.06.2009	Department stores	86,830
45. Wells Fargo	Wellsfargo	01.03.2007	Banks	78,747
46. Wendy's	Wendys	01.07.2009	Restaurants	636,179

REFERENCES

Aksoy, Lerzan, Bruce Cooil, Christopher Groening, Timothy L. Keiningham, and Atakan Yalçın (2008), "The Long-Term Stock Market Valuation of Customer Satisfaction," *Journal of Marketing*, 72 (4), 105-22.

Anderson, Eugene W., Claes Fornell, and Donald R. Lehmann (1994), "Customer Satisfaction, Market Share, and Profitability: Findings from Sweden," *Journal of Marketing*, 58 (3), 53-66.

Anderson, Eugene W., Claes Fornell, and Sanal K. Mazvancheryl (2004), "Customer Satisfaction and Shareholder Value," *Journal of Marketing*, 68 (4), 172-85.

Anderson, Eugene W., Claes Fornell, and Roland T. Rust (1997), "Customer Satisfaction, Productivity, and Profitability: Differences Between Goods and Services," *Marketing Science*, 16 (2), 129-45.

Anderson, Eugene W. and Vikas Mittal (2000), "Strengthening the Satisfaction-Profit Chain," *Journal of Service Research*, 3 (2), 107-20.

Arndt, Johan (1967), "Role of Product-Related Conversations in the Diffusion of a New Product," *Journal of Marketing Research*, 4 (3), 291-95.

Cooil, Bruce, Timothy L. Keiningham, Lerzan Aksoy, and Michael Hsu (2007), "A Longitudinal Analysis of Customer Satisfaction and Share of Wallet: Investigating the Moderating Effect of Customer Characteristics," *Journal of Marketing*, 71 (1), 67-83.

De Haan, Evert, Peter C. Verhoef, and Thorsten Wiesel (2015), "The Predictive Ability of Different Customer Feedback Metrics for Retention," *International Journal of Research in Marketing*, 32 (2), 195-206.

De Vries, Lisette, Sonja Gensler, and Peter S.H. Leeflang (2012), “Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing,” *Journal of Interactive Marketing*, 26 (2), 83-91.

Fornell, Claes (1992), “A National Customer Satisfaction Barometer: The Swedish Experience,” *Journal of Marketing*, 56 (1), 6-21.

Fornell, Claes, Michael D. Johnson, Eugene W. Anderson, Jaesung Cha, and Barbara Everitt Bryant (1996), “The American Customer Satisfaction Index: Nature, Purpose, and Findings," *Journal of Marketing*, 60 (4), 7-18.

Fornell, Claes, Sunil Mithas, Forrest V. Morgeson III, and Mayuram S. Krishnan (2006), “Customer Satisfaction and Stock Prices: High Returns, Low Risk,” *Journal of Marketing*, 70 (1), 3-14.

Granger, Clive W.J. (1969), “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods,” *Econometrica: Journal of the Econometric Society*, 37 (3), 424-38.

Gruca, Thomas S. and Lopo L. Rego (2005), “Customer Satisfaction, Cash Flow, and Shareholder Value,” *Journal of Marketing*, 69 (3), 115-30.

Guardian, The (2019), “Gillette #MeToo Razors ad on 'Toxic Masculinity' Gets Praise – and Abuse,” (accessed August 20, 2019), https://www.theguardian.com/world/2019/jan/15/gillette-metoo-ad-on-toxic-masculinity-cuts-deep-with-mens-rights-activists.

Gupta, Sunil, Donald R. Lehmann, and Jennifer Ames Stuart (2004), “Valuing Customers” *Journal of Marketing Research*, 41 (1), 7-18.

Gupta, Sunil and Valarie Zeithaml (2006), “Customer Metrics and Their Impact on Financial Performance,” *Marketing Science*, 25 (6), 718-39.

Hanssens, Dominique M. (2015), *Empirical Generalizations About Marketing Impact: What We Have Learned from Academic Research*, 2d ed. Cambridge, MA: Marketing Science Institute.

Hu, Minqing and Bing Liu (2004), “Mining and Summarizing Customer Reviews,” in *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 168-77.

Ittner, Christopher D. and David F. Larcker (1998), “Are Nonfinancial Measures Leading Indicators of Financial Performance? An Analysis of Customer Satisfaction,” *Journal of Accounting Research*, 36, 1-35.

Luo, Xueming (2009), “Quantifying the Long-Term Impact of Negative Word of Mouth on Cash Flows and Stock Prices,” *Marketing Science*, 28 (1), 148-65.

Malkiel, Burton G., and Eugene F. Fama (1970), "Efficient capital markets: A review of theory and empirical work." *The Journal of Finance*, 25 (2), 383-417.

Marketing Science Institute (2018), “2018-2020 Research Priorities,” http://www.msi.org/research/2018-2020-research-priorities/.

Mizik, Natalie and Robert Jacobson (2008), “The Financial Value Impact of Perceptual Brand Attributes,” *Journal of Marketing Research*, 45 (1), 15-32.

Morgan, Neil A. and Lopo Leotte Rego (2006), “The Value of Different Customer Satisfaction and Loyalty Metrics in Predicting Business Performance,” *Marketing Science*, 25 (5), 426-39.

Pew Research Center (2019), “Social Media Fact Sheet,” (June 12), (accessed August 20, 2019), https://www.pewinternet.org/fact-sheet/social-media/.

Rego, Lopo L., Neil A. Morgan, and Claes Fornell (2013), “Reexamining the Market Share–Customer Satisfaction Relationship,” *Journal of Marketing*, 77 (5), 1-20.

Reichheld, Frederick F. (2003), “The One Number You Need to Grow,” *Harvard Business Review*, 81 (12), 46-55.

Rinker, Tyler (2018), “Package ‘qdap’,” (accessed August 10, 2019), https://cran.r-project.org/web/packages/qdap/qdap.pdf.

Skowronski, John J. and Donal E. Carlston (1989), “Negativity and Extremity Biases in Impression Formation: A Review of Explanations,” *Psychological Bulletin*, 105 (1), 131-42.

Srinivasan, Shuba, Oliver J. Rutz, and Koen Pauwels (2016), “Paths to and Off Purchase: Quantifying the Impact of Traditional Marketing and Online Consumer Activity,” *Journal of the Academy of Marketing Science*, 44 (4), 440-53.

Srinivasan, Shuba, Marc Vanhuele, and Koen Pauwels (2010), “Mind-Set Metrics in Market Response Models: An Integrative Approach,” *Journal of Marketing Research*, 47 (4), 672-84.

Tamrakar, Chanchal, Tae-Hyung Pyo, and Thomas S Gruca (2018), “Social Media Sentiment and Firm Value,” working paper, Georgia Southern University.

Tirunillai, Seshadri and Gerard J. Tellis (2012), “Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance,” *Marketing Science*, 31 (2), 198-215.

Van Doorn, Jenny, Peter S.H. Leeflang, and Marleen Tijs (2013), “Satisfaction as a Predictor of Future Performance: A Replication,” *International Journal of Research in Marketing*, 30 (3), 314-18.

Verhoef, Peter C., Philip Hans Franses, and Janny C. Hoekstra (2001), “The Impact of Satisfaction and Payment Equity on Cross-Buying: A Dynamic Model for a Multi-Service Provider,” *Journal of Retailing*, 77 (3), 359-78.

Wagenmakers, Eric-Jan and Simon Farrell (2004), “AIC Model Selection Using Akaike Weights,” *Psychonomic Bulletin &* Review, 11 (1), 192-96.

You, Ya, Gautham G. Vadakkepatt, and Amit M. Joshi (2015), “A Meta-Analysis of Electronic Word-of-Mouth Elasticity,” *Journal of Marketing*, 79 (2), 19-39.

TABLES

TABLE 1
Selective Literature Overview and Contribution

Study	Satisfaction Data	eWOM Data	Firm Performance	Sample
Anderson, Fornell, and Lehmann (1994)	SCSB (levels and growth)	N/A	ROA, market share	77 firms, 2 years
Anderson, Fornell, and Rust (1997)	SCSB (log)	N/A	ROA	4 years (n = 126–170 firm-year obs., depending on the analysis)
Ittner and Larcker (1998)	ACSI and related measures (levels and growth)	N/A	Revenue, expenses, margins, return on sales, retail customers, business and professional customers, abnormal stock return	138–140 firms, 2 years
Anderson, Fornell, and Mazvancheryl (2004)	ACSI (levels and growth)	N/A	Tobin's q, equity price, price-to-book ratio	4 years (n = 216-330 firm-year obs., depending on the analysis)
Gruca and Rego (2005)	ACSI (levels)	N/A	Cash flow growth and variability	9 years (n = 840 firm-year obs.)
Morgan and Rego (2006)	ACSI (levels, various measures)	N/A	Tobin's q, cash flow, shareholder return, sales growth, gross margins, market share	80 firms, 7 years
Fornell et al. (2006)	ACSI (levels and growth)	N/A	Stock return	20–26 firms, 7 years
Aksoy et al. (2008)	ACSI (growth and relative)	N/A	Stock return	151, 39 quarters

Luo (2009)	N/A	Volume of (unexplained) negative WOM	Cash flow, stock price, stock volatility,	1 industry, 9 firms, 7 years
Tirunillai and Tellis (2012)	N/A	Volume, star rating, positive vs. negative sentiment	Abnormal stock return, idiosyncratic risk, trading volume	6 industries, 15 firms, 4 years
Rego, Morgan, and Fornell (2013)	ACSI (levels and relative)	N/A	Market share	23 industries, 104 firms, 13 years
Van Doorn, Leeflang, and Tijs (2013)	Various satisfaction related measures	N/A	Sales revenue growth, gross margin, cash flow	4 industries, 46 firms, 3 years
Srinivasan, Rutz, and Pauwels (2016)	N/A	Facebook likes + unlikes	Sales volume	1 firm, 40 weeks
This study	ACSI (levels, growth, and relative)	Volume and sentiment (various measures) of branded tweets (levels, growth and relative)	Sales revenue, revenue growth gross margin, operating margin, EBT margin, cash flows, ROA, return on equity, market share, market value, trading volume, time traded, stock return	11 industries, 46 firms, 8 years, (n = 344 firm-year obs.)

Note: ACSI = American Customer Satisfaction Index, eWOM = electronic Word-of-Mouth, ROA = Return on Assets , SCSB = Swedish Customer Satisfaction Barometer

TABLE 2
Strengths and Weaknesses of the Data Sources

<table>
<tr><th></th><th>Satisfaction Data (and Other Survey-Based Measures)</th><th>eWOM Data</th></tr>
<tr><td>Strengths</td><td>• Provides structured information that can be directly analyzed
• Control of who to ask
• Can be made representative
• Control of sample size
• Control of when to ask
• Control of what to ask
• Strong empirical evidence of usability</td><td>• Actual outspoken opinions
• Can be observed by and influence other consumers
• Not priory restricted in what is measured
• Richer data give potential to further analyze:
  • Exact content, words and topic, reach (e.g., # of readers and replies)
• Continuous (real-time) feedback
• Amount of eWOM also provides potential information</td></tr>
<tr><td>Weaknesses</td><td>• Costly and time consuming to collect
• Usually measures specific moments (e.g., annually with the ACSI)
• Expressed opinions by consumers in surveys might not be expressed in public and not reach/influence other consumers
• Low dimensionality of data (restricted in what is being included in the survey)
• Satisfaction data are lagged current affairs
• Surveys can annoy customers (i.e., they must free up time to fill out the survey)
• Nonresponse bias
• Higher dropout rates for longer surveys</td><td>• Less (or not at all) representative:
  • Very (dis)satisfied consumers are more likely to engage in eWOM
  • Consumers who are very active on social media are not the “average” consumer
• eWOM can be scarce at moments:
  • Highly depends on size of the firm
  • Highly depends on industry
  • Highly depends on whether something is going on with the firm (e.g., a campaign, scandal)
• Less straightforward to analyze
• Less control of who, when and what to measure</td></tr>
</table>

TABLE 3
Descriptive Statistics

	N	M	Median	SD	Min.	Max.
ACSI	344	76.83	78.00	5.40	56.00	88.00
Net tweet sentiment	344	4.52%	4.52%	5.13%	-16.67%	40.00%
Share positive tweets	344	8.60%	8.14%	4.36%	.00%	60.00%
Share negative tweets	344	4.08%	3.78%	2.21%	.00%	20.00%
SD tweet sentiment	344	.33	.33	.03	.23	.68
Tweet volume	344	24,524.01	10,401.50	40,324.40	2.00	441,480.00
Revenue (in millions of $)	344	$ 40,480.62	$ 30,967.50	$ 34,375.43	$ 658.00	$ 201,159.00
Revenue growth	344	4.37%	4.02%	12.07%	-52.01%	59.76%
Gross margin	273	36.02%	31.06%	16.07%	12.56%	84.04%
Operating margin	273	9.64%	7.70%	9.94%	-18.66%	50.37%
EBT margin	344	9.24%	7.29%	10.61%	-18.10%	40.66%
Cash flow (in millions of $)	344	$ 5,140.94	$ 1,361.50	$ 12,992.23	$ (13,858.00)	$ 107,953.00
ROA	343	5.42	5.10	8.41	-29.64	35.79
Return on equity	317	24.26	14.53	75.02	-426.90	814.93
Market share	344	25.58%	24.17%	21.40%	.58%	100.00%
Market value (in millions of $)	329	$ 61,653.41	$ 27,943.24	$ 96,454.89	$ 50.70	$ 741,593.62
Trading volume (in millions)	329	2584.01	1253.15	5711.53	.00	53645.52
Times traded	329	2.93	2.17	2.71	.00	23.50
Stock return	328	21.79%	18.05%	40.20%	-76.09%	272.06%

Note: ACSI = American Customer Satisfaction Index, , S.D. = Standard Deviation

TABLE 4

Correlation Table (n = 317-344)

	ACSI	Net Tweet Sentiment	Share Pos. Sentiment	Share Neg. Sentiment	SD Tweet Sentiment	Tweet Volume
ACSI	1.000					
Net tweet sentiment	.250****	1.000				
Share pos. sentiment	.131**	.905****	1.000			
Share neg. sentiment	-323****	-.539****	-.128**	1.000		
SD tweet sentiment	-.135**	.344****	.645****	.473****	1.000	
Tweet volume	-.175***	-.089*	-.093*	.023	-.084	1.000
Revenue	-.154***	-.276****	-.136**	.375****	.169***	-.090*
Revenue growth	.081	.008	-.012	-.043	.015	.007
Gross margin	.115***	-.008	-.037	-.050	-.154***	.086
Operating margin	-.085	-.067	-.062	.055	-.042	.322****
EBT margin	-.051	-.185****	-.123**	.189****	.021	.154***
Cash flow	-.174***	-.053	.109**	.338****	.298****	-.114**
ROA	.056	-.013	-.021	-.013	-.020	.198****
Return on equity	-.009	-.127**	-.091	.114**	-.004	.106**
Market share	-.118**	-.319****	-.229****	.289****	.048	-.028
Market value	.034	-.186****	-.094*	.247****	.083	.075
Trading volume	-.266****	.036	.228***	.373****	.378****	-.093*
Times traded	-.120**	.140**	.133**	-.063	.081	-.030
Stock return	-.061	.032	.003	-.068	-.042	.044

Notes: ACSI = American Customer Satisfaction Index, Pos. = positive, Neg. = negative, S.D. = Standard Deviation.

****p < .001, ***p < .01, **p < .05, *p < .10.

TABLE 5
Model Estimates with Predictors as Dependent Variable (n = 344)

	ACSI t + 1	ACSI t + 1	Pos. Sent. T + 1	Pos. Sent. T + 1	Neg. Sent. T + 1	Neg. Sent. T + 1	Tweet Vol. T + 1	Tweet Vol. T + 1
Intercept	79.826****	16.673****	-.032	.050*	.123****	.052****	10.526****	3.391****
ACSI		.793****	.001****	.000	-.001****	-.000**	-.014	-.007
Pos. sent.	7.151	-4.181		.225****		-.073****	-.207	1.146
Neg. sent.	-85.158****	-11.906		-.554****		.470****	-4.211	.156
Tweet vol.	.038	.023	.001	-.000	.000	.000		.712****
R^2	.128	.729	.064	.281	.110	.409	.004	.731
Adjusted R^2	.119	.725	.057	.272	.104	.401	-.006	.728
Incl. lag DV		√		√		√		√

Notes: ACSI = American Customer Satisfaction Index, Pos. = positive, Neg. = negative, sent. = sentiment of tweets, vol. = volume.

****$p < .001$, ***$p < .01$, **$p < .05$, *$p < .10$.

TABLE 6
Model Estimates with Firm Performance as Dependent Variable (n = 317-344)

	Revenue t + 1	**Revenue Growth T + 1**	**Gross Margin T + 1**	**Operating Margin T + 1**	**EBT Margin T + 1**
ACSI	0.001 (0.01%)	0.107 (0.00%)	0.197 (0.00%)	-0.039 (0.00%)	-0.082* (0.00%)
Δ ACSI	-0.004 (0.00%)	-0.342 (0.87%)	0.065 (0.24%)	-0.208* (0.87%)	0.049 (0.12%)
Relative ACSI	0.002 (0.02%)	0.171 (0.00%)	-0.254 (0.00%)	0.005 (0.00%)	-0.015 (0.00%)
Net sentiment	-0.088 (1.14%)	-9.647 (0.00%)	0.250 (0.00%)	-3.416 (0.00%)	-12.834** (0.00%)
Δ net sentiment	0.056 (0.00%)	5.275 (16.53%)	-0.505 (12.96%)	-5.721 (14.23%)	-0.335 (5.87%)
Rel. net sentiment	0.055 (1.21%)	0.314 (0.00%)	-1.084 (0.00%)	0.571 (0.00%)	-14.299** (0.00%)
Positive sentiment	-0.253* (0.00%)	-26.241* (0.00%)	-1.616 (0.00%)	-6.265 (0.00%)	-14.982** (0.00%)
Δ positive sentiment	0.033 (0.00%)	3.830 (16.36%)	-0.073 (15.21%)	-8.088 (19.12%)	-4.392 (7.80%)
Rel. positive sentiment	0.001 (1.15%)	-4.207 (0.00%)	6.044 (0.00%)	2.434 (0.00%)	-16.840** (0.00%)
Negative sentiment	**-0.915** (60.31%)**	-93.957*** (0.00%)	-9.673 (0.00%)	-0.169 (0.00%)	4.890 (0.00%)
Δ negative sentiment	-0.182 (0.00%)	**-12.135 (42.90%)**	**5.470 (46.16%)**	**-1.145 (35.66%)**	**-26.709* (66.44%)**
Rel. neg. sentiment	-0.570 (6.13%)	-47.856 (0.00%)	47.275 (0.00%)	9.839 (0.00%)	-12.621 (0.00%)
SD sentiment	-0.505** (27.58%)	-51.667*** (0.00%)	1.059 (0.00%)	-7.833 (0.00%)	-9.252 (0.00%)
Δ SD sentiment	-0.104 (0.00%)	-6.742 (22.92%)	-6.704 (24.83%)	-12.592 (29.88%)	-10.805 (19.18%)
Rel. SD sentiment	-0.147 (1.93%)	-18.495 (0.00%)	10.280 (0.00%)	0.737 (0.00%)	-17.541* (0.00%)
Tweet volume	0.012** (0.38%)	1.091** (0.00%)	0.170 (0.00%)	0.197 (0.00%)	0.056 (0.00%)
Δ tweet volume	-0.001 (0.00%)	-0.073 (0.42%)	0.186 (0.59%)	0.032 (0.25%)	-0.259 (0.58%)
Rel. tweet volume	0.012 (0.13%)	1.074 (0.00%)	0.991 (0.00%)	0.135 (0.00%)	-0.071 (0.00%)
Year dummy included	√	√	√	√	√
Lag of DV included	√	√	√	√	√

(Continued on the next page)

	Cash Flow T + 1	ROA T + 1	Return on Equity T + 1	Market Share T + 1	Market Value T + 1
ACSI	-575*** (0.00%)	-0.005 (0.00%)	1.489* (0.00%)	0.000 (0.00%)	0.016 (0.00%)
Δ ACSI	-376 (0.00%)	-0.198 (0.50%)	-4.129** (4.75%)	-0.001 (0.01%)	0.000 (0.26%)
Relative ACSI	39 (0.00%)	-0.002 (0.00%)	1.919* (0.00%)	-0.001 (0.00%)	-0.016 (0.00%)
Net sentiment	45726**** (0.00%)	0.378 (0.00%)	-25.298 (0.00%)	0.051 (0.00%)	-0.734 (0.00%)
Δ net sentiment	46850**** (36.67%)	2.112 (15.50%)	9.630 (15.67%)	0.018 (0.53%)	0.021 (13.99%)
Rel. net sentiment	49507*** (0.00%)	-5.911 (0.00%)	-19.195 (0.00%)	-0.018 (0.00%)	-1.365* (0.00%)
Positive sentiment	59066**** (0.00%)	4.462 (0.00%)	-6.366 (0.00%)	0.040 (0.00%)	-0.930 (0.00%)
Δ positive sentiment	**50178**** (58.47%)**	2.011 (19.22%)	8.656 (15.96%)	0.079 (1.56%)	-0.051 (14.58%)
Rel. positive sentiment	52351*** (0.00%)	-2.749 (0.00%)	-4.329 (0.00%)	0.005 (0.00%)	-1.213* (0.00%)
Negative sentiment	68822* (0.00%)	10.510 (0.00%)	110.130 (0.00%)	-0.063 (0.00%)	-1.472 (0.00%)
Δ negative sentiment	-484 (0.22%)	**-4.731 (36.99%)**	**-9.137 (41.25%)**	**0.393*** (89.12%)**	**-0.496 (41.38%)**
Rel. neg. sentiment	22631 (0.00%)	35.330 (0.00%)	151.392 (0.00%)	0.257 (0.00%)	0.785 (0.000%)
SD sentiment	82480**** (0.00%)	15.686 (0.00%)	19.393 (0.00%)	0.135 (0.00%)	-1.743** (0.00%)
Δ SD sentiment	50882*** (4.63%)	-9.363 (27.16%)	19.196 (21.92%)	0.162** (8.76%)	-0.539 (29.36%)
Rel. SD sentiment	56328** (0.00%)	19.642 (0.00%)	41.866 (0.00%)	0.047 (0.00%)	-1.537 (0.00%)
Tweet volume	-1443** (0.00%)	0.312 (0.00%)	-1.082 (0.00%)	0.002 (0.00%)	0.027 (0.00%)
Δ tweet volume	-175 (0.00%)	-0.256 (0.63%)	-0.603 (0.45%)	0.000 (0.01%)	0.005 (0.43%)
Rel. tweet volume	-420 (0.00%)	-0.036 (0.00%)	-3.118 (0.00%)	0.001 (0.00%)	0.053 (0.00%)
Year dummy included	√	√	√	√	√
Lag of DV included	√	√	√	√	√

(Continued on the next page)

	Trading Volume T + 1	Times Traded T + 1	Stock Return T + 1	Total Significant	Mean Akaike Weight
ACSI	-0.003 (0.00%)	0.005* (0.06%)	-0.009** (0.00%)	5/13	0.01%
Δ ACSI	0.016 (0.39%)	0.011* (0.00%)	0.001 (0.14%)	3/13	0.63%
Relative ACSI	-0.001 (0.00%)	0.007* (0.07%)	-0.007 (0.00%)	2/13	0.01%
Net sentiment	-0.149 (0.00%)	0.658** (20.79%)	-1.004** (0.00%)	4/13	1.69%
Δ net sentiment	-0.169 (14.84%)	0.214 (0.00%)	-0.588 (13.75%)	1/13	12.35%
Rel. net sentiment	-0.400 (0.00%)	0.440 (2.80%)	-0.932 (0.00%)	3/13	0.31%
Positive sentiment	-0.439 (0.00%)	0.459 (3.39%)	-1.178** (0.00%)	5/13	0.26%
Δ positive sentiment	-0.412 (16.16%)	0.098 (0.00%)	-0.773 (21.46%)	1/13	15.84%
Rel. positive sentiment	-0.477 (0.00%)	0.336 (1.99%)	-0.895 (0.00%)	3/13	0.24%
Negative sentiment	-1.165 (0.00%)	**-1.577*** (59.47%)**	0.414 (0.00%)	4/13	9.21%
Δ negative sentiment	**-1.585 (45.88%)**	-0.864 (0.00%)	-1.042 (25.53%)	2/13	**36.27%**
Rel. neg. sentiment	-0.480 (0.00%)	-1.224 (7.33%)	0.903 (0.00%)	0/13	1.04%
SD sentiment	-0.781 (0.00%)	-0.342 (2.07%)	-1.080 (0.00%)	4/13	2.28%
Δ SD sentiment	-0.599 (22.33%)	0.122 (0.00%)	**-1.179 (38.89%)**	2/13	19.22%
Rel. SD sentiment	-1.094 (0.00%)	-0.069 (1.86%)	-0.660 (0.00%)	2/13	0.29%
Tweet volume	-0.042 (0.00%)	-0.008 (0.05%)	0.010 (0.00%)	3/13	0.03%
Δ tweet volume	0.000 (0.40%)	-0.010 (0.00%)	-0.008 (0.23%)	0/13	0.31%
Rel. tweet volume	0.030 (0.00%)	-0.022 (0.13%)	-0.008 (0.00%)	0/13	0.02%
Year dummy included	√	√	√		
Lag of DV included	√	√	√		

Notes: Each estimate is based on a separate regression model as shown in Equation (2). The parameter is provided, with the Akaike weights shown in parentheses. ACSI = American Customer Satisfaction Index, Δ = growth in measure over the previous year, Rel. = relative measure compared with industry average in year t., S.D. = Standard Deviation, The best predictor based on the Akaike weight is shown in boldface.

****$p < .001$, ***$p < .01$, **$p < .05$, *$p < .10$.

TABLE 7
Combination of Predictors

	Revenue t + 1	Gross Margin t + 1	Cash Flow t + 1	ROA t + 1
ACSI		0.648****	-842****	
Δ ACSI				-0.191
Relative ACSI		-0.848****	725**	
Positive sentiment	-0.920***		500051****	
Δ positive sentiment				
Rel. positive sentiment	1.011***			
Negative sentiment	-0.792**			
Δ negative sentiment				
Rel. neg. sentiment				
SD sentiment			131937***	
Δ SD sentiment				
Rel. SD sentiment			-197416***	
Tweet volume	0.011**			0.570**
Δ tweet volume				
Rel. tweet volume				
Year dummies included	√	√		
Lag of DV included	√	√		√

Notes: ACSI = American Customer Satisfaction Index, Δ = growth in measure over the previous year, Rel. = relative measure compared to industry average in year t., S.D. = Standard Deviation

****$p < .001$, ***$p < .01$, **$p < .05$, *$p < .10$.

TABLE 8
Hypotheses Testing Models

	Revenue t + 1	Revenue Growth T + 1	Gross Margin T + 1	Operating Margin T + 1	EBT Margin T + 1
Intercept	-0.030 (0.092)	1.278 (1.401)	13.793**** (1.927)	1.065* (0.601)	1.464** (0.568)
NTS	-0.075 (0.280)	-7.903 (26.637)	-10.047 (16.836)	-4.903 (12.571)	-0.548 (11.843)
Tweet volume	0.006 (0.009)	0.494 (0.820)	1.641*** (0.595)	0.531 (0.344)	0.138 (0.309)
Year	-0.009** (0.004)	-0.777** (0.353)	0.062 (0.226)	0.087 (0.153)	0.050 (0.151)
Year × NTS	-0.041 (0.081)	-3.421 (7.847)	-5.953 (4.699)	-1.763 (3.723)	1.447 (3.574)
Year × NTS	-0.001 (0.002)	-0.099 (0.167)	0.319*** (0.109)	0.055 (0.082)	-0.010 (0.074)
NTS × Tweet vol.	0.068 (0.059)	5.553 (5.582)	4.594 (5.143)	3.326 (3.317)	2.230 (2.373)
Lag DV	1.004**** (0.009)	0.153*** (0.052)	0.580**** (0.041)	0.904**** (0.030)	0.916**** (0.026)

	Cash Flow t + 1	ROA t + 1	Return on Equity t + 1	Market Share t + 1	Market Value t + 1
Intercept	6275*** (2034)	2.486*** (0.814)	8.813 (8.130)	0.131**** (0.020)	0.232* (0.121)
NTS	74281*** (26024)	-9.474 (16.712)	-149.768 (176.514)	-0.120 (0.133)	-2.263** (0.937)
Tweet volume	-614 (965)	0.936** (0.464)	1.778 (4.681)	0.006 (0.006)	-0.025 (0.023)
Year	80 (367)	0.096 (0.212)	3.604 (2.288)	-0.002 (0.002)	-0.024** (0.012)
Year × NTS	14318* (7546)	-4.905 (4.974)	-40.621 (52.872)	-0.030 (0.038)	-0.378 (0.277)
Year × NTS	36 (167)	0.144 (0.106)	0.765 (1.094)	0.001 (0.001)	-0.015*** (0.006)
NTS × Tweet vol.	-15205*** (5732)	3.235 (4.721)	12.806 (34.934)	-0.002 (0.029)	0.256 (0.180)
Lag DV	-0.120* (0.062)	0.564**** (0.052)	1.232**** (0.121)	0.444**** (0.039)	0.983**** (0.011)

(Continued on the next page)

	Trading Volume T + 1	Times Traded T + 1	Stock Return T + 1	Total Significant	Hypothesis
Intercept	1.776**** (0.298)	0.122*** (0.037)	0.178**** (0.049)	10/13 (10 pos.)	
NTS	1.277 (2.092)	1.542** (0.59)	-1.189 (1.098)	3/13 (2 pos.)	
Tweet volume	0.012 (0.053)	0.003 (0.015)	-0.040 (0.028)	2/13 (2 pos.)	
Year	0.032 (0.027)	0.016** (0.008)	-0.026* (0.014)	4/13 (3 pos.)	
Year × NTS	0.568 (0.630)	0.297* (0.178)	-0.092 (0.332)	2/13 (2 pos.)	**H1a**
Year × NTS	0.018 (0.013)	0.004 (0.004)	-0.018*** (0.007)	3/13 (1 pos.)	**H1b**
NTS × Tweet vol.	-0.382 (0.415)	-0.113 (0.116)	0.079 (0.218)	1/13 (0 pos.)	**H2**
Lag DV	0.916**** (0.014)	0.917**** (0.025)	-0.005 (0.064)	12/13 (11 pos.)	

Notes: NTS = net tweet sentiment, DV = dependent variable, pos. = number of positive and significant effects. Standard errors are in parentheses.

****$p < .001$, ***$p < .01$, **$p < .05$, *$p < .10$.