

# Earnings Conference Call Content and Stock Price: The Case of REITs

James S. Doran · David R. Peterson ·  
S. McKay Price

Published online: 18 August 2010  
© Springer Science+Business Media, LLC 2010

**Abstract** Using computer based content analysis, we quantify the linguistic tone of quarterly earnings conference calls for publicly traded Real Estate Investment Trusts (REITs). After controlling for the earnings announcement, we examine the relation between conference call tone and the contemporaneous stock price reaction. We find that the tone of the conference call dialogue has significant explanatory power for the abnormal returns at and immediately following quarterly earnings announcements. The question and answer portion of the conference calls dominates prepared managerial introductory remarks in explanatory significance. Furthermore, an overall positive tone in the conference call discussion between management and analysts is found to nearly offset the damaging effects of a negative earnings surprise.

**Keywords** REITs · Conference calls · Content analysis · Tone

## Introduction

Earnings conference calls have become an increasingly important medium through which management may disseminate information to investors. Kimbrough (2005) finds a reduction in the post-earnings-announcement drift for firms who initiate conference calls and contends that conference calls provide corporate managers with a forum through which they can emphasize specific aspects of recent performance and highlight their implications for future financial performance. Frankel et al. (1999) analyze stock volatility and trading volume levels during the conference calls and find that both are elevated. We dissect conference calls into their primary

---

J. S. Doran · D. R. Peterson  
Department of Finance, College of Business, Florida State University, 821 Academic Way,  
P.O. Box 3061110, Tallahassee, FL 32306–1110, USA

S. M. Price (✉)  
Perella Department of Finance, Lehigh University, 621 Taylor Street, Bethlehem, PA 18015, USA  
e-mail: smp210@lehigh.edu

components through linguistic analysis to better understand this voluntary disclosure mechanism.

Using computer based content analysis to quantify the linguistic tone of earnings conference calls, we analyze the information content of earnings conference calls for a sample of Real Estate Investment Trusts (REITs). We examine the conference call language using both the voluminous Harvard Psychosocial dictionary as in Tetlock (2007) and Tetlock et al. (2008), as well as a smaller custom dictionary with terms relevant to earnings announcements, conference calls, and finance, as advocated by Henry (2008) and Loughran and McDonald (2009). We look at whether conference call linguistic information becomes embedded into stock prices in the context of the accompanying earnings announcement and, also, whether detection of such relations is better accomplished using generic or context specific dictionaries.

The unique characteristics of REITs provide a natural setting in which to study the relation between conference call content and stock returns. REITs are constantly engaged in asset acquisition and disposition activities. As properties move in and out of REIT portfolios, their underlying revenue generating asset bases change substantially. REITs are also eager to realize appreciation profits as documented by Muhlhofer (2008).<sup>1</sup> However, transaction due diligence and negotiation efforts usually span a period of several months and the progress of such transactions is not reflected in quarterly earnings. The question and answer portion of earnings conference calls provides an ideal venue for information regarding transaction progress to come to light.

In this study we analyze the content of the quarterly earnings conference calls for all Real Estate Investment Trusts included in the FTSE NAREIT All REITs Index at any time between January 2004 and December 2007.<sup>2</sup> We do so using a hand collected sample of over 1,700 conference call transcripts spanning sixteen consecutive quarters in the post Regulation FD era.<sup>3</sup> This is the first study of which we are aware to utilize tone oriented content analysis to examine either REITs or the component of quarterly earnings conference calls that deals with the unscripted question and answer session between management and analysts.

Other factors besides the management tone of conference calls, however, may be simultaneously affecting stock returns. Thus, we take three steps to determine if management tone provides new information. First, we control for the magnitude of the actual earnings surprise, which has clearly been shown in prior research to influence returns. Second, we separate the conference call into two parts, the prepared remarks by management and the following question and answer session with analysts. Third, the tone of the components of the conference call may be

<sup>1</sup> REITs are required to hold individual properties for at least 4 years and can sell up to 10% of their asset bases in a given year. The intent of the rule is to keep REITs from buying property primarily for resale. Muhlhofer (2008) demonstrates that REITs would buy and sell assets more frequently if they were not subject to the holding period requirement. Still, there are no restrictions on the number of properties they can acquire in a given year. With new acquisitions and the potential to sell up to 10% of their portfolios, REITs are able to substantially alter their asset bases each year.

<sup>2</sup> By definition this includes all publicly traded REITs.

<sup>3</sup> Regulation Fair Disclosure, effective October 23, 2000, opened up conference calls to the public. The final SEC rule, widely known as "Regulation FD", is available at <http://www.sec.gov/rules/final/33-7881.htm>.

proxies for new information that are actually attributable to other factors. To account for this possibility, we add numerous control variables to our analysis.

Our results show that the overall tone of REIT earnings conference calls as a whole has significant explanatory power for the accompanying abnormal returns, even after controlling for the numerical value of earnings per share and associated announcement surprises. This is consistent with the behavior of typical equities as documented in studies that examine earnings announcements (Davis et al. 2007; Henry 2008; Sadique et al. 2008). More importantly, we find that after controlling for the both the earnings surprise and the tone of the earnings announcement, the tone of the conference call dialogue between management and analysts has even stronger explanatory significance. These results are robust to the inclusion of control variables.

Somewhat surprisingly, we find that a positive conference call tone nearly offsets the damaging effects of a negative earnings surprise, suggesting that the information revealed in the conference calls as reflected in the linguistic tone is nontrivial. In other words, conference call participants are not merely going through the motions. Rather, they are actively engaged to the extent that new and meaningful information comes to light. This supports the findings of Frankel et al. (1999) who suggest that conference calls provide information over and above the earnings announcement itself, and that more information about firms' future prospects becomes known during conference calls.

The results are robust to various constructions of the linguistic tone measure and to the type of dictionary used. However, the results are significantly stronger for measures which utilize the customized dictionary, providing out of sample support in a context which has been previously untested for the argument of Henry (2008) and Loughran and McDonald (2009) that custom dictionaries may be preferable. As such, this paper contributes to the new and growing content analysis literature in finance, accounting, and economics related fields, in addition to increasing our understanding of corporate disclosure in the unique context of the real estate investment trust tax structure.

This study is developed in the following sections. We review the literature and develop our hypotheses in the next section. The third section describes our sample. The fourth section outlines the empirical methods, including a discussion of how we quantify the tone of a conference call. Results are presented in the fifth section. The final section concludes.

## Literature Review

Starting with Beaver (1968), a long stream of literature has established that capital markets react to earnings news. Ball and Brown (1968) find evidence that stock prices continue to drift upward (downward) after initial positive (negative) earnings announcements, rendering the initial stock price reaction to the earnings news incomplete and raising questions of market efficiency. Fama (1998) points out that this post-earnings-announcement drift anomaly has survived over the years across samples and refinements in abnormal return estimation methods (Bernard and Thomas 1990; Chan et al. 1996).

Bernard (1992) shows that there is much discussion in the finance and accounting literature regarding potential explanations for the incomplete market reaction to unexpected earnings. One camp provides evidence that the drift may be due to underreaction (Bernard and Thomas 1989, 1990; Freeman and Tse 1989; Mendenhall 1991; Wiggins 1991). That is, following an earnings announcement the stock price fails to fully impound information about future firm potential. The drift then becomes the result of information slowly making its way into the stock price over time. Another camp suggests that stock prices seem to overreact to earnings surprises. DeBondt and Thaler (1987) assert that the overreaction theory implies that investors myopically place too much weight on recent information and fail to fully account for the future implications of the information. As a result, extreme stock price increases and decreases are predictive of subsequent earnings reversals due to the high correlation between stock returns and mean reverting company earnings (Brooks and Buckmaster 1976; DeBondt and Thaler 1985, 1987).<sup>4</sup> One commonality in all of these studies, however, is the exclusive focus on the numerical value of the earnings.

Abbarbanell and Bernard (1992) state that the literature on both underreaction and overreaction provides evidence suggesting that stock price behavior around earnings announcements may be caused by the failure of market participants to fully appreciate the information content of current earnings. This implies that we need to further our understanding of contemporaneous stock price reactions to the information content associated with earnings announcements. Such information is not limited to the numerical content of an announcement alone, but includes statements that typically accompany the announcement. An excellent example is the conference calls that usually immediately follow the release of earnings.

The National Investor Relations Institute (NIRI 2004) indicates that next to press releases, fully accessible webcast conference calls are the most widely used means for disseminating corporate information to the investment community.<sup>5</sup> Kimbrough (2005) contends that conference calls are a voluntary disclosure mechanism of increasing importance that provide corporate managers with a forum in which they can emphasize specific aspects of recent performance and highlight their implications for future financial performance. Following opening remarks by management, analysts are allowed to make comments and pose questions. Kimbrough notes that during this somewhat informal exchange, details not contained in the earnings press release are often disclosed.

Frankel et al. (1999) empirically examine conference calls as a voluntary disclosure medium by analyzing stock volatility and trading volume levels at the time of the conference calls. They find that return variances and trading volume are both elevated during conference calls, suggesting that conference calls provide information to the market beyond that which is found in the press release alone. In other words, the calls contain material information and investors trade on this information in real time.<sup>6</sup>

<sup>4</sup> See Bernard (1992) for a thorough review of possible explanations for the anomalous evidence for stock price reactions to earnings announcements.

<sup>5</sup> This is also noted in Frankel et al. (1999).

<sup>6</sup> Sunder (2002) and Irani (2004) find that this holds both before and after passage of Regulation Fair Disclosure.

Kimbrough (2005) examines the effect of conference calls on post-earnings-announcement drift in order to provide insight into whether conference calls improve the efficiency of the market reaction to earnings announcements. He finds that the initiation of conference calls is associated with a significant reduction in the post-earnings announcement drift. This reduction is concentrated in small firms, the companies for which prior literature has determined the drift to be most severe.<sup>7</sup>

All of these studies of conference calls add to our understanding of how earnings information and other information related to firm prospects affects stock prices. Recently, however, a technique used in other disciplines for roughly four decades, called content analysis, has been employed to provide insights into how specific wording in statements may affect stock prices.<sup>8</sup>

Content analysis enables quantification of text based information. A recent stream of literature utilizing tools relatively new to finance, but well established in a wide variety of other disciplines, allows a better understanding of the information content of words.<sup>9</sup> Tetlock (2007) is the first finance study to incorporate text based content analysis. In his effort to explore the link between media reports and the stock market, Tetlock captures the sentiment of the media by applying content analysis to the daily Wall Street Journal column, *Abreast of The Market*. He finds that high media pessimism predicts downward pressure on market prices and subsequent reversion. Engelberg (2008) applies the method outlined in Tetlock (2007) to measure the qualitative content of Dow Jones News Service stories about firms' earnings announcements to analyze the link between "soft" information and equity prices. He finds support for Tetlock (2007) and demonstrates a relation between linguistic media content and the stock market.

Davis et al. (2007), Henry (2008), and Sadique et al. (2008) analyze the information content of the wording of official earnings announcements. Henry (2008) quantifies tone using computer-based analytical tools to measure the frequency of positive and negative words found in earnings press releases. She finds that the tone of earnings press releases influences investors' reaction to earnings in a sample of technology firms. Davis et al. (2007) confirm Henry's findings across a broad sample of firms, with a significant positive association between levels of optimistic tone in earnings press releases and both future return-on-assets and the initial market response. They suggest that managers use tone in their press releases to provide investors with information about expected future firm performance.

Sadique et al. (2008) look at the impact of tone on volatility in addition to stock returns using earnings press releases and the financial news coverage of those releases with a sample of S&P 100 firms. They find that positive tone is directly related to returns, and negatively related to volatility. Taken together, Davis et al. (2007), Henry (2008), and Sadique et al. (2008) demonstrate that the information content of the words accompanying an earnings announcement say much regarding the outlook of the participants beyond that learned by scrutinizing the earnings in

<sup>7</sup> This is also noted in Fama (1998).

<sup>8</sup> Content analysis techniques, developed over 40 years ago, were originally applied in journalism, psychology, communications, and other social sciences.

<sup>9</sup> See Krippendorff (2004) for a complete history of content analysis, including its development and application across disciplines.

isolation. However, there is a much larger and broader information content in conference calls than in the more narrow earnings press release.

Studies have only recently begun to incorporate the richness of the linguistic information content of conference calls. In their investigation of investor relations costs, Frankel et al. (2010) examine the stock market reaction to conference call tone measures as a robustness test. They find a positive relation between linguistic tone and conference call returns. Mayew and Venkatachalam (2009) incorporate the linguistic content of conference calls as a control in their analysis of managerial affective states, utilizing audio files and vocal emotion analysis software, during earnings conference calls.

We add a new dimension to the discussion by analyzing the content of quarterly earnings conference calls for a sample of Real Estate Investment Trusts (REITs). The REIT market provides an ideal setting in which to examine the relation between conference call content and stock returns given the unique characteristics of the REIT structure. When REITs were established they were intended to be pass-through entities. Accordingly, REITs are required to pay out at least 90% of their taxable earnings in the form of dividends.<sup>10</sup> Danielsen et al. (2009) argue that the high payout provides an unusually strong incentive for REIT managers to be transparent, compared to non-REITs, since they are forced to turn to the capital markets repeatedly in order to capitalize on growth opportunities.<sup>11</sup>

Kallberg et al. (2003) note that the extensive dividend payout requirement for REITs removes much of their discretion over dividend payout, and should make the dividend pricing/present value relation more certain for REITs than for typical equities. Theoretically, the value of a stock is the present value of discounted future cash flows. As such, the high required payout renders the relation between REIT prices and earnings much more unambiguously tied to theory than it is for other equities.

Kallberg et al. (2003) find dividend pricing/present value models fit REITs well and show a wide disparity between the earnings payout for REITs and typical equities. They note that while REITs are required to meet the 90% minimum payout, typical equities paid out 35%, on average, of pretax profits from 1973 to 2000. The National Association of Real Estate Investment Trusts (NAREIT) data show that, as a whole, REITs paid out roughly \$16 billion in dividends in 2006 alone.<sup>12</sup> The NAREIT data further show that from 1986 through September 2008 the income component comprised two-thirds (7.5%) of the average annual total return (11.5%) for the industry.

Liu and Mei (1994) analyze the returns of both REITs and typical industrial equities. Using variance decomposition they show that earnings related cash flows are the primary driver of stock return variability for REITs and that this relation is significantly stronger than that which we observe for typical equities. Cash flow risk is the dominant risk for REIT returns.

<sup>10</sup> Internal Revenue Code Sect. 857(a)(1).

<sup>11</sup> A growing body of empirical research documents REIT transparency (Hardin et al. 2005; Devos et al. 2007; Hardin and Hill 2008; Blau et al. 2009; Dolvin and Pyles 2009). See Blau et al. (2009) for a review of this work.

<sup>12</sup> <http://www.reit.com/InstitutionalInvestors/Dividends/tabid/101/Default.aspx>

Given the pass through nature of the REIT structure and corresponding importance of REIT cash flows, changes in the revenue generating asset bases for REITs can have a meaningful impact on stock returns. Muhlhofer (2008) shows that REIT portfolios are constantly changing as managers eagerly work to realize appreciation profits. However, commercial real estate transactions are often heavily negotiated and require extensive due diligence efforts, taking upwards of several months to come to fruition. Seeing as the progress of, and future outlook for, asset acquisitions and dispositions cannot be determined by analyzing quarterly earnings figures, the question and answer portion of earnings conference calls provides an ideal venue for information regarding potential changes of the revenue generating asset base to come to light. This becomes particularly relevant given REIT structural incentives to be informationally transparent.

## Data

The National Association of Real Estate Investment Trusts (NAREIT) produces an industry performance index that tracks all of the tax qualified REITs which are publicly traded on one of the three major US Exchanges (NYSE, AMEX, or NASDAQ), called the FTSE NAREIT All REITs Index (NAREIT index). Our sample contains all stocks included in the calculation of the NAREIT index at any point during the post Regulation FD four-year period 2004–2007.<sup>13</sup> Prior to this time, conference call transcript availability is sporadic. Since earnings conference calls are usually held no earlier than a few weeks following the close of each quarter, calls held during the sample period discuss firm performance starting with the fourth quarter of 2003 and run through the third quarter of 2007.

While some firms remain in the sample throughout, many firms enter, exit, or both enter and exit the index during the sample period. Only conference calls for those firms that are included in the NAREIT index calculation at the time of the call are included in the sample. A total of 233 firms appear in the NAREIT index at some point during the sample period, although transcripts are not available for all firms.

Conference call transcripts are obtained from two sources, Fair Disclosure Wire and The American Intelligence Wire. There are 1,865 total calls for which conference call transcripts are available out of 2,955 potential firm quarters, making the hit rate 63%. Of course, the real percentage is probably higher since the denominator represents total potential calls when, in fact, conference calls are voluntary.

Call dates generally accompany the transcripts. In cases of ambiguity, however, accurate dates are confirmed through Thomson Streetevents. Table 1 provides detail on the relation between earnings announcements and conference call timing. As shown in Table 1, Panel A, 99% of the conference calls are held either on the same day as the earnings announcement (41%), or on the following trading day (58%). In

<sup>13</sup> Because the NAREIT index tracks all publicly traded tax qualified REITs, there is no index inclusion effect such as is seen when firms are included in, for example, the S&P 500 Index, which could lead to increased exposure with higher levels of analyst coverage.



**Table 1** Earnings announcement and conference call timing

	Number of Trading Days After Announcement	Number of Calls	Percent of Observations	Cumulative Percent
Panel A	0	714	40.68	40.68
	1	1,021	58.18	98.86
	2	14	0.80	99.66
	3	3	0.17	99.83
	4	1	0.06	99.89
	5	1	0.06	99.94
	7	1	0.06	100.00
	Total	1,755	100.00	
Panel B	Quarter	Number of Calls	Percent of Observations	Cumulative Percent
	2003 Q4	94	5.36	5.36
	2004 Q1	106	6.04	11.40
	2004 Q2	105	5.98	17.38
	2004 Q3	111	6.32	23.70
	2004 Q4	107	6.10	29.80
	2005 Q1	110	6.27	36.07
	2005 Q2	117	6.67	42.74
	2005 Q3	116	6.61	49.34
	2005 Q4	109	6.21	55.56
	2006 Q1	113	6.44	61.99
	2006 Q2	119	6.78	68.77
	2006 Q3	118	6.72	75.50
	2006 Q4	112	6.38	81.88
	2007 Q1	107	6.10	87.98
	2007 Q2	103	5.87	93.85
	2007 Q3	108	6.15	100.00
	Total	1,755	100.00	

Panel A shows the timing relation between earnings announcements and conference calls. Most of the calls take place on either the same day as the announcement or on the following day. Panel B illustrates the fairly even distribution of conference call observations by quarter across the sample period of sixteen quarters

only six cases are the conference calls held more than two trading days after the earnings announcement.

Returns for individual firms and the value-weighted market portfolio are obtained from CRSP. Earnings data comes from Compustat. The Dow Jones REIT Composite Index returns are downloaded from the Dow Jones Indexes website.<sup>14</sup> The Dow Jones REIT Composite (DJRC) Index is similar to the FTSE NAREIT All REITs

<sup>14</sup> <http://www.djindexes.com/>



Index in that it contains all the publicly traded tax qualified U.S. REITs. However, the DJRC index is readily available on a daily basis which is conducive for use in our study in the estimation of abnormal returns.<sup>15</sup>

A total of 1,755 conference calls have sufficient corresponding data on both CRSP and Compustat to be included in the final sample. At least 165 REITs have one or more calls that meet data sufficiency requirements. The final sample of earnings conference calls are well distributed across the sixteen quarters in the sample period, as shown in Table 1, Panel B.

## Empirical Methods

Most of content analysis involves word tabulation through the aid of computer software that applies predefined rules to categorize root words according to their meaning. Stone (1966) created the first computer program to aid in this effort, the General Inquirer (GI). The GI has been used extensively for decades across disciplines and is a powerful tool for quantifying text based information.

When analyzing text, the GI first runs disambiguation rules as part of the word recognition process and then applies a predefined dictionary in order to produce a tabulation of raw counts of words that fall within each dictionary category. The GI automatically applies the Harvard IV-4 Psychosocial Dictionary to the text, although custom dictionaries of one's own creation can be used as well. The Harvard dictionary contains 77 categories within which a word can be classified. We employ eight of these categories that are relevant to this study, including positive, negative, active, passive, strong, weak, overstatement, and understatement.

In recent work, Tetlock (2007) and Tetlock et al. (2008) utilize the GI and the Harvard IV-4 Psychosocial Dictionary to quantify the tone of written material in the media.<sup>16</sup> Other studies such as Davis et al. (2007), Henry (2008), and Sadique et al. (2008) quantify the tone of earnings announcements using the Diction software package and a custom dictionary as found in Henry (2008). The Henry dictionary is a simple list of "domain relevant" words that are categorized as either positive or negative.<sup>17</sup>

We use the GI for word recognition and apply both the Harvard and Henry dictionaries for comparison and robustness.<sup>18</sup> After categorization, all raw word

<sup>15</sup> Differences between the FTSE NAREIT All REITs Index and the Dow Jones REIT Composite Index are minimal. The NAREIT index is set to 100 as of December 31, 1971, while DJRC index pegs the index at 100 on the base date of December 31, 1991. Additionally, the exact date a firm is added to, or removed from, index calculation may vary slightly between the two. On a monthly basis, returns between the NAREIT and DJRC indices have a correlation coefficient of 0.9988 during the sample period.

<sup>16</sup> See Tetlock et al. (2008) for a fairly detailed listing of work involving the application of content analysis in finance.

<sup>17</sup> See the Appendix for further detail on content analysis, a description of the categories, and word lists.

<sup>18</sup> Loughran and McDonald (2009) create a customized word list dealing with financial terms as well. However, we use the Henry (2008) word list because we are dealing specifically with earnings and earnings announcement related terminology, consistent with Davis et al. (2007), Henry (2008), Sadique et al. (2008).

counts are divided by the total words in their respective conference calls to produce a relative measure ( $REL\_M_{i,j}$ ) for each category  $i$  and conference call  $j$ :

$$REL\_M_{i,j} = \frac{\text{Raw Word Count for Category}_{i,j}}{\text{Total Call Word Count}_j} \quad (1)$$

where the  $REL\_M_{i,j}$  categories are *Positive<sub>j</sub>*, *Negative<sub>j</sub>*, *Active<sub>j</sub>*, *Passive<sub>j</sub>*, *Strong<sub>j</sub>*, *Weak<sub>j</sub>*, *Overstated<sub>j</sub>*, and *Understated<sub>j</sub>*.

The eight relevant Harvard dictionary categories comprise four opposing pairs. Expressed in ratio form we can tell, for example, how positive a call is relative to how negative it is. To arrive at an overall numerical representation of conference call tone for each conference call  $j$  ( $TONE_{1,j}$ ), we take the first principal component of the four relevant category ratios:

$$TONE_{1,j} = \text{First Principal Component} \left\{ \frac{\text{Positive}_j}{\text{Negative}_j}, \frac{\text{Active}_j}{\text{Passive}_j}, \frac{\text{Strong}_j}{\text{Weak}_j}, \frac{\text{Overstated}_j}{\text{Understated}_j} \right\} \quad (2)$$

Taking the first principal component (the eigenvector with the largest eigenvalue) allows collapsing of the variation across the four ratios into one simple measure of overall conference call tone, consistent with Tetlock (2007). This measure is used in place of simply isolating the positive-to-negative ratio since investors will likely see words that fall within the active (passive), strong (weak), or overstated (understated) categories as “positive” (“negative”). Thus, we treat the  $TONE$  variable as a metric of relative positivity.

For robustness, we also construct the tone variable using an alternative form of the Harvard dictionary category ratios described above. First we take the difference in the opposing  $REL\_M_{i,j}$  categories and divide by the sum of the two. Specifically, for each conference call  $j$  we calculate the following four ratios using the  $REL\_M_{i,j}$  categories:

$$PN_j = \frac{\text{Positive}_j - \text{Negative}_j}{\text{Positive}_j + \text{Negative}_j} \quad (3)$$

$$AP_j = \frac{\text{Active}_j - \text{Passive}_j}{\text{Active}_j + \text{Passive}_j} \quad (4)$$

$$SW_j = \frac{\text{Strong}_j - \text{Weak}_j}{\text{Strong}_j + \text{Weak}_j} \quad (5)$$

$$OU_j = \frac{\text{Overstated}_j - \text{Understated}_j}{\text{Overstated}_j + \text{Understated}_j} \quad (6)$$

In this form the ratios are bounded between  $-1$  and  $1$ , consistent with the tone measures of Uang et al. (2006) and Henry (2008). For the second overall numerical representation of conference call tone for each conference call  $j$  ( $TONE_{2,j}$ ), we take the first principal component of the four relevant category ratios:

$$TONE_{2,j} = \text{First Principal Component} \{PN_j, AP_j, SW_j, OU_j\} \quad (7)$$

where  $TONE_{2,j}$  is not bounded between  $-1$  and  $1$ .

For the Henry measure for each conference call  $j$  ( $H-TONE_j$ ), we express the  $REL\_M_{i,j}$  categories  $Positive_j$  and  $Negative_j$ , as defined in the Henry dictionary, in the two ratio forms described previously:

$$H - TONE_{1,j} = \frac{Positive_j}{Negative_j} \quad (8)$$

and

$$H - TONE_{2,j} = \frac{Positive_j - Negative_j}{Positive_j + Negative_j} \quad (9)$$

where  $H-TONE_{2,j}$  is bounded between  $-1$  and  $1$ .

To isolate the stock price reaction to the earnings announcements and corresponding conference calls, we calculate cumulative abnormal returns, where daily abnormal returns are estimated using two different benchmark models. We use returns on the Dow Jones REIT Composite Index (DJRC) as the explanatory variable in the first model, and both the DJRC Index returns and the CRSP Value-Weighted returns as explanatory variables in the second model.

We define the conference call date as day  $t=0$ . Using daily returns data from days  $t=-100$  to  $t=-11$ , we estimate intercept and slope coefficients from the regression equation:

$$R_{j,t} = \hat{\alpha}_j + \hat{\beta}_j R_{djrc,t} + \varepsilon_{j,t} \quad (10)$$

and

$$R_{j,t} = \hat{\alpha}_j + \hat{\beta}_{1,j} R_{djrc,t} + \hat{\beta}_{2,j} R_{mkt,t} + \varepsilon_{j,t} \quad (11)$$

where  $R_{j,t}$  is the return for the underlying stock associated with conference call  $j$  on day  $t$ ,  $R_{djrc,t}$  is the return on the DJRC Index on day  $t$ , and  $R_{mkt,t}$  is the return on the CRSP Value-Weighted Index on day  $t$ . Abnormal returns ( $AR\_DJRC_{j,t}$ ,  $AR\_BOTH_{j,t}$ ) for firm  $j$  on day  $t$ , from days  $t=-10$  to  $t=10$ , are then calculated using the parameter estimates from the two benchmark models:

$$AR\_DJRC_{j,t} = R_{j,t} - \hat{\alpha}_j - \hat{\beta}_j R_{djrc,t} \quad (12)$$

$$AR\_BOTH_{j,t} = R_{j,t} - \hat{\alpha}_j - \hat{\beta}_{1,j} R_{djrc,t} - \hat{\beta}_{2,j} R_{mkt,t} \quad (13)$$

where the estimated parameters are obtained from Eqs. 10 and 11.

We cumulate abnormal returns over two different windows, days zero through one and days two through ten. We analyze the effect of conference call tone on the contemporaneous stock price reaction using CARs from the initial two-day period, while the subsequent nine-day CARs allow us to check for short-term persistence. Thus, we use four cumulative abnormal returns measures for each stock associated with conference call  $j$  ( $CAR(0,1)DJRC_j$ ,  $CAR(0,1)BOTH_j$ ,  $CAR(2,10)DJRC_j$ , and  $CAR(2,10)BOTH_j$ ), where:

$$CAR(0,1)DJRC_j = \sum_{t=0}^1 AR\_DJRC_{j,t} \quad (14)$$

$$CAR(0, 1)BOTH_j = \sum_{t=0}^1 AR\_BOTH_{j,t} \quad (15)$$

$$CAR(2, 10)DJRC_j = \sum_{t=2}^{10} AR\_DJRC_{j,t} \quad (16)$$

$$CAR(2, 10)BOTH_j = \sum_{t=2}^{10} AR\_BOTH_{j,t} \quad (17)$$

After developing the cumulative abnormal returns measures, we sort the sample into tone quintiles and check for differences in CARs between the high and low tone portfolios. We compare the CAR means and medians of the high and low tone quintiles using t-tests and Wilcoxon rank-sum tests, respectively. The comparison is done for both tone measures and all four series of estimated CARs. To further illustrate the differences across tone portfolios, we plot the cumulative abnormal returns from 10 days prior to the conference call through 10 days after the conference call for firms in the high  $TONE_2$  portfolio, firms in the low  $TONE_2$  portfolio, and an average of all firms in the sample across the  $TONE_2$  quintiles.

We examine the separate effects of tone and earnings on returns in two ways. Our first analysis is at the firm level. To determine the effect of conference call tone on stock price reaction, we separately regress the CARs on the different tone measures, controlling for the unexpected earnings from the announcement, by running cross-sectional regressions of the following form:

$$CAR_j = \gamma_0 + \gamma_1 SURP_j + \gamma_2 TONE_{i,j} + \varepsilon_j \quad (18)$$

where  $CAR_j$  represent the cumulative abnormal return measures for conference call  $j$  defined above;  $TONE_{i,j}$  is interchangeable with  $H-TONE_{i,j}$ , both as defined above and where  $i=1$  or  $2$ ; and  $SURP_j$  is the earnings surprise calculated using the difference between the earnings-per-share (EPS) and the EPS lagged four quarters, scaled by the stock price at the close of the lagged quarter:<sup>19</sup>

$$SURP_j = \frac{(EPS_j - EPS_{j,q-4})}{Stock\ Price_{j,q-4}} \quad (19)$$

We use EPS rather than funds from operations (FFO) in calculating  $SURP_j$  because the literature has been unable to establish FFO as conveying more useful information to investors over EPS (Fields et al. 1998; Gore and Stott 1998; Vincent 1999). Most notably, Fields et al. (1998) find that the superiority of one measure

<sup>19</sup> We run our analysis using an analyst EPS forecast based surprise measure as well. While results are unchanged, incorporating analyst forecasts causes substantial sample attrition which can potentially lead to a large firm bias. Furthermore, Ljungqvist et al. (2009) show analyst data sources to be unreliable and Graham et al. (2005) find that 85.1% of CFO's consider earnings in the same quarter of the prior year to be the most important earnings benchmark. Accordingly, analyst forecast based measure results are untabulated.

over the other is highly contextual and that net income (EPS) explains more variation in current stock price than does FFO.<sup>20</sup>

Regression standard errors are corrected for potential heteroscedasticity following White (1980). However, since White's standard errors are not robust to dependence within observations of a firm across time, or within a quarter across firms, we cluster by firm (165) and by quarter (16) to adjust for potential firm- and time- effects, respectively, using the methodology outlined in Petersen (2009).

We expect the coefficient on the earnings surprise control variable to be positive and significant during the contemporaneous reaction period ( $t=0$  to  $t=1$ ), consistent with prior studies (Davis et al. 2007; Henry 2008; Sadique et al. 2008). Intuitively, we anticipate the unexpected earnings to account for much of the abnormal returns at the time of the conference call. We expect a positive coefficient on the tone measures, indicating a direct relation between the CARs and tone. The efficient markets hypothesis suggests that the full impact of both the earnings surprise and the tone measures should be completely incorporated into the initial response. However, in light of the post-earnings-announcement drift literature (Bernard and Thomas 1990; Chan et al. 1996), the coefficients on both the earnings surprise measure and the tone measures may be significant over the subsequent nine-day horizon ( $t=2$  to  $t=10$ ).

Our second analysis is at the portfolio level. To further determine the extent of the relation between abnormal returns and conference call tone, we conduct two-way sequential sorts on earnings surprise and tone across all four CAR representations and both tone measures. We first sort into terciles on earnings surprise and subsequently sort by tone to form nine portfolios that contain mean CARs. The mean differences between the high and low portfolios are compared using *t*-tests.

## Results

Table 2 provides descriptive statistics (Panel A) and correlations (Panel B) for the CARs, tone measures, and the earnings surprise. There is very little difference between the CARs estimated using the two different benchmark models, indicating that the DJRC Index is a reasonable benchmark for REIT stocks. The descriptive statistics for the two-day CARs in particular are nearly identical, ranging from  $-.3980$  and  $-.3990$  to  $.3354$  and  $.3445$  for the DJRC model and the model that includes both the DJRC and the CRSP Value Weighted Index, respectively. The median in both cases is  $-.0007$ .

With a median earnings surprise of  $0.000$ , the sample appears evenly split between the realization of positive and negative unexpected earnings, although there are some negative surprises of slightly greater magnitude (in absolute value) than the most extreme positive surprises. The tone measures indicate that the overall tone of the conference calls in the sample is positive. The mean  $H\text{-TONE}_1$  level of  $4.2$  shows that earnings conference call dialogue is on average more than four times more positive than negative. It is not surprising that the general tone of conference calls is positive, reflecting executive effort to present their case in an upbeat manner and their discretionary nature. What is important here is that we find substantial

<sup>20</sup> Although Downs and Guner (2006) find that analyst REIT FFO forecast quality is higher than for EPS, Baik et al. (2008) show that the usefulness of FFO is dependent on its reconciliation to GAAP EPS.

**Table 2** Descriptive statistics and correlations

	CAR (0,1) DJRC	CAR (0,1) BOTH	CAR (2,10) DJRC	CAR (2,10) BOTH	SURP	TONE <sub>1</sub>	TONE <sub>2</sub>	H-TONE <sub>1</sub>	H-TONE <sub>2</sub>
<b>Panel A</b>									
Mean	-0.0023	-0.0022	0.0013	0.0015	-0.0023	0.00	0.00	4.16	0.58
Min	-0.3980	-0.3990	-0.5677	-0.5408	-1.2241	-3.24	-5.88	1.09	0.04
P25	-0.0161	-0.0155	-0.0202	-0.0209	-0.0034	-0.92	-0.88	3.03	0.50
P50	-0.0007	-0.0007	-0.0007	-0.0012	0.0000	-0.16	0.05	3.87	0.59
P75	0.0125	0.0121	0.0188	0.0200	0.0032	0.76	0.93	4.88	0.66
Max	0.3354	0.3445	0.7667	0.4866	0.4234	7.42	4.42	24.50	0.92
Std.Dev.	0.0390	0.0393	0.0533	0.0509	0.0535	1.34	1.35	1.72	0.13
N	1755	1755	1755	1755	1755	1755	1755	1755	1755
<b>Panel B</b>									
CAR (0,1) DJRC	1								
CAR (0,1) BOTH	0.9940	1							
CAR (2,10) DJRC	0.0297	0.0324	1						
CAR (2,10) BOTH	0.0825	0.0887	0.9581	1					
SURP	0.1676	0.1686	-0.2609	-0.1526	1				
TONE <sub>1</sub>	0.0776	0.0742	0.0090	0.0168	0.0518	1			
TONE <sub>2</sub>	0.0872	0.0836	-0.0026	0.0055	0.0724	0.9617	1		
H-TONE <sub>1</sub>	0.0923	0.0870	-0.0457	-0.0396	0.0874	0.4954	0.4896	1	
H-TONE <sub>2</sub>	0.1237	0.1153	-0.0682	-0.0674	0.1253	0.4876	0.5258	0.8825	1

This table provides descriptive statistics (Panel A) and correlations (Panel B) for the full sample of 1,755 conference calls. CAR(0,1)DJRC is the two-day cumulative abnormal return, starting the day of the conference call, where the abnormal returns are estimated using a benchmark model with returns on the Dow Jones REIT Composite Index as the explanatory variable. CAR(0,1)BOTH is the two-day cumulative abnormal return, starting the day of the conference call, where the abnormal returns are estimated using a benchmark model with both the Dow Jones REIT Composite Index returns and the CRSP Value Weighted returns included as explanatory variables. CAR(2,10)DJRC is calculated the same way as CAR(0,1)DJRC, only it is cumulated from days two through ten. CAR(2,10)BOTH is calculated in the same manner as CAR(0,1)BOTH, only it is cumulated from days two through ten. SURP is the earnings surprise calculated as  $\{[\text{EPS}(\text{qtr}) - \text{EPS}(\text{qtr}-4)] / \text{Stock Price (end of qtr-4)}\}$ . TONE<sub>1</sub> is the first principal component of the four ratios Positive/Negative, Strong/Weak, Active/Passive, and Overstated/Understated and TONE<sub>2</sub> is the first principal component of the four ratios (Positive-Negative)/(Positive+Negative), (Strong-Weak)/(Strong+Weak), (Active-Passive)/(Active+Passive), and (Overstated-Understated)/(Overstated+Understated), where each category reflects the proportion of words in a given call as defined by the Harvard IV-4 Psychosocial Dictionary. H-TONE<sub>1</sub> is the ratio of Positive words to Negative words and H-TONE<sub>2</sub> is the ratio (Positive-Negative)/(Positive+Negative) where each category reflects the proportion of words in a given call as defined by Henry (2008)

dispersion in the tone measures. The H-TONE<sub>2</sub> measure, which by construction is bounded between  $-1$  and  $1$ , has a minimum value of  $0.04$  and a maximum of  $0.92$ . While the minimum value is near the point of tone neutrality, the highest observation at the other end of the distribution represents a conference call that is almost completely void of negative words. Descriptive statistics for the TONE measures tell the same story; however, the numerical representations are slightly different since we use the first principal component of the four ratios. The first principal component, or first eigenvalue, is mean centered by construction. With a minimum value of  $-3.24$  and a maximum of  $7.42$ , we see the same kind of dispersion in the distribution with the TONE measures as with the H-TONE measures.<sup>21</sup>

In Table 2, Panel B, there is near perfect correlation between the CARs estimated using the two benchmark models over both two- and nine-day windows. The tone measures defined using the Harvard and Henry dictionaries are highly correlated as well, with coefficients ranging from  $0.49$  to  $0.53$ . We observe relatively low levels of correlation between the other variables although, interestingly, the earnings surprise and some of the tone measures are negatively correlated with the CARs calculated over the  $t=2$  to  $t=10$  window. The negative correlations are weaker for the tone measures than the earnings surprise measure when viewed relative to their respective positive correlation with CARs over days  $0$  and  $1$ .

Table 3 shows the differences in CARs when the sample is sorted into quintile portfolios based on the TONE<sub>1</sub> and H-TONE<sub>1</sub> measures. Mean and median CARs for each quintile are reported in separate sorts for both TONE<sub>1</sub> and H-TONE<sub>1</sub>. For both TONE<sub>1</sub> and H-TONE<sub>1</sub>, mean and median Q5 minus Q1 two-day CARs are positive and significant at the  $1\%$  level. The two-day CAR difference of roughly  $1\%$  ( $0.96\%$  for TONE<sub>1</sub> and  $1.33\%$  for H-TONE<sub>1</sub>) represents the abnormal return to a portfolio long stocks in the high tone quintile and short stocks in the low tone quintile. When sorted on H-TONE<sub>1</sub> the results are slightly stronger for both mean and median two-day CAR differences. Across the TONE<sub>1</sub> quintiles we find a monotonic ordering of the mean two-day CARs, with the highest returns associated with the most positive earnings conference calls. The ordering of means is nearly monotonic across the H-TONE<sub>1</sub> quintiles. Mean and median Q5 minus Q1 differences for the nine-day CARs are insignificant at the  $10\%$  level, except for one median with TONE<sub>1</sub>.

Table 4 is similar to Table 3, only in Table 4 the CARs are sorted into quintiles formed using the TONE<sub>2</sub> and H-TONE<sub>2</sub> measures. Results for TONE<sub>2</sub> are very similar to those for TONE<sub>1</sub> in Table 3. The H-TONE<sub>2</sub> results in Table 4 are identical to those for H-TONE<sub>1</sub> in Table 3, reflecting the fact that the sorting break points do not change for H-TONE based on how the ratio is constructed. Results in Tables 3 and 4 strongly support a positive relation between the tone of conference calls and the immediate, two-day stock price response. The results also suggest there is no subsequent relation between the tone and abnormal returns.

Figure 1 provides a graphical representation of the cumulative abnormal returns, in percent, from day  $t=-10$  through  $t=10$ . Cumulative abnormal returns for the highest and lowest TONE<sub>2</sub> quintile portfolios are plotted in addition to the mean

<sup>21</sup> Although not shown, the descriptive statistics for the various tone measures broken down by quarter show little variation through time.



**Table 3** Test of differences, means and medians, by TONE<sub>1</sub> and H-TONE<sub>1</sub> quintiles

TONE <sub>1</sub> Quintiles		CAR (0,1) DJRC	CAR (0,1) BOTH	CAR (2,10) DJRC	CAR (2,10) BOTH	H-TONE <sub>1</sub> Quintiles	CAR (0,1) DJRC	CAR (0,1) BOTH	CAR (2,10) DJRC	CAR (2,10) BOTH
1 (Low)	Mean	-0.0069	-0.0066	0.0020	0.0017	1 (Low)	-0.0101	-0.0093	0.0089	0.0084
	Median	-0.0028	-0.0021	-0.0013	-0.0033		-0.0038	-0.0040	-0.0009	-0.0016
2	Mean	-0.0046	-0.0046	-0.0011	-0.0003	2	-0.0018	-0.0018	0.0000	0.0004
	Median	-0.0013	-0.0019	-0.0011	-0.0019		-0.0021	-0.0021	-0.0014	-0.0015
3	Mean	-0.0022	-0.0022	0.0008	-0.0002	3	-0.0045	-0.0045	-0.0011	-0.0008
	Median	-0.0004	-0.0007	-0.0018	-0.0015		-0.0022	-0.0025	0.0002	-0.0011
4	Mean	-0.0004	-0.0002	0.0014	0.0023	4	0.0016	0.0016	-0.0019	-0.0015
	Median	-0.0006	-0.0003	0.0017	0.0003		0.0016	0.0020	-0.0021	-0.0028
5 (High)	Mean	0.0026	0.0026	0.0033	0.0041	5 (High)	0.0032	0.0032	0.0004	0.0010
	Median	0.0021	0.0021	0.0018	0.0005		0.0025	0.0028	0.0014	-0.0001
Mean Q5-Q1		0.0096***	0.0092***	0.0013	0.0023		0.0133***	0.0125***	-0.0084	-0.0075
t-statistic		(3.25)	(3.10)	(0.32)	(0.57)		(3.56)	(3.32)	(-1.56)	(-1.49)
Wilcoxon rank-sum test										
Median Q5-Q1		0.0049***	0.0043***	0.0031	0.0039*		0.0064***	0.0068***	0.0023	0.0015
z-statistic		(3.23)	(2.96)	(1.56)	(1.77)		(4.54)	(4.53)	(0.00)	(0.00)

This table shows the differences in CARs when sorted into TONE<sub>1</sub> and H-TONE<sub>1</sub> quintile portfolios. Test statistics (t and z) are in parentheses. CAR(0,1)DJRC is the two-day cumulative abnormal return, starting the day of the conference call, day 0, where the abnormal returns are estimated using a benchmark model with returns on the Dow Jones REIT Composite Index as the explanatory variable. CAR(0,1)BOTH is the two-day cumulative abnormal return, starting the day of the conference call, where the abnormal returns are estimated using a benchmark model with both the Dow Jones REIT Composite Index returns and the CRSP Value Weighted returns included as explanatory variables. CAR(2,10)DJRC is calculated the same way as CAR(0,1)DJRC, only it is cumulated from days two through ten. CAR(2,10)BOTH is calculated in the same manner as CAR(0,1)BOTH, only it is cumulated from days two through ten. TONE<sub>1</sub> is the first principal component of the four ratios Positive/Negative, Strong/Weak, Active/Passive, and Overstated/Understated, where each category reflects the proportion of words in a given call as defined by the Harvard IV-4 Psychosocial Dictionary. H-TONE<sub>1</sub> is the ratio of positive words to negative words as defined by Henry (2008). Q5-Q1 is the difference in CARs between quintile 5 and quintile 1.

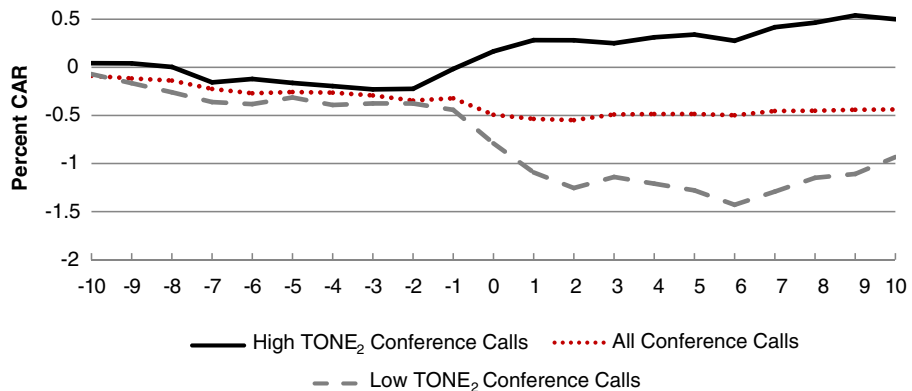
\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

**Table 4** Test of differences, means and medians, by TONE<sub>2</sub> and H-TONE<sub>2</sub> quintiles

TONE <sub>2</sub> Quintiles		CAR (0,1) DJRC	CAR (0,1) BOTH	CAR (2,10) DJRC	CAR (2,10) BOTH	H-TONE <sub>2</sub> Quintiles	CAR (0,1) DJRC	CAR (0,1) BOTH	CAR (2,10) DJRC	CAR (2,10) BOTH
1 (Low)	Mean	-0.0065	-0.0061	0.0012	0.0012	1 (Low)	-0.0101	-0.0093	0.0089	0.0084
	Median	-0.0034	-0.0026	-0.0013	-0.0030		-0.0038	-0.0040	-0.0009	-0.0016
2	Mean	-0.0066	-0.0066	0.0015	0.0007	2	-0.0018	-0.0018	0.0000	0.0004
	Median	-0.0018	-0.0023	-0.0017	-0.0020		-0.0021	-0.0021	-0.0014	-0.0015
3	Mean	-0.0010	-0.0009	-0.0003	0.0002	3	-0.0045	-0.0045	-0.0011	-0.0008
	Median	0.0010	0.0009	-0.0007	-0.0015		-0.0022	-0.0025	0.0002	-0.0011
4	Mean	0.0000	0.0000	0.0011	0.0020	4	0.0016	0.0016	-0.0019	-0.0015
	Median	-0.0013	-0.0006	0.0021	0.0016		0.0016	0.0020	-0.0021	-0.0028
5 (High)	Mean	0.0026	0.0026	0.0027	0.0035	5 (High)	0.0032	0.0032	0.0004	0.0010
	Median	0.0024	0.0023	0.0013	0.0003		0.0025	0.0028	0.0014	-0.0001
Mean Q5-Q1		0.0091***	0.0088***	0.0015	0.0023		0.0133***	0.0125***	-0.0084	-0.0075
t-statistic		(3.07)	(2.95)	(0.35)	(0.55)		(3.56)	(3.32)	(-1.56)	(-1.49)
Wilcoxon rank-sum test										
Median Q5-Q1		0.0058***	0.0049***	0.0027	0.0033		0.0064***	0.0068***	0.0023	0.0015
z-statistic		(3.43)	(3.24)	(1.39)	(1.56)		(4.54)	(4.53)	(0.00)	(0.00)

This table shows the differences in CARs when sorted into TONE<sub>2</sub> and H-TONE<sub>2</sub> quintile portfolios. Test statistics (t and z) are in parentheses. CAR(0,1)/DJRC is the two-day cumulative abnormal return, starting the day of the conference call, day 0, where the abnormal returns are estimated using a benchmark model with returns on the Dow Jones REIT Composite Index as the explanatory variable. CAR(0,1)/BOTH is the two-day cumulative abnormal return, starting the day of the conference call, where the abnormal returns are estimated using a benchmark model with both the Dow Jones REIT Composite Index returns and the CRSP Value Weighted returns included as explanatory variables. CAR(2,10)/DJRC is calculated the same way as CAR(0,1)/DJRC, only it is cumulated from days two through ten. CAR(2,10)/BOTH is calculated in the same manner as CAR(0,1)/BOTH, only it is cumulated from days two through ten. TONE<sub>2</sub> is the first principal component of the four ratios (Positive-Negative)/(Positive+Negative), (Strong-Weak)/(Strong+Weak), (Active-Passive)/(Active+Passive), and (Overstated-Understated)/(Overstated+Understated), where each category reflects the proportion of words in a given call as defined by the Harvard IV-4 Psychosocial Dictionary. H-TONE<sub>2</sub> is the ratio of (Positive-Negative)/(Positive+Negative) where each category reflects the proportion of words in a given call as defined by Henry (2008). Q5-Q1 is the difference in CARs between quintile 5 and quintile 1

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$



**Fig. 1** Cumulative abnormal returns across all conference calls and by high and low TONE<sub>2</sub> quintiles. Cumulative abnormal returns (CAR) are shown in percent, where the abnormal returns are estimated using a benchmark model with returns on the Dow Jones REIT Composite Index as the explanatory variable. High TONE<sub>2</sub> Conference Calls shows the cumulative abnormal returns for the highest TONE<sub>2</sub> quintile conference calls, where TONE<sub>2</sub> is the first principal component of the four ratios (Positive–Negative)/(Positive+Negative), (Strong–Weak)/(Strong+Weak), (Active–Passive)/(Active+Passive), and (Overstated–Understated)/(Overstated+Understated), where each category reflects the proportion of words in a given call as defined by the Harvard IV-4 Psychosocial Dictionary. Low TONE<sub>2</sub> Conference Calls shows the cumulative abnormal returns for the lowest TONE<sub>2</sub> quintile conference calls. All Conference Calls represents the cumulative abnormal returns for all conference calls in the sample. Day zero is the date of the earnings conference call

abnormal returns for all firms in the sample.<sup>22</sup> Figure 1 clearly shows that firms whose conference calls strike a highly positive tone have higher abnormal returns than the sample as a whole, while firms whose calls have a low TONE<sub>2</sub> measure experience much lower abnormal returns. The jump takes place contemporaneously (days  $t=0$  and  $t=1$ ) with abnormal returns for conference calls in the low TONE<sub>2</sub> portfolio partially reversing their initial drop over the next 9 days ( $t=2$  through  $t=10$ ). For the high TONE<sub>2</sub> portfolio there is only a slight upward response after day  $t=1$ .

Table 5 presents the firm-level results from regressing the CAR measures on tone measures and unexpected earnings. The explanatory tone measures in Panels A, B, C, and D are TONE<sub>1</sub>, H-TONE<sub>1</sub>, TONE<sub>2</sub>, and H-TONE<sub>2</sub>, respectively. As expected, the earnings surprise coefficients are highly significant across all four CAR measures in panels A through D. With two-day CARs as the dependent variables, the surprise coefficients are positive, but they turn negative in cases where we cumulate abnormal returns over the subsequent nine-day window. The tone variable coefficients are positive and highly significant when two-day CARs are the dependent variable, but are small and insignificant when the subsequent nine-day period is the dependent variable.<sup>23</sup>

<sup>22</sup> Again, results are identical for both H-TONE<sub>1</sub> and H-TONE<sub>2</sub> based abnormal return portfolios.

<sup>23</sup> Feldman et al. (2009) advocate using the change-in-tone from the most recent past rather than tone levels. They argue that change-in-tone helps mitigate potential problems associated with recurring word usage which may be unique to a particular industry or firm. Additionally, the unexpected component of tone may convey the most relevant information to investors. Davis et al. (2007) and Engelberg (2008) use change-in-tone in addition to tone levels as alternative tone measures and find their results remain qualitatively unchanged. Following Davis et al. (2007), we take the difference in conference call tone in the current and immediately preceding quarter as an alternative specification. Although not shown, we find similar results when the tone measures are calculated as the change-in-tone from the preceding quarter.

**Table 5** Regression results of cumulative abnormal returns on earnings surprise and tone measures

	CAR (0,1) DJRC	t-statistic	CAR (0,1) BOTH	t-statistic	CAR (2,10) DJRC	t-statistic	CAR (2,10) BOTH	t-statistic
Panel A								
SURP	0.1200***	(6.18)	0.1210***	(6.18)	-0.2610***	(-2.75)	-0.1460***	(-3.26)
TONE <sub>1</sub>	0.0020***	(3.15)	0.0019***	(3.17)	0.0009	(0.86)	0.0009	(0.97)
Constant	-0.0020**	(-2.16)	-0.0019	(-1.63)	0.0007	(0.45)	0.0012	(0.64)
R-squared	0.033		0.033		0.069		0.024	
Panel B								
SURP	0.1170***	(5.74)	0.1190***	(5.77)	-0.2580***	(-2.75)	-0.1430***	(-3.34)
H-TONE <sub>1</sub>	0.0018***	(3.50)	0.0017***	(3.20)	-0.0007	(-0.76)	-0.0008	(-0.77)
Constant	-0.0094***	(-3.67)	-0.0088***	(-3.08)	0.0036	(0.76)	0.0044	(0.80)
R-squared	0.034		0.034		0.069		0.024	
Panel C								
SURP	0.1180***	(6.69)	0.1200***	(6.65)	-0.2610***	(-2.74)	-0.1460***	(-3.24)
TONE <sub>2</sub>	0.0022***	(3.05)	0.0021***	(3.06)	0.0006	(0.60)	0.0006	(0.60)
Constant	-0.0020**	(-2.15)	-0.0019	(-1.63)	0.0007	(0.45)	0.0012	(0.64)
R-squared	0.034		0.034		0.068		0.024	
Panel D								
SURP	0.1130***	(6.40)	0.1150***	(6.45)	-0.2550***	(-2.74)	-0.1390***	(-3.46)
H-TONE <sub>2</sub>	0.0322***	(4.25)	0.0297***	(3.52)	-0.0152	(-0.92)	-0.0197	(-1.08)
Constant	-0.0206***	(-4.29)	-0.0190***	(-3.44)	0.0094	(0.90)	0.0125	(1.06)
R-squared	0.039		0.037		0.069		0.026	

This table provides results for cross-sectional regressions of CARs on SURP and TONE<sub>1</sub> (Panel A), on SURP and H-TONE<sub>1</sub> (Panel B), on SURP and TONE<sub>2</sub> (Panel C), and on SURP and H-TONE<sub>2</sub> (Panel D). Each regression has 1,755 observations. Standard errors are adjusted for heteroscedasticity following White(1980) and for clustering by firm and quarter following Petersen (2009). t-statistics are in parentheses. CAR(0,1)DJRC is the two-day cumulative abnormal return, starting the day of the conference call, where the abnormal returns are estimated using a benchmark model with returns on the Dow Jones REIT Composite Index as the explanatory variable. CAR(0,1)BOTH is the two-day cumulative abnormal return, starting the day of the conference call, where the abnormal returns are estimated using a benchmark model with both the Dow Jones REIT Composite Index returns and the CRSP Value Weighted returns included as explanatory variables. CAR(2,10)DJRC is calculated the same way as CAR(0,1)DJRC, only it is cumulated from days two through ten. CAR(2,10)BOTH is calculated in the same manner as CAR(0,1)BOTH, only it is cumulated from days two through ten. SURP is the earnings surprise calculated as  $\{[EPS(qtr) - EPS(qtr-4)] / \text{Stock Price (end of } qtr-4)\}$ . TONE<sub>1</sub> is the first principal component of the four ratios Positive/Negative, Strong/Weak, Active/Passive, and Overstated/Understated, where each category reflects the proportion of words in a given call as defined by the Harvard IV-4 Psychosocial Dictionary. H-TONE<sub>1</sub> is the ratio of positive words to negative words as defined by Henry (2008). TONE<sub>2</sub> is the first principal component of the four ratios (Positive–Negative)/(Positive+Negative), (Strong–Weak)/(Strong+Weak), (Active–Passive)/(Active+Passive), and (Overstated–Understated)/(Overstated+Understated), where each category reflects the proportion of words in a given call as defined by the Harvard IV-4 Psychosocial Dictionary. H-TONE<sub>2</sub> is the ratio of (Positive–Negative)/(Positive+Negative), where each category reflects the proportion of words in a given call as defined by Henry (2008)

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

The positive and significant coefficients on the tone measures in the two-day CAR regressions are in line with our prior expectations that the tone variables are positive predictors of contemporaneous abnormal returns around earnings conference call events. That is, there is a direct relation between CARs and Tone.<sup>24</sup> This confirms the results in Tables 3 and 4, after controlling for the earnings surprise. The negative coefficients on earnings surprise over the  $t=2$  to  $t=10$  period are consistent with the correlation coefficients in Table 2. They suggest that the stock market may have overreacted over the  $t=0$  to  $t=1$  period to the information unique to the numerical earnings release.<sup>25</sup>

The same is not true, however, with respect to information unique to the conference call. Over the initial two-day window there is a strong and direct effect of the tone on CARs. For the subsequent 9 days, though, there are only insignificant coefficients for the tone variables. Thus, there is an efficient response to the information unique to the conference call. The observation of a more efficient response to the information in the conference call than in the earnings release is not surprising since a detailed discussion is likely less ambiguous than a brief earnings release. This is also consistent with Kimbrough (2005), who finds that conference calls improve the efficiency of the market response to earnings announcements.<sup>26</sup>

Next we present portfolio-based analyses of the extent to which conference call tone influences stock returns, after controlling for the degree of the earnings surprise, by the level of the tone. In Tables 6 and 7 we conduct two-way sequential sorts on earnings surprise and tone across all four CAR representations, (CAR(0,1)DJRC, CAR(0,1)BOTH, CAR(2,10)DJRC, CAR(2,10)BOTH). Within earnings surprise terciles (low, medium, and high) we sort the mean CARs into tone terciles (low, medium, and high) to create nine portfolios. We also construct H-L portfolios that are formed by going long in a given high portfolio and shorting the corresponding low portfolio.

Although untabulated, the number of observations in each of the nine portfolios is fairly well balanced. For both Tables 6 and 7 the smallest bin contains 188 observations (10.7% of total), while the largest contains 202 observations (11.5% of total).  $TONE_1$  is the basis for the tone-dimensional sort in Table 6 and  $TONE_2$  is used in Table 7.<sup>27</sup>

The H-L return differences in means between the high and low earnings surprise (SURP) portfolios for the two-day CARs show strong evidence that higher abnormal returns are associated with higher levels of earnings surprise. That is, holding the

<sup>24</sup> While we find statistical significance, the proportion of unexplained variation in abnormal returns remains large as seen with the low R-squared measure. However, this is entirely consistent with the literature which generally finds low R-squared measures when working with CARs.

<sup>25</sup> Recalling the relation between earnings announcement and conference call dates displayed in Table 1, the two-day conference call period includes or immediately follows the earnings announcement date. Thus, the nine-day significantly negative coefficients on earnings surprise suggest correction to initial overreaction.

<sup>26</sup> The abnormal return pattern in Fig. 1 reflects the combined influences of the conference call and the earnings release.

<sup>27</sup> Sorts based on H-TONE<sub>1</sub> and H-TONE<sub>2</sub>, in addition to sorts using medians, are untabulated since the results are qualitatively the same as those presented. Furthermore, results for H-TONE<sub>1</sub> and H-TONE<sub>2</sub> are identical to one another since the sort breakpoints are the same for the two tone constructs.

**Table 6** CARs for sorts by earnings surprise and TONE<sub>1</sub>

TONE <sub>1</sub>	SURP			TONE <sub>1</sub>			SURP			H - L	t-statistic
	Low (L)	Medium	High (H)	H - L	t-statistic	Low (L)	Medium	High (H)			
Panel A											
	CAR (0,1) DJRC						CAR (0,1) BOTH				
Low (L)	-0.0161	-0.0020	0.0002	0.0164	(3.62)***	Low (L)	-0.0159	-0.0020	0.0007	0.0166	(3.64)***
Medium	-0.0075	-0.0008	0.0035	0.0110	(2.28)**	Medium	-0.0074	-0.0007	0.0033	0.0107	(2.22)**
High (H)	-0.0013	-0.0003	0.0043	0.0057	(1.45)	High (H)	-0.0013	-0.0004	0.0047	0.0060	(1.54)
H-L	0.0148	0.0016	0.0041			H - L	0.0146	0.0016	0.0040		
t-statistic	(3.13)***	(0.61)	(1.12)			t-statistic	(3.07)***	(0.61)	(1.07)		
Panel B											
	CAR (2,10) DJRC						CAR (2,10) BOTH				
Low (L)	0.0017	0.0002	0.0009	-0.0007	(-0.11)	Low (L)	0.0022	0.0011	0.0004	-0.0019	(-0.27)
Medium	-0.0002	0.0016	0.0040	0.0042	(0.68)	Medium	-0.0026	0.0018	0.0043	0.0069	(1.37)
High (H)	-0.0001	0.0002	0.0031	0.0032	(0.61)	High (H)	0.0005	0.0012	0.0049	0.0044	(0.83)
H-L	-0.0017	0.0000	0.0022			H-L	-0.0017	0.0002	0.0045		
t-statistic	(-0.25)	(-0.01)	(0.42)			t-statistic	(-0.24)	(0.05)	(0.88)		

This table shows (Panel A) mean two-day CARs sorted sequentially by SURP and TONE<sub>1</sub>, and (Panel B) mean nine-day CARs sorted sequentially by SURP and TONE<sub>1</sub>. H-L is the difference between the highest and lowest tercile. t-statistics are in parentheses. CAR(0,1)DJRC is the two-day cumulative abnormal return, starting the day of the conference call, where the abnormal returns are estimated using a benchmark model with returns on the Dow Jones REIT Composite Index as the explanatory variable. CAR(0,1)BOTH is the two-day cumulative abnormal return, starting the day of the conference call, where the abnormal returns are estimated using a benchmark model with both the Dow Jones REIT Composite Index returns and the CRSP Value Weighted returns included as explanatory variables. CAR(2,10)DJRC is calculated the same way as CAR(0,1)DJRC, only it is cumulated from days two through ten. CAR(2,10)BOTH is calculated in the same manner as CAR(0,1)BOTH, only it is cumulated from days two through ten. SURP is the earnings surprise calculated as  $\{EPS(qtr) - EPS(qtr-4)\} / \text{Stock Price (end of } qtr-4\})$ . TONE<sub>1</sub> is the first principal component of the four ratios Positive/Negative, Strong/Weak, Active/Passive, and Overstated/Understated, where each category reflects the proportion of words in a given call as defined by the Harvard IV-4 Psychosocial Dictionary.

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

**Table 7** CARs for sorts by earnings surprise and TONE<sub>2</sub>

TONE <sub>2</sub>	SURP			TONE <sub>2</sub>			SURP				
	Low (L)	Medium	High (H)	H - L	t-statistic	Low (L)	Medium	High (H)	H - L	t-statistic	
Panel A											
	CAR (0,1) DJRC					CAR (0,1) BOTH					
Low (L)	-0.0159	-0.0018	0.0003	0.0161	(3.17)***	Low (L)	-0.0156	-0.0019	0.0008	0.0164	(3.18)***
Medium	-0.0061	-0.0004	0.0041	0.0102	(2.76)***	Medium	-0.0061	-0.0004	0.0040	0.0101	(2.71)***
High (H)	-0.0030	-0.0009	0.0036	0.0066	(1.51)	High (H)	-0.0029	-0.0008	0.0040	0.0069	(1.58)
H-L	0.0129	0.0010	0.0034			H-L	0.0127	0.0011	0.0032		
t-statistic	(2.29)**	(0.37)	(0.92)			t-statistic	(2.24)**	(0.40)	(0.86)		
Panel B											
	CAR (2,10) DJRC					CAR (2,10) BOTH					
Low (L)	0.0047	0.0003	0.0004	-0.0043	(-0.54)	Low (L)	0.0026	0.0012	-0.0001	-0.0027	(-0.39)
Medium	-0.0021	0.0011	0.0038	0.0059	(1.36)	Medium	-0.0017	0.0015	0.0042	0.0059	(1.30)
High (H)	-0.0013	0.0006	0.0039	0.0052	(0.94)	High (H)	-0.0008	0.0014	0.0055	0.0063	(1.14)
H-L	-0.0060	0.0004	0.0035			H-L	-0.0035	0.0001	0.0056		
t-statistic	(-0.73)	(0.10)	(0.68)			t-statistic	(-0.48)	(0.04)	(1.09)		

This table shows (Panel A) mean two-day CARs sorted sequentially by SURP and TONE<sub>2</sub>, and (Panel B) mean nine-day CARs sorted sequentially by SURP and TONE<sub>2</sub>. H-L is the difference between the highest and lowest tercile. t-statistics are in parentheses. CAR(0,1)DJRC is the two-day cumulative abnormal return, starting the day of the conference call, where the abnormal returns are estimated using a benchmark model with returns on the Dow Jones REIT Composite Index as the explanatory variable. CAR(0,1)BOTH is the two-day cumulative abnormal return, starting the day of the conference call, where the abnormal returns are estimated using a benchmark model with both the Dow Jones REIT Composite Index returns and the CRSP Value Weighted returns included as explanatory variables. CAR(2,10)DJRC is calculated the same way as CAR(0,1)DJRC, only it is cumulated from days two through ten. CAR(2,10)BOTH is calculated in the same manner as CAR(0,1)BOTH, only it is cumulated from days two through ten. SURP is the earnings surprise calculated as  $\{[\text{EPS}(\text{qtr}) - \text{EPS}(\text{qtr-4})] / \text{Stock Price (end of qtr-4)}\}$ . TONE<sub>2</sub> is the first principal component of the four ratios (Positive-Negative)/(Positive+Negative), (Strong-Weak)/(Strong+Weak), (Active-Passive)/(Active+Passive), and (Overstated-Understated)/(Overstated+Understated), where each category reflects the proportion of words in a given call as defined by the Harvard IV-4 Psychosocial Dictionary.

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$



tone constant, a positive (negative) earnings surprise is rewarded (punished) by the stock market. This holds for the low and medium  $TONE_1$  and  $TONE_2$  portfolios in Tables 6 and 7. However, there are no significant H-L differences for earnings surprises in Tables 6 and 7 for nine-day returns. Thus, the effect on returns of earnings surprises is concentrated around the times of the earnings announcement and conference call.

The much more interesting result in Tables 6 and 7, however, is found analyzing the tone sort dimension for two-day CARs at the time of the conference call. Specifically, for those firms in the low earnings surprise terciles, a higher level of earnings conference call tone is significantly related to higher abnormal returns.<sup>28</sup> The H-L difference for the low earnings surprise tercile is positive and significant at the 1% level in Table 6 and at the 5% level in Table 7. Further, in both tables the high tone, low earnings surprise, portfolio has two-day CARs close to zero. This suggests that a positive tone can completely offset a negative earnings surprise. For firms in the medium and high earnings surprise terciles, the two-day H-L CAR differences across tone are always positive, but are insignificantly different than zero in both tables. Finally, for the subsequent nine-day CARs there are no significant H-L differences across tone in Tables 6 and 7.

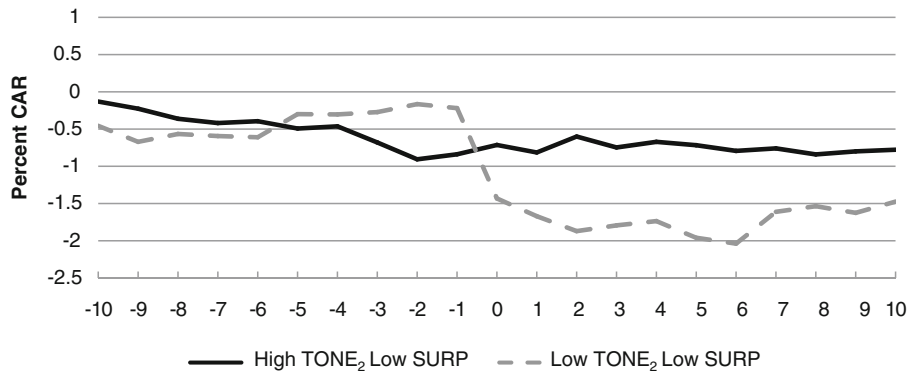
For two-day CARs, results in Tables 6 and 7 indicate that tone is extremely important when the earnings surprise is negative. However, for favorable earnings surprises, tone appears to be less relevant. We can infer from the sorts that a favorable tone is able to offset a bad earnings report, while nothing a firm says seems to significantly affect a good reported earnings number.

We gain further insight by plotting the cumulative abnormal returns when the earnings surprise is negative from 10 days prior to the conference call through 10 days after the conference call. Figure 2 shows the cumulative abnormal returns for two  $TONE_2$  sorted portfolios. The abnormal returns, shown in percent, are estimated using the benchmark model shown in Eq. 12. Figure 2 shows how returns drop precipitously for those firms with a negative earnings surprise and low conference call tone, while returns hold fairly steady for those firms with a negative earnings surprise and high earnings conference call tone.

As has been shown, the results are robust to the choice of benchmark models used in abnormal return estimation, the particular dictionary used to categorize words, and the construction of the tone variable. However, it is possible that we are simply picking up the results of Davis et al. (2007), Henry (2008), and Sadique et al. (2008) in that the conference call tone measures may be proxying for the tone of the press releases. To this end, we control for the tone of the press releases by separating the conference calls into two sections as in Matsumoto et al. (2007). Matsumoto et al. (2007) examine the relation between firm performance and the length and type of disclosures made in the managers' presentations and analyst discussion periods of earnings conference calls.

To a large degree the prepared statements by management at the beginning of a call reiterate the carefully crafted press release. As such, we use the tone of the

<sup>28</sup> Across both tables, 12 out of 12 two-day CARs in the low earnings surprise tercile are negative, indicating that these earnings releases tend to be negative surprises.



**Fig. 2** Cumulative abnormal returns by earnings surprise and TONE<sub>2</sub> Terciles. Cumulative abnormal returns (CAR) are shown in percent, where the abnormal returns are estimated using a benchmark model with returns on the Dow Jones REIT Composite Index as the explanatory variable. Cumulative abnormal returns are plotted for High TONE<sub>2</sub> Low SURP conference calls and Low TONE<sub>2</sub> Low SURP conference calls. High TONE<sub>2</sub> Low SURP contains the conference calls within the highest TONE<sub>2</sub> tercile and lowest SURP tercile. Low TONE<sub>2</sub> Low SURP contains the conference calls within the lowest TONE<sub>2</sub> tercile and lowest SURP tercile. TONE<sub>2</sub> is the first principal component of the four ratios (Positive–Negative)/(Positive+Negative), (Strong–Weak)/(Strong+Weak), (Active–Passive)/(Active+Passive), and (Overstated–Understated)/(Overstated+Understated), where each category reflects the proportion of words in a given call as defined by the Harvard IV-4 Psychosocial Dictionary. SURP is the earnings surprise calculated as  $\{[\text{EPS}(\text{qtr}) - \text{EPS}(\text{qtr}-4)] / \text{Stock Price (end of qtr-4)}\}$ . Day zero is the date of the earnings conference call

introductory statements prepared by management as a proxy for the tone of the earnings announcements. To accomplish this we split the TONE<sub>2</sub> and H-TONE<sub>2</sub> measures into separate components (INTRO TONE<sub>2</sub> and Q&A TONE<sub>2</sub>, as well as INTRO H-TONE<sub>2</sub> and Q&A H-TONE<sub>2</sub>). INTRO TONE<sub>2</sub> is constructed exactly as the TONE<sub>2</sub> variable above, only it is limited to the managers' introductory presentation text. Likewise, Q&A TONE<sub>2</sub> represents the TONE<sub>2</sub> measure for the question and answer portion of the call. INTRO H-TONE<sub>2</sub> and Q&A H-TONE<sub>2</sub> follow the same pattern using the Henry measure.<sup>29</sup>

In light of previous studies which demonstrate that conference calls convey information beyond that contained in the earnings announcements (Frankel et al. 1999; Kimbrough 2005), we expect that, after controlling for the tone of the press releases using the conference call introductory statements, the tone of the question and answer dialogue between management and analysts will contain explanatory power for the cumulative abnormal returns around the conference call date. However, it is plausible that the question and answer tone may act as a proxy for new information conveyed during the conference call that is attributable to other

<sup>29</sup> Regression results are consistent across all tone measures and choice of benchmark model used to estimate abnormal returns. For brevity, results using the TONE<sub>1</sub> and H-TONE<sub>1</sub> tone measures, as well as the benchmark model that incorporates both the Dow Jones REIT Composite Index returns and the CRSP Value Weighted returns as explanatory variables in abnormal returns estimation, are not tabulated. Additionally, untabulated descriptive statistics show that the tone of the introductory portion of the calls is slightly more positive than discussion portion in every quarter of the sample, with little difference in both relative and absolute magnitudes of the different tone measures over time.

factors. In order to control for this possibility, we incorporate several additional independent variables in the regression as follows:<sup>30</sup>

$$CAR_j = \gamma_0 + \gamma_1 SURP_j + \gamma_2 INTRO\ TONE_{i,j} + \gamma_3 Q\&A\ TONE_{i,j} + CONTROLS_j + \varepsilon_j \quad (20)$$

where  $INTRO\ TONE_{i,j}$  and  $Q\&A\ TONE_{i,j}$  are interchangeable with  $INTRO\ H-TONE_{i,j}$  and  $Q\&A\ H-TONE_{i,j}$ , both as defined above and where  $i=1$  or  $2$ ; and  $[CONTROLS_j]$  includes measures of call length, firm size, book-to-market equity, profitability, leverage, trading volume, returns volatility, and analyst coverage for firm  $j$ . Call length provides an indication of the quantity of information contained in the conference call and is measured for the two separate call components as their respective word counts, in thousands (INTRO COUNT and Q&A COUNT). SIZE is the log of firm market capitalization at the end of the previous quarter. The ratio of book-to-market equity (BM) as of the end of the previous quarter, scaled by one thousand, picks up disclosure differences associated with growth opportunities. Return on assets (ROA), calculated as net income divided by total assets, accounts for increased investor information demand when firms are not profitable. Similarly, the ratio of total liabilities to total assets (LEVERAGE) controls for increased information demand when firms are experiencing financial distress. VOLUME is the total share trading volume on the day of the conference call, in millions. VOLATILITY is calculated as the standard deviation of daily returns for the ninety trading day period ending 10 days prior to the conference call since investors may require more information during periods of greater uncertainty. The data for SIZE, BM, ROA, LEVERAGE, VOLUME, and VOLATILITY are from CRSP and Compustat. Analyst coverage data (ANALYST) are from Thomson Reuters' First Call Historical Database and calculated as the log of the number of analysts covering a given firm immediately prior to the announcement.

Panel A of Table 8 shows the results of regressing CARs on the earnings surprise (SURP), split tone measures ( $INTRO\ TONE_2$ ,  $Q\&A\ TONE_2$ ), and the additional control variables (CONTROLS). The tone from both the prepared introductory statements and the discussion portion of the calls are significant when specified individually. The significance of the former confirms the findings of Davis et al. (2007), Henry (2008), and Sadique et al. (2008) in finding a relation between the level of earnings announcement tone, or optimism, and the contemporaneous stock market reaction. Finding significance for REITs suggests that REITs behave similar to the equities of industrial firms in this regard.<sup>31</sup>

<sup>30</sup> We thank an anonymous referee for suggesting this addition. The additional variables are selected following related work by Davis et al. (2007), Tetlock (2007), Tetlock et al. (2008), Engelberg (2008), and Frankel et al. (2010), which control for both the disclosure of additional information and other factors that are known to affect returns.

<sup>31</sup> The literature shows that REIT stocks and industrial equities behave in a similar manner in some cases and differently in others. For example, when analyzing REIT IPOs, Hartzell et al. (2005) and Buttner et al. (2005) find that, in contrast to typical equities, REITs do not experience short term underpricing and long run underperformance. However, other studies find that REITs tend to behave like other stocks following structural changes to the industry in the early 1990s. Chan et al. (2005) find that REITs exhibit the weekend effect similar to typical stocks. In a similar fashion, Redman et al. (1997) show that the January effect is present in REIT returns like other equities.

**Table 8** Regression results of cumulative abnormal returns on earnings surprise, split tone measures, and controls

	CAR (0,1) DJRC	t-statistic	CAR (0,1) DJRC	t-statistic	CAR (0,1) DJRC	t-statistic	CAR (0,1) DJRC	t-statistic
Panel A								
SURP	0.1180***	(6.65)	0.1210***	(6.12)	0.1190***	(6.70)	0.1370***	(4.04)
INTRO TONE <sub>2</sub>	0.0015*	(1.78)			0.0010	(0.97)	0.0008	(0.83)
Q&A TONE <sub>2</sub>			0.0023***	(3.07)	0.0020**	(2.19)	0.0017**	(2.01)
INTRO COUNT							-0.0004	(-0.33)
Q&A COUNT							-0.0014***	(-3.56)
SIZE							0.0058***	(3.92)
BM							-0.0125	(-0.11)
ROA							0.1070	(0.95)
LEVERAGE							0.0082	(1.41)
VOLUME							-0.0063***	(-2.85)
VOLATILITY							0.8820**	(2.03)
ANALYST							0.0023	(1.10)
Constant	-0.0020**	(-2.19)	-0.0020**	(-2.10)	-0.0020**	(-2.11)	-0.0520***	(-3.27)
Observations	1755		1749		1749		1747	
R-squared	0.031		0.031		0.035		0.075	
Panel B								
SURP	0.1170***	(6.04)	0.1140***	(6.59)	0.1120***	(6.67)	0.1310***	(3.90)
INTRO H-TONE <sub>2</sub>	0.0154***	(2.68)			0.0077	(1.01)	0.0064	(0.89)
Q&A H-TONE <sub>2</sub>			0.0338***	(6.02)	0.0313***	(4.40)	0.0294***	(5.04)
INTRO COUNT							-0.0003	(-0.24)
Q&A COUNT							-0.0011***	(-2.67)
SIZE							0.0052***	(3.48)

Table 8 (continued)

	CAR (0,1) DJRC	t-statistic	CAR (0,1) DJRC	t-statistic	CAR (0,1) DJRC	t-statistic
BM						
ROA					-0.0143	(-0.13)
LEVERAGE					0.1320	(1.19)
VOLUME					0.0081	(1.25)
VOLATILITY					-0.0062***	(-2.93)
ANALYST					0.8970**	(2.00)
Constant	-0.0111***	(-3.05)	-0.0212***	(-6.42)	-0.0701***	(-4.15)
Observations	1755		1742		1740	
R-squared	0.032		0.044		0.084	

This table provides results for cross-sectional regressions of CARs on SURP, INTRO TONE<sub>2</sub>, Q&A TONE<sub>2</sub>, and controls (Panel A), as well as CARs on SURP, INTRO H-TONE<sub>2</sub>, Q&A H-TONE<sub>2</sub>, and controls (Panel B). Standard errors are adjusted for heteroscedasticity following White (1980) and for clustering by firm and quarter following Petersen (2009). t-statistics are in parentheses. CAR(0,1)DJRC is the two-day cumulative abnormal return, starting the day of the conference call, where the abnormal returns are estimated using a benchmark model with returns on the Dow Jones REIT Composite Index as the explanatory variable. SURP is the earnings surprise calculated as  $\{EPS(qtr) - EPS(qtr-4)\}/\text{Stock Price (end of } qtr-4)\}$ . TONE<sub>2</sub> is the first principal component of the four ratios (Positive-Negative)/(Positive+Negative), (Strong-Weak)/(Strong+Weak), (Active-Passive)/(Active+Passive), and (Overstated-Understated)/(Overstated+Understated), where each category reflects the proportion of words in a given call as defined by the Harvard IV-4 Psychosocial Dictionary. H-TONE<sub>2</sub> is the ratio of (Positive-Negative)/(Positive+Negative), where each category reflects the proportion of words in a given call as defined by Henry (2008). The variable prefixes INTRO and Q&A indicate that the referenced tone measures are for the introductory managerial statement and the question and answer portions of the conference calls, respectively. COUNT is the number of words, in thousands, for a given portion of a conference call (INTRO and Q&A). SIZE is the log of firm market capitalization from the previous quarter. BM is the ratio of book-to-market equity as of the end of the previous quarter, scaled by one thousand. ROA is return on assets, calculated as net income divided by total assets. LEVERAGE is the ratio of total liabilities to total assets. VOLUME is the total share trading volume on day zero. VOLATILITY is calculated as the standard deviation of daily returns for the ninety trading-day period ending 10 days prior to the conference call. ANALYST is the log of the number of analysts who cover a given firm. When Q&A TONE<sub>2</sub> is included in the regression specification, the number of observations is reduced by six because six calls do not open up for discussion. Seven additional observations are dropped when Q&A H-TONE<sub>2</sub> is specified because seven calls do not contain any positive or negative words, as defined by the narrow Henry dictionary, in either the introductory or Q&A sections. An additional two observations are dropped (both panels) when control variables are included due to data availability.

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

Interestingly, when the tone measures for both the introductory and discussion portions of the call are included in the regression, the question and answer tone is significant at the 5% level (panel A) or 1% level (Panel B), while the introductory statement tone loses significance. In other words, after controlling for the introductory comments, the question and answer tone has significant explanatory power. But, once we control for the question and answer tone, the introductory tone lacks significant explanatory power for the contemporaneous stock market reaction; the introductory tone provides no information beyond that in the question and answer tone. This holds even after including numerous controls.<sup>32</sup> Our results are generally consistent with Frankel et al. (1999) and Kimbrough (2005) who show that conference calls provide information to the stock market above and beyond the information content of the earnings announcements. However, our results deepen the understanding of conference call content by demonstrating that the linguistic tone of the question and answer portion of conference calls dominates that of the introductory management comments.

An important finding in Tables 6 and 7 is that for low earnings surprise firms, positive tone can completely offset a negative earnings surprise. As a further robustness check of this important finding, we regress the CARs of REITs with high and low tone and low earnings surprise on the full set of control variables following the specification in the last regression of Table 8, Panels A and B. We then compare the mean tone coefficients (high tone minus low tone) using t-tests.<sup>33</sup> In untabulated results, we find that the introductory tone mean coefficient differences are small and insignificant. The question and answer tone mean coefficient differences of  $-0.0057$  (Q&A TONE<sub>2</sub>) and  $-0.1063$  (Q&A H-TONE<sub>2</sub>) have the expected signs and the latter difference is significant at the 1% level. Taken together, this suggests that the positive tone in the question and answer part of the conference calls has the greatest effect in offsetting the negative earnings surprise.

## Conclusion

We empirically examine the linguistic tone of quarterly earnings conference calls and its relation to stock returns for publicly traded Real Estate Investment Trusts in the post Regulation FD period, where conference calls are open to the public. The results, which are robust to variation in tone definition and construction, show that conference call tone conveys information to market participants. All tone measures have significant explanatory power for the contemporaneous stock price reaction, as measured by the two-day cumulative abnormal return (days zero through one), even after controlling for the earnings surprise. However, the tone measures based on a customized, domain relevant dictionary provide noticeably stronger results than tone measures derived from the more general Harvard dictionary. We find no significance for the nine-day cumulative abnormal returns (days two through ten) that follow the initial reaction period.

<sup>32</sup> Furthermore, in untabulated results we find that our inferences remain unchanged when the tone measures are calculated as the change in tone from the previous quarter.

<sup>33</sup> We thank an anonymous referee for suggesting this robustness check.

When analyzing the tone of the earnings announcement and the conference call discussion separately, we find each has a strong individual effect on contemporaneous abnormal returns. The relation with the former is consistent with the results of Davis et al. (2007), Henry (2008), and Sadique et al. (2008), which suggests that REIT return behavior is consistent with the behavior of typical equities with respect to the incorporation of the information content of quarterly earnings announcements. However, unique characteristics of the REIT tax structure provide a good setting to move beyond simple earnings announcements and analyze the incrementally informative value of earnings conference calls, particularly the question and answer discussion portion. When tone variables for both the introductory comments and discussion are included in our regressions, the information content of the discussion section dominates that of the prepared statement.

Interestingly, when earnings are less than expected, the tone of the accompanying conference call dialogue between management and analysts plays a significant role in shaping the contemporaneous stock price reaction. That is, when the numbers are poor, management needs to speak for them. We're not suggesting that a smooth talking group of managers can dupe analysts and investors into feeling good about a company whose performance is disappointing. In fact, since the question and answer discussion dominates prepared remarks, the positive tone of the prepared remarks cannot by itself mislead investors. Rather, we interpret our results as being consistent with Frankel et al. (1999) who find that conference calls provide material information to the market over and above the earnings numbers contained in the corresponding press release. When the earnings surprise is negative, conference call tone conveys information about a firm's future prospects to the market. Those firms with highly positive overall call tone do not see contemporaneous negative abnormal returns, while the stock prices of those firms with lower tone levels experience a precipitous drop.

Conversely, our statistical tests show that when the earnings numbers are good, they speak for themselves. We do not find conference call tone to be a determinant of cumulative abnormal returns when a firm's unexpected earnings are highly positive, suggesting that investors may place more weight on numerically represented information or that they only turn to alternate means of gaining insight into a firm's future prospects when the earnings numbers are less than what they had hoped. A complete investigation of this conjecture, however, is a matter for future study.

**Acknowledgements** The authors acknowledge the helpful comments and suggestions of Thomas Barkley, Bruce Billings, Dean Gatzlaff, David Geltner, Elaine Henry, James Kau, Bill Mayew, Rick Morton, Milena Petrova, CF Sirmans, Stacy Sirmans, an anonymous referee, and workshop participants at MIT, the ARUEUA 2009 Mid-Year Conference, and the 2009 FMA conference. We thank Brad Case for providing FTSE NAREIT All REITs Index firm data and Aleksander Wawer for General Inquirer technical support. All remaining errors are our own.

## Appendix

The term *content analysis* appears in the Webster's Dictionary of the English Language defined as, "analysis of the manifest and latent content of a body of communicated material through classification, tabulation, and evaluation of its key



symbols and themes in order to ascertain its meaning and probable effect.” This study uses the General Inquirer (GI), a computer based application that identifies, classifies and tabulates the words in a written document. Developed by Harvard researcher, Philip J. Stone, this package has been widely used across various disciplines for over 40 years. Its application can be found in psychology, communication, journalism, and other social sciences (Krippendorf 2004).

The GI applies over 6,300 disambiguation rules in order to properly identify words within a body of text. For example, the program is able to identify *Accomplished*, *Accomplishes*, and *Accomplishing* as having the same common root, *Accomplish*, which it then classifies according to the definitions or categorizations of the dictionary applied to the analysis. The GI will recognize *Accomplishments*, as a separate word with the root *Accomplishment*, which then receives its own categorization based on the dictionary.

The default dictionary is the Harvard IV-4 Psychosocial Dictionary, a broad collection of roughly 12,000 words that fall within one or more of 77 separate categories. The categories include not only positive and negative words, but cover diverse areas such as political, religious, academic, gender specific, or even race related words. We identified eight relevant categories for the purpose of establishing an overall measure of positive and negative tone. Category descriptions, the number of root words, and a corresponding description of the type of words found in each category are in Table 9. The eight categories form four opposing pairs. Ultimately, it doesn’t matter how positive a body of text is considered in isolation since the positivity can be offset by an equal or greater amount of negativity. To account for this we use ratios in our analysis to create a relative measure.

Henry (2008) argues that the Harvard IV-4 Psychosocial Dictionary is too broad to be properly applied within the realm of finance and related disciplines. As such she creates her own list of “domain relevant” words that connote either a positive or negative outlook. The list is shown in Table 10. Of the collection of 189 words in the Henry list, all are found within the Harvard dictionary categorized under one of the

**Table 9** Harvard dictionary categories, description, and word counts

Select Harvard IV-4 Psychosocial dictionary categories	Number of words and description
Positive:	1,915 words of positive outlook
Negative:	2,291 words of negative outlook
Strong:	1,902 words implying strength
Weak:	755 words implying weakness
Active:	2,045 words implying an active orientation
Passive:	911 words indicating a passive orientation
Overstated:	696 words indicating emphasis in realms of speed, frequency, causality, inclusiveness, quantity or quasi-quantity, accuracy, validity, scope, size, clarity, exceptionality, intensity, likelihood, certainty, and extremity
Understated:	319 words indicating de-emphasis and caution in these realms

**Table 10** Henry word list

Positive Tone			Negative Tone		
Above	Expands	Rewarded	Below	Fell	Weakness
Accomplish	Expansion	Rewarding	Challenge	Hurdle	Weaknesses
Accomplished	Good	Rewards	Challenged	Hurdles	Worse
Accomplishes	Greater	Rise	Challenges	Least	Worsen
Accomplishing	Greatest	Risen	Challenging	Less	Worsening
Accomplishment	Grew	Rises	Decline	Low	Worsens
Accomplishments	Grow	Rising	Declined	Lower	Worst
Achieve	Growing	Rose	Declines	Lowest	
Achieved	Grown	Solid	Declining	Negative	
Achievement	Grows	Strength	Decrease	Negatives	
Achievements	Growth	Strengthen	Decreased	Obstacle	
Achieves	High	Strengthened	Decreases	Obstacles	
Achieving	Higher	Strengthening	Decreasing	Penalties	
Beat	Highest	Strengthens	Depressed	Penalty	
Beating	Improve	Strengths	Deteriorate	Risk	
Beats	Improved	Strong	Deteriorated	Risks	
Best	Improvement	Stronger	Deteriorates	Risky	
Better	Improvements	Strongest	Deteriorating	Shrink	
Certain	Improves	Succeed	Difficult	Shrinking	
Certainty	Improving	Succeeded	Difficulty	Shrinks	
Definite	Increase	Succeeding	Disappoint	Shrunk	
Deliver	Increased	Succeeds	Disappointed	Slump	
Delivered	Increases	Success	Disappointing	Slumped	
Delivering	Increasing	Successes	Disappointment	Slumping	
Delivers	Larger	Successful	Disappoints	Slumps	
Encouraged	Largest	Up	Down	Smaller	
Encouraging	Leader		Downturn	Smallest	
Enjoy	Leading		Drop	Threat	
Enjoyed	More		Dropped	Threats	
Enjoying	Most		Dropping	Uncertain	
Enjoys	Opportunities		Drops	Uncertainty	
Exceed	Opportunity		Fail	Under	
Exceeded	Pleased		Failing	Unfavorable	
Exceeding	Positive		Fails	Unsettled	
Exceeds	Positives		Failure	Weak	
Excellent	Progress		Fall	Weaken	
Expand	Progressing		Fallen	Weakened	
Expanded	Record		Falling	Weakening	
Expanding	Reward		Falls	Weakens	

eight categories listed in Table 9, except four words—*Deteriorate*, *Downturn*, *Hurdle*, and *Hurdles*.

## References

- Abarbanell, J. S., & Bernard, V. L. (1992). Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *Journal of Finance*, 47, 1181–1207.
- Baik, B., Billings, B. K., & Morton, R. M. (2008). Reliability and transparency of Non-GAAP disclosures by Real Estate Investment Trusts (REITs). *Accounting Review*, 83(2), 271–301.
- Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 6, 159–178.
- Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of Accounting Research*, 6, 67–92.
- Bernard, V. (1992). Stock price reactions to earnings announcements: A summary of recent anomalous evidence and possible explanations. In R. Thaler (Ed.), *Advances in behavioral finance*. New York: Russell Sage Foundation.
- Bernard, V., & Thomas, J. (1989). Post-earnings-announcement drift: delayed price response or risk premium? *Journal of Accounting Research*, 27, 1–36.
- Bernard, V., & Thomas, J. (1990). Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics*, 13, 305–340.
- Blau, B. M., Hill, M. D., & Wang, H. (2009). REIT Short sales and return predictability. *Journal of Real Estate Finance and Economics*. Forthcoming.
- Brooks, L., & Buckmaster, D. (1976). Further evidence on the time series properties of accounting income. *Journal of Finance*, 31, 1359–1373.
- Buttimer, R. J., Hyland, D. C., & Sanders, A. B. (2005). REITs, IPO waves and long-run performance. *Real Estate Economics*, 33, 51–88.
- Chan, L., Jegadeesh, N., & Lakonishok, J. (1996). Momentum strategies. *Journal of Finance*, 51, 1681–1713.
- Chan, S. H., Leung, W. K., & Wang, K. (2005). Changes in REIT structure and stock performance: evidence from the monday stock anomaly. *Real Estate Economics*, 33, 89–120.
- Danielsen, B. R., Harrison, D. M., & Van Ness, R. A. (2009). REIT auditor fees and financial market transparency. *Real Estate Economics*, 37, 515–557.
- Davis, A. K., Piger, J. M., & Sedor, L. M. (2007). Beyond the numbers: Managers' use of optimistic and pessimistic tone in earnings Press Releases, AAA 2008 Financial Accounting and Reporting Section (FARS) Paper.
- DeBondt, W. F. M., & Thaler, R. H. (1985). Does the stock market overreact? *Journal of Finance*, 40, 793–805.
- DeBondt, W. F. M., & Thaler, R. H. (1987). Further evidence of investor overreaction and stock market seasonality. *Journal of Finance*, 42, 557–582.
- Devos, E., Ong, S., & Spieler, A. (2007). Analyst activity and firm value: evidence from the REIT sector. *Journal of Real Estate Finance and Economics*, 35, 333–356.
- Dolvin, S., & Pyles, M. (2009). REIT IPOs and the cost of going public. *Journal of Real Estate Finance and Economics*, 39, 92–106.
- Downs, H. D., & Guner, Z. N. (2006). On the quality of FFO forecasts. *Journal of Real Estate Research*, 28(3), 259–274.
- Engelberg, J. (2008). Costly Information processing: Evidence from earnings announcements. AFA 2009 San Francisco Meetings Paper.
- Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49, 283–306.
- Feldman, R., Bovindaraj, S., Livnat, J., & Segal, B. (2009). Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, Forthcoming.
- Fields, D. T., Rangan, S., & Thiagarajan, R. S. (1998). An empirical evaluation of the usefulness of Non-GAAP accounting measures in the real estate investment industry. *Review of Accounting Studies*, 3, 103–130.
- Frankel, R. M., Johnson, M., & Skinner, D. J. (1999). An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research*, 37, 133–150.
- Frankel, R. M., Mayew, W. J., & Sun, Y. (2010). Do pennies matter? Investor relations consequences of small negative earnings surprises. *Review of Accounting Studies Journal*, 15(1), 220–242.

- Freeman, R., & Tse, S. (1989). The multi-period information content of earnings announcements: rational delayed reactions to earnings news. *Journal of Accounting Research*, 27, 49–79.
- Gore, R., & Stott, M. D. (1998). Toward a more informative measure of operating performance in the REIT industry: net income vs. funds from operations. *Accounting Horizons*, 12(4), 323–339.
- Graham, J., Harvey, C., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics*, 40, 3–73.
- Hardin, W., III, & Hill, M. (2008). REIT dividend determinants: excess dividends and capital markets. *Real Estate Economics*, 36, 349–369.
- Hardin, W., III, Liano, K., & Huang, G. (2005). REIT stock splits and market efficiency. *Journal of Real Estate Finance and Economics*, 30, 297–315.
- Hartzell, J. C., Kallberg, J. G., & Liu, C. H. (2005). The role of the underlying real asset market in REIT IPOs. *Real Estate Economics*, 33, 27–50.
- Henry, E. (2008). Are investors influenced by how earnings press releases are written? *The Journal of Business Communication*, 45(4), 363–407.
- Irani, A. J. (2004). The effect of regulation fair disclosure on the relevance of conference calls to financial analysts. *Review of Quantitative Finance and Accounting*, 22, 15–28.
- Kallberg, J. G., Liu, C. H., & Srinivasan, A. (2003). Dividend pricing models and REITs. *Real Estate Economics*, 31, 435–450.
- Kimbrough, M. D. (2005). The effect of conference calls on analyst and market underreaction to earnings announcements. *The Accounting Review*, 80, 189–219.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks: Sage.
- Liu, C. H., & Mei, J. (1994). An analysis of real-estate risk using the present value model. *Journal of Real Estate Finance and Economics*, 8, 5–20.
- Ljungqvist, A., Malloy, C., & Marston, F. (2009). Rewriting history. *Journal of Finance*, 64, 1935–1960.
- Loughran, T., & McDonald, B. (2009). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, forthcoming.
- Matsumoto, D., Pronk, M., & Roelofsens, E. (2007). Conference call disclosures and firm performance: An empirical analysis of the length and content of earnings-related conference calls. Working Paper. University of Washington.
- Mayew, W. J., & Venkatachalam, M. (2009). The power of voice: Managerial affective states and future firm performance. Working Paper. Duke University.
- Mendenhall, R. (1991). Evidence of possible underweighting of earnings-related information. *Journal of Accounting Research*, 29, 170–180.
- Muhlhofer, T. (2008). They would if they could: Assessing the bindingness of the property holding constraints for REITs. Working Paper. Indiana University Bloomington.
- NIRI. (2004). *Standards of practice for investor relations*. Vienna: National Investor Relations Institute.
- Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: comparing approaches. *Review of Financial Studies*, 22(1), 435–480.
- Redman, A. L., Manakyan, H., & Liano, K. (1997). Real estate investment trusts and calendar anomalies. *Journal of Real Estate Research*, 14, 19–28.
- Sadique, S., In, F., & Veeraraghavan, M. (2008). The impact of spin and tone on stock returns and volatility: Evidence from firm-issued earnings announcements and the related press coverage. Working Paper. Monash University.
- Stone, P. J. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge: MIT.
- Sunder, S. V. (2002). Investor access to conference call disclosures: Impact of regulation fair disclosure on information asymmetry. Working Paper. New York University.
- Tetlock, P. C. (2007). Giving content to investor sentiment: the role of media in the stock market. *Journal of Finance*, 62, 1139–1168.
- Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: quantifying language to measure firms' fundamentals. *Journal of Finance*, 63, 1437–1467.
- Uang, J. Y., Citron, D. B., Sudarsanam, S., & Taffler, R. J. (2006). Management going-concern disclosures: impact of corporate governance and auditor reputation. *European Financial Management*, 12, 789–816.
- Vincent, L. (1999). The information content of Funds From Operations (FFO) for Real Estate Investment Trusts (REITs). *Journal of Accounting and Economics*, 26, 69–104.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test of heteroskedasticity. *Econometrica*, 48, 817–838.
- Wiggins, J. B. (1991). Do misperceptions about the earnings process contribute to post-announcement drift? Working Paper. Cornell University.