Marriott Student Review

Volume 2
Issue 2 *A Bright Future*

Article 8

August 2018

Predictive Power? Textual Analysis in Mergers & Acquisitions

Philip E. Morgan
Brigham Young University, philipellisonmorgan@gmail.com

Marriott Student Review is a student journal created and published as a project for the Writing for Business Communications course at Brigham Young University (BYU). The views expressed in Marriott Student Review are not necessarily endorsed by BYU or The Church of Jesus Christ of Latter-day Saints.

Predictive Power? Textual Analysis in Mergers & Acquisitions

Cover Page Footnote

I would like to express my gratitude to both the university for the opportunity to perform this research as well as the many individuals who have provided support and assistance. I would like to appropriately thank Dr. Rob Schonlau, who has worked tirelessly and patiently to teach me the intricacies of finance research. Specifically, I would like to thank Dr. Bronson Argyle, who originally encouraged me to perform this research, and who has made a profound impact on my life as a teacher, mentor, and friend.

Honors Thesis

PREDICTIVE POWER? TEXTUAL ANALYSIS IN MERGERS & ACQUISITIONS

by

Philip Morgan

Submitted to Brigham Young University in partial fulfillment of graduation requirements for University Honors

Finance Department

Brigham Young University

April 2018

Honors Advisor: Dr. Bronson Argyle

Signature: ________________________________

Honors Reader: Dr. Rob Schonlau

Signature: ________________________________

Honors Coordinator: Dr. Hal Heaton

Signature: ________________________________

ABSTRACT

PREDICTIVE POWER? TEXTUAL ANALYSIS IN MERGERS & ACQUISITIONS

Philip Morgan

Finance Department

Bachelor of Science

Modern corporations utilize mergers and acquisitions as strategies to develop shareholder value today more than ever before, yet the need for understanding firms' rationale and strategy is critical in predicting post-merger stock performance for all investors. I apply the interpretive power of textual analysis and regression to a corpus of SEC mergers and acquisitions public company filings between 1994-2017. Not only do I challenge the statistically significant correlation between word content and post-transaction abnormal stock returns, but I also characterize the effects of time segments, transaction size, and industry variation across time. As a final application, I consider sentiment analysis using Diction software packages across the corpus, and measure correlation across economic cycles to assess post-filing stock performance.

ACKNOWLEDGMENTS

I would like to express my gratitude to both the university for the opportunity to perform this research as well as the many individuals who have provided support and assistance. I would like to appropriately thank Dr. Rob Schonlau, who has worked tirelessly and patiently to teach me the intricacies of finance research. Specifically, I would like to thank Dr. Bronson Argyle, who originally encouraged me to perform this research, and who has made a profound impact on my life as a teacher, mentor, and friend.

TABLE OF CONTENTS

I. Introduction

According to data from Thomson Reuters, mergers & acquisitions (M&A) represent a vital mechanism for consolidation and restructuring within industries and carry significant influence over rapid changes in market prices. As such, any insight into the unique ability of investors to more successfully utilize publicly available information to identify firms in M&A that are likely to have large price movements is valuable. This paper seeks to answer the question, “How does a textual analysis of mergers & acquisitions SEC filings, segmented by industry, size, and time, predict long-term, cumulative abnormal post-merger-filings returns for an acquirer?” My hypothesis is built on the premise that the soft, qualitative text scribed by corporate boards into M&A proxy letters for shareholders can provide insight into the future success of acquirers’ stock that other quantitative data does not. I hypothesize that using regression analyses with textual measures from the Diction software program I can gain insights to answer this question.

This paper is not only important within academia, but is also viable as a strategy from which investors may achieve superior returns. Huge amounts of capital are involved with merger arbitrage investment firms and in M&A activity generally. For example, U.S. companies alone engage, on average, in more than $1.5 trillion (WSJ Dealogic) of M&A activity each year since 2014. But many of these deals result in lost value for shareholders. As such, studies show the failure rate of M&A (i.e., negative post-transaction stock returns) is somewhere between 70% and 90% (Christensen et al. (2011)). Many researchers have tried to explain this astonishing number by analyzing aspects of transactions that worked and those that did not work. However, even after a myriad of quantitative factors tested, researchers have had limited success in identifying the potential causes of these successes and failures. Despite the losses associated with many mergers, some other deals have generated significant economic gains and hence investors frequently make speculative investments in M&A transactions.

A majority of publicly traded companies, once targeted for M&A activity, involve company shareholders to vote by proxy in the approval process for the transaction. As part of this process companies are required to make specific SEC filings. These filings include, but are not limited to, the DEFM14A, DEFM14C, and 8K. By examining a variety of possible explanatory variables related to these filings, I hope to find that as firms publish longer filings, or filings that have more certain or optimistic sentiments, they are more likely to experience heightened post-filing acquirer stock returns. This idea builds on work previously conducted by Steven Fortney and Karl Diether at Brigham Young University, using a similar methodology but with improved measures of textual sentiment analysis.

The main underlying idea for my thesis is that firms with more specific and concrete strategic rationales for M&A activity are more likely to generate economic wealth via the transaction leading to better future returns. This more robust rationale is likely to be reflected in differently written or more extensive SEC filings related to the deal. In contrast, deals where the firms might be merging for less robust reasons (e.g., quick fixes or dramatic attempts to shore future performance and compensate management) are likely to be associated with SEC filings whose language reflects these shorter-term rationales. Thus, this work seeks to use textual analysis to identify the soft perspective and intentions of corporate boards and attorneys toward M&A activity through the text in the SEC filings and hopefully be able to identify which firms are most likely to experience superior subsequent returns.

My approach to this textual analysis utilizes the pre-determined sentiment libraries in Diction software packages to measure the sentiment reflected in the M&A public filings. I then regress announcement, and subsequent returns, on measures of text length and sentiment to find out how much of the post-transaction acquirer performance might be explained by the soft information in the filing text. In order for my analysis to find significant results that would be useful to investors it must be the case that (1) the SEC filings tend to differ in systematic ways between deals that are made for strong versus questionable strategic reasons, (2) these differences must be detectable using the type of textual analysis provided by the Diction software, and (3) the soft information being assessed by the textual analysis measures must not already be reflected in the price. If the market already understands the information then these measures will not be able to predict successful M&A returns.

The idea that important soft information is available to investors via textual analysis is not new to this paper. For example, Cohen, Malloy, and Nguyen (2016) examined quarterly and annual SEC filings, and found that changes to the language within filings have strong implications for future returns. Li (2010) also examines risk sentiment within filings' text, and finds that these reports can predict certain future returns.[1] Closest to this paper is perhaps research done by Yan (2015), who focuses on textual sentiment in merger-related corporate filings and finds that overly optimistic acquirers experience worse long-term post-transaction returns. Given these papers, I was hopeful that textual analysis of merger filings would provide useful information to investors seeking to invest around these transactions.

As described above, the literature has shown evidence that textual analysis can provide guidance in which filings are most likely to be followed by statistically significant returns. However, none of this literature specifically

[1] See also Muslu et al. (2009); and Li (2011) for a survey of various textual analysis approaches.

tests whether the various measures of the M&A filing's sentiment can explain future stock performance. I hypothesize that the post-filing returns of acquiring companies can be explained by the M&A filings in forms DEFM14A and DEFM14C. The acceptance or rejection of this hypothesis will provide a unique perspective into the potential predictive success that textual analysis of SEC proxy letters and other M&A-related SEC filings can have in explaining M&A success as measured by post-merger returns.

As a result of this paper, I identify a generalized lack of statistically significant correlation between textual sentiment and future returns, confirming general market efficiency. However, results in this paper indicate specific instances of significant positive correlation between the sentiment optimism and future returns. Still, this strong positive correlation may be a function of latent variables. Nonetheless, specific market inefficiencies exist in the pricing of acquiring firms' stock following the public release of corporate M&A filings. As such, investors who utilize this strategy of textual analysis followed by investing in acquirers expressing strong levels of optimism and shorting acquirers expressing weak levels of optimism may experience superior returns – and thus, text may be used in specific instances to predict successful returns to investors in M&A transactions.

II. Literature Review and Hypothesis Development

Various strands of literature exist in the field of M&A research that attempt to explain variation in success and failure of corporate transactions. Many identify certain deal attributes (e.g., firm sizes, firm industries, deal structure, anticipated synergies) and seek to find correlation between particular attributes and M&A success (Hitt et al. (1998); Schoenberg (2006); Homburg et al. (2006)). Additionally, theories have been proposed to explain differences in M&A success/failure that veer away from the historically transaction attribute-focused approach, focusing instead on a fundamental understanding of target companies' business models and how to integrate business combinations (Christensen et al. (2011)). Each research project has contributed to the general body of M&A knowledge, shedding incremental light on the growing number of M&A transactions and the declining number of value-creating transactions, but for the most part have not provided feasible investment approaches or strategies that investors can follow around M&A transactions.

As the number of M&A-related corporate SEC filings increase over time, and as tools become more available to analyze the soft information in the filings, new approaches and test are possible. Specifically, with the emergence of textual analysis as a tool to easily analyze sentiment and tone within financial documents through the application of built-in dictionaries, deeper insights might be revealed about the incentives and perspective underlying managers and boards to engage in M&A activity (Tudor (2014)). Academics argue that this "soft information" of qualitative text explains more of the abnormal returns than the "hard [quantitative] information" (Brockman & Cicon (2007)). As such, textual analysis has been used in a number of recent papers. For example a growing body of papers have used textual analysis to analyze 10K reports (Cohen et al. (2016); Loughran & McDonald (2011)), IPO prospectuses (Brau et al. (2016)), earnings announcements (Brockman et al. (2007)), and M&A reports (Yan (2015); Fortney & Diether (2016)).

The use of business–related word lists in the data libraries available within Diction software (www.dictionsoftware.com) has increased dramatically over time. These Diction libraries have been used to assess managerial performance (Kahveci (2016)), experience versus ability in board appointees (Harford & Schonlau (2013)), equity incentives and sentiment in earnings releases

(Arslan-Ayaydin et al. (2014)), and managerial confidence on corporate debt maturities (Ataullah et al. (2017)). Cohen, Malloy, and Nguyen (2016) results suggest that textual analyses of corporate filings could be used as an investment strategy. For this to be the case, the soft information in the filings must not have been reflected immediately in the price. But, as investors increasingly become aware of the importance of the soft information in filings and as textual analysis becomes more common, market efficiency will presumably improve. Assuming that markets are not yet efficient with respect to all the soft information in the filings, and drawing on the intuition from the above papers, my project seeks to apply the power of textual analysis on a collection of M&A filings to distill qualitative information on strategy and rationale. In contrast with the papers above, I focus on the DEFM14A and DEFM14C filings as the source of merger-related soft information. Unlike other M&A-related SEC filings, these are the first official documents released to the public (following press releases) explaining M&A rationale and strategy; thus, the first opportunity for investors to utilize textual analysis in predicting future returns.

The DEFM filings are proxy letters filled by or on behalf of a registrant at the SEC when a shareholder vote is required on an issue related to M&A. The form's purpose is "to provide security holders with sufficient information to allow them to make an informed vote at an upcoming security holders' meeting or to authorize a proxy to vote on their behalf" (Investopedia). The documents frequently include information regarding the location and time of the vote, information about proxy voting and revocability of proxy vote, questions and answers, merger summaries and details, financial statements, voting procedures, and other details.

Traditionally, M&A transactions requiring shareholder vote – thus, all DEFM-related transactions – are associated with lower synergistic gains, underperformance in the long run, and a lower probability of deal completion according to Hsieh (2008). The efficacy hypothesis states that the mere existence of a shareholder voting requirement should prevent management from making deals that engage in value-destroying acquisitions to shareholder wealth (Karpoff et al. (1996)). Their research states that "a central tenet of shareholder activism holds that shareholder proposals ameliorate the shareholder-manager agency conflict and pressure managers to adopt value-increasing policies." As such, the majority of all firms have shareholder voting provisions as fundamental pieces of their strategy. This background of shareholders' involvement in transactions underlies the filing of these documents.

Academics have noted a lack of standardization throughout these corporate filings, which presents a challenge for systematic analysis given differences in size and the quality of information. Many of these DEFM14A and DEFM14C filings are created by M&A lawyers who often rely on precedent work performed within the firm rather than a modular process provided by governing bodies. In 2011, the ABA Business Law Section's Model Agreement for the Acquisition of a Public Company was published to guide disclosure around these events; however, it seems that the majority of law firms have ignored this standard as a source of precedent, as only 1 agreement in 3,500+ had over a 50% similarity with the model agreement (Anderson & Manns (2017)). This creates significant challenges and inefficiencies in dissecting and understanding underlying attitudes, sentiment, or strategy in M&A filings, and delegitimizes aspects of a textual analysis application to the field of M&A.

Despite these challenges, textual analysis within M&A corporate filings has increased and resulted in statistically significant findings, as indicated by Yan (2015) and Fortney et al. (2016). Unlike previous research focused on positivity/optimism within SEC M&A related documents by Yan (2015), where textual analysis was performed with a simple, popular negative sentiment measure as in previous research by Tetlock (2007) and Loughran et al. (2011) to measure post-transaction outcomes, I sought to utilize the Diction software to do the textual analysis. The Diction software has been widely used by academics in several hundred books and research articles.[2] It has recently been used in several articles that appear in well-known finance journals including the Journal of Financial Economics, the Journal of Finance, and the Journal of Banking and Finance. Unlike some of these prior studies, I also focused on DEFM filings as opposed to using 10Ks, 8Ks, or 10Qs.

Additionally, I build upon previous research focused exclusively on specificity v. generic terminology in M&A filings as conducted by Fortney et al. (2016), where the same proxy letters' text was analyzed through a vector space analysis employed by Hoberg et al. (2010), resulting in correlation between specificity and post-transaction performance. Both Yan and Fortney's research provided intuition for my specific hypotheses that numerous sentiment-based factors may exist within the text of M&A filings.

My project also differs from other existing literature in that my thesis considers variation across SEC M&A filings through a variety of segmentations focused on time, the size of acquirers, the relevant size of targets to acquirers, and industries These segmentations help remove certain macro-economic trends (e.g., economic cycles) that potentially could prevent the detection of statistically significant results. By removing these factors by segmenting data into batches, industries, and sizes, I hope to better predict improved post-merger returns and extract meaningful results that could be applied in the future for better investment fund performance.

[2] For a list of articles and books that have used Diction see: http://www.dictionsoftware.com/published-studies/#peerarticles.

III. Dataset and Methodology

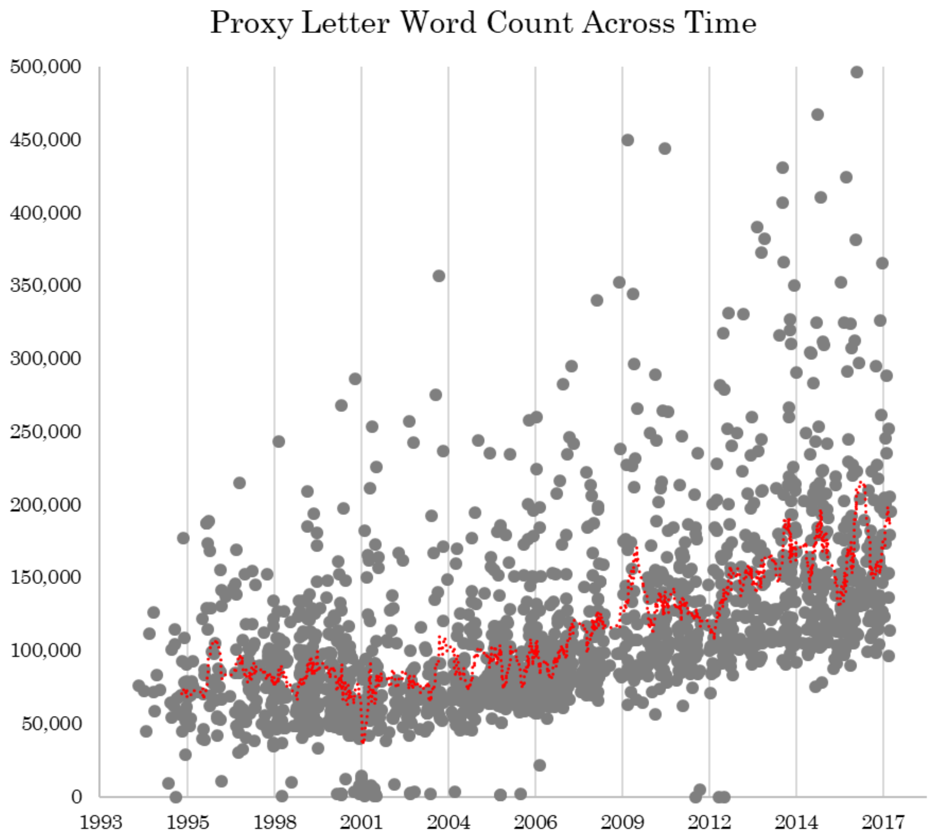
The dataset analyzed in this thesis consists of information on ~5,000 public U.S. M&A transactions requiring shareholder votes between 1994 and 2017. To collect and analyze the data, I worked with Rob Schonlau, Thaddeus Crockett, Oliver Morgan, Tanner Thompson, and David Lowe. We indexed, cleaned (stripped HTML), and matched documents in the SEC EDGAR database to the Compustat database, and then matched each document to the SDC M&A database. After matching these databases, we had definitive M&A announcement transaction dates. We then linked the data to the CRSP database for stock prices in order to be able to measure post-transaction abnormal returns. Finally, we processed each transaction using the textual-analysis program Diction to produce word count and various measures of sentiment analysis. These databases are described in more detail below.

Since 1994, the U.S. Securities and Exchange Commission (SEC) has posted information in the online EDGAR quarterly and annual indexes related to public firms' corporate filings to provide both institutional and lay investors with detailed reports of financial performance, strategy, and capital markets activity. M&A activity is reported publicly in annual 10Ks, quarterly 10Qs, and 8Ks, yet is systematically disclosed to shareholders of companies engaged in M&A in either a proxy statement on Schedule 14-A (DEFM14A) or an information statement on Schedule 14-C (DEFM14C). DEFM14A filings contain requests for the target company's shareholders to vote in favor or against the pending transaction, contain FAQs and answers for shareholders, and write out thorough rationale and strategy underlying the transaction. DEFM14C filings contain the same information with the exception of requiring a formal vote. The form's purpose is "to provide security holders with sufficient information to allow them to make an informed vote at an upcoming security holders' meeting or to authorize a proxy to vote on their behalf" (Investopedia).

Figure 1 provides information on the number of filings across time in the sample as well as the number of words in each filing. The number of words in the filings has tended to increase over time.

Figure 1: Textual Filing History

Figure 1 shows increasing text length since 1994 in all proxy letter filings. Note that length roughly follows economic cycles, with increased word lengths in the 2000 Dot.com bubble and 2008-09 recession; and short gaps in time without filings immediately following sharp economic declines. The red line tracks a weekly average across time.



Each M&A-applicable DEFM filing in the EDGAR index contains a CIK code for the filing target company, the date of the filing, and the file name. We used this CIK code to identify firm CUSIPs from the Compustat database. The CUSIPs and dates were then used to identify targets in the SDC database. Before matching databases and performing any textual analysis, the body of DEFM filings included 5,231 transactions (92% being DEFM14A filings requiring shareholder proxy vote, 8% DEFM14C; see Table 2). Some of the CIK and CUSIP codes did not have corresponding matches across database so the final sample size is less than 5,231.[3]

Compustat is a database that contains annual financial information for most public US firms. The CIK-CUSIP match also allowed us to identify each firm's Global Industry Classification Codes (GIC code) from Compustat. The CUSIP identifier was used to match with target firms in the Securities Data Company (SDC) database.

The SDC database provides acquirer and target information as well as M&A deal information. This information includes the acquirer company name, acquirer CUSIP, transaction announcement date (historically, DEFM filings were often the initial public announcements of M&A transactions, while more recently these filings are registered with the SEC weeks or months following the initial press release or leak), acquirer market capitalization (if applicable), transaction enterprise value, etc.

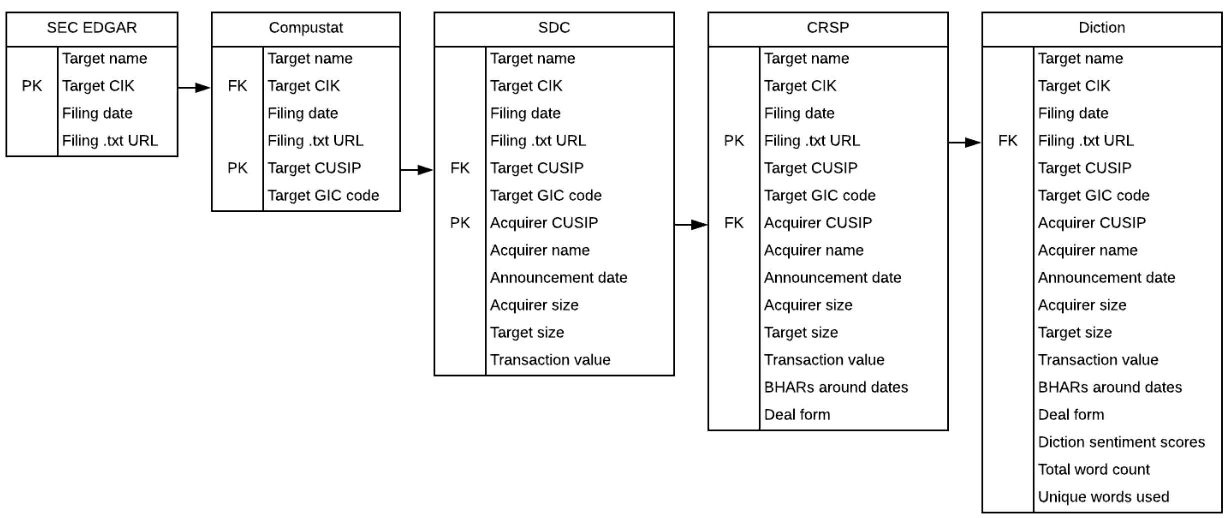
In scenarios where multiple potential acquirers submitted bids in the transaction process, SDC had multiple acquirer CUSIPs for each target CUSIP, making for potentially inaccurate matches in the dataset. However, this only occurred in a very small number of cases. In each, we selected the SDC record which was closest in time to the EDGAR filing but that preceded the announcement date. After review, these CUSIP identifiers, CIK codes, and transaction dates were consistent with the filing company, and were then matched with the CRSP database using acquirer CUSIP codes.

[3] The date of the filing tends to occur ~60 days after the SDC announcement date. The "close date" or day the transaction is consummated tends to occur ~60 days after the filing date. The average length of DEFM14A (soliciting vote) is 113,000 words, as the average length of DEFM14C (not soliciting vote) is 93,000 words.

The Center for Research in Security Prices (CRSP) database by Chicago Booth provides historical security prices and return information. Given the acquirer and target CUSIPs, CRSP identified the acquirer's corresponding PERMNO identifier, which allowed use to calculate the acquirer's stock price changes around the filing dates. Figure 2 summarizes the databases and variables described above.

Figure 2: Database Relationships

Figure 2 below indicates the sources and connections between various sources used in my research. The PKs (Primary Key) and FKs (Foreign Key) indicate the connections between databases used to develop the dataset. I then confirmed information was consistent throughout by checking random target names and acquirer names across the dataset before performing analysis. In the process we started with the target name, CIK, filing date, and text as described in the SEC EDGAR box. Reading the diagram from left-to-right the aggregated set of variables is shown. For this reason each new list going from left-to-right encompasses the variables from the prior list.



In the empirical tests that follow, we use cumulative abnormal returns to measure the stock price reaction to M&A announcements and DEFM filings. To calculate the abnormal returns we first find each firm's beta during the year before the deal by regressing the firm's daily stock returns on the market returns over the year prior to the transaction as shown in Formula (1):

$$R_{u,k} = \alpha_{t,k} + \beta_{t,k} R_{u,m} \tag{1}$$

where $R_{u,k}$ is the return of company k on day u. $R_{u,m}$ is the market return on day u. Beta is estimated over the year before the announcement day. We then use the estimated beta to determine what the firm's returns would have been going forward based on the market's performance over the same period; to do this we essentially take the estimated beta and multiply it by the observed market return.

Following the literature, in calculating the regression from 1-year into the past, we include data up through 15 days prior to the announcement to avoid the potential effects of leakage in the days before the announcement.[4] (Schwert (1996) showed evidence that such a run-up in prices started occurring as early as 15 days before the public announcement.)

The abnormal return (AR) is then calculated as the firm's actual return minus the estimated return calculated using estimated beta ($\hat{\beta}$), or, in other words, the excess return above what the Capital Asset Pricing Model (CAPM) expects is outlined in Formula (2) :

$$AR_{u,k} = R_{u,k} - \hat{\beta}_{t,k} R_{u,m} \tag{2}$$

AR is the difference between the observed return of the firm on that day and the return we expected of the firm if the stock price had continued to move relative to the market in the same way as it had historically over the prior year.

We calculated buy-and-hold abnormal returns (BHAR) using this model to summarize the cumulative abnormal returns over the intervals under

[4] Schwert (1996) showed evidence that such a run-up in prices started occurring as early as 15 days before the public announcement.

consideration. A buy-and-hold strategy calculates an investor's return experience if the investor had held the acquiring company's stock over a variety of time periods without trading. These BHARs include the 3-day returns around the transaction announcement date and the filing date, and the 30-day, 60-day, 90-day, and 1-year returns following the DEFM filing date. With these post-filing returns, we start the return from two days after the filing date, as this would give a prospective investor adequate time to process the filing text and make the necessary trades. Using the approach described by Dellavigna and Pollet (2009), the BHARs capture the cumulative returns experienced over each time frame in excess of market movements over the same time period. Thus, the returns for an investor starting two days after the filing date and holding for 30 days would be calculated as follows in Formula (3):

$$\mathrm{BHAR}\,(+2, 30) = \left[\prod_{j=+2}^{30}(1 + \mathrm{R}_{j,k})\right] - 1 - \hat{\beta}_{t,k}\left[\prod_{j=+2}^{30}(1 + \mathrm{R}_{j,m}) - 1\right] \qquad (3)$$

Where $R_{j,k}$ is the return of company k at day j starting two days after the filing date. As shown in the formula, the returns we analyze in relation to the DEFM company filings are adjusted according to market returns over the same period of time. The example above focuses on the 30 days following the filing date but the same approach is used over the 30-day, 60-day, 90-day, and 1-year returns.

IV. Textual Analysis

I used the Diction software program to perform textual analysis on all the filings. As described on the Diction program webpage, Diction is a "computer-aided text analysis program for determining the tone of a verbal message" (Harp (2010)) and has been used throughout many peer-reviewed articles and books. Processing text through built-in libraries, the software searches for words that correspond to certain tones within subject-specific libraries, while also generating general word count, unique words used, and average word length for analysis. These tones, or the character or attitude toward an M&A transaction, together form a sentiment, or a more generalized view toward the event. I utilized the "Business: Corporate Public Relation" built-in library package to evaluate the filings. The Diction software then provided numeric measures of the filings' optimism, certainty, and 38 other sentiment variables. I then regressed the BHARs described above on these measures to test whether future returns are related to the soft information in the M&A filings.

By utilizing Diction, I was able to move beyond the more simplified dichotomies (good/bad, happy/sad) that are focused and prevalent throughout the field of textual analysis, and view text through the lens of 40 different tones and sentiments, identifying patterns within phrases to determine how a certain corporate filings might conform to norms. Given the large number of potential independent variables, I identified and analyzed each one, but narrowed down the results around the five "master variables" of:

(1) Activity (movement, change, and implementation of ideas)
(2) Optimism (satisfaction, praise, endorsements, highlight positives)
(3) Certainty (resoluteness, completeness, tenacity, avoid ambivalence)
(4) Realism (tangible, familiar, and recognizable to people)
(5) Commonality (agreed-upon values, cooperation, rejecting exclusion)

These five master variables are identified by Diction and "provide the most general understanding of a given text and any study" (Diction (2013)). From this set, I particularly focused my analysis around the independent variables of optimism and certainty, but first sought to understand any correlation between these independent variables.

Table 1 provides information about how these five variables correlate. As noted in the table several of the variables are highly correlated. For example

Certainty and Optimism have a .42 correlation, which is considered moderate yet not extreme.

Table 1: Correlation Matrix of Independent Variables

Table 1 below depicts correlation between the five master variables identified by Diction. I identified the threshold of significant correlation at .30, where the strength of correlation between individual variables could statistically alter the coefficient results. From that point, I could identify which variables would "tell the same story" and narrow in for further research. The number of *s indicate the strength of correlation between these five master independent variables, not statistical significance.

		(1)	(2)	(3)	(4)	(5)
Activity	(1)	1.00	-0.17	-0.05	-0.09	-0.06
Optimism	(2)	-0.17	1.00	0.42*	0.56**	0.39*
Certainty	(3)	-0.05	0.42*	1.00	0.76***	0.72***
Realism	(4)	-0.09	0.56**	0.76***	1.00	0.62**
Commonality	(5)	-0.06	0.39*	0.72***	0.62**	1.00

In order to accurately understand which independent variables may have predictive power, or rather, correlation with the return variables, I sought to separate variables from one another which express significantly high correlation. Including these variables in the same regression would likely lead to collinearity issues. Collinearity, as explored in many introductory statistics textbooks, is a phenomenon in which a predictor (or independent) variable in a regression model can be linearly predicted from another predictor variable – i.e., high correlation with one another variable. In a situation where variables exhibit high correlation between one another, the regression coefficients may change erratically for small changes in the data or depending on which variables are included in the model. As such, I narrowed in on the included set of independent variables in consideration in the model to avoid the potential collinearity in the analysis (Dormann et al. (2013)). Also to avoid potential data errors, all transactions were cleaned to represent actual M&A activities, while excluding stock buybacks, board appointments, etc. that may require proxy letters to shareholders.

V. Data Description

The final dataset considered in my research contains 1,900+ transaction filings published in each year since 1994, and ranges in both transaction size and industry. By considering the dataset all together first, I began to better understand where particular correlation or significance existed and where it did not. Table 2 below summarizes statistics across the two types of documents (DEFM14A – voting proxy letter; DEFM14C – letter without proxy vote request).

Table 2: DEFM Document Summary Statistics Comparison

Table 2 displays summarized information about the two documents that compose the sample dataset. The vast majority of transactions considered required shareholder votes (DEFM14As) and as such the combined averages are significantly closer to the DEFM14 statistics. The small differences (with the exception of the certainty sentiment score) indicate overall document similarity; thus, I considered all of these in the following regression analyses. Column (4) has statistics on the entire dataset as a whole, while columns (2-3) contain document-specific summary statistics.

(1)	(2) DEFM14A	(3) DEFM14C	(4) Combined
Average Word Count	113,352	92,669	112,498
Average Word Size	4.9	5.0	4.9
Unique Words Used	40,282	33,753	40,013
Average Optimism Score	49.5	48.6	49.4
Average Certainty Score	27.9	36.3	28.3

Table 3 below shows the proportion of positive vs. negative BHARs in each year across differing timelines.

Table 3: Sample Set by Year

Table 3 displays descriptive information about the sample dataset across time. The number of transactions in column (1) reports the number of transactions remaining after the matching process. Fewer filings exist from the early 90's, as the SEC at the time had yet not required companies to file proxy shareholder voting letters. Columns (2) – (7) report the percentage of all transactions in each respective year which had positive abnormal returns above the CAPM-predicted return. Lastly, the *Average BHARs* row shows the historical average returns for each time series.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Announcement Date		Filing Date			
	Number of Transaction Filings	Positive 3-Day BHAR	Negative 3-Day BHAR	Positive 3-Day BHAR	Negative 3-Day BHAR	Positive 90-Day BHAR	Negative 90-Day BHAR
1994	9	33%	67%	83%	17%	83%	17%
1995	34	48%	52%	32%	68%	33%	67%
1996	40	44%	56%	63%	37%	56%	44%
1997	58	50%	50%	57%	43%	59%	41%
1998	72	39%	61%	36%	64%	42%	58%
1999	123	48%	52%	42%	58%	47%	53%
2000	98	33%	67%	41%	59%	59%	41%
2001	99	46%	54%	38%	63%	53%	47%
2002	48	58%	42%	45%	55%	39%	61%
2003	56	43%	57%	47%	53%	63%	37%
2004	74	51%	49%	57%	43%	46%	54%
2005	117	49%	51%	45%	55%	48%	52%
2006	156	42%	58%	44%	56%	51%	49%
2007	152	52%	48%	37%	63%	42%	58%
2008	78	34%	66%	60%	40%	46%	54%
2009	45	43%	57%	57%	43%	48%	52%
2010	87	43%	57%	43%	57%	50%	50%
2011	83	56%	44%	56%	44%	56%	44%
2012	73	54%	46%	61%	39%	54%	46%
2013	76	50%	50%	59%	41%	45%	55%
2014	84	64%	36%	46%	54%	52%	48%
2015	95	51%	49%	55%	45%	38%	62%
2016	106	42%	58%	49%	51%	67%	33%
2017	76	41%	59%	61%	39%	37%	63%
Total	**1,939**	**46.9%**	**53.1%**	**48.1%**	**51.9%**	**49.5%**	**50.5%**
Average BHARs		***4.58%***	**-4.88%**	***2.54%***	***-2.67%***	***15.43%***	***-12.65%***

Table 3 summarizes the returns around the announcement and filing dates. The table corroborates what previous literature explains: the majority of transactions adversely affect shareholders. On average, >50% of transactions since 1994 have resulted in negative returns across a variety of time periods, naturally peaking at recessionary periods (2000, 2008-09). Additionally, both the positive and negative returns of an acquirer are more significant at the time of public announcement than at the time of the target company's proxy shareholder vote filing. Given the existence of abnormal returns due to filings, the market continues to process a transaction's rationale months following an announcement. As the market processes a company's strategy behind M&A activity, I see how the average negative return for acquirers diminishes over time. This indicates an area of future interesting research.

Table 4 below provides detailed information about the sample and reports how many transactions correspond to each industry group. Throughout the paper, I use Global Industry Classification Codes (GIC Codes) to make industry assignments. These codes were supplied by Compustat through the dataset merging process described above. I chose to use GIC codes over the formerly used SIC and NAICS industry codes, as past research has shown that GIC codes are better at identifying firms in today's age and reflect groups of firms that exhibit similar valuation multiples and stock price co-movements than other classification approaches (Bhojraj (2003)).

Table 4: Sample Description by GIC Code

Table 4 gives a general description of the Global Industry Classification Standard Codes (GIC Codes) and the applicable number of transactions in the dataset that correspond to each industry group. This provides a broad outline of the industries analyzed in this paper.

GIC Industry Group Description	Number of Transactions	GIC Industry Group Description	Number of Transactions
Energy	106	Food, Beverage & Tobacco	36
Materials	67	Household & Personal Products	10
Capital Goods	95	Health Care Equipment Supplies	165
Real Estate	8	Pharma, Biotech & Life Sciences	103
Commercial Services & Supplies	82	Banks	192
Transportation	32	Diversified Financials	57
Automobiles & Components	6	Insurance	60
Consumer Durables & Apparel	51	Software & Services	253
Consumer Services	77	Technology Hardware & Equipment	147
Media	62	Semiconductors & Equipment	45
Retailing	83	Telecommunication Services	57
Food & Staples Retailing	20	Utilities	47

Total 1,861

The total number of transactions considered decreased from Table 3 to Table 4 as Compustat did not report or incorrectly keyed the GIC Codes for some firms. Nonetheless, the dataset considers a broad representation throughout various transactions and confirms the increased number of transactions in the respective industries. Lastly, in Table 5 below, I provide a summary description of the dataset. As the data merging left holes in some of the data, the total number of observations in Table 5 vary.

Table 5: Summary Description

Table 5 gives a description of the most frequently considered independent variables and the applicable number of transactions in the dataset that correspond to each variable in the dataset. Column (1) contains the total number of observations differing in certain variables. Columns (2 – 7) give scale to the Diction sentiment variables and provide context to the dataset.

($ in millions)	(1) Observations	(2) Min	(3) 25th Perc.	(4) Median	(5) Mean	(6) 75th Perc.	(7) Max
Acquirer Value ($)	1,205	14	1,040	3,668	21,210	17,047	458,780
Target Value ($)	1,939	20	128	420	2,112	1,624	72,671
Relative Size	1,205	0%	6%	20%	33%	53%	100%
Certainty	1,939	15	15	20	28	48	58
Optimism	1,939	38	49	49	49	50	58
Activity	1,939	16	47	47	47	47	56
Realism	1,939	24	43	44	45	47	80
Commonality	1,939	40	49	49	50	51	65
Total Words	1,939	125	74,185	97,839	112,498	135,066	592,694

As my main analysis, I tested the sensitivity of the BHARs to measures of sentiment analysis across different time intervals (1-yr intervals) and industry groups (GIC code industries). I did this to allow for the possibility that a relation exists but only within certain time periods or industries. For example, the relation may be sensitive to recessions or bubbles. Additionally, specific industry segments across time (e.g., technology, oil & gas), or specific outliers such as word size, target value, or acquirer value could affect the overall conclusions but become evident as I focus on different subsets. After analyzing these batches across time, I note that correlation (both positive and negative) exists in very specific, one-off situations between sentiment and returns. However, as a whole, the segmented approach indicates that the vast majority of industries and time intervals show no statistically significant correlation.

Using the statistical analysis software programs R Studio and Stata, I report the regression results in the following section.

VI. Regression Results

I performed regression analyses with these variables to find statistically significant correlation between individual transaction attributes/sentiments and returns across five different time periods: (1) the 3-day BHAR around the transaction announcement, (2) the 3-day BHAR around the transaction filing date (one trading day before, one trading day after, (3) the 30-day BHAR following two days after the filing date, (4) the 60-day BHAR following two days after the filing date, (5) the 90-day BHAR following two days after the filing date, and (6) the 1-year BHAR following two days after the filing date. The description and background behind is given in Section III (Dataset and Methodology).

Analyzing potential correlation across a series of days allowed us to understand the longevity of this correlation and its influence at different snapshots through time. Given the 3-day BHARs around the announcement date provide unique insights, they are not actionable as an investor; thus, I will not consider these returns moving forward. Additionally, in each BHAR analyzed after the filing date, returns are measured beginning two days after the filing date; I estimate this time would be sufficient in applying this trading strategy in practice and investing in the stocks. Both Tables 6 and 7 help indicate whether or not future returns are a function of sentiment and if this text-based market inefficiency truly exists. Table 6 first analyzes each individual sentiment across the two separate time horizons; this takes out any impact of collinearity between independent variables. Table 7 then considers the five master sentiments together.

Table 6: Total Sentiment Analysis (Univariate Regression)

Table 6 displays regression coefficients, p-values for the five "master variables," and Adjusted R-Squares for each univariate regression. There are no industry, time, or size controls. Columns (1 – 4) indicate the acquirer returns across different horizons. Significance at the 10%, 5%, and 1% level is shown with *, **, ***, respectively. P-values are shown below the regression coefficients. Corresponding Adjusted R-Squares appear to the right of the regression coefficients.

Univariate	(1) 30-Day BHAR	(2) Adjusted R-Square	(3) 90-Day BHAR	(4) Adjusted R-Square
Certainty	0.000	0.000	0.001***	0.006
	(0.324)		(0.008)	
Optimism	0.011***	0.010	0.0279***	0.016
	(0.001)		(0.000)	
Activity	-0.001	-0.001	-0.007*	0.002
	(0.792)		(0.087)	
Realism	0.002*	0.002	0.007***	0.008
	(0.054)		(0.002)	
Commonality	0.003*	0.002	0.010***	0.006
	(0.083)		(0.005)	
Observations	1,086		1,081	

The results in Table 6 indicate primarily positive statistically significant individual longer-term correlation between the master sentiment variables and returns across two different horizons. However, these results do not take into consideration the influences and relationships between variables as Table 7 does. Given the results in Table 6, specifically regarding the strongest Adjusted R-Square for the sentiment optimism, I determined to narrow in on this variable as a potential predictor.

Table 7: Total Sentiment Analysis (Multivariate Regression)

Table 7 displays regression coefficients and p-values for the five "master variables" for all transactions available between '94 -'17. There are no industry, time, or size controls. Columns (1-5) indicate the acquirer returns around (1) and following the filling date (2-5). Significance at the 10%, 5%, and 1% level is shown with *, **, ***, respectively. P-values are shown below the regression coefficients.

	(1) 3-Day BHAR	(2) 30-Day BHAR	(3) 60-Day BHAR	(4) 90-Day BHAR	(5) 1-Year BHAR
Certainty	-0.000	-0.000	-0.000	-0.000	0.003*
	(0.580)	(0.270)	(0.450)	(0.998)	(0.061)
Optimism	0.002	0.011***	0.018***	0.022***	-0.005
	(0.258)	(0.007)	(0.004)	(0.006)	(0.758)
Activity	-0.001	0.001	0.002	-0.004	-0.001
	(0.302)	(0.691)	(0.512)	(0.355)	(0.875)
Realism	-0.000	0.001	0.002	0.001	0.001
	(0.516)	(0.713)	(0.594)	(0.779)	(0.935)
Commonality	-0.000	0.003	0.006	0.004	-0.018
	(0.620)	(0.271)	(0.156)	(0.430)	(0.106)
Intercept	0.001	-0.719***	-1.371***	-1.168**	1.107
	(0.992)	(0.002)	(0.000)	(0.016)	(0.271)
Observations	1,089	1,086	1,083	1,081	1,065
Adjusted R-Square	0.001	0.008	0.015	0.015	0.000

Looking at the regression coefficients in Tables 6 and 7, indications of negative statistically significant correlation at the 10% level between Optimism and 30-day returns exist, confirming prior literature (Yan (2015)) which concluded that increased optimism has *negative* correlation with post-filing returns. These results therefore do not conflict with prior literature in regard to short-term returns; however, these results indicate that *positive* significant correlation at the 1% level exists in the longer term periods. Thus, as managers express higher optimism in transaction filings (as measured by Diction software's built-in library), the potential returns for an investor in an acquirers' stock increases.

Only optimism registers as a significant predictor of returns throughout the near and mid-term horizons. Thus, I narrow in on optimism as the variable of choice moving forward. None of the others are close to significance. I recognize that the R-Square is very low; however, this is understandable as these returns are abnormal, and anything indicates returns above what is expected as accounted for above.

I then ask if post-filing returns are a function of total word count, and thus of file size. Table 8 below contains the results of this regression analysis. Considering outliers in size (see Figure 1), I have normalized the distribution in understanding the predictability of word count and long-term returns in Table 8 below.

Table 8: Word Count Correlation

Table 8 displays regression coefficients and p-values for word count (and thus, file length) for all transactions available between '94 - '17. This table measures if total word count is a predictor of future returns. Columns (1-5) measure returns correlation around and following various filing returns. Column (6) scales (normalizes) the regression analysis by dividing the independent variable (total word count) by the standard deviation of total words. There are no industry, time, or size controls. Significance at the 10%, 5%, and 1% level is shown with *, **, ***, respectively. P-values are shown below the regression coefficients.

	(1) 3-Day BHAR	(2) 30-Day BHAR	(3) 60-Day BHAR	(4) 90-Day BHAR	(5) 1-Year BHAR	(6) 1-Year BHAR
Total Word Count	0.000*	-0.000	-0.000	-0.000	-0.000**	
	(0.083)	(0.258)	(0.185)	(0.182)	(0.049)	
Total Word Count /						-0.027**
Total Word StDev						(0.049)
Intercept	-0.005**	0.009	0.024**	0.030**	0.056*	0.056*
	(0.028)	(0.215)	(0.042)	(0.044)	(0.058)	(0.058)
Observations	1,089	1,086	1,083	1,081	1,065	1,065
Adjusted R-Square	0.002	0.000	0.001	0.001	0.003	0.003

I note the lack of statistically significant correlation at various levels between word length and returns across the same time horizons, despite the p-value approaching 15%. In the long-term (1-year BHAR), I am beginning to see statistically significant negative correlation between word count and returns. This may indicate how longer returns might indicate additional risks associated with the M&A transactions, following previous literature. Given the negative significant correlation between word count and long-term returns (Column 6), I recognize a vertical for further interesting research. However, I made a decision no longer to analyze file length as a predictor of returns given the weak correlation and the necessity of normalization to realize any significant coefficients over the long-term.

After narrowing down specifically on the sentiment variable of optimism, while excluding word count from predicting returns, I then split the dataset into annual batches in order to analyze optimism and the potential correlation differences throughout various time periods. After controlling for dates, does optimism as a predictor of returns differ across certain time periods? This potential interaction between variables could help explain results and tests whether the correlation is stable across different years and accounts for the effects that time, economic cycles, and period-relevant trends might have on the results. Tables 9a and 9b display these results below as I do not control for base effects in Table 9a, but do control for base effects in Table 9b.

Table 9a below measures any potential year effect in predicting future returns, only individually with certain years and alone with optimism. The results confirm the hypothesis that year effects have no impact on post-filing returns.

Table 9b then measures any potential year effect in predicting future returns, when considering the interaction effects between optimism and each individual year. This test explores if certain years, when combined with the leading sentiment, have year effects on results.

Table 9a: Regressions across Time

Table 9a displays regression coefficients and p-values for the sentiment optimism across each year (Columns (1-5). No industry or size controls. Significance at the 10%, 5%, and 1% level is shown with *, **, ***, respectively.

	(1)	(2)	(3)	(4)	(5)
	3-Day BHAR	30-Day BHAR	60-Day BHAR	90-Day BHAR	1-Year BHAR
1995	-0.041**	0.017	-0.078	-0.098	-0.129
	(0.023)	(0.404)	(0.357)	(0.364)	(0.554)
1996	-0.031*	0.000	0.060	0.004	0.027
	(0.082)	(0.991)	(0.474)	(0.973)	(0.901)
1997	-0.029*	0.008	0.018	0.018	-0.210
	(0.090)	(0.661)	(0.820)	(0.859)	(0.310)
1998	-0.052***	-0.007	-0.015	-0.042	-0.142
	(0.002)	(0.695)	(0.854)	(0.678)	(0.491)
1999	-0.038**	0.009	0.068	0.073	0.012
	(0.023)	(0.631)	(0.382)	(0.462)	(0.951)
2000	-0.042**	-0.005	0.007	0.026	0.172
	(0.012)	(0.771)	(0.932)	(0.791)	(0.393)
2001	-0.047***	0.000	-0.001	-0.023	0.055
	(0.005)	(0.981)	(0.991)	(0.820)	(0.786)
2002	-0.031*	0.006	-0.003	-0.015	-0.037
	(0.079)	(0.778)	(0.973)	(0.885)	(0.860)
2003	-0.031*	0.008	0.113	0.112	0.301
	(0.075)	(0.689)	(0.174)	(0.284)	(0.154)
2004	-0.031*	0.010	0.005	-0.014	-0.117
	(0.067)	(0.609)	(0.950)	(0.892)	(0.567)
2005	-0.035**	0.007	-0.009	-0.025	-0.055
	(0.035)	(0.707)	(0.908)	(0.804)	(0.785)
2006	-0.037**	0.003	-0.015	-0.026	-0.055
	(0.024)	(0.881)	(0.849)	(0.790)	(0.785)
2007	-0.039**	0.008	-0.013	-0.03	-0.073
	(0.019)	(0.664)	(0.874)	(0.762)	(0.716)
2008	-0.033*	-0.002	-0.067	-0.112	0.034
	(0.058)	(0.924)	(0.413)	(0.278)	(0.871)
2009	-0.036**	0.013	0.053	0.031	-0.085
	(0.045)	(0.526)	(0.535)	(0.774)	(0.695)
2010	-0.037**	0.010	0.000	-0.015	-0.001
	(0.030)	(0.583)	(0.998)	(0.881)	(0.996)
2011	-0.027	0.008	0.026	-0.009	-0.041
	(0.132)	(0.692)	(0.759)	(0.929)	(0.849)
2012	-0.032*	0.013	-0.001	-0.026	-0.010
	(0.058)	(0.493)	(0.994)	(0.801)	(0.963)
2013	-0.029*	0.021	0.000	-0.001	-0.049
	(0.090)	(0.263)	(0.996)	(0.994)	(0.811)
2014	-0.029*	0.025	0.005	-0.009	-0.069
	(0.079)	(0.188)	(0.955)	(0.926)	(0.735)
2015	-0.029*	0.010	-0.029	-0.074	-0.091
	(0.081)	(0.608)	(0.719)	(0.464)	(0.656)
2016	-0.033**	0.004	0.030	0.037	0.046
	(0.050)	(0.842)	(0.703)	(0.717)	(0.822)
2017	-0.024	-0.002	-0.016	-0.032	-0.119
	(0.151)	(0.902)	(0.844)	(0.755)	(0.562)
Optimism Alone	0.001	-0.002	0.016***	0.020***	-0.001
	(0.421)	(0.133)	(0.005)	(0.006)	(0.923)
Intercept	-0.014	-0.451**	-0.777***	-0.955***	0.102
	(0.819)	(0.012)	(0.008)	(0.010)	(0.894)
Observations	1,089	1,086	1,083	1,081	1,065
Adjusted R-Square	0.009	0.017	0.030	0.030	0.020

Table 9b: Regressions across Time with Optimism Interaction

Table 9b displays regression coefficients and p-values for the sentiment optimism across each year considered in the data set (Columns (1-5). The results were again not material. No industry or size controls. Significance at the 10%, 5%, and 1% level is shown with *, **, ***, respectively.

	(1)	(2)	(3)	(4)	(5)
	3-Day BHAR	30-Day BHAR	60-Day BHAR	90-Day BHAR	1-Year BHAR
1995 x Optimism	-0.011	-0.001	0.095	0.102	0.281
	(0.592)	(0.984)	(0.311)	(0.388)	(0.241)
1996 x Optimism	-0.003	-0.016	0.084	0.022	0.132
	(0.889)	(0.545)	(0.390)	(0.857)	(0.601)
1997 x Optimism	-0.008	0.001	0.077	0.075	0.238
	(0.696)	(0.974)	(0.409)	(0.520)	(0.315)
1998 x Optimism	-0.004	-0.006	0.100	0.074	0.221
	(0.841)	(0.825)	(0.286)	(0.532)	(0.362)
1999 x Optimism	-0.005	-0.008	0.113	0.122	0.250
	(0.787)	(0.740)	(0.219)	(0.290)	(0.285)
2000 x Optimism	-0.005	-0.015	0.087	0.125	0.314
	(0.786)	(0.558)	(0.346)	(0.280)	(0.183)
2001 x Optimism	-0.012	0.005	0.088	0.086	0.259
	(0.535)	(0.851)	(0.344)	(0.463)	(0.273)
2002 x Optimism	-0.007	0.004	0.038	0.028	0.184
	(0.737)	(0.888)	(0.695)	(0.819)	(0.453)
2003 x Optimism	-0.013	-0.004	-0.005	-0.068	0.093
	(0.610)	(0.877)	(0.965)	(0.646)	(0.756)
2004 x Optimism	-0.007	0.005	0.049	0.033	0.195
	(0.726)	(0.852)	(0.613)	(0.783)	(0.430)
2005 x Optimism	-0.023	0.001	0.091	0.094	0.573
	(0.436)	(0.984)	(0.507)	(0.587)	(0.102)
2006 x Optimism	-0.005	0.003	0.058	0.047	0.219
	(0.819)	(0.908)	(0.546)	(0.701)	(0.374)
2007 x Optimism	-0.006	-0.004	0.081	0.079	0.250
	(0.764)	(0.884)	(0.383)	(0.496)	(0.290)
2008 x Optimism	0.013	0.001	0.046	0.041	0.197
	(0.563)	(0.981)	(0.671)	(0.766)	(0.476)
2009 x Optimism	-0.005	0.026	0.049	0.099	0.174
	(0.863)	(0.414)	(0.696)	(0.527)	(0.584)
2010 x Optimism	-0.001	-0.005	0.086	0.068	0.272
	(0.955)	(0.860)	(0.447)	(0.633)	(0.348)
2011 x Optimism	-0.053	0.001	0.148	0.246	0.599
	(0.338)	(0.980)	(0.571)	(0.454)	(0.369)
2012 x Optimism	0.001	-0.032	0.142	0.123	0.167
	(0.972)	(0.277)	(0.191)	(0.369)	(0.547)
2013 x Optimism	0.006	-0.009	0.065	0.084	0.113
	(0.826)	(0.799)	(0.622)	(0.613)	(0.737)
2014 x Optimism	0.015	0.007	0.093	0.112	0.369
	(0.483)	(0.803)	(0.374)	(0.392)	(0.167)
2015 x Optimism	-0.007	0.003	0.034	0.036	0.184
	(0.761)	(0.912)	(0.752)	(0.789)	(0.498)
2016 x Optimism	-0.019	-0.004	0.111	0.137	0.049
	(0.506)	(0.891)	(0.420)	(0.428)	(0.890)
2017 x Optimism	-0.002	-0.004	-0.151	-0.180	-0.026
	(0.974)	(0.882)	(0.495)	(0.518)	(0.963)
Intercept	-0.322	-0.118	3.569	3.531	12.512
	(0.737)	(0.926)	(0.432)	(0.537)	(0.281)
Base Effects	Yes	Yes	Yes	Yes	Yes
Observations	1,089	1,939	1,083	1,081	1,065
Adjusted R-Square	0.000	0.013	0.024	0.031	0.010

I find that there is no significant correlation between any year and long-term filing returns. Additionally, no interactions between optimism and year provide significant correlation. The same is true in years of recessions (2001, 2008-09), which indicates there is no specific effect that recessionary years have on predicting correlation with future returns.

Next, I sought to answer the questions: does the relation between optimism and returns differ by industry? What interaction exists between certain industries and sentiment? Using the GIC codes for each transaction, I performed regression analyses for each industry group individually and together with optimism. The results are contained in Tables 10a and 10b.

Table 10a: Regression by GIC Industry Groups

Table 10a explores if certain industry groups are better predictors of future returns. Regression coefficients and p-values for the sentiment optimism and each individual GIC industry group are shown in Columns (1-5). Significance at the 10%, 5%, and 1% level is shown with *, **, ***, respectively.

	(1) 3-Day BHAR	(2) 30-Day BHAR	(3) 60-Day BHAR	(4) 90-Day BHAR	(5) 1-Year BHAR
Optimism	0.000	-0.003***	0.021***	0.027***	0.001
	(0.848)	(0.010)	(0.000)	(0.000)	(0.960)
Materials	0.002	0.011	0.014	0.042	0.152
	(0.838)	(0.192)	(0.706)	(0.375)	(0.112)
Industrials	-0.002	0.014**	0.060**	0.095***	0.132*
	(0.677)	(0.033)	(0.028)	(0.006)	(0.060)
Consumer Discretionary	0.003	0.017***	0.053**	0.104***	0.154**
	(0.539)	(0.007)	(0.048)	(0.002)	(0.026)
Consumer Staples	0.003	0.012	0.072*	0.119**	0.188*
	(0.698)	(0.164)	(0.073)	(0.018)	(0.067)
Healthcare	-0.002	0.011*	0.026	0.052*	0.100
	(0.675)	(0.076)	(0.291)	(0.096)	(0.118)
Financials	0.001	0.004	0.062***	0.100***	0.151**
	(0.919)	(0.531)	(0.009)	(0.001)	(0.013)
Information Technology	-0.003	0.000	0.047**	0.071**	0.040
	(0.501)	(0.988)	(0.045)	(0.016)	(0.503)
Telecom	-0.008	0.003	0.078**	0.123**	-0.027
	(0.321)	(0.709)	(0.049)	(0.014)	(0.792)
Utilities	0.007	0.005	0.038	0.072	0.106
	(0.449)	(0.621)	(0.421)	(0.229)	(0.382)
Real Estate	0.007	-0.022	0.053	0.061	-0.026
	(0.659)	(0.260)	(0.502)	(0.536)	(0.896)
Intercept	-0.011	0.135**	-1.084***	-1.380***	-0.125
	(0.836)	(0.016)	(0.000)	(0.000)	(0.852)
Observations	1,089	1,936	1,083	1,081	1,065
Adjusted R-Square	-0.005	0.011	0.017	0.023	0.005

The results in Table 10a indicate that multiple unique industries (specifically Information Technology, Telecom, and Consumer Discretionary Goods) are significant at explaining returns at the 30-day, 60-day, 90-day, and 1-year horizons alone. Thus, certain industry groups on average tend to have statistically significant returns nothing to do with optimism. These significant correlations are widespread throughout various industries; however, the results present a difficult strategy in practice to utilize.

Table 10b: Regression by GIC Industry Groups - Optimism Interaction

Table 10b answers the question if certain industry groups are better predictors of future returns (BHARs across different horizons) when considering the interaction effects between individual years and the sentiment optimism. Results indicate certain industries are better predictors and when combined with optimism the returns are amplified. The regression coefficients and p-values for the sentiment optimism and each individual GIC industry group are shown in columns (1-5). No time or size controls. Significance at the 10%, 5%, and 1% level is shown with *, **, ***, respectively.

	(1) 3-Day BHAR	(2) 30-Day BHAR	(3) 60-Day BHAR	(4) 90-Day BHAR	(5) 1-Year BHAR
Materials x Optimism	0.003	-0.005	0.010	0.027	0.105
	(0.655)	(0.480)	(0.743)	(0.462)	(0.160)
Industrials x Optimism	-0.001	-0.006	0.008	0.011	0.046
	(0.860)	(0.298)	(0.743)	(0.737)	(0.469)
Consumer Discretionary x Optimism	0.002	-0.007	0.026	0.057*	0.220***
	(0.776)	(0.163)	(0.318)	(0.087)	(0.001)
Consumer Staples x Optimism	0.001	-0.009	-0.021	0.017	0.090
	(0.845)	(0.179)	(0.483)	(0.664)	(0.240)
Healthcare x Optimism	-0.004	-0.001	0.003	-0.01	-0.045
	(0.421)	(0.801)	(0.888)	(0.744)	(0.481)
Financials x Optimism	0.002	-0.003	0.02	0.027	0.070
	(0.677)	(0.519)	(0.375)	(0.342)	(0.228)
Information Technology x Optimism	0.000	-0.007	0.069***	0.071***	-0.007
	(0.986)	(0.152)	(0.001)	(0.008)	(0.899)
Telecom x Optimism	-0.001	-0.007	0.037	0.085**	0.002
	(0.882)	(0.293)	(0.282)	(0.050)	(0.980)
Utilities x Optimism	0.012	-0.004	-0.006	-0.018	0.043
	(0.246)	(0.544)	(0.894)	(0.767)	(0.723)
Real Estate x Optimism	-0.277	0.305	0.009	0.09	0.716
	(0.372)	(0.467)	(0.995)	(0.961)	(0.847)
Intercept	-0.005	-0.110	0.240	0.296	1.862
	(0.982)	(0.586)	(0.799)	(0.804)	(0.440)
Base Effect	Yes	Yes	Yes	Yes	Yes
Observations	1,089	1,936	1,083	1,081	1,065
Adjusted R-Square	-0.009	0.009	0.036	0.035	0.020

Table 10b indicates the interaction effect between industry group and optimism as predictors of returns. Here, information technology and telecommunications, together with optimism, each express significant correlation with future returns. Technology, in particular, exhibits strong positive correlation in the long-term horizons.

This technology correlation is a result of managers' enthusiasm and optimism regarding transactions, but may have more to do with understanding what makes this industry group more difficult to understand. Managers may express more optimism regarding a technology transaction when the technology itself or the integration is complex, in an attempt to calm investors who may not perfectly understand the fundamental technology itself. Thus, latent variables exist here. Nonetheless, I note strong correlation between optimism and the technology industry as a predictor of future returns. This industry specific correlation only occurs in 5 of 59 sub-industry groups, equating to ~8.5% of specific industries wherein this tool could be successfully used. Given the prevalence of transactions in these sub-industries, this equates to ~27.0% of transactions in the dataset that exhibit statistically significant positive correlation between sentiment and returns.

Lastly, I tested to see if outliers in size could have an undue impact on predicting future returns. Eliminating the top 5% and bottom 5% of total observations according to acquirer size, I analyzed if optimism's strong positive long-term correlation was being influenced unduly. The results are found in Table 11 that follows.

Table 11: Regression without Size Outliers

Table 11 seeks at answer the question if certain outliers of acquirer size significantly affect the results between optimism and returns across different horizons. The table displays regression coefficients and p-values for the sentiment optimism across time with all observations in Columns (1-5). Columns (6-10) contain the regression coefficients and p-values for optimism with specific size controls, namely by removing the bottom 5% of acquirers by size (micro-cap stocks) and the top 5% of acquirers by size (mega-cap stocks). Significance at the 10%, 5%, and 1% level is shown with *, **, ***, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	All Observations Considered					Without Top and Bottom 5%				
	3-Day BHAR	30-Day BHAR	60-Day BHAR	90-Day BHAR	1-Year BHAR	3-Day BHAR	30-Day BHAR	60-Day BHAR	90-Day BHAR	1-Year BHAR
Optimism	0.001	0.009***	0.016***	0.020***	-0.003	0.001	0.005	0.006	0.012*	0.006
	(0.463)	(0.007)	(0.005)	(0.006)	(0.830)	(0.396)	(0.138)	(0.242)	(0.062)	(0.635)
Intercept	-0.011	-0.480***	-0.834***	-1.030***	0.071	-0.021	-0.296	-0.435	-0.709**	-0.479
	(0.864)	(0.008)	(0.005)	(0.006)	(0.926)	(0.758)	(0.105)	(0.113)	(0.034)	(0.509)
Year controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,089	1,936	1,083	1,081	1,065	919	917	914	912	897
Adjusted R-Square	0.003	0.018	0.031	0.036	0.027	0.012	0.017	0.027	0.032	0.023

By eliminating the bottom and top 5% of observations by acquirer size in Table 11, I find that optimism is still positively correlated with long-term filing returns, though these results are weaker. This suggests that some of the correlation results are being driven by extremely large or small acquirers. As such, the original significant correlation between optimism and post-filing returns is not as strong as initially supposed from Table 7. Though the total number of observations decreases in Columns (6-10) from (1-5), the sample size remains large enough for adequate statistical inference. Additionally – though not displayed in tables in this paper – I find that neither small cap nor large cap acquirers exhibit unique significant correlation between optimism and returns, and as such I do not further analyze if this investment strategy is more applicable in large v. small cap stocks.

Lastly, in result of decreased correlation between optimism and future returns when analyzed without outliers of acquirer size, I sought to better understand how not only how acquirer size might influence results, but also how a target's relative size with its acquirer might interact with optimism. This size effect in M&A transactions was previously studied by Moeller (2004). These relative size effect results are displayed in Table 12.

Table 12: Size Effect Application in Transactions

Table 12 answers the question that the relevant size of a target to its acquirer might influence optimism as a predictor of returns (BHARS across different horizons). Table 12 displays regression coefficients and p-values for the sentiment optimism across time. Optimism and relative size are initially tested individually. Then, the two variables are considered interacting with one another. Significance at the 10%, 5%, and 1% level is shown with *, **, ***, respectively.

	(1) 3-Day BHAR	(2) 30-Day BHAR	(3) 60-Day BHAR	(4) 90-Day BHAR
Optimism	0.001	0.000	0.008	0.008
	(0.599)	(0.876)	(0.306)	(0.409)
Relative Size	-0.009	0.154	-0.729	-1.062
	(0.937)	(0.216)	(0.158)	(0.104)
Optimism x Relative Size	0.000	-0.003	0.015	0.022*
	(0.952)	(0.202)	(0.147)	(0.096)
Intercept	-0.008	-0.034	-0.446	-0.475
	(0.923)	(0.759)	(0.256)	(0.337)
Year controls	Yes	Yes	Yes	Yes
Industry controls	Yes	Yes	Yes	Yes
Observations	1,089	1,936	1,083	1,081
Adjusted R-Square	0.002	0.019	0.034	0.039

In the test results in Table 12, relative size of target to the acquirer is very close to significant negative correlation at the 10% level. This indicates along with prior literature that as the relative size goes up (the target company is larger in comparison to the acquirer), the future returns are negatively impacted. However, when considered in conjunction with optimism, the coefficient becomes significant, being influenced strongly by the optimism sentiment. These results show how the larger the transaction relative to the acquirer, the more difficult abnormal returns become. However, if optimism is truly a significant predictor of long-term filing returns, I expect the returns to be amplified on the larger relative transactions, as indicated by these results.

The R-square variable, measuring the percent of variation in the response variable explained by changes in the independent variable remains very small in each calculation. However, given that I am modeling *abnormal* as opposed to *raw* returns, having a small R-square value is consistent with the finance literature.

VII. Conclusion

Textual analysis in M&A related SEC filings has been performed historically in finance literature. This paper asks if particular "soft" qualitative data within the text of DEFM14A and DEFM14C documents can predict post-filing abnormal returns. The main underlying idea for my thesis is that firms with more specific and concrete strategic rationales for M&A activity are more likely to generate economic wealth via the transaction leading to better future returns.

To answer this question I created a comprehensive dataset of these M&A filings and collected information about each transaction along with the estimated returns for acquiring firms' returns. I then utilized the pre-determined sentiment libraries of the Diction software to identify key significant predictors of future returns. These statistically significant results pointed toward, and then focused around positive correlation between optimism and long-term returns. As management expressed more optimistic language in these SEC filings, returns improved across the 60-day and 90-day returns. These firms which express optimism toward transactions – most specifically in the technology and telecommunication industry groups – are most likely to experience superior subsequent returns to the M&A filings.

A general lack of statistically significant correlation in textual analysis indicates market efficiency on a broad scale, as a significant portion of the market's reaction to M&A transactions is processed in the returns immediately following the announcement. However, the specific results indicated in this paper display market inefficiencies that exist in the pricing of acquiring firms' stock following the public release of corporate M&A filings. I also presume this lack of significance may be due to the sheer number and diversity of factors considered in M&A transactions. M&A strategy may be delineated in proxy letters to shareholders, while other managers choose to refrain from long-winded explanations of transaction rationale, instead inferring that the shareholders already see crossover and logic behind the merger or acquisition immediately following the transaction announcement. Many institutional shareholders may have already discussed merger particulars with company management, factoring this strategic information and the management sentiment into the stock prices.

Importantly, I note that this study speculates a relationship between sentiment variables and returns across a number of investment horizons. This study has not determined causality, nor professes that it does. Rather, this study explores the relationships between these sentiment variables and post-filing returns to determine correlation. The true abnormal returns associated with M&A transactions are more likely driven by the skill, ability, or experience with M&A strategy that management at the acquiring and target firms have. These transactions may experience more accurate forecasted synergies, or be rooted in economically sound, strategic reasons. At face value there is clearly a statistically significant relationship between optimism and returns, but I observe this is likely driven by latent variables.

Previous applications of textual analysis in M&A filings have been applied sporadically to various texts, including documents outside of the DEFM filings. Methodologies used have varied from the method I used, and thus, these results are varied. The results described in this paper indicate general weak and infrequent correlation with M&A documents; however, in particular settings and over different horizons, statistically significant correlation exists between optimism and post-filing returns. As such, an investor who utilizes this strategy of textual analysis followed by investing in acquirers who express the strongest optimism may experience superior returns – and thus, text may be used to predict successful returns to investors in M&A transactions. Applications of textual analysis in M&A filings will continue to improve as software programs improve. There is still much more to be studied to gain a complete picture of all the causes of positive post-filing returns, yet strategy and intention revealed through sentiment will continue to remain a viable method in investing.

References

Anderson, R., & Manns, J. (2017). Engineering Greater Efficiency in Mergers and Acquisitions. *Business Lawyer*, *72*(3).

Arslan-Ayaydin, Ö., Boudt, K. & Thewissen, J. (2016), Managers Set the Tone: Equity Incentives and the Tone of Earnings Press Releases. *Journal of Banking and Finance*, 72: S132-S147.

Ataullah, A., Vivian, A., & Xu, B. (2017). Time-varying managerial overconfidence and corporate debt maturity structure. *The European Journal of Finance*, *24*(2), 157-181.

Bhojraj, S., Lee, C., & Oler, D. K. (2003). What's my line? A comparison of industry classification schemes for capital market research. *Journal of Accounting Research*, *41*(5), 745-774.

Brau, J. C., Cicon, J., & McQueen, G. (2016). Soft strategic information and IPO underpricing. *Journal of Behavioral Finance*, *17*(1), 1-17.

Brockman, P. and Cicon, J. (2013), The Information Content of Management Earnings Forecasts: An Analysis of Hard Versus Soft Information. *Journal of Financial Research*, 36: 147–174.

Christensen, C. M., Alton, R., Rising, C., & Waldeck, A. (2011). The big idea: The new M&A playbook. *Harvard Business Review*, *89*(3), 48-57.

Cohen, L., Malloy, C. J., & Nguyen, Q. H. (2016). Lazy prices. *Harvard Business School Faculty Research*. Accessed 15 Apr. 2017.

DellaVigna, S., & Pollet, J. (2009). Investor inattention and Friday earnings announcements. *Journal of Finance, forthcoming*.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... & Münkemüller, T. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27-46.

Fortney, Steven, Karl Diether. (2016) Mergers and Acquisitions Document Vector Space Analysis. *BYU Orca Grant*s. Accessed 14 Apr. 2017. <http://jur.byu.edu/?p=20084>.

Harford, J., & Schonlau, R. J. (2013). Does the director labor market offer ex post settling-up for CEOs? The case of acquisitions. *Journal of Financial Economics*, *110*(1), 18-36.

Harp, R. P., C. E. Carroll. (2010). Diction Manual Help Guide. *Digitext*. Accessed 16 Feb. 2018.

Hitt, M., Harrison, J., Ireland, R. D., & Best, A. (1998). Attributes of successful and unsuccessful acquisitions of US firms. *British Journal of Management*, *9*(2), 91-114.

Hoberg, G., & Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. *The Review of Financial Studies*, *23*(10), 3773-3811.

Homburg, C., & Bucerius, M. (2006). Is speed of integration really a success factor of mergers and acquisitions? An analysis of the role of internal and external relatedness. *Strategic Management Journal*, *27*(4), 347-367.

Hsieh, J., & Wang, Q. (2008). Shareholder voting rights in mergers and acquisitions. *Georgia Institute of Technology working paper*.

Investopedia. (2018). SEC Form DEFM14A. Accessed 4 Apr. 2017. <https://www.investopedia.com/terms/s/sec-form-defm14a.asp>

Johnson, M. W., Christensen, C. M., & Kagermann, H. (2008). Reinventing your business model. *Harvard Business Review*, *86*(12), 57-68.

Kahveci, E. (2016). The Tone of Annual Reports as a Strategic Performance Management Tool: Application on Turkey's Borsa Istanbul Corporate Governance Index Firms. *Journal of Economics Finance and Accounting*, *3*(3).

Karpoff, J. M., Malatesta, P. H., & Walkling, R. A. (1996). Corporate governance and shareholder initiatives: Empirical evidence. *Journal of financial economics*, *42*(3), 365-395.

Li, F. (2006). Do stock market investors understand the risk sentiment of corporate annual reports? *University of Michigan working paper*.

Li, F. (2010). Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature*, *29*, 143-165.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35-65.

Moeller, S. B., Schlingemann, F. P., & Stulz, R. M. (2004). Firm size and the gains from acquisitions. *Journal of financial economics*, *73*(2), 201-228.

Muslu, V., Radhakrishnan, S., Subramanyam, K. R., & Lim, D. (2009). Forward-Looking Disclosures in the MD&A and the Financial Information Environment. *Working paper, University of Texas at Dallas and University of Southern California.*

Routledge, B. R., Sacchetto, S., & Smith, N. A. (2013). Predicting Merger Targets and Acquirers from Text. *Working Paper, Carnegie Mellon University*.

Schoenberg, R. (2006). Measuring the performance of corporate acquisitions: An empirical comparison of alternative metrics. *British Journal of Management*, *17*(4), 361-370.

Schwert, G. W. (1996). Markup pricing in mergers and acquisitions. *Journal of Financial economics*, *41*(2), 153-192.

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

Tudor, C. L., & Vega, C. (2014). A Review of Textual Analysis in Economics and Finance. *Communication and Language Analysis in the Corporate World*, *2010*, 122-139.

WSJ, Dealogic. (2017) Global Full Year M&A Review Presentation. *Dealogic.com*. Accessed 5 Apr. 2017. <https://publishing.dealogic.com/ib/DealogicGlobalMAReviewFullYear2016FINALMEDIA.pdf>

Yan, S. (2015). Managerial attitudes and takeover outcomes: Evidence from corporate filings. *Journal of International Financial Markets, Institutions and Money*, *35*, 30-44.