Using Machine Learning to Analyze Merger Activity

Tiffany Jiang

March 16, 2018

Abstract

In this paper, I use machine learning techniques in order to use a new dataset to analyze merger activity - a firm's annual 10-k SEC statement. In this statement contains description about firm and firm product that I believe will capture harder to observe variables that could do well with explaining mergers. I find that the lasso and the ridge regularization techniques have found not only words that align with previous merger theory, but also some interesting variables that were not previously considered. Using this new technique, I obtained a predictive model that yields an R-squared that ranges from 0.02 to 0.07.

1 Introduction

Merger activity is defined as the consolidation of companies or assets and has been hard to predict in the past. Trying to systematically analyze the patterns of merging companies could assist in understanding growth, economies of scale, or perhaps in identifying collusion motives. Understanding merger patterns could not only prove beneficial for the savvy investor, but also help shape anti-trust policy.

Anti-trust economists are generally interested in collusion because it unfairly tips the scales for a firm to gain more market power and harm consumers. For instance, a firm's growing market power could drive up prices and lower wages and standards of living. It is in societal interest that antitrust authorities are able to identify and punish collusion in order to promote competition. Vertical mergers are often seen as societally desirable, while horizontal mergers are not.

Previously, firm characteristics have often been analyzed using financial data and stock prices. While this may prove to be accurate in capturing a firm's worth or success, there are many characteristics in a firm that are not just determined through pure numerical values alone. There are many aspects to a firm such as recent events, competition, regulations, special operating costs, seasonal factors, labor issues, and insurance matters. In order to measure harder to observe variables, I look to use a text document and examine a firm's 10-K SEC filing. The 10-K is an annual document required by the government for all firms to complete and submit to release to the public.

By taking advantage of the large amount of data available in public records, I estimate the effects of words on predicting merger activity using machine learning techniques. I use words as predictor variables to fit a regression and then use machine learning in two ways. First, natural language processing is used to identify words in the text documents. Second, I fit a sparse model using the combined regularization techniques "ridge" and "lasso" regularization known as "elastic net". The hyper-parameters that the lasso and ridge use are then selected through a list that best minimizes the mean squared error, "trained", and then "tested" to verify accuracy.

Statistical learning problems fall mostly in two categories: supervised or unsupervised learning. Unsupervised learning does not fit a linear regression model as there is no response variable to predict. Logistic regression, which I use as a classification method, estimates a qualitative, binary response. My training dataset is then used to test the accuracy of the predictions that I obtain when I apply my method to previously unseen test data.

Historically, attempts to empirically estimate effects of merger activity have had little success. These attempts have been primarily captured using financial variables and numbers. Lawyers and economists who follow merger incentives, however, point to merger factors that may not be easy to capture through numbers alone. For instance, Lipton (2013) describes reasons such as "technological advancements" or more critically "ego and the desire for size and diversity without regard to profitability." I make use of a large set of annual reports required by the U.S. Securities and Exchange Commission (SEC) in attempt to verify existing theories on mergers through the use of words. Many investment bankers and investors cite reading 10-K reports (Kennon 2017) as a way to examine a company while considering business opportunities. I use words as proxies for hard to observe variables such as merger types or technological change. I hypothesize that innovation is a key predictor of merger activity, and I test for its success as well as verify whether these characteristics can be systematically analyzed in a company's own filing.

2 Understanding Mergers and Acquisitions

Three common ways of looking at merger types are to categorize them into horizontal, vertical and conglomerate mergers. Merger motives are primarily due to financial considerations for the profit maximizing firm, and these considerations are driven by increasing market power, exploiting economies of scale, and eliminating managerial inefficiency. Other motivations include risk reduction by diversifying activity, government policy, and principal-agent problems in which company managers have different interests from the stakeholders and prefer to instead maximize their own income. To examine theoretical models of mergers we have three groups. The first are neoclassical models, which propose that merger waves come from political, economic, industrial, or regulatory shocks. The second are models that demonstrate herding, hubris or agency problems and propose takeovers are led by managerial inefficiency. The third are models that reflect capital market development and attribute mergers to market timing. The second may be hard to measure through words alone.

2.1 Neoclassical models

Coase (1937) is an early proponent of the model suggesting that takeover activity is driven by technological change. A later model by Gort (1969) claims that economic disturbances, such as market disequilibrium, may cause wholesale industry restructuring.

Jovanovic and Rousseau (2001, 2002) builds on Gort's theory, and developed the Q-theory of takeovers, which posits that economic and technological changes cause a higher degree of corporate growth opportunities. Such changes may cause capital to be reallocated to more productive and efficient firms.

2.2 Empirical Verifications

Gentzkow and Shapiro measured a previously difficult to measure variable, *media slant*, by using words from newspapers. They determined whether a newspaper was more Republican or Democratic, and then used this specifically to incorporate in a demand function that maximizes newspaper profits to predict consumer behavior. This was compared with with an actual profit maximizing choice to validate economic preference, and it was found that consumers had a preference for newspapers that were like-minded.

Hoberg and Phillips (2010) determine firm similarity and product differentiation through textual analysis and found that estimating patterns of similarities in this method performed better than SIC or NAICS codes. SIC and NAICS codes had drawbacks of being too broad. For example, numerous technology and web-based firms are under the "business services" industry. Hoberg and Phillips further examined asset complementaries as a way to analyze merger pairs and further predict merger activity.

Bhagwat, Dam, and Harford (2016) use market conditions to test the effects of uncertainty on acquisition by using a firm's financial data that includes size, stock returns, and dividends to find that assets and being in a high acquisition industry increases the probability of being acquired. The model yields R-squared of around 0.02.

Previous attempts to explain merger activity have yielded a psuedo R-squared that ranges from 0.01 to 0.09. (Hasbrouck (1985), Palepu (1986), Morck, Shleifer, and Vishny (1988), Ambrose and Megginson (1992), Shivdasani (1993), Comment and Schwert (1995), Cremers, Nair, and John (2009), Edmans, Goldstein, and Jiang (2012), Chatterjee, John, and Yan (2012), Cocco and Volpin (2013))

Routledge, Sacchetto, and Smith (2017) are the first to use words alone to predict merger activity. The authors cite a number of papers that use various financial information to predict merger activity and note "predicting target firms with any accuracy has proven difficult" (Betton, Eckbo, and Thorburn, 2008). In their study, they use one specific section of a company's 10-k SEC filing, the firm's Management Discussion and Analysis and two-word phrases to fit their model. Their text uses the frequency of words appearing on a document and transform the counts with a logarithm function to account for any right-hand skew. Their main discovery supports the Q-theory of takeovers and find that firms that are struggling financially are more likely to be acquired. Their models range from a 0.01 to a 0.07 R-squared.

Theoretically, neoclassical models have pointed to the ways that increase the likelihood of a firm merging with another. These changes have been hard to measure in the past, and accordingly have not had confident estimates about the magnitude of an effect or a conclusion on the sign. Thus, we turn to a different empirical investigation.

3 Data

The estimates presented below are based on US SEC filings for the periods of 2013-2016. Merger activity are based on aggregate US data from 2013-2017.

3.1 10-k Filings

I start off with textual data because of the use of 10-k's by investment bankers and investors to determine company value. The 10-k differs from other documents such the annual shareholders report in length, detail, and scrutiny and are meant to be lengthy, detailed, and not easily digestible. Successful fund managers have cited reading the 10-k as a way to gauge worthwhile investments and have listed notable sections in the Management Discussion and Analysis, the chairman's letter, the risk factor analysis, proxy statements, earnings adjustments and even footnotes. See table 1 for the full 10-k description.

I use the entire document in my findings. Total observations are 22,418. I try to test whether there is a way to automate the human process determining the characteristics of a company and then use these new predictors to predict merger activity. Natural language processing techniques are ways to quantify business phrases such as "synergies" in a more robust manner or at least in a different way compared to previous methods.

Merger events are drawn through Thomson Reuters' SDC Platinum database. Firms that have been recorded as involved with merger activity are labeled, and those that have not are also marked accordingly.

Observation Count	
Involved in a Merger (Total)	Not Involved (Total)
5,379	17,073
Training Observations	Test Observations
4,303	13,658

4 Model

To study the phenomenon of neoclassical models, one would like information about the characteristics of a firm. To study whether firms abide by diversification incentives or horizontal and vertical merger incentives, one would prefer information on the specifics on a firm such as product descriptions and equipment. While examining a principal agent problem, one would look for information detailing management and executive leadership.

The 10-k documents contain a business description of who and what the company does, subsidiaries it owns, and what markets it operates in, recent events, competition, regulations, and labor issues, operating costs, season factors, and insurance matters as well as a section describing the properties and physical assets of the company. To aid with examining management concerns, the documents contain two pertinent sections: certain relationships and related transactions and director independence and directors, executive officers and corporate governance.

Traditionally, machine learning estimation works best in creating a predictive model. The trade-offs of creating a flexible, nonparametric predictive model are that causal interpretations are often lost. Linear regression is relatively inflexible approach but easy to interpret. Flexible models avoid assumptions of a particular functional form for a model, but require a large number of observations and are more difficult to interpret. A "lasso", what I use to fit my model, relies on

10-k Section Descriptions	
Name	Section Description
Item 1 – Business	This describes the business of the company: who and what the company does, what subsidiaries it owns, and what markets it operates in. It may also include recent events, competition, regulations, and labor issues. (Some industries are heavily regulated, have complex labor requirements, which have significant effects on the business.) Other topics in this section may include special operating costs, seasonal factors, or insurance matters.
Item 1A – Risk Factors	Here, the company lays anything that could go wrong, likely external effects, possible future failures to meet obligations, and other risks disclosed to adequately warn investors and potential investors.
Item 1B – Unresolved Staff Comments	
Item 2 – Properties	This section lays out the significant properties, physical assets, of the company. This only includes physical types of property, not intellectual or intangible property.
Item 3 – Legal Proceedings	Here, the company discloses any significant pending lawsuit or other legal proceeding. References to these proceedings could also be disclosed in the Risks section or other parts of the report.
Item 4 – Mine Safety Disclosures	This section requires some companies to provide information about mine safety violations or other regulatory matters.
Item 5 – Market	Gives highs and lows of stock, in a simple statement. Market for Registrant's Common Equity, related stockholder matters and issuer purchases of equity securities.
Item 6 – Consolidated Financial Data	In this section Financial Data showing consolidated records for the legal entity as well as subsidiary companies.
Item 7 – Management's Discussion and Analysis of Financial Condition and Results of Operations	Here, management discusses the operations of the company in detail by usually comparing the current period versus prior period. These comparisons provide a reader an overview of the operational issues of what causes such increases or decreases in the business.
Item 8 – Financial Statements	Here, also, is the going concern opinion. This is the opinion of the auditor as to the viability of the company. Look for "unqualified opinion" expressed by auditor. This means the auditor had no hesitations or reservations about the state of the company, and the opinion is without any qualifications (unconditional). 1. Independent Auditor's Report 2. Consolidated Statements of Operation 3. Consolidated Balance Sheets 4. Other accounting reports and notes

10-k Section Names - Items 9-15		
Item 9. Changes in and Disagreements With Accountants on Accounting and Financial Disclosure	Item 9A. Controls and Procedures	Item 9B. Other Information
Item 10. Directors, Executive Officers and Corporate Governance	Item 11. Executive Compensation	Item 12. Security Ownership of Certain Beneficial Owners and Management and Related Stockholder Matters
Item 13. Certain Relationships and Related Transactions, and Director Independence	Item 14. Principal Accounting Fees and Services	Item 15. Exhibits, Financial Statement Schedules Signatures

a linear model but uses an alternative fitting producing for estimating the coefficients. However, it is more interpretable than linear regression because it creates a “sparse” regression where only a small subset of variables are selected.

Machine learning prediction methods are often a “black box”, but I attempt to find causality rather than the best predictive model. ** ADD MORE Thus, I use a logistic regression and regularization techniques. I assume that phrases such as “technological”, “commercial”, “marketing”, “integrated”, “data”, “development”, “electronically”, “technical”, and “support” are proxies for technological change as proposed by Coase (1937). “Liabilities”, “loan”, “losses”, “expense”, “adversely”, “adverse”, “negatively”, “fail”, “deteriorate”, “risk”, and “depreciation” are used to test Q-theory of takeovers. “Global”, “established”, “minimum”, “competitive”, “holders”, “forward looking”, “comparable”, “health”, “international”, “respect”, “power”, “properties”, “longterm”, “exceed”, and “trends” are used to test for market power theory. “Promotion”, “training”, “managerial”, “finance” are phrases that test for management inefficiency. I also test directly for mention of acquisition with “acquired”, “consolidated”, “accumulated”, “aggregate”, “integrated”, “cumulative”, “portfolio”, “spread”. For market conditions and any anti-trust considerations, I examine words such as “fluctuations” and “sarbanesoxley”. These proxy variables have limitations of having poor correlation with my intended variable of interest. As these are one-word phrases, it is difficult to directly measure the extent in which it describes the characteristic I wish to examine. For instance, “risk” is used to test for signs of a financially deteriorating firm, but perhaps the phrase “risk” was used in the phrase “little risk”.

As increasingly flexible methods are used, variance will increase and bias decreases. Traditional models want no bias. Machine learning allows some bias and reduces variability (e.g., Lasso, Ridge). The model is penalized for size, i.e., how many coefficients we put into the equation. To adjust, we use a different sample once a model is selected to test for goodness of fit. In order to ensure that the model I created has external validity, I use cross-validation training techniques and separate the dataset to a “training set” and a “test set” where I minimize the error rate for the training observations and then obtain a test error rate. My training set is the two previous years before my "test" year, where I test for the predictive power of the model I estimate two years before. For example, I fit a model with estimated coefficients from years 2013-2014. I then test this on the subsequent year, year 2015, to see if the previous words from the past two years predict year 2015 mergers correctly at all.

The second way I attempt to reduce variance is to use shrinkage methods. The combination of the lasso and ridge regression, or the ‘elastic net’, performs variable selection and shrinks coefficient estimates to zero. With this I assume that the observations are uncorrelated and independent and identically distributed variables. Because I cannot observe the error terms of my model, it is difficult to verify this assumption, which is another potential limitation to this model. The lasso’s L1 penalty (Tibshirani 1996) is extremely popular: it yields sparse solutions (some estimated coefficients will be exactly zero) with a number of desirable properties (e.g., Bickel et al. 2009; Wainwright 2009; Belloni et al. 2011; Buhlmann and van de Geer 2011), and the number of nonzero estimated coefficients is an unbiased estimator of the regression degrees of freedom (which is useful in model selection; see Zou et al. 2007). The number of nonzero estimated coefficients is an unbiased estimator of the regression degrees of freedom (which is useful in model selection; see Zou et al. 2007).

To translate the words into variables to perform a regression on, I run the documents into R and transform the words into a matrix where a vector contains the information about the frequency that words appear in the document, a dummy of “0” or “1” for each observation on each word variable.These vectors become the data that we fit the model into. This is another limitation of my model, where I hope to fit the proportion of words relative to the number of documents I have and use words that are more relatively significant to form my matrix.

Given our explanatory variables, our prediction y will have a binary value: 1 for a prediction that a firm will be involved in the merger as either a target or be the takeover company, 0 for a prediction that will not. The prediction formula is:

$$\hat{y} = \arg\max_{y} p(y|\mathbf{x}) \tag{1}$$

To model the relationship between $p(X) = Pr(Y{=}1/X)$ and X, we use the logistic regression:

$$p(y = 1|\mathbf{x}) = \frac{\exp(\beta_0 + \beta^{\tau} x)}{1 + \exp(\beta_0 + \beta^{\tau} x)} = \frac{1}{\exp(-\beta_0 - \beta^{\tau} x)} \tag{2}$$

Table: Words Chosen	
Theory	Words
Technological Change	“technological”, “commercial”, “marketing”, “integrated”, “data”, “development”, “electronically”, “technical”, and “support”
Q-theory	“Liabilities”, “loan”, “losses”, “expense”, “adversely”, “adverse”, “negatively”, “fail”, “deteriorate”, “risk”, “depreciation”
Market Power	“Global”, “established”, “minimum”, “competitive”, “holders”, “future”, “comparable”, “health”, “international”, “respect”, “power”, “properties”, “longterm”, “exceed”, and “trends”
Management Inefficiency	“Promotion”, “training”, “managerial”, “finance”
Direct Signals	“acquired”, “consolidated”, “accumulated”, “aggregate”, “integrated”, “cumulative”, “portfolio”, “spread”

To fit this logistic regression, the parameters are fit through a maximum likelihood function:

$$l(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i':y_i'=0} (1 - p(x_i\prime)) \tag{3}$$

The final equation is:

$$\hat{\beta} = \arg\underset{\beta}{max} \sum_{n=1}^{N} \log p(y_n|x_n) + \lambda_1||\beta||_1 + \lambda_2||\beta||_2^2 \tag{4}$$

5 Results

I found that the words chosen by the lasso are statistically significant and align with the theory. Traditionally, standard errors are not provided for a logistic regression that uses the regularization techniques. However, a popular way to determine some sort of significance is to bootstrap the standard errors. The bootstrap is a way to obtain the standard errors of coefficients for a wide range of different models. The variance of coefficients in your model is minimized in an equation, and is calculated repeatedly in a sample from the data set and averaged. A massive loop over the lasso routine is performed, where a large number of samples is drawn with replacement and implemented with the lasso on each sample. The standard errors can be retrieved directly by observing the variation across coefficient estimates in each sample draw.

Another way of validation for my model was run automatically through my glm-net R package. This technique is called the k-fold cross validation, where observations are randomly divided into sets or groups of approximately equal size. The first fold is a validation set and then used to fit the remaining folds. The mean squared error is computed for each fold, and then a final CV estimate is computed by averaging the values.

I use cross-validation techniques not to directly compute the mean squared error (in this case, the minimum MSE is 1.249354), but to help find the lambda, or the penalty, that is the minimizes the mean square error for my coefficient. This penalty that I find is then used into my model in order to determine the best coefficients to use to predict.

I have listed the most frequent terms in the documents below. It is to note that in choosing words, you cannot pick words that appear too frequently across all documents, nor can you pick words that are too few. However, this is what I view to be the summary statistics.

6 Discussion of Results

I had two major results in mind at the start of my model. The first was to think about which words would potentially be very predictive, and the words chosen were based on the theories that were

Table 1: Words Picked by the Lasso

test	2007	unrecognized	improvements	2005	testing
−0.086	−0.080	−0.079	−0.069	−0.065	−0.064
out	2006	2018	accounted	allocated	component
−0.062	−0.061	−0.057	−0.044	−0.041	−0.041
contributions	competition	respective	longlived	locations	102
−0.039	−0.038	−0.035	−0.032	−0.031	−0.030
discounted	combined	next	areas	strategy	yield
−0.030	−0.030	−0.029	−0.029	−0.027	−0.025
then	trends	retirement	earned	environment	par
−0.025	−0.025	−0.023	−0.023	−0.023	−0.023
treasury	protection	final	investing	cumulative	low
−0.023	−0.022	−0.022	−0.020	−0.020	−0.020
performed	impaired	several	maturities	leased	supplemental
−0.020	−0.020	−0.019	−0.018	−0.018	−0.018
weightedaverage	assumed	reflected	outside	primary	directly
−0.017	−0.017	−0.017	−0.017	−0.016	−0.015
domestic	registration	acquired	summary	competitors	institutions
−0.013	−0.013	−0.011	−0.011	−0.010	−0.010
termination	taxable	active	developed	north	resulted
−0.009	−0.009	−0.008	−0.007	−0.007	−0.006
designed	generated	classified	delivery	measures	therefore
−0.006	−0.005	−0.004	−0.004	−0.004	−0.003
excluding	although	sources	highly	law	building
−0.003	−0.003	−0.003	−0.002	−0.002	−0.001
restrictions	completion	lives	projected	recovery	
−0.001	−0.001	−0.001	−0.0003	−0.0001	−0.0001
right	professional	covered	authorized	consist	need
−0.00003	−0.00002	0.0001	0.0003	0.001	0.002
individual	reflect	10q	1934	application	dependent
0.003	0.005	0.007	0.007	0.009	0.009
contained	unless	actions	treatment	organizations	ending
0.010	0.010	0.011	0.011	0.012	0.016
approach	proxy	major	recognize	name	proprietary
0.016	0.017	0.018	0.022	0.022	0.026
timing	variable	strategic	gaap	transfer	standard
0.027	0.032	0.035	0.045	0.048	0.059
evaluate	2017	entitled	adoption	early	fasb
0.075	0.113	0.136	0.142	0.155	0.290
goods	2019				
0.204	0.604				

Table 2: Goodness of Fit, Years 2013-2015

	Dependent variable:
	Actual firms that merged in 2015
Lasso Model based on 2013-2014 Observations	-0.196^{***}
	(0.041)
Constant	0.634^{***}
	(0.032)
Observations	8,655
R^2	0.003
Adjusted R^2	0.002
Residual Std. Error	0.499 (df = 8653)
F Statistic	22.649^{***} (df = 1; 8653)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3: Goodness of Fit, Years 2014-2016

	Dependent variable:
	Actual firms that merged in 2016
Lasso Model based on 2014-2015 Observations	-0.288^{***}
	(0.036)
Constant	0.742^{***}
	(0.020)
Observations	9,468
R^2	0.007
Adjusted R^2	0.007
Residual Std. Error	0.490 (df = 9466)
F Statistic	64.577^{***} (df = 1; 9466)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: With just a Lasso, significant variables

(Intercept)	acquired
0.236	0.411

Table 5: With Ridge, Technological Growth

(Intercept)	technology	commercial	marketing	international
0.236	0.008	−0.001	−0.001	−0.030
data	development	foreign	technical	plans
0.002	0.010	0.007	0.008	0.004

Table 6: With Ridge, Q-theory

Liabilities	loan	losses	expense	adversely
0.011	−0.008	0.010	0.008	0.017
adverse	negatively	fail	deteriorate	depreciation
0.012	0.010	0.007	0.008	0.004

Table 7: With Ridge, Market Power

Growth	established	market	competitive	holders
0.025	−0.008	0.0001	0.008	0.017
future	comparable	longterm	international	power
0.002	0.010	0.001	0.042	0.012

described above. The second result is where the machine learning comes into play and allows me to understand what the best predictor variables are amongst the ones that are given to the model. For instance, I had around 312 variables that I fit into the model, and then the Lasso picked out the following variables as the best, unbiased estimates of predictor variables for merger activity.

I find some interesting results. Of course, as I had no prior hypothesis except for words that could align with previous merger theory, I can only speculate to the reasons why a merger occurred with the variables at hand. For instance, the most significantly positive variable that drove up the probability of a firm merging was the variable "2019". Perhaps firms that plan for future events and consistently talk about forward-looking events are more likely to seek to merge in order to expand upon firm objectives and create opportunity for advancement of services or business products, or gain an edge on competition. Other variables such as "impaired" or "unrecognized" which have a negative influence on merger activity also make sense. Those that seem to be more of an unsuccessful endeavor perhaps may not be as an attractive as a merger target.

There are obvious drawbacks to this model. Nonsensical that may not perhaps provide much explanatory power are picked out by the lasso - for instance, there is "then" and "unless" that were included. However, although the words may not have not much explanatory power, the words still provide a valid statistically significant predictive model. In the same way economic models include controls in a regression that may not be of much interest, "control" words can be included in order to assist in helping the model out.

Lastly, when all documents were picked out by the lasso, there was only one seemingly significant variable - "acquired". However, this can be seen as a sort of validation test that the statistical methods implemented were valid - as it makes sense. Firms who say that they will acquire or were acquired will include that description in their annual statement, and thus, it seemingly appears that the word "acquired" has much predictive power.

7 Conclusion and Next Steps

There is enormous potential for machine learning given the amount of new data language and words can provide as a new form of information. Traditionally, machine learning has been popularized as a unsupervised learning technique, where there is no requirement of a parameter to predict upon. For understanding document similarity and the value of words as predictive power, a linear discriminant analysis is often used. This is where the first steps are to use machine learning techniques to determine how similar documents that merged are to each other, and how similar documents that didn't merge are to each other. What can then be used is to look at words in terms of 2-tokens- that is, examine words as phrases rather than just singular tokens.

Another step could to be separate the predictor variable and determine whether or not a firm was the acquired or the doing the acquiring. This can later be used to better separate the theory - in what position is a firm better to do the acquiring, and what type of firms are attractive merger targets?

However, I find that the words do have predictive power. Although the model fit is not drastically accurate, I think that my very basic machine learning model provides a critical first step in using machine learning to answer how to better analyze merger activity.

8 Appendix

Table 8: Top 100 words and their frequency count

word	freq
company	6,546,647
financial	4,905,807
december	4,815,425
million	4,578,425
may	4,110,697
2013	4,075,164
stock	3,842,785
2012	3,553,304
income	3,496,941
2014	3,473,933
net	3,439,453
value	2,977,859
cash	2,951,022
assets	2,949,461
these	2,685,970
year	2,636,903
interest	2,519,998
under	2,477,473
statements	2,470,777
operations	2,457,814
tax	2,450,091
common	2,436,549
business	2,398,305
ended	2,383,669

9 Works Cited

2017, Routledge, Sachetto, and Smith, “Predicting Merger Targets and Acquirers from Text.”
2017, Kennon, "What is a 10-K and Why Should an Investor Read It?"
2010, Gentzkow and Shapiro, "What Drives Media Slant?"
2010, Hoberg and Phillips, "Text-based Network Industry Classifications"
2017, Gentzkow, Kelly, and Taddy, "Using Text as Data"
2013, Lipton, "Predicting Future Merger Activity"
Gregoriou and Renneboog, "Understanding mergers and acquisitions: activity since 1990"

Table 9: Top 100 words and their frequency count, continued

word	freq
2015	$2,307,452$
total	$2,162,221$
shares	$2,122,668$
consolidated	$2,000,399$
related	$1,959,298$
2011	$1,938,866$
fair	$1,914,157$
securities	$1,911,431$
market	$1,901,193$
agreement	$1,869,478$
costs	$1,843,889$
years	$1,839,842$
loss	$1,834,817$
operating	$1,827,386$
including	$1,772,055$
management	$1,752,407$
certain	$1,752,086$
table	$1,663,531$
products	$1,659,522$
capital	$1,641,630$
future	$1,612,867$
services	$1,592,832$
sales	$1,591,853$
during	$1,586,554$
credit	$1,583,294$
based	$1,572,340$
rate	$1,527,627$
results	$1,500,229$
period	$1,454,946$
form	$1,439,973$
equity	$1,430,217$
plan	$1,422,192$

Table 10: Count continued

word	freq
inc	1,421,218
notes	1,420,993
share	1,383,101
due	1,380,390
information	1,378,197
amount	1,375,947
over	1,370,765
new	1,368,598
expense	1,351,481
price	1,326,812
liabilities	1,307,753
current	1,287,339
compensation	1,278,168
revenue	1,274,276
report	1,234,421
increase	1,231,662
expenses	1,228,223
accounting	1,227,824
fiscal	1,227,756
result	1,204,940