Content of Annual Reports as a Predictor for Long Term Stock Price Movements

John A. Doucette and Robin Cohen
Cheriton School of Comp. Sci.
University of Waterloo
200 University Avenue West
Waterloo, ON N2L 3G1 Canada
{j3doucet, rcohen}@cs.uwaterloo.ca

Abstract

This paper examines the possibility of automatic extraction of future stock price information from the annual Form 10-K produced by publicly traded companies in the United States of America. While previous approaches to automatically interpreting corporate documents have tended to utilize extensive expert knowledge to preprocess and analyze documents, our approach inputs documents verbatim to a compression classifier. We demonstrate the effectiveness of the new approach on a newly constructed dataset based around the Dow Jones Industrial Average over the period 1994-2009. We find statistically significant increase in average returns of stocks recommended by the new system as compared with the Dow as a whole. Also examined are two hypotheses regarding the predictive power of 10-K reports. First, whether congressional attempts to make Form 10-K filings more informative had a measurable impact, and second, whether the filings have long-term predictive value in a dynamically changing market.

Introduction

This work examines several hypothesises regarding the application of Dynamic Markov Compression (DMC) to the prediction of long term stock price movements, based on the information present in annual regulatory filings. This problem has recently seen a substantial amount of interest following Li's 2006 findings that the level of linguistic complexity in such documents was a good predictor of company quality (Li 2008). Currently, most approaches incorporate considerable expert knowledge and preprocessing of the data in order to produce good results. In contrast, we examine the performance of a simplistic trading strategy, based on a classifier which is provided with raw (i.e. unprocessed) regulatory filings as input. We find that the resulting system is able to reliably differentiate between profitable and unprofitable stocks. Further experiments show that the systems performance improves as additional data is provided, and that attempts to regulate the language present in such reports do not appear to produce measurable effects.

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Background

We begin by examining related work using the annual regulatory filings (specifically Form 10-K) of publicly trade companies for automated analysis of stock prices and company outlooks. While there has been considerable recent work, much of it utilizes copious expert knowledge, and experimental datasets are largely unstandardized.

We follow this review with a brief discussion of Dynamic Markov Compression and its application to classification tasks, especially spam detection. Parallels are drawn between spam and the reports of under-performing companies.

Related Work

Form 10-K is a document which the United States Securities and Exchange Commission (SEC) requires publicly traded companies to file on an annual basis, providing a comprehensive picture of the company's performance over the past year. Information provided in the form includes 3rd-party audited financial statements, comprehensive analysis of the companies' exposure to risk factors (e.g. which commodities are most important to production; foreseeable factors that could dramatically change demand etc.), any legal filings, and future performance forecasts written by upper management.

Form 10-K, and the annual reports to shareholders which present simplified information from the corresponding 10-K filings, are widely considered by investors to be a good source of predictive information with respect to a company's future performance. In a large international study of the reports' perceived usefulness, they were consistently ranked as the most important item used to make investment decisions (Chang, Most, and Brain 1983). Annual reports were also rated more useful by professional investors and analysts than private investors. One possible reason for this is found in more recent research which suggests annual reports should be classified as technical literature, putting them outside the ready understanding of the general public (Jones and Shoemaker 1994). A related 1993 study of investors in the Netherlands produced similar findings, including showing that professional investors and analysts considered previous years' reports to have valuable information as well (Vergoossen 1993). Following rapid improvements in natural language processing (NLP) and information retrieval

(IR) techniques over the last 20 years, there has been increasing interest in the automatic processing and classification of these reports as a means of both eliminating human bias in their evaluation and improving the speed with which they can be processed. With respect to the later point, the savings could be considerable: professionals often take more than 2 hours to parse a single report (Vergoossen 1993).

Recent work in this area begins with Li's 2006 paper, which revealed a correlation between the readability of section 7 (future performance forecasts) in annual reports and future earnings(Li 2008). Reports which had a higher reading level (i.e. longer or more complex words and sentences) and/or comprised more pages tended to produce lower earnings, this effect being attributed to the tendency of management to avoid the disclosure of negative information (Bloomfield 2002). Butler and Keselj built on Li's work from a machine learning prospective by attempting to classify future stock performance using the readability of their annual reports and n-grams derived from the reports as inputs to a support vector machine. They report exceptional annualized returns of 80% over the period 2003-2008, during which the S&P500 (a common performance index) *lost* around 20% of its value (Butler and Kešelj 2009). Balakrishnan et. al. performed a similar experiment using a regression model and analyze the performance of different categories of stocks, reporting annualized returns of 12.16% per year (Balakrishnan, Qiu, and Srinivasan 2010). Feldman et. al. attempted to measure the change in linguistic tone of section 7 over time for single companies(Feldman et al. 2010). They found that changes in tone had a significant impact on stock prices over the 2 days following the SEC filing date. Tangentially, very recent work by Cecchini et al and Humpherys et al attempt to detect the presence of fraudulent information in reports using IR techniques to identify predictive keywords (Cecchini et al. 2010), and linguistic changes in the tones of reports which contained fraudulent information (Humpherys et al. 2011) respectively.

One factor common in the previous approaches to this area is their underlying complexity. While good results are often produced, the steps involved in obtaining them do not serve to make the reports' information content more accessible to the general public. The measures of linguistic tone used in (Humpherys et al. 2011)(Feldman et al. 2010) for example, have been shown to be dependent on expert knowledge insofar as financial jargon tends to use words in ways which differ substantially from their conventional meanings (Loughran and McDonald 2011). For example, while "liability" might indicate negative tone in an ordinary document, it is unlikely, by itself, to do so in a financial document. All companies have liabilities. Similarly, even the steps involved in preprocessing an annual report or 10-K filing in order to make it suitable for use with a conventional classification system involves considerable effort. Annual reports are not required to conform to a particular data format, and appear in forms as diverse as plain text and image files. Since use of these systems requires a domain expert to facilitate preprocessing, and said systems can produce significant benefits, they are unlikely to see public use.

A second criticism of previous work is their tendency to use unpublished and imprecisely described datasets, constructed specifically for the work in question. This is probably due to the considerable difficulties in obtaining and prepreprocessing historical reports and pricing data, especially for bankrupt or defunct companies, but it does present a problem in reproducing their results. With that in mind, the dataset constructed for evaluation in this paper is designed as an objective benchmark to facilitate reproduction of results.

Dynamic Markov Compression for Classification

Dynamic Markov Compression (DMC) is a compression algorithm developed by Cormack and Horspool (Cormack and Horspool 1987), which functions by constructing a Markov model of a bitwise representation of the input message. The model is used to inform the choice of compression coding scheme for the next bit, providing a highly flexible and effective compression method . Like other compression algorithms, DMC is readily converted into a classification algorithm. A compression model is derived separately for each class in the training data. Test data is then compressed under each model separately, and the relative compression levels are indicators of the test data's similarity to each of the various classes (Büttcher, Clarke, and Cormack 2010).

DMC has been applied with success in the domain of spam classification (Bratko et al. 2006), which suggests it might also be well suited to classification of regulatory filings. As discussed above, managers of under-performing companies, consciously or unconsciously, attempt to obfuscate regulatory filings (Bloomfield 2002)(Li 2008). Consequently, classification of such reports is an *adversarial* task, much like spam classification.

Experiment

Three hypotheses are evaluated. First, we test whether a compression classifier based around DMC can make better than chance predictions about the performance of various stocks over a one year period, based on previous years' Form 10-K filings filings. Second, the effect of the SEC "Regulation Fair Disclosure" (RFD) on the model's performance is evaluated. RFD mandated that publicly traded companies must disclose investment information simultaneously to all investors, rather than providing it in private meetings with select professional investors. Previous work has found measurable decrease in performance in classification systems based around the writings of professional investors and analysts(Huang, Zang, and Zheng 2014) following the 2001 implementation of RFD. Thus, we are interested in whether the information content of annual reports increased as a result, indicating that companies were now sharing information more broadly, or remained unchanged. In the event that reports did contain more explicit information, we would expect to see a decrease in the performance of DMC, since obvious information will be capitalized on more rapidly by traders, reducing the potential for DMC to exploit hidden information. Finally, we examine whether DMC improves its performance as more historical data is provided to the model. While one might intuitively suppose this final hypothesis to be true, older reports could be misleading. For

example, a classifier with 10 reports from a company during a period of booming business might take several years to start detecting that the company is headed for bankruptcy.

Data

There is, to the best knowledge of these authors (and others: (Butler and Kešelj 2009)), no publicly available dataset of the sort required to test the hypotheses outlined above. Probably this is due to the high demand for such information, which makes it undesirable to share free of charge. Services like Thomson Reuters run profitable businesses providing this data. Further, stock market data is often difficult to collect, and is not well behaved. For example, most public databases of stock information are indexed by a stock's "ticker symbol" - a unique identifier which is used when trading on an exchange. No constraints exist to prevent companies from adopting defunct ticker symbols as their own. A good example of this is Bell Telephone's symbol 'T'. When Bell's monopoly was broken up, the symbol was transfered to child company AT&T, because it was considered a prestigious item. When AT&T was sold to SBC, SBC adopted 'T' as its own. SBC later changed its name to AT&T, while retaining the same symbol. Consequently, querying for information on the symbol 'T' may bring up information on any subset of these companies. Especially confusing is that when a company changes symbols, its older records are likely to be stored under the new symbol. Thus a query for reports on 'T' in 1999 could bring up reports for SBC, labeled with its present day name (AT&T), as well as reports for the now defunct company which traded as AT&T in 1999. Finally, we are aware of no publicly available database which stores historical pricing information for stocks which are no longer publicly traded, likely due in part to the prodigious amounts of storage space required for such a task.

As a result a new dataset was constructed specifically for this task. Previous authors (e.g.(Butler and Kešelj 2009),(Balakrishnan, Qiu, and Srinivasan 2010)) have used a somewhat arbitrary approach to the construction of their datasets, focusing on specific industries or time periods without ample justification or description, and so their results are difficult to reproduce. Consequently, this dataset is constructed to be a representative sample of the stock market as a whole, over the full period for which data is readily attainable. The set was based around the stocks comprising the Dow Jones Industrial Average (hereafter, the DJIA), a basket of 30 stocks thought to be representative of the American economy as a whole (Website 2014a). The stocks comprising the DJIA are periodically adjusted, so that companies whose fortunes are no longer representative of their industry, or industries that are no longer representative of the market as a whole, are replaced by representative competitors. For example, Walmart was added to the average as a replacement for the department stores it displaced, including Walgreens and Sears. The DJIA was selected for use in this work both because of its widespread usage (Website 2014a)(OConnor and Madden 2006), and because its comparatively small size better facilitated the largely manual data collection than other indexes such as the S&P500.

The new dataset was constructed in three phases. First, the set of stocks comprising the DJIA was determined for each year in the period 1994-2010. This period was selected because 1994 is the earliest year for which companies began to make electronic filings with the SEC. Consequently, 1994 is the first year for which 10-K filings are available through the SEC's public database, EDGAR (Securities and Website 2014). The historical composition of the index was obtained from the Dow Jones Indexes, the holding company which maintains the index (Website 2014b). Second, the EDGAR database was searched by hand to obtain the Form 10-K filings of each company during the period which they were part of the DJIA. The reports were downloaded in their entirety, and any attachments to the filings were appended. Finally, historical data was obtained from the publicly accessible Yahoo Finance database [1] for each company during the period of interest. The data collected during this phase included monthly stock prices, dividend payments, and stock splits. This data was then processed to produce a simple scalar value indicating the year-over-year returns for each company during the period in which they were on the DJIA. Years were assumed to start and end on April 1st.

This date is somewhat arbitrary, but provides consistency to the dataset. The only deadline the SEC provides for 10-K filings is 90 days following the end of the company's fiscal year. Fortunately, although companies are not required to use any particular dates as the start and end of a fiscal year, the companies comprising the DJIA overwhelming use the calender year as their fiscal year, with virtually all reports in the dataset appearing during January and March of the following year. A very small number appear before that time, but none earlier than November. Consequently, the April 1st deadline ensures that all reports have been filed. This approach to computing returns is conservative with respect to the hypothesis of interest, because stock movements in response to positive reports will likely have taken place well in advance of the time at which our models are allowed to invest. Therefore, our models' profits ought to be lower on this dataset than on one where purchases of a particular company's stock are instead permitted on the day its report is released.

Unfortunately, the data set is incomplete, due to the incompleteness of databases from which it is derived. In particular, the lack of price, dividend, and split information for companies that went bankrupt or were involved in mergers or acquisitions during the period of interest was difficult or impossible to obtain from public sources. The omissions fall disproportionately heavily on the earliest years in the data, and that many of the companies in question were not performing especially well (as evidenced by the number of bankruptcies and acquisitions). Consequently, we expect that the returns on this data set will tend to be higher in early years than in later years. This is compensated for in our experiments by using baseline performance values derived from the dataset, rather than from the historic DJIA performance. While the missing data could be collected from newspaper archives, that process was not pursued in this work.

[1] http://finance.yahoo.com/

All told, the collected data comprised some 2.5 gigabytes. For each year, we split the data into two labeled sets. Over-performing stocks were those which had returns greater than the median return for that year. Under-performing stocks, which returned less than the median percentage over a one year period, formed the balance of the data. The reason that median returns are used to split the set is that DMC is sensitive to class balance in the training data. Models which have substantially more training data will tend to compress any data better than models which have substantially less. Consequently, using absolute returns would tend to produce degenerate models for this algorithm. Additionally, we justify this division by noting that investors want to invest in the best performing stocks, not just stocks with positive returns.

The problem examined with respect to this dataset is the classification of stocks into over- vs. under-performing categories on a yearly basis. A secondary criteria of interest is to select stocks such that yearly returns are maximized.

Experimental Setup

The basic experimental setup was as follows. For a given year in the dataset, two DMC compression models were constructed, based on the 10-K filings of all over- and under-performing stocks from all prior years respectively. The filings for each company in the present year where then compressed by each model. Those that were compressed to a smaller size under the over-performing model were labeled "buy", with the rest being labeled "ignore". This approach mirrors one used by other authors in spam detection (Frank, Chui, and Witten 2000)(Teahan 2000). Data is gathered for each year in the period 1995-2009, and then pooled for analysis. The pooling of data ignores the effects due to having larger and larger amounts of data present in the model, but this effect is evaluated independently in a later hypothesis. The purpose of this experiment is simply to determine if the recommendations made by the system over the period of interest are better than chance, and hence, whether there is usable information in the reports.

To test the two remaining hypotheses small modifications were made to the basic experimental setup. For the second hypothesis, one model was evaluated as described above on each year from 1995 to 2000. A second model was evaluated on each year from 2002-2007, and allowed only to use reports from 2001 onwards in its training. The two models were compared on the basis of their over- or under-performance against the DJIA as a whole in each of their respective years. The third hypothesis was evaluated similarly, with DMC being trained each year only on the filings from the immediately preceding year. The data were compared on a year by year basis with the data from experiment 1 to determine the relevance of historical filings to the model.

The labels returned by the models were evaluated on the basis of the returns of the corresponding stocks. Stock returns were simply the percentage gain which an investor would receive if they purchased the stock on April 1st of the year in question, and then sold it on the following March 31st, including all dividends and stock splits. This represents a very passive investment strategy. The choice of a passive strategy is deliberate, both because such a strategy is easy to implement, and because the strategy can reveal the existence of hidden information in reports which is not properly priced into the underlying stock by the market. If DMC earns a profit on out of date information, then the information is not accurately reflected in the stock price.

Results and Discussion

Information Content of Annual Reports

The data from the first experiment consisted of 407 data points, which were used to test the hypothesis that the reports contained useful data. To facilitate statistical analysis, 13 points with yearly returns of over 100% were removed as outliers. The returns of remaining data were found to be statistically consistent with the assumption of normality using a Shapiro-Wilk test (p = 0.08) (Shapiro and Wilk 1965). A Student t-test found that a report labeled by DMC as over-performing (i.e. one which compressed better under the over-performing model) had significantly better expected yearly returns than one which was labeled as under-performing (p ¡ 0.001). The 95% confidence interval for difference in mean return of stocks labeled by DMC as over-performing was versus under-performing was 6.07-18.95%, with a mean of 12.5%.

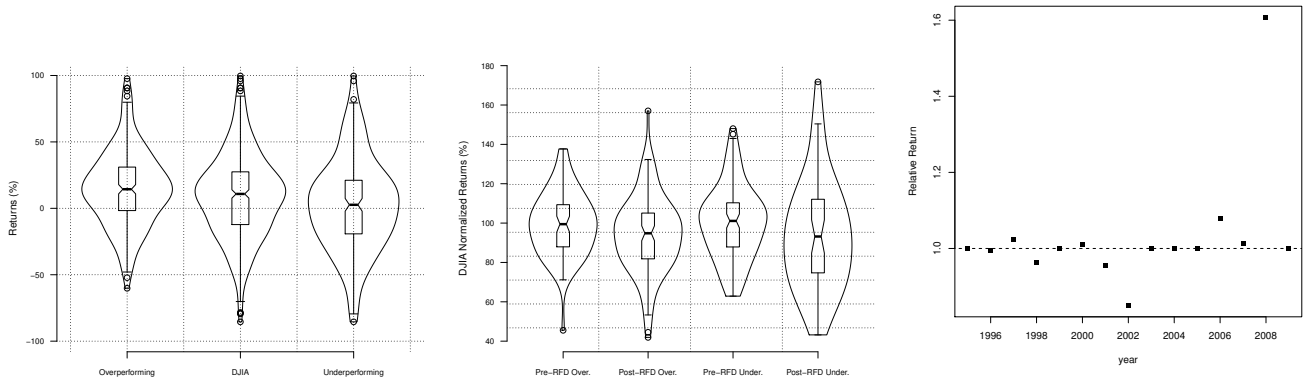
Further, stocks picked by DMC as over-performing were found to have mean returns higher than the DJIA as a whole. A t-test found that the mean performance of stocks picked by DMC was significantly better than investing in the DJIA over the period of interest (p ¡0.05), with a mean improvement of 4.98%, and 95% confidence interval of 0.58-9.38%. Figure 1a summarizes these results with a violin plot[2].

In spite of these promising findings, the models produced were not especially good at partitioning stocks into their "true" class, i.e. at separating them about the median value in each year. The accuracy of the compression classifier in producing the labels assigned in the dataset was only 52%, hardly better than random guessing.

From these results, we conclude that the hypothesis that annual reports contain no useful information for predicting the movements of stock prices should be rejected. Following the recommendations made by a classifier trained on the annual reports produced a significant improvement in the average returns generated over simply buying all stocks in the DJIA. A possible reason for the poor predictive accuracy is that reports during a good year (i.e. a Bull market) might be much more similar to those generated by top performing companies in the past than by poorly performing companies. Consequently, the model will tend to label them all as over-performers and so will be "wrong" about half of them.

The number of companies which are rated over-performing weakly correlated with the movements of the DJIA, with a correlation coefficient of 0.3. Of particular interest though is that the model does exhibit a tendency to rate more companies over-performing when an especially strong

[2]Violin plots show a box plot with a rotated kernel density (smoothed histogram) plot overlaid. This gives a good overview of the distribution of the data. Plots were generated with the vioplot library in R (Team and others 2012)(Adler 2005).



(a) Distribution of returns for stocks labeled by the model as over or under-performing.

(b) Distribution of stock returns with respect to model label, and year.

(c) Scatter plot of relative return of the baseline model over the forgetful model.

Figure 1

bull market is present, and to behave oppositely during a market crash.

Impact of Regulation on Information in Form 10-K

The second hypothesis was evaluated on data gathered as described above. The same outlier removal was performed. Stock returns were normalized against the DJIA in the corresponding year. ANOVA was used for analysis, with the source of the data (years 2002-2007 vs. 1995-2000) and the labels produced by DMC used as explanatory variables for stock return. The interaction term in this model represents the combined effect of report type (pre/post RFD) and model label (over- or under-performing). A significant interaction term would indicate that the change in report type produces a change in the predictive power of the labels produced by DMC. The p-value associated with this term by ANOVA was insignificant (p ¿ 0.8). Consequently, we reject the hypothesis that Regulation Fair Disclosure improved the information content of the Form 10-K filings of DJIA components. Any change in performance is attributable to the differing time periods studied This is particularly obvious in the summary of the results found in figure 1b. The post RFD period contains a market crash, while the pre RFD period does not, which manifests as increased volatility in the returns as well as lower average returns overall.

More Data, More Profits

The third hypothesis was evaluated by constructing a new classification model. This model (hereafter the "forgetful" model) was trained only on the previous years' annual reports, rather than all previous years' reports. Like the baseline model from section , the model was trained and evaluated separately on each year of the data set. After removal of outliers, as discussed in , a Student's t-test was applied to determine whether the classification labels applied by the model were a good predictor of stock returns. The test found no significant difference in performance between stock labeled as over-performers by the model, and those labeled as under-performers (p ¿ 0.29). In fact, the observed mean of returns for the over-performing group was *lower* than the observed mean for the under-performing group, suggesting that the model would have performed better if its output was inverted.

The forgetful classification model was also compared directly with the model model from experiment 1 (the baseline model). ANOVA using the interaction between classification model and label selection as an explanatory variable and returns as the response variable found that the interaction term for classification model and labels had a significant effect on returns (p ¡ 0.005). Consequently, we reject the hypothesis that providing the classifier with more historical training data has no positive effect on its performance. Increasing year was also found to be positively correlated with the ratio of the baseline model's returns to the forgetful model's returns ($\rho = 0.39$). A year-by-year comparison of average returns for the forgetful and baseline models can be seen in figure 1c.

One of the most visible effects in figure 1c is the behavior of the models during the recent stock market crash in 2008. While the baseline model recommended purchase of only Walmart shares that year, the forgetful model suggested buying all 29 stocks presented to it. This suggests that the baseline model could be recalling the earlier market crash in 2002 when making decisions in 2008.

Conclusion and Future Work

Several contributions are worth highlighting in the above passages. First, the creation of a new dataset for stock market prediction could provide a useful benchmark for future researchers. We intend to make the set publicly available in the near future. Second, to the best of our knowledge, no previous work has examined the use of compression classifiers for market prediction based on annual reports and company filings, with previous approaches adopting more widely used machine learning methods like Support Vector Machines and Bayesian classifiers. We show that compression clas-

sifiers can be applied with statistically significant benefit to a comparatively small sample of data (around 10% of the size used in previous work (Butler and Kešelj 2009)(Balakrishnan, Qiu, and Srinivasan 2010)), and that larger amounts of data leads to improved performance. Tentatively, we also claim that more data allows the model to recall periods of high volatility in the market, based on its performance during the first market crash in the data set, but much improved performance during the more recent 2008 crash.

The new system is also interesting in that it utilizes no preprocessing of annual reports at all, predicting their content directly from the raw text. This could be a considerable advantage for average users, who may lack the technical expertise to process reports into formats suitable for a conventional classifier (primarily creating features using expert knowledge). This could reduce the benefits derived by domain experts from the subtextual information present in annual reports, by making it more available to typical users.

Further work in this domain with DMC would first seek to construct an expanded dataset, perhaps based around the S&P500 index, in order to expand upon the findings presented here. It might also be interesting so test the impact of the variety of file formats used by different companies in their filings on DMC's performance, or the performance of a hierarchical classifier which incorporates DMC's recommendations into other data, like current market prices, volatility, and company history. It might also be interesting to investigate the use of DMC in fraud detection by comparing the literal sentiment of reports to the recommendations made by DMC.

Acknowledgments

The authors gratefully acknowledge finacial support from the Natural Sciences and Engineering Research Council of Canada.

References

Adler, D. 2005. vioplot: Violin plot. *R package version 0.2, URL http://CRAN. R-project. org/package= vioplot*.

Balakrishnan, R.; Qiu, X. Y.; and Srinivasan, P. 2010. On the predictive ability of narrative disclosures in annual reports. *European Journal of Operational Research* 202(3):789–801.

Bloomfield, R. J. 2002. The incomplete revelation hypothesis and financial reporting. *Accounting Horizons* 16(3):233–243.

Bratko, A.; Filipič, B.; Cormack, G. V.; Lynam, T. R.; and Zupan, B. 2006. Spam filtering using statistical data compression models. *The Journal of Machine Learning Research* 7:2673–2698.

Butler, M., and Kešelj, V. 2009. Financial forecasting using character n-gram analysis and readability scores of annual reports. In *Advances in Artificial Intelligence*. Springer. 39–51.

Büttcher, S.; Clarke, C. L.; and Cormack, G. V. 2010. *Information retrieval: Implementing and evaluating search engines*. Mit Press.

Cecchini, M.; Aytug, H.; Koehler, G. J.; and Pathak, P. 2010. Making words work: Using financial text as a predictor of financial events. *Decision Support Systems* 50(1):164–175.

Chang, L. S.; Most, K. S.; and Brain, C. W. 1983. The utility of annual reports: An international study. *Journal of International Business Studies* 63–84.

Cormack, G. V., and Horspool, R. 1987. Data compression using dynamic markov modelling. *The Computer Journal* 30(6):541–550.

Feldman, R.; Govindaraj, S.; Livnat, J.; and Segal, B. 2010. Managements tone change, post earnings announcement drift and accruals. *Review of Accounting Studies* 15(4):915–953.

Frank, E.; Chui, C.; and Witten, I. H. 2000. Text categorization using compression models. In *Proceedings of the Conference on Data Compression*, 555. IEEE Computer Society.

Huang, A.; Zang, A.; and Zheng, R. 2014. Evidence on the information content of text in analyst reports. *The Accounting Review, Forthcoming*.

Humpherys, S. L.; Moffitt, K. C.; Burns, M. B.; Burgoon, J. K.; and Felix, W. F. 2011. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems* 50(3):585–594.

Jones, M. J., and Shoemaker, P. A. 1994. Accounting narratives: a review of empirical studies of content and readability. *Journal of Accounting Literature* 13(1):142–184.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and economics* 45(2):221–247.

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

OConnor, N., and Madden, M. G. 2006. A neural network approach to predicting stock exchange movements using external factors. *Knowledge-Based Systems* 19(5):371–378.

Securities, U. S., and Website, E. C. O. 2014. Edgar, the electronic data gathering, analysis, and retrieval system.

Shapiro, S. S., and Wilk, M. B. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 591–611.

Teahan, W. J. 2000. Text classification and segmentation using minimum cross-entropy. In *RIAO, 6th International Conference Recherche dInformation Assistee par Ordinateur*, 943–961.

Team, R. C., et al. 2012. R: A language and environment for statistical computing.

Vergoossen, R. G. 1993. The use and perceived importance of annual reports by investment analysts in the netherlands. *European Accounting Review* 2(2):219–244.

Website, D. J. I. O. 2014a. Dow jones industrial average history.

Website, D. J. I. O. 2014b. Dow jones industrial average: Overview.