INTELLIGENT SYSTEMS IN ACCOUNTING, FINANCE AND MANAGEMENT
*Intell. Sys. Acc. Fin. Mgmt.* **19,** 75–89 (2012)
Published online in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/isaf.1326

# BEYOND THE NUMBERS: MINING THE ANNUAL REPORTS FOR HIDDEN CUES INDICATIVE OF FINANCIAL STATEMENT FRAUD

SUNITA GOEL[a]* AND JAGDISH GANGOLLY[b]
[a] *Siena College, Accounting & Law, Loundonville, NY, USA*
[b] *State University of New York at Albany, Department of Informatics, Albany, NY, USA*

SUMMARY

Unlike previous fraud detection research, a vast majority of which has focused primarily on the use of quantitative financial information to predict fraud, in this study we examine qualitative textual content in annual reports to predict fraud and see whether there are discernible differences in the writing and presentation style between companies that committed fraud and those that did not. We believe that while numeric financial information in the annual reports can hide details of fraud, textual information relating to writing and presentation styles in such reports provides valuable clues pertaining to the existence of fraud. In this study we use the chi-square test to analyse our data and test hypotheses about predictors of fraud that may explain linguistic feature variations in fraudulent and nonfraudulent annual reports. We provide new results on the usefulness of the qualitative content of annual reports in detecting fraud. Copyright © 2012 John Wiley & Sons, Ltd.

**Keywords:** financial statement fraud; fraud detection; textual analysis; linguistic markers; chi-square test

## 1. INTRODUCTION

Corporate reports are the primary means of communicating information regarding past performance as well as prospects of future performance to all interested parties. In the USA, to comply with the securities laws, companies whose stock is traded on the US stock exchanges, unless they are exempt from the reporting requirements, must file with the Securities and Exchange Commission (SEC) such annual reports prepared in conformity with the Generally Accepted Accounting Principles (GAAP) and audited by accountants registered with the SEC. GAAP is a framework of guidelines and includes standards, conventions and rules that must be followed in the preparation of financial disclosures in the corporate financial reports. The financial statements and the disclosures contained in such corporate financial reports are expected to be free of material errors, be not misleading and include disclosures relating to all material events affecting the financial condition of the company. In other words, corporate reports must present a fair representation of the financial position of the company and must contain meaningful disclosures about the company's past, present and future direction.

Companies can, however, be very innovative in camouflaging the true financial situation by painting a favourable picture by deploying the linguistic arsenal at their disposal. Annual reports are typically written by the management and corporate attorneys. Corporate writers carefully craft the language of all corporate communication to maintain a company's positive image. Furthermore, management is aware that users of these annual reports view them credibly due to the fact that these are audited reports. As a result, management strategically fabricates the content of annual reports. We argue that the textual

* Correspondence to: Sunita Goel, Siena College, Accounting & Law, Loundonville, NY, USA. E-mail: sgoel@siena.edu

documents released by companies contain linguistic cues that can be examined to assess how true and honest these annual reports are. We believe that the quantitative financial numbers contain redundant information that does not change when a company is committing a fraud but the writing style and the presentation style employed to communicate financial information changes. For example, most of the fraudulent companies were effective at hiding their fraudulent activities with an erroneous perception of strong financial performance to forestall shareholder investigation into management actions and discover performance shortfall.

The term fraudulent financial reporting, also known as financial statement fraud or management fraud, refers to fraud committed when the financial reports contain misrepresentation of material facts or such material facts have not been adequately disclosed in such reports. Since fraud involves deliberate wrongful acts performed with an intent to gain unfair advantage by deceiving another person, fraudulent financial reporting involves intentional misstatements or omission of material facts affecting the financial condition of the companies. Fraud typically exhibits three characteristics: motive, opportunity and rationalization, usually referred to as the 'fraud triangle' (Cressey, 1950). These three characteristics can be thought of as preconditions for fraud. Montgomery *et al*. (2002) explain these three characteristics of the fraud triangle in the context of fraudulent financial reporting as follows. For fraud to occur: first, there must be incentives or pressures to materially misstate or to omit material facts from the financial statements; second, there must be opportunities to carry out such material misstatements or omissions; and third, there must be values, beliefs and attitudes that allow one to knowingly and intentionally commit a dishonest act, and rationalize committing such a dishonest act. In this study, fraud is construed to be consistent with how financial statement fraud or management fraud is defined in the literature.

The remainder of the paper is organized as follows. In Section 2 we provide results of prior studies that have utilized the textual content of annual reports. Section 3 discusses methodology and hypotheses development. Section 4 gives results, explaining whether or not the hypotheses can be supported, and Section 5 summarizes the findings of this study.

## 2. RELATED LITERATURE

Even though fraud detection is a fertile research area, the use of qualitative content in annual reports as opposed to the use of quantitative financial information in annual reports to detect fraud is a relatively new phenomenon. In general, annual reports contain both qualitative and quantitative information. Most of the prior studies on fraud detection have used a variety of quantitative financial metrics to build their fraud detection models (Dopouch *et al*., 1987; Persons, 1995; Ragothaman *et al*., 1995; Beasley, 1996; Hansen *et al*., 1996; Eining *et al*., 1997; Green and Choi, 1997; Fanning and Cogger, 1998; Summers and Sweeney, 1998; Beneish, 1999; Lee *et al*., 1999; Abbot *et al*., 2000; Bell and Carcello, 2000; Lennox, 2000; Spathis, 2002; Kaminski *et al*., 2004; Hoogs *et al*., 2007; Kirkos *et al*., 2007; Dikmen and Küçükkocaoglu, 2010; Dechow *et al*., 2011). The major thrust of many of these studies has been on empirically examining the relationship between fraudulent financial reporting and quantitative indicators such as composition of boards of directors, insider trading, auditor rotation or financial restatements. However, Kaminski *et al*. (2004) provided evidence that financial metrics such as ratios have limited ability to predict fraud accurately. The limitation of quantitative information to predict fraud correctly, as noted by others, is likely attributable to high rates of false positives (nonfraudulent document classified as fraudulent) and false negatives (fraudulent document classified as nonfraudulent) reported in earlier studies.



DOI: 10.1002/isaf

We next discuss significant studies in the domains of accounting and finance that have used textual analysis for measuring features such as sentiment, tone and readability of a company's disclosures, as there is insufficient crossover research utilizing textual content of annual reports in fraud detection. This is then followed by the review of recent studies (Cecchini *et al*., 2010; Goel *et al*., 2010; Humpherys *et al*., 2011) that have used textual content of annual reports in fraud detection.

Li (2006) examined the information in the text of annual reports to determine the risk sentiment of annual reports for predicting future earnings and stock returns. He measured the risk sentiment of annual reports by counting the frequency of words related to risk or uncertainty in the 10-K filings. He found that a stronger emphasis on risk in the annual report is associated with lower future earnings. The evidence in that paper suggests the stock market underreacts to public information disclosed in the text portion of annual reports. Tetlock (2007) explored the interactions between media content and stock market activity by examining investor sentiment extracted from a popular *Wall Street Journal* column. He measured the pessimism index that is composed of mostly negative and weak words from the General Inquirer (GI) dictionary. He found that high media pessimism predicts downward pressure on market prices followed by a reversion to fundamentals, and unusually high or low pessimism predicts high market trading volume.

Li (2008) examined the relationship between annual report readability and firm performance and earnings persistence. He measured the readability of annual reports using both the fog index from computational linguistics and the length of the document. He found that the annual reports of firms with lower earnings are harder to read (i.e. they have higher fog and are longer). Moreover, the positive earnings of firms with annual reports that are easier to read are more persistent. This suggests that managers may be opportunistically choosing the readability of annual reports to hide adverse information from investors.

Kothari *et al*. (2009) examined whether favourable and unfavourable disclosures in a quarter affect firms' costs of capital, return volatility and forecast dispersions. They found that when content analysis indicates favourable disclosures, the firm's cost of capital, stock return volatility and dispersion in analysts' earnings forecasts decline significantly. In contrast, unfavourable disclosures are accompanied by significant increases in the cost of capital, stock return volatility and analysts' earnings forecast dispersion. Loughran *et al*. (2009) examined the occurrence of ethics-related terms in 10-K annual reports over 1994–2006 and found that firms using ethics-related terms are more likely to be 'sin' stocks, are more likely to be the object of class-action lawsuits, and are more likely to score poorly on measures of corporate governance. They further suggest that managers who portray their firm as 'ethical' in 10-K reports are more likely to be systematically misleading the public.

Li (2010) examined the information content of the forward-looking statements (FLSs) in the Management Discussion and Analysis (MD&A) section of 10-K and 10-Q filings using a naive Bayesian machine-learning algorithm. He found that firms with better current performance, lower accruals, smaller size, lower market-to-book ratio, less return volatility, lower MD&A fog index and longer history tend to have more positive FLSs.

Recently, researchers have used the qualitative textual content of annual reports in fraud detection (Cecchini *et al*., 2010; Goel *et al*., 2010; Humpherys *et al*., 2011). Cecchini *et al*. (2010) analysed MD&A sections of annual reports to predict fraud and bankruptcy. They created a dictionary (ontology) of discriminating concepts from MD&As which could discriminate fraudulent from nonfraudulent firms 75 % of the time and bankrupt from nonbankrupt firms 80 % of the time. After combining text data with quantitative data they were able to increase prediction accuracy to 81.97 % for fraud and 83.87 % for bankruptcy.

Goel *et al*. (2010) examined the qualitative portion of the annual reports using natural language processing tools and explored linguistic features that distinguish fraudulent from nonfraudulent annual reports. They created a methodology to proactively identify means to detect fraud using support vector



DOI: 10.1002/isaf

machines, a supervised machine-learning technique. Their results indicated that employment of linguistic features is an effective means for detecting fraud. They were able to improve the prediction accuracy of the fraud detection model from initial baseline results of 56.75 % accuracy, using a ‘bag of words’ approach, to 89.51 % accuracy when they incorporated linguistically motivated features. They found systematic differences in communication and writing style of fraudulent annual reports.

Humpherys *et al*. (2011) examined 202 publicly available financial disclosures to investigate linguistic cues that can identify use of deceptive language in managerial financial fraud. They found that fraudulent disclosures use more activation language, words, imagery, pleasantness, group references and less lexical diversity than nonfraudulent ones. They suggested that the writers of fraudulent disclosures might write more to appear credible, while communicating less in actual content. Their results support the potential use of linguistic analyses by auditors to flag questionable financial disclosures and to assess fraud risk under Statement on Auditing Standards No. 99.

The current study complements previous research that sought to uncover hidden indicators in annual reports through text mining. The consequences of financial statement fraud are enormous in several respects, including direct costs, indirect costs, low investor confidence and sometimes even terminal in the case of some companies (e.g. the demise of Enron and Arthur Anderson LLP). The examination of textual content of annual reports might prove to be informative to fraud examiners, auditors, analysts and standard setters because little is known about its usefulness in fraud detection. For instance, there has not been much research that directly utilizes the qualitative portion of annual reports to detect fraud. With the exception of Cecchini *et al*. (2010), Goel *et al*. (2010) and Humpherys *et al*. (2011), no other study has utilized the qualitative portion of annual reports in detection of fraud. In particular, we are interested in investigating if linguistic feature variations in fraudulent and nonfraudulent annual reports can be associated with existence of fraud. We also wanted to find out if differences among these features are large enough to motivate future work in this area. Here, we use the chi-square test of significance to test relationships between variables and examine linguistic feature variations and provide new evidence on the usefulness of the qualitative content of annual reports in detecting fraud.

## 3. METHODOLOGY AND HYPOTHESES DEVELOPMENT

### 3.1. Sample

We selected fraud companies from various sources, including the LexisNexis database, Compustat, *Wall Street Journal* (*WSJ*) Index and Accounting and Auditing Enforcement Releases (AAERs) issued by the SEC over the period 1993–2006. The sample began with all companies accused of financial-reporting-related enforcement actions involving fraud. We then examined all of these cases to make sure that they involved violation of GAAP and alleged companies misstated their 10-Ks. From this sample, companies that lacked documented evidence of fraud affecting 10-Ks and whose 10-Ks were not available for download from the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) database were dropped. The resultant fraud dataset contained 126 alleged companies where corroborating evidence of fraud was subsequently found.

We noticed that in our sample of 126 fraud companies several fraud companies were accused of filing fraudulent statements for more than one fiscal year (company’s reporting period); as a result, we collected 10-Ks for each year of alleged fraud for a fraud company and one year immediately preceding this alleged fraud period, resulting into 405 fraud company years. This is done to ensure that fraud company years correspond to all the years of the company’s performance when it was alleged to have



DOI: 10.1002/isaf

perpetrated fraud and of issuing materially false and misleading statements. Since, all US publicly listed companies are required to file their 10-Ks with the SEC within 90 days of the end of their fiscal year, the first year of the alleged fraud period typically includes statements for the prior year. For example, the time period of alleged fraud for Computer Associates is found to range from 1999 to 2002 and during this time it issued materially false and misleading statements. In this case, the fraud company years not only include 10-Ks for its fiscal years ending December 1999, 2000, 2001, and 2002, but also for the fiscal year ending 31 December 1998.

Figure 1 shows the yearly distribution of the 126 fraud companies over the sample period (1993–2006). It provides the number of observations by the last year of the fraud period. In cases of multiple frauds, the last year of the most recent fraud period was included. In our sample of fraud companies there were seven companies where multiple frauds were identified at different time periods. In such cases, multiple events were included in the fraud period for that company, but the company was counted only once in the sample. For example, for Bristol Myers Squibb, two frauds (one in each period) were identified for the periods 1999–2003 and 2006; as a result, both time periods were included in the fraud company years, but the company was counted only once in the sample. We can see from this chart that a majority of high-profile accounting scandals occurred during this period, which eventually led to the enactment of the Sarbanes–Oxley (SOX) legislation in 2002. Owing to the infrequency of financial statement fraud, this time period (1993–2006) was selected because a relatively large number of fraud observations could be identified. For example, the number of fraud companies almost doubled from 1998 and hit its peak in 2002 at 28 observations and then declined to about 10 companies per year.

For each fraud company we selected multiple no-fraud companies during the sample period (1993–2006) where fraud had never been reported and whose 10-Ks were available for download from EDGAR. No-fraud companies were selected from Compustat via Research Insight on the basis of year, industry classification and firm size (total assets). We selected multiple no-fraud companies to simulate the actual population distribution of the fraud and no-fraud observations. Since incidence of fraud is a rare event, the relative frequency of fraud observations is low in reality. Given the infrequency of fraud, a peer set of control companies was used in the current study as opposed to a matched sample of control companies.

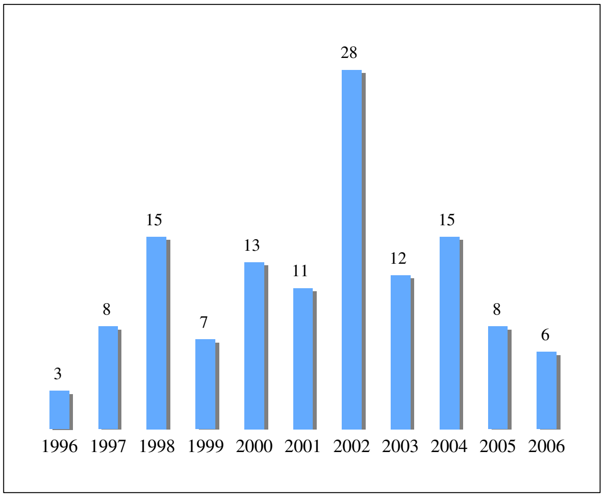


Figure 1. Distribution of fraud companies by last year of fraud period.



DOI: 10.1002/isaf

For the no-fraud dataset we created two different versions. For the first version, we collected an average of five no-fraud companies for each fraud company, resulting in 622 no-fraud company years that correspond to 622 no-fraud companies. This approach is consistent with previous studies using the unmatched sample approach. For example, Bell and Carcello (2000) selected about four nonfraud engagements for each fraud engagement in a sample of 77 fraud and 305 nonfraud engagements and Hoogs *et al*. (2007) selected about six control companies for each target company in a sample of 51 fraud and 339 no-fraud companies. For the second version of the no-fraud dataset we collected as many no-fraud companies as we could find that matched the selection criteria for each fraud company. This resulted in a dataset of 6741 no-fraud company years that corresponds to the 6741 no-fraud companies.

The collected data for the final sample are described in Table I. The first version of the dataset included 405 fraud company years for the 126 fraud companies paired with 622 no-fraud company years that correspond to the 622 no-fraud companies where no documented evidence of fraud was discovered. The second version of the dataset included same 405 fraud company years of 126 fraud companies used in the first version paired with 6741 no-fraud company years of 6741 no-fraud companies. For our sample, we obtained the text of original annual reports (10-Ks) corresponding to the fraud and no-fraud company-year observations for the respective fraud and no-fraud companies from the EDGAR database available on the SEC website (www.sec.gov). We carried out our hypotheses tests with both versions of the no-fraud datasets and report our results with them (see Section 4).

Table II shows the distribution of fraud and no-fraud companies split by industrial sector for the first version of the dataset. The 126 fraud companies were unevenly distributed across eight industrial

Table I. Final dataset

| Dataset | Fraud companies | Fraud company years | No-fraud companies | No-fraud company years |
|---|---|---|---|---|
| Version 1 | 126 | 405 | 622 | 622 |
| Version 2 | 126 | 405 | 6741 | 6741 |

Table II. Distribution of fraud and no-fraud companies by industrial sector for first version

| | Fraud companies | | No-fraud companies | |
|---|---|---|---|---|
| Industrial sector (industry codes) | Count | % | Count | % |
| Mining (1000–1499) | 2 | 1.6 | 10 | 1.6 |
| Construction (1500–1799) | 2 | 1.6 | 10 | 1.6 |
| Manufacturing (2000–3999) | 47 | 37.3 | 232 | 37.3 |
| Transportation, communication, electric, gas & sanitary services (4000–4999) | 17 | 13.5 | 84 | 13.5 |
| Wholesale trade (5000–5199) | 6 | 4.8 | 30 | 4.8 |
| Retail trade (5200–5999) | 5 | 3.9 | 24 | 3.9 |
| Finance, insurance & real estate (6000–6799) | 16 | 12.7 | 79 | 12.7 |
| Services (7000–8999) | 31 | 24.6 | 153 | 24.6 |
| Totals | 126 | 100.0 | 622 | 100.0 |



DOI: 10.1002/isaf

sectors. With 47 fraud companies, the manufacturing sector had the highest percentage (37.3 %) of fraud companies. The services sector, with 31 fraud companies, was second at 24.6 % of identified fraud companies. Among the eight sectors, the fraud companies were widely distributed across 31 two-digit SIC industries. Most prominent were 39 companies with SIC codes from 3000 to 3999 and 29 companies with SIC codes from 7000 to 7999. Most of these were technology companies.

### 3.2. Methodology

We used the chi-square test to analyse our data as it is the most commonly used significance test in corpus linguistics (Williams, 1986; Smith, 1988). The chi-square analyses employed were able to determine the probability of the association between variables and helped us conclude whether this association was real or due to chance. Each 10-K in the fraud corpus was preprocessed, in which we removed all the quantitative information from it. After collecting 10-Ks, we performed a linguistic analysis of the data where we extracted linguistic markers from the textual content of annual reports and tested hypotheses about predictors of fraud that explained if linguistic feature variations in such reports were statistically significant.

We used DICTION 5.0 (Hart, 2000), STYLE (Cherry and Vesterman, 1991) and LIWC (Pennebaker *et al*., 2007) tools to extract linguistic markers. We drew on the prior literature (Goel *et al*., 2010; Humpherys *et al*., 2011) to identify linguistic markers that might be associated with fraud. We also added additional linguistic indicators of fraud that were not identified in earlier studies. Furthermore, we also examined if the differences in the proportions of linguistic markers between fraudulent and nonfraudulent annual reports were statistically relevant or not. When comparing the word usage, we found that the 10-K filings for later years (post-SOX period) had more total number of words than the 10-K filings for earlier years did (pre-SOX period). The median number of words in 10-K filings increased from 15,991 in 1994 to over 55,000 in 2007. This increase may be due to the fact that companies were required to provide more information in their 10-K reports in the post-SOX period.

### 3.3. Hypotheses

We believe that, in cases of fraud, annual reports often contain informative descriptive statements obscured by a pile of uninformative verbiage; nonfraudulent annual reports, on the other hand, are usually expressed clearly and succinctly. It is difficult for users of the annual reports to sift through all this information to understand the real motive of the management when companies are committing fraud. Our main hypothesis, stated next, is of interest as it is believed that, in order to conceal fraud, companies resort to many creative techniques to proactively construct a meaning rather than reveal 'what was there'. There is wide consensus among researchers that the annual report is nothing but a highly sophisticated product of the corporate design environment. For example, Griffiths (1986) suggests that companies present annual reports in ways that conceal as much as they reveal. Companies use both selective accounting language and presentation style features to hide fraud; therefore, text contains more diverse and dense information than numbers do.

*Main hypothesis*: In cases of fraud, companies try to misrepresent information and therefore employ writing techniques that differ from the companies that do not commit fraud.



DOI: 10.1002/isaf

In order to test this main hypothesis, we develop five related sub-hypotheses that are discussed next.

*H1*: The greater the use of complex sentential structures in the qualitative content of a company's annual report, the greater the likelihood that there will be fraud.

Our first hypothesis, H1, was based on the notion that companies committing fraud tend to present their annual reports in a convoluted style, consciously employing means to obscure real information that might expose fraud. This hypothesis rests on the idea that fraudulent annual reports make greater use of complex sentential structures and, thus, exhibit high levels of both lexical and structural ambiguity. Conversely, clear succinct language is employed to convey information more clearly, which would require short and direct sentential constructs. For H1, the null hypothesis that the count of complex sentential structures indicated by features such as ambiguity index and type/token ratio is similar in both fraudulent and nonfraudulent annual reports was tested.

Related to hypothesis H1, we also defined and tested a sub-hypothesis, H1a, as follows:

*H1a*: The more difficult it is to read and understand a company's annual report, the greater the likelihood that there is fraud.

Hypothesis H1a is of interest as it is believed that, in cases of fraud, companies deliberately employ tools to make annual reports difficult to read and comprehend. Therefore, fraudulent annual reports show greater use of longer sentences, difficult words and complex syntactic sentence structures. Conversely, to make reports easier to understand, management makes precise, molecular statements by making greater use of shorter sentences with simple words so that the readability of such reports is improved. It is particularly important to do readability analysis in the case of a qualitative examination of annual reports, as these reports are read by a wide number of users and are the most important vehicle through which companies communicate their past performance and future prospects. For H1a, the null hypothesis that the readability scores are similar in both fraudulent and nonfraudulent annual reports was tested.

*H2*: The greater the use of negative words in a company's annual report, the greater the likelihood that there is fraud.

Hypothesis H2 is of interest as it is believed that management consciously chooses selective accounting language to conceal fraud. The annual report typically reflects on a company's strategy and, when things are not going well, its numbers do not change but qualitative content changes. For H2, the null hypothesis that negative and positive words are used similarly in the qualitative content of fraud and no-fraud annual reports was tested.

*H3*: The greater the use of passive voice in a company's annual report, the greater the likelihood that there is fraud.

Hypothesis H3 is of interest as it is believed that, in cases of fraud, management consciously tries to shift responsibility away from itself and makes greater use of the passive voice rather than using active voice constructions. As a result, a sentence becomes wordy, unclear, vague and, hence, misleading. The use of the passive voice creates an uncertain meaning in the minds of readers. Linguistic studies have shown that people are better able to remember material they read in the active voice than the same



DOI: 10.1002/isaf

material in the passive voice. They noted that, in the case of the active voice, the mind remains geared toward a 'subject–verb–object' pattern, whereas passive voice sentences derail that mental process of retention (Coleman, 1965). The passive voice is typically used when the speaker wants to hide the agent or obscure what occurs, and as a result the sentence becomes unclear and misleading. In this type of situation, the passive voice is used in a deceptive way to avoid disclosing uncomfortable news. Zhou *et al*. (2004; Zhou and Zhang, 2008) found that deceivers engage in depersonalism manipulations and use the passive voice to disassociate themselves from their messages and the content of those messages and despite their efforts to the contrary reveal their deceptive intent. Porter and Yuille (1996) found that use of the passive voice is often associated with lack of personal responsibility. Conversely, in cases of no-fraud, in order to take credit for positive outcomes and attribute these outcomes to its own actions, management will use active voice constructions. The use of the active voice makes a sentence more direct, short and concise. As a result, there is no ambiguity in its meaning and a writer is able to clearly express their ideas. For H3, the null hypothesis that there is no difference in the use of voice in the qualitative content of fraudulent and nonfraudulent annual reports was tested.

*H4*: The greater the use of uncertainty markers in a company's annual report, the greater the likelihood that there is fraud.

Hypothesis H4 is of interest as it is believed that, in cases of fraud, management deliberately employs more uncertainty markers to make reports unclear and ambiguous. The management may rationalize this act by arguing that it is better to introduce uncertainty than to make false statements. Conversely, in cases of no-fraud, to provide a definite and clear picture, management will avoid using uncertainty markers. In linguistics, uncertainty markers are often associated with weakening and deintensification of the message. These words have been shown to function as a subtle means to avoid responsibility and evade the truth. For H4, the null hypothesis that the use of uncertainty markers is similar in the qualitative content of both fraudulent and nonfraudulent annual reports was tested.

*H5*: The greater the use of adverbs in a company's annual report, the greater the likelihood that there is fraud.

Hypothesis H5 is of interest as it is believed that, in cases of fraud, management deliberately uses more adverbs so as to lead users of annual reports to draw wrong inferences about the true state of affairs of the company. Adverbs have been shown to capture the intensity of messages and qualify the meaning in statements. Many studies in linguistics have successfully used adverbs to examine truthful manipulations associated with deception (Buller and Burgoon, 1994; Zhou *et al*., 2002). Conversely, in cases of no-fraud, to provide a true and fair picture, management will use fewer adverbs. For H5, the null hypothesis that the use of adverbs is similar in the qualitative content of both fraudulent and nonfraudulent annual reports was tested.

*H6*: The greater the use of formatting styles such as use of caps or use of punctuation in a company's annual report, the greater the likelihood that there is fraud.

Hypothesis H6 is of interest as it is believed that, in cases of fraud, aesthetics play a role. When management wants less attention to the underlying issues, it will make use of the cosmetic arsenal at its disposal to distract the attention of users. Conversely, in cases of no-fraud, to provide a clear and actual picture, management will use simple designs with less formatting variations. For H6, the null



DOI: 10.1002/isaf

hypothesis that the use of caps and punctuation is similar in both fraudulent and nonfraudulent annual reports was tested.

## 4. RESULTS AND DISCUSSION

In this section we present the results of the hypothesis tests and discuss their implications. As explained earlier, all the hypotheses were evaluated using a chi-square test of significance to measure linguistic feature variations in the fraud and no-fraud corpus (see Section 3).

For H1, the null hypothesis that the count of complex sentential structures as indicated by an ambiguity index obtained through DICTION, the frequency of sentences beginning with a subordinating conjunction, the frequency of different word types and the frequency of function words is similar in both fraudulent and nonfraudulent annual reports was tested. We examined the frequency of function words, as these words are typically the most frequent words. Prior studies in linguistics have shown that documents containing more frequent words are easy to read when everything else is same (Graham *et al*., 2005). For H1a, the null hypothesis that the readability index is similar in both fraudulent and nonfraudulent annual reports was tested. We used STYLE to compute several readability indices for all the 10-Ks in the two corpora (fraud, no-fraud). The results of chi-square tests for hypotheses H1 and H1a indicate that we could reject the null hypotheses and conclude that the variation between the fraudulent and nonfraudulent annual reports is due to systematic alterations and not due to chance: $p < 0.001$ and $p < 0.05$ for H1 and H1a respectively for the first version of the fraud dataset and $p < 0.001$ and $p < 0.001$ for H1 and H1a respectively for the second version of the fraud dataset.

For testing H2, the null hypothesis that there exists no relationship between the polarity of tone and the outcome of fraud was tested. We used DICTION 5.0 to compute frequency of occurrence of tone category words for all the 10-Ks in the two corpora (fraud, no-fraud) by creating custom dictionaries in DICTION. Table III presents the chi-square results of testing the null hypothesis for H2. The results of chi-square tests for hypothesis H2 indicate that we could reject the null hypothesis and conclude that the variation in the use of positive and negative tone category words between the fraudulent and nonfraudulent annual reports is significant ($p < 0.001$). Furthermore, the relative distribution of negative and positive category words between the two corpora showed that there is not much difference. For example, negative and positive words have similar percentages: 45.93 % (negative) and 54.07 % (positive) out of total of 125 591 tone category words in the fraud corpus compared with 45.59 % (negative) and 54.41 % (positive) out of total of 178 841 tone category words in the no-fraud corpus for the first version of the fraud dataset. Nevertheless, the normalized counts of the negative and the positive category words showed that the fraud corpus has 5804.16 negative and 6831.70 positive instances per 1 000 000 words, compared with 5675.49 negative and 6774.19 positive instances in the no-fraud corpus for the first version of the fraud dataset.

Table III. Chi-square results for the tone category words

| Fraud datasets | Degrees of freedom | $\chi^2$ | $p$-value |
|---|---|---|---|
| Version 1[a] | 2 | 20.04 | ≤0.001 |
| Version 2[b] | 2 | 276.94 | ≤0.001 |

[a]405 fraudulent 10-Ks versus 622 nonfraudulent 10-Ks.
[b]405 fraudulent 10-Ks versus 6741 nonfraudulent 10-Ks.

 

For H3, the null hypothesis that there is no difference in the use of voice in the qualitative content of fraudulent and nonfraudulent annual reports was tested. We used STYLE for gathering voice statistics for all the 10-Ks in the two corpora (fraud, no-fraud). Table IV presents the chi-square results of testing the null hypothesis for H3. The results of chi-square tests for hypothesis H3 indicate that we could reject the null hypotheses with very high confidence and conclude that the variation in the use of passive and active voice sentences between the fraudulent and nonfraudulent annual reports is due to systematic alterations and not due to chance ($p < 0.001$). A closer examination of the distribution of passive/active voice construction in the two corpora showed a high incidence of passive instead of active voice sentences in the fraud corpus. This explanation is supported by the predominance of the passive voice sentences found in the fraudulent annual reports, whereas examination of the nonfraudulent annual reports, on the other hand, revealed high figures for active voice sentences.

Table IV. Chi-square results for use of passive voice and active voice sentences

| Fraud datasets | Degrees of freedom | $\chi^2$ | $p$-value |
|---|---|---|---|
| Version 1[a] | 1 | 339.44 | ≤0.001 |
| Version 2[b] | 1 | 1354.06 | ≤0.001 |

[a]405 fraudulent 10-Ks versus 622 nonfraudulent 10-Ks.
[b]405 fraudulent 10-Ks versus 6741 nonfraudulent 10-Ks.

For H4, the null hypothesis that the use of uncertainty markers is similar in the qualitative content of both fraudulent and nonfraudulent annual reports was tested. We used DICTION 5.0 to extract the frequency of uncertainty markers in our datasets by creating a custom dictionary in DICTION. Table V presents the chi-square results of testing the null hypothesis for H4. The results of chi-square tests for hypothesis H4 indicate that we could reject the null hypotheses with very high confidence and conclude that the variation in the use of uncertainty markers between the fraudulent and nonfraudulent annual reports is due to systematic alterations and not due to chance ($p < 0.001$). Furthermore, the normalized counts of uncertainty markers showed that the fraud corpus has 11 642.93 instances per 1 000 000 words, compared with 10 994.29 instances in the no-fraud corpus for the first version of the fraud dataset.

Table V. Chi-square results for use of uncertainty markers

| Fraud datasets | Degrees of freedom | $\chi^2$ | $p$-value |
|---|---|---|---|
| Version 1[a] | 1 | 222.02 | ≤0.001 |
| Version 2[b] | 1 | 7110.77 | ≤0.001 |

[a]405 fraudulent 10-Ks versus 622 nonfraudulent 10-Ks.
[b]405 fraudulent 10-Ks versus 6741 nonfraudulent 10-Ks.

For H5, the null hypothesis that there is no difference in the use of adverbs in the qualitative content of fraudulent and nonfraudulent annual reports was tested. We used LIWC for gathering frequency of adverbs for all the 10-Ks in the two corpora (fraud, no-fraud). Table VI presents the chi-square results of testing the null hypothesis for H5. The results of chi-square tests for hypothesis H5 indicate that we could reject the null hypotheses with very high confidence and conclude that the variation in the use of adverbs between the fraudulent and nonfraudulent annual reports is due to systematic alterations and not due to chance ($p < 0.001$). For H6, the null hypothesis that the use of capitals and punctuation is



DOI: 10.1002/isaf

Table VI. Chi-square results for use of adverbs

| Fraud datasets | Degrees of freedom | $\chi^2$ | $p$-value |
|---|---|---|---|
| Version 1[a] | 1 | 406.58 | 0.001 |
| Version 2[b] | 1 | 616.13 | 0.001 |

[a]405 fraudulent 10-Ks versus 622 nonfraudulent 10-Ks.
[b]405 fraudulent 10-Ks versus 6741 nonfraudulent 10-Ks.

similar in both fraudulent and nonfraudulent annual reports was tested. The results of chi-square tests for hypotheses H6 indicate that we could not reject the null hypotheses ($p > 0.10$) for both versions of the fraud dataset. Therefore, we conclude that the variation between the fraudulent and nonfraudulent annual reports is not significant.

### 4.1. Further Tests

As an additional analysis we wanted to examine whether our results of hypotheses testing would differ if we just examine MD&A section as opposed to the entire annual report. For this, we extracted the MD&A section for all the 10-Ks that we have in our two versions of the dataset. Some of the MD&As of the 10-Ks had references to the annual reports for some of its subsections. In order to ensure that we use the same approach for identifying the MD&A section for both the fraud and no-fraud corpus, we included only the content that was directly included under MD&A sections. Table VII presents the results of hypotheses testing. The results of chi-square tests for hypotheses H3, H4 and H5 indicate that we could reject the null hypotheses ($p < 0.001$) for both versions of the fraud dataset. However, we could not reject the null hypotheses for H1, H1a, H2 and H6 ($p > 0.10$). Consequently, we conclude that the variation between the fraudulent and nonfraudulent annual reports remains significant for usage of passive voice, uncertainty markers and adverbs.

Table VII. Hypotheses test results for MD&A section of the 10-Ks

| Hypothesis | Fraud datasets | $p$-value |
|---|---|---|
| H1 | Version 1[a] | >0.10 |
| | Version 2[b] | >0.10 |
| H1a | Version 1[a] | >0.10 |
| | Version 2[b] | >0.10 |
| H2 | Version 1[a] | >0.10 |
| | Version 2[b] | >0.10 |
| H3 | Version 1[a] | ≤0.001 |
| | Version 2[b] | ≤0.001 |
| H4 | Version 1[a] | ≤0.001 |
| | Version 2[b] | ≤0.001 |
| H5 | Version 1[a] | ≤0.001 |
| | Version 2[b] | ≤0.001 |
| H6 | Version 1[a] | >0.10 |
| | Version 2[b] | >0.10 |

[a]405 fraudulent MD&As versus 622 nonfraudulent MD&As.
[b]405 fraudulent MD&As versus 6741 nonfraudulent MD&As.



DOI: 10.1002/isaf

We also carried out $Z$-tests with our data, as it has been proposed by researchers for use in testing the difference between two proportions (Coates and Fant, 1993; O'Leary, 1998). Our results of $Z$-tests were not much different from the results that we had using the chi-square test. For example, for the hypothesis relating to use of negative tone for the first version of the dataset the $z$-score equals 4.13 and the $p$-value equals $3.591\,973\,9 \times 10^{-5}$ and for the second version of the dataset the $z$-score equals 10.37 and the $p$-value equals $3.501\,972 \times 10^{-25}$. Thus, we could reject the hypothesis relating to use of negative tone. For the hypothesis relating to the use of passive voice for the first version of the dataset the $z$-score equals 18.42 and the $p$-value equals $9.058\,918 \times 10^{-76}$ and for the second version of the dataset the $z$-score equals 36.80 and the $p$-value equals zero. Here, we could reject the hypothesis relating to use of the passive voice. For the hypothesis relating to use of uncertainty markers for the first version of the dataset the $z$-score equals 14.90 and the $p$-value equals $3.442\,05 \times 10^{-50}$ and for the second version of the dataset the $z$-score equals $-84.33$ and the $p$-value equals zero. Here, we could reject the hypothesis relating to uncertainty markers. For the hypothesis relating to use of adverbs for the first version of the dataset the $z$-score equals 20.16 and the $p$-value equals $2.050\,405 \times 10^{-90}$ and for the second version of the dataset the $z$-score equals $-24.82$ and the $p$-value equals zero. Here, we could reject the hypothesis relating to use of adverbs.

## 5. CONCLUSION

In this study, we empirically examined the implications of linguistic markers of annual reports for likelihood of fraud. The results of this study shed light on the usefulness of textual content of annual reports in fraud detection. In particular, the current study suggests that six categories of linguistic cues have been associated with fraudulent financial reporting: (a) use of complex sentential structures, (b) difficulty of reading and comprehension as measured by readability index, (c) use of positive tone, (d) use of passive voice, (e) use of uncertainty markers and (f) use of adverbs. Furthermore, we noticed that the textual portion allowed companies much flexibility in portraying themselves in terms of the linguistic markers, but there were not significant differences in the use of formatting styles, such as use of capitals or use of punctuation in fraudulent and nonfraudulent annual reports, which indicated that the companies were limited in formatting options. Analysis of these hypotheses also helped us identify whether certain linguistic cues are used more or less frequently in fraudulent annual reports.

Furthermore, results of this study can aid auditors in fraud risk assessment as auditors face adverse legal and regulatory consequences if fraud goes undetected. Annual reports accompanied by high frequency of linguistic markers identified in this study would indicate an increased likelihood of fraud, thus signalling the auditor to companies that are at higher risk for fraud. Our results can also help regulators such as the SEC in assessing the likelihood of fraud by flagging potential fraudulent companies that exhibit these linguistic markers and the SEC can pursue these suspicious companies for further investigation.

The current study is limited by several factors. The results of this study may be sample specific. In addition, some of the fraud companies may not have been included into our sample because there was no documented evidence of fraud. Likewise, some of the control companies may have also committed fraud but such fraud may have gone undetected. In summary, the current study provides new results on the usefulness of the qualitative content of annual reports in fraud detection. Furthermore, results of the current study strongly corroborate findings of the earlier studies that suggest that systematic analysis of linguistic information is useful in fraud detection. As such, more research in this area seems to be warranted. Future research would continue to discover novel linguistic constructs and novel approaches to fraud detection.



DOI: 10.1002/isaf

ACKNOWLEDGEMENTS

We thank Daniel E. O'Leary (editor), two anonymous reviewers and conference participants at the AAA Midyear Meeting of the Information Systems Section for helpful comments and suggestions on early drafts of this paper. We also acknowledge helpful comments and suggestions of Sue R. Faerman and Ozlem Uzuner in completing this research.

REFERENCES

Abbot JL, Park Y, Parker S. 2000. The effects of audit committee activity and independence on corporate fraud. *Managerial Finance* **26**(11): 55–67.

Beasley MS. 1996. An empirical analysis of the relation between the board of director composition and financial statement fraud. *The Accounting Review* **71**(4): 443–465.

Bell TB, Carcello JV. 2000. A decision aid for assessing the likelihood of fraudulent financial reporting. *Auditing: A Journal of Practice and Theory* **19**(1): 169–175.

Beneish M. 1999. The detection of earnings manipulation. *Financial Analysts Journal* **55**: 24–36.

Buller DB, Burgoon JK. 1994. Deception: strategic and nonstrategic communication. In *Strategic Interpersonal Communication*, Daly JA, Wiemann JM (eds). Erlbaum: Hillsdale, NJ; 191–223.

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

Cherry LL, Vesterman W. 1991. Writing tools – the STYLE and DICTION programs: 4.3BSD Unix system documentation, University of California, Berkley.

Coates P, Fant L. 1993. Recognizing financial distress patterns using a neural network tool. *Financial Management* (Autumn): 142–155.

Coleman EB. 1965. Learning of prose written in four grammatical transformations. *Journal of Applied Psychology* **49**(5): 332–341.

Cressey DR. 1950. The criminal violation of financial trust. *American Sociological Review* **15**(6): 738–743.

Dechow PM, Ge W, Larson CR, Sloan RG. 2011. Predicting material accounting misstatements. *Contemporary Accounting Research* **28**(1): 17–82.

Dikmen B, Küçükkocaoglu G. 2010. The detection of earnings manipulation: the three-phase cutting plane algorithm using mathematical programming. *Journal of Forecasting* **29**(5): 442–466.

Dopouch N, Holthausen R, Leftwich R. 1987. Predicting audit qualifications with financial and market variables. *The Accounting Review* **62**(3): 431–454.

Eining MM, Jones DR, Loebbecke JK. 1997. Reliance on decision aids: an examination of auditors' assessment of management fraud. *Auditing: A Journal of Practice and Theory* **16**(2): 1–19.

Fanning K, Cogger K. 1998. Neural network detection of management fraud using published financial data. *International Journal of Intelligent Systems in Accounting, Finance and Management* **7**: 21–24.

Goel S, Gangoly J, Faerman S, Uzuner O. 2010. Can linguistic predictors detect fraudulent financial filings? *Journal of Emerging Technologies in Accounting* **7**: 25–46.

Graham N, Hirst G, Marthi B. 2005. Segmenting documents by stylistic character. *Natural Language Engineering* **11**(4): 397–415.

Green BP, Choi JH. 1997. Assessing the risk of management fraud through neural-network technology. *Auditing: A Journal of Practice and Theory* **16**: 14–28.

Griffiths I. 1986. *Creative Accounting: How to Make Your Profits What You Want Them to Be*. Unwin Hyman: London.

Hansen JV, McDonald JB, Messier WF, Bell TB. 1996. A generalized qualitative-response model and the analysis of management fraud. *Management Science* **42**(7): 1022–1033.

Hart RP. 2000. Diction 5.0: The Text Analysis Program, Computer Software. Sage: Thousand Oaks, CA.

Hoogs B, Kiehl T, Lacomb C, Senturk D. 2007. A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud. *International Journal of Intelligent Systems in Accounting, Finance and Management* **15**: 41–56.

Humpherys S, Moffitt KC, Burns MB, Burgoon JK, Felix WF. 2011. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems* **50**: 585–594.



DOI: 10.1002/isaf

Kaminski KA, Wetzel TS, Guan L. 2004. Can financial ratios detect fraudulent financial reporting? *Managerial Auditing Journal* **19**(1): 15–28.
Kirkos E, Spathis C, Manolopoulos Y. 2007. Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications* **32**: 995–1003.
Kothari SP, Li X, Short J. 2009. The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: a study using content analysis. *The Accounting Review* **84**(5): 1639–1670.
Lee T, Ingram R, Howard T. 1999. The difference between earnings and operating cash flow as an indicator of financial reporting fraud. *Contemporary Accounting Research* **16**: 749–786.
Lennox C. 2000. Do companies successfully engage in opinion-shopping? Evidence from the UK. *Journal of Accounting and Economics* **29**(3): 321–337.
Li F. 2006. Do stock market investors understand the risk sentiment of corporate annual reports? Working paper, University of Michigan.
Li F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* **45**(2–3): 221–247.
Li F. 2010. The information content of forward-looking statements in corporate filings – a naïve Bayesian machine learning approach. *Journal of Accounting Research* **48**(5):1049–1102.
Loughran T, McDonald B, Yun H. 2009. A wolf in sheep's clothing: the use of ethics-related terms in 10-K reports. *Journal of Business Ethics* **89**(1): 39–49.
Montgomery DD, Beasley MS, Menelaides SL, Palmrose Z-V. 2002. Auditors' new procedures for detecting fraud. *Journal of Accountancy* (May): 39–46.
O'Leary DE. 1998. Using neural networks to predict corporate failure. *International Journal of Intelligent Systems in Accounting, Finance and Management* **7**(3): 187–197.
Pennebaker JW, Booth RJ, Francis ME. 2007. Linguistic inquiry and word count: LIWC 2007. LIWC: Austin, TX (www.liwc.net).
Persons O. 1995. Using financial statement data to identify factors associated with fraudulent financial reporting. *Journal of Applied Business Research* **11**: 38–46.
Porter S, Yuille JC. 1996. The language of deceit: an investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior* **20**(4): 443–458.
Ragothaman S, Carpenter J, Buttars T. 1995. Using rule induction for knowledge acquisition: an expert systems approach to evaluating material errors and irregularities. *Expert Systems with Applications* **9**(4): 483–490.
Smith MJ. 1988. *Contemporary Communication Research Methods*. Wadsworth Publishing Company: Belmont, CA.
Spathis C. 2002. Detecting false financial statements using published data: some evidence from Greece. *Managerial Auditing Journal* **17**: 179–191.
Summers SL, Sweeney JT. 1998. Fraudulently misstated financial statements and insider trading: an empirical analysis. *The Accounting Review* **73**(1): 131–146.
Tetlock P. 2007. Giving content to investor sentiment: the role of media in the stock market. *Journal of Finance* **62**: 1139–1168.
Williams F. 1986. *Reasoning with Statistics*. Holt, Rinehart and Winston: New York.
Zhou L, Zhang D. 2008. Following linguistic footprints: automatic deception detection in online communication. *Communications of the ACM* **51**: 119–122.
Zhou L, Twitchell D, Qin T, Burgoon J, Nunamaker J. 2002. An exploratory study into deception detection in text-based computer-mediated communication. In *Proceedings of the 36th Hawaii International Conference on System Sciences*.
Zhou L, Burgoon JK, Nunamaker JF, Twitchell D. 2004. Automated linguistics based cues for detecting deception in text-based asynchronous computer-mediated communication: an empirical investigation. *Group Decision and Negotiation* **13**(1): 81–106.



DOI: 10.1002/isaf