Title
Textual Classication of SEC Comment Letters

Author
Ryans, James Patrick

Publication Date
2016

Peer reviewed|Thesis/dissertation

Textual Classification of SEC Comment Letters

by

James Patrick Ryans

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Patricia M. Dechow, Chair
Assistant Professor Alastair Lawrence
Assistant Professor Panos N. Patatoukas
Professor Richard G. Sloan
Professor Stephen M. Solomon

Spring 2016

Textual Classification of SEC Comment Letters

Abstract

Textual Classification of SEC Comment Letters

by

James Patrick Ryans

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor Patricia M. Dechow, Chair

The purpose of this study is to identify important SEC comment letters and examine the mechanisms by which they affect firm value. The SEC periodically reviews public-company financial statements, issuing comment letters in response to disclosure deficiencies, to ensure that investors are provided with material information, and to prevent fraud. Given that comment letters consist of unstructured text, statistical text classification may be an effective technique to identify comment letter importance. The information in comment letters is distributed over several separate filings and they are not widely cited by the press or analysts as information sources, which may result in investor inattention and underreaction to their disclosure. I utilize negative abnormal returns following comment letter disclosure as the primary indicator of comment letter importance, and develop a Naive Bayesian classification model that signals important comment letters from their text features that are associated with the indicator. In a holdout sample, the text classification model correctly identifies important comment letters between 10 and 40 percent better than chance. The average out-of-sample abnormal return for firms with signaled comment letters is -5.8 percent during the 90 days post-disclosure, but only when the comment letters were viewed on EDGAR. Signaled comment letters are associated with lower persistence of profits and increased material restatements in the year following comment letter disclosure.

+

For Sara
Charles, Felix, and Cecilia
Mum and Dad

+

Contents

List of Figures

List of Tables

Acknowledgments

I especially thank my dissertation and oral examination committee members: Patricia Dechow (dissertation chair), Alastair Lawrence (oral examination chair), Robert Bartlett, Panos Patatoukas, Richard Sloan, and Stephen Davidoff Solomon. I also thank John Barrios, Stefano DellaVigna, Paul Fischer, Miles Gietzmann, Mark Huson, Lillian Mills, Miguel Minutti-Meza, Reining Petacchi, Gordon Phillips, Anastasia Zakolyukina, Luigi Zingales, and workshop participants at Cornell University, Dartmouth College, IESE, London Business School, the University of California at Los Angeles, the University of Texas at Austin, the University of Toronto, Yale University, the 2015 AAA FARS Conference, 2015 EAA Annual Congress, and the 2015 JAR/PCAOB Conference.

Chapter 1

Introduction

This dissertation examines Securities and Exchange Commission (SEC) comment letters, specifically those correspondences between the SEC and public issuers in the US, which occur as a result of periodic reviews of issuers' annual financial statements. The SEC's examinations aim to reduce disclosure deficiencies and ensure that investors are "...provided with material information and to prevent fraud and manipulation..."(SEC 2001). The primary purpose of this dissertation is to develop a method to identify comment letters that are material to investors, and to examine possible mechanisms by which comment letters affect firm value.

Section 408 of the Sarbanes Oxley Act of 2002 requires that the SEC examine all public issuers, including an review of the annual financial statements, at least once every three years (SOX 2002). During these reviews, SEC examiners issue written questions to management, and management provides written responses. These questions and responses are collectively referred to as "comment letters". A "conversation" consists of several separate letters from the SEC to the company and the company's corresponding written responses. The median conversation has four letters, and the 90th percentile has eight, issued over a median of 54 days. Comment letters and associated company responses are not publicly disclosed until after a review is complete. Once the review is complete, all comment letters in the conversation are disclosed after a 20 business day waiting period (45 calendar days prior to 2012), on the SEC's Electronic Data Gathering and Retrieval (EDGAR) system.

These SEC reviews, and the comment letters they generate, are an important monitoring requirement of SOX, and the examination and support staff represent a significant portion of the SEC's budget. At least a subset of comment letters should be expected to be important, even if not all are material. Furthermore, the review process itself may play an important role in monitoring financial reporting, deterring fraud and abuse, and improving the informational efficiency of the stock market.

Important comment letters could encourage managers to reveal strategically withheld information and could identify firms with inadequate financial reporting capabilities or insufficient internal controls to comply with disclosure requirements. Important comment letters may also cause managers and auditors to revise their assumptions and estimates in subsequent reporting periods, resulting in disclosure changes and changes to reported financial results. Textual analysis techniques are well-suited for the comment letter setting, because the letters consist of unstructured text, without consistent quantitative information or summary statistics. For the main results of this dissertation, I train and validate a Naive Bayesian classification model using post-disclosure returns as the measure of comment letter importance, and examine the relation between signaled-important comment letters, financial performance, and financial reporting quality in a holdout sample. Another benefit of the textual classification model is that it provides feedback as to the specific text features, i.e., keywords or phrases, which may differentiate important from unimportant comment letters, giving potential insight into the mechanisms at work.

There is little evidence that comment letters are commonly used by investors, although the presence of commercial comment letter data vendors, such as Audit Analytics, indicates that some stakeholder demand exists, whether from investors, auditors, corporate users, or researchers.[1] This apparent investor inattention is supported by prior research, which finds that downloads of comment letters occur at approximately 1 percent of the rate of downloads of the associated 10-K report (Dechow, Lawrence, and Ryans 2016). The CFA Institute does not identify comment letters as an information source in financial analyst training materials (CFA Institute 2014), nor do widely used textbooks on financial analysis (e.g., Revsine, Collins, Johnson, and Mittelstaedt 2011). The financial press also makes very little use of comment letters as news sources.[2] The most prominent users of comment letters appear to be short sellers (e.g., Sandler 2013), who have the most incentive to identify negative information and publicize their results (Ljungqvist and Qian 2014).[3]

[1] In a conversation with Audit Analytics, it was revealed that few investors are customers of this data, which is primarily accessed by accounting firms and large corporate clients.

[2] Although there are infrequent examples of media articles sourced from comment letters (e.g., Gilbert 2014). A Factiva search of the Wall Street Journal during calendar 2013 reveals just five articles reporting on an SEC comment letter conversation with an individual company.

[3] Examples of short-oriented research that makes use of issues raised in comment letters include

If it is costly to process the information content in comment letters, or if investors simply pay little attention to comment letters in general, prices may underreact to their disclosure. Each letter in the conversation is filed separately on EDGAR, according to the date the original letter was issued, which is generally months prior to the date the letter is actually disclosed, thus making it difficult for investors to identify recently disclosed comment letters and gather all components of the conversation.[4] Because the conversation is comprised of separate communications, and because the subject matter may be both lengthy and technical, it is reasonable to suppose that comment letters are costly for investors to process.

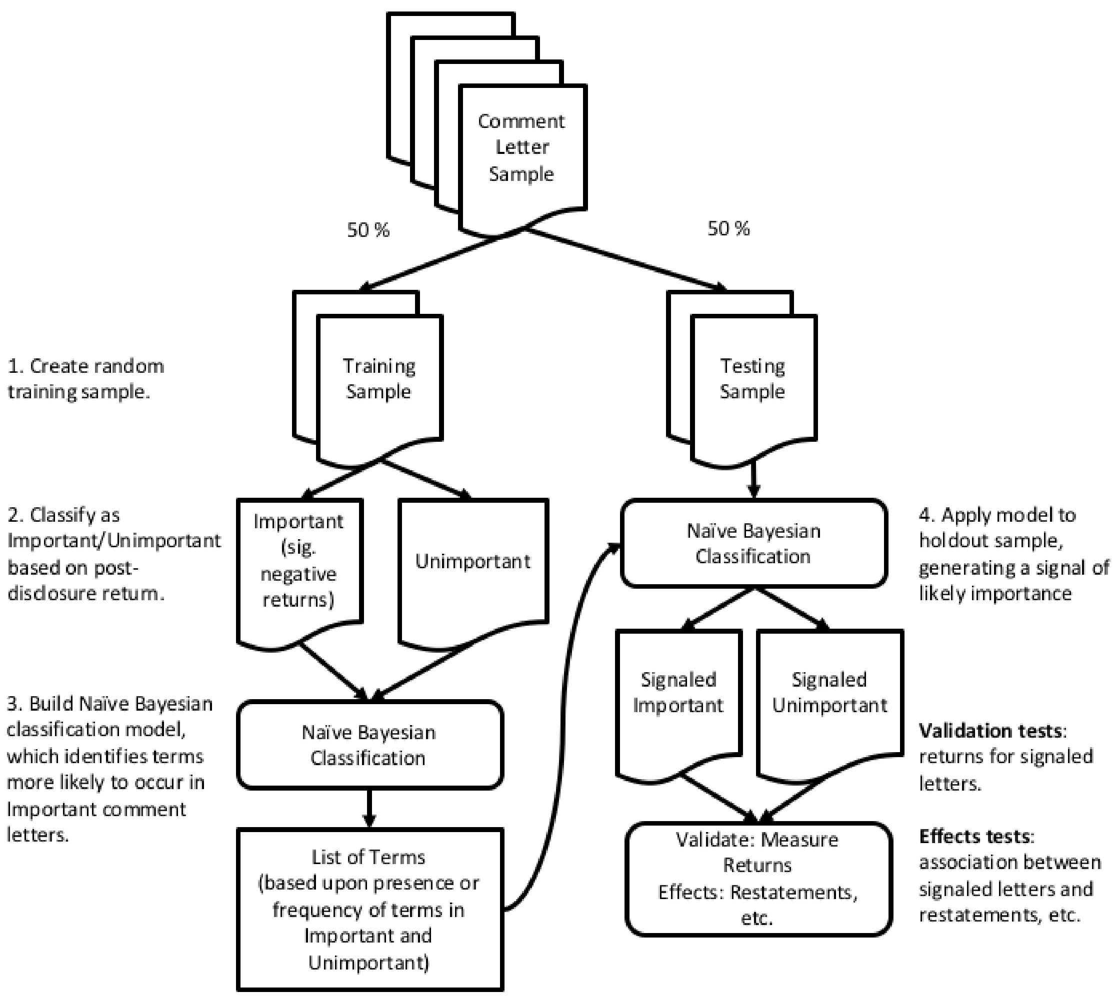
There are stakeholders whose actions indicate that comment letters are important. Public accountants are heavily involved in the comment letter process, as they assist client responses, and because comments made regarding their clients' financial reports reflects negatively on both financial reporting and audit quality. Accounting firms produce commentaries on comment letter trends, though these tend to be compilations of frequently-raised issues and sample responses, as opposed to economic analyses of implications for issuers or financial statement users (e.g., Deloitte & Touche LLP 2014). Insiders are another stakeholder group who are well-placed to be aware of important comment letters, and Dechow et al. (2016) show increased insider sales at comment letter disclosure, especially for those related to more important topics and in the presence of high short interest. Other studies examine the causes and consequences of comment letters, and use comment letters as a proxy for financial reporting and audit quality (e.g., Ertimur and Nondorf 2006; Cassell, Dreher, and Myers 2013; Hribar, Kravet, and Wilson 2014; Johnston and Petacchi 2016).

In this dissertation, I collect a comprehensive set of 10-K-related comment letters and company responses from the EDGAR web site, then I build a textual classification model to signal important comment letters using a four-step approach. Figure 1.1 provides a graphical representation of the classification and analysis process. First, I create a random training sample of comment letter conversations. Second, I classify comment letters in the training sample as important if the firms' abnormal returns are in the bottom quartile of abnormal returns following the comment letters' disclosure. I focus on negative returns as the signal of importance, because the SEC primarily

presentations by Greenlight Capital on Green Mountain Coffee (Greenlight Capital 2011), Pershing Square on Herbalife (Pershing Square 2013), and Prescience Point on Boulder Brands (Prescience Point 2013).

[4]Disclosure services such as FactSet and Morningstar Document Research allow investors to set up "alerts" to notify investors when new filings are disclosed, which partially reduces this burden.

Figure 1.1: Comment Letter Textual Classification Process



This figure illustrates the analysis process used in this study.

aims to identify disclosure weakness in their reviews. Third, I use the training sample to build a Naive Bayesian classification model that identifies the text features (words or short phrases) most associated with important comment letters. This model is then used to generate a *signal* for the importance a new comment letter, based on its text features. Thus, I use the term *signaled comment letters* to indicate that the textual classification model predicts a comment letter conversation is important. Fourth, I validate the effectiveness of the model to predict returns in a holdout sample that was not used to fit the model. The classification model detects important comment letters in the holdout sample by identifying those with subsequent price declines up to 40 percent more accurately than chance.

Within the holdout sample, I examine the relation between signaled comment letters and firm performance and measures of financial reporting and audit quality. I investigate underreaction to signaled comment letters by examining how the market response to signaled comment letters varies based on EDGAR views, finding that the signal is a significant predictor of negative post-disclosure returns only when the comment letters are viewed. For comment letters with above-median views in the three days post-disclosure, the signal is associated with abnormal returns of -1.2 percent three days, and -5.8 percent over 90 days. I examine the association between signaled comment letters and earnings, earnings persistence, material restatements, and internal control weaknesses, and find that firms with signaled comment letters have lower future persistence of profits and increases in material restatements. Signaled comment letters are associated with an increase in material restatements of 47 percent in the year following the comment letter conversation, indicating that important comment letters provide evidence of financial reporting weaknesses and lower audit quality.

This study extends the comment letter literature by examining the association between comment letters and future earnings, earnings persistence, material restatements, and internal control weaknesses, indicating that comment letters can be used to identify companies with weaker financial reporting and audit quality. This study also extends the literature relating to textual analysis of accounting disclosures by classifying large passages of text, and avoids typical hand-coding of training documents, since using tock market response to a document's disclosure as a signal of importance unaffected by researcher bias.

A limitation of this study is that textual analysis techniques distill large amounts of text into broad signals, and the underlying mechanisms that relate these signals to observed characteristics such as stock returns or material restatements cannot be precisely determined. In this setting, I expect the mechanisms to be diverse: the SEC may comment about a wide variety of issues and the firm may preempt the comment letter disclosure by restating prior financials during the comment letter responses, or may provide limited information in the comment letter response and instead making more substantial disclosures in subsequent filings. Furthermore, textual analysis techniques involve subjective model parameter selections, so similar results may not hold in different settings or for different research design choices. I attempt to address these issues by illustrating model performance across a range of parameters and providing the specific text features that the classification model associates with important comment letters.

Overall, this study suggests that comment letters do contain useful information, and that textual analysis techniques can be useful for analyzing larger passages of unstructured financial disclosures.

Chapter 2

Background

2.1 SEC Comment Letters

For decades, the SEC has conducted reviews of the disclosures of public issuers, and when these reviews require the SEC to ask questions of the issuer, the SEC does so via comment letters. These letters may request explanations or modifications to financial disclosures, and the issuers respond with their own letters as well as by amending past filings or modifying future filings. These reviews have been conducted as part of special SEC-designed projects (e.g., Schroeder and Gibson 1990), or in the course of reviewing filings in connection with the SEC's normal ongoing regulatory activities, such as during the security registration process (e.g., Ertimur et al. 2006). Historically, the comment letters and company responses were not publicly disclosed.

The issuance of comment letters became much more systematic for registrants as a result of legislation enacted as a result of the bankruptcies and frauds in the early 2000s. Per the requirements of Section 408 of the Sarbanes-Oxley Act of 2002, the SEC now reviews the financial reports of every public issuer at least once every three years, with the specific goal of investor protection (SOX 2002). If a review identifies issues that warrant additional disclosure, correction, or clarification, the examiner issues a comment letter, and a written correspondence with the issuer proceeds until the SEC is satisfied that all questions are resolved.

Beginning with comments on filings made after August 1, 2004, the SEC began posting all comment letters and the issuer's responses on the EDGAR web site for public dissemination 45 calendar days after the review completion. In 2012, the disclosure delay was reduced to 20 business days. The SEC, companies, and public accounting firms expend considerable resources reviewing firms, issuing comment letters, and responding to them: in 2014, the SEC conducted 4,350 reviews, an activity that represented the significant majority of the Division of Corporation Finance's headcount and $135 million budget (SEC 2015).

Table 2.1 provides summary statistics illustrating the estimated rate at which the SEC issues comment letters, given that a review was completed, by comparing the number of comment letters issued that reference an annual report filing to the number of reviews completed according to the SEC's budget reports. The result is that approximately 86 percent of reviews result in a comment letter (e.g., SEC 2015). There appears to be some variation in the rate of issuance over time, with comment letter issuance rates each year ranging between 65 percent and 91 percent over this period. Cassell et al. (2013) report that from 2006 to 2009, which includes two years for which I do not have comparable data, 23-37 percent of companies do not receive any comment letter, and since reviews happen at least once every three years and on average every two years, this statistic indicates an upper bound on the comment letter issuance rate of 63 to 77 percent during this time period, a rate that is not inconsistent with my tabulation. These estimated comment letter issuance rates are also comparable to the statistics reported in Schroeder et al. (1990), where the SEC conducted a focused review project of MD&A disclosures and issued 345 comment letters after conducting 362 reviews, an issuance rate of 95 percent.

Since the majority of reviews generate comment letters, a comment letter by itself does not necessarily indicate below-average financial reporting quality. The important question for financial statement preparers and users, when a comment letter is issued, is not whether the comment letter per se indicates deficient financial reporting, instead, the question should be whether a particular comment letter identifies important issues.

An underlying null hypothesis in studies of comment letters, especially those relating to the market's response to comment letters, is that they contain little useful information. On the one hand, most comment letters are generated by a mandatory review process of public filings, and examiners have much less information than other disclosure reviewers, such as auditors. As Johnston et al. (2016) note, comment letters are issued based on a review of already-public filings, and if markets are efficient at incorporating publicly available information, it is not clear why the questions of an albeit knowledgeable analyst, should have information content. The process also has poorly defined goals and outcomes, with SOX requiring simply that a review be conducted on a systematic basis "for the protection of investors" (SOX 2002). On

Table 2.1: Comment Letter Issuance Rate

Year	Percent of Issuers Reviewed	Number of Reviews Conducted	Comment Letters Issued	Comment Letter Issuance Rate
2014	52%	4,350	2,808	65%
2013	52%	4,500	3,556	79%
2012	48%	4,380	3,566	81%
2011	48%	4,773	4,342	91%
2010	44%	5,167	4,658	90%
2009	40%	5,152	4,613	90%
2008	39%	5,300	4,057	77%

This table presents an estimate of the rate at which the SEC issues comment letters based upon the number of reviews disclosed and the number of comment letters actually issued on annual reports. The average issuance rate from 2008-2012, the years with available statistics which overlap with this study's period, is 86 percent. Percent of issuers reviewed is a statistic reported by the SEC, defined as the number of reviews conducted, divided by the number of active public issuers. The number of reviews conducted is either reported directly by the SEC, or is calculated by multiplying the percent of public issuers reviewed by the number of issuers filing annual reports with the SEC. The number of comment letters issued is from the Audit Analytics Comment Letter database. The issuance rate is the nummber of comment letters issued divided by the number of reviews conducted. Source documents: SEC Annual Report and Congressional Budget Justifications for the years 2008-2016.

the one hand, this setting is notably different from analysis conducted by sell-side equity analysts, who may not be incentivized to discover or highlight accounting or other disclosure deficiencies, and such reviews may prompt the disclosure of material information and highlight reporting deficiencies. On the other hand, this setting may simply represent a bureaucratic process by which reviewers issue comment letters highlighting unimportant technical issues, with little benefit to investors. In short, the focus of this dissertation is: do any comment letters contain material information, and can these comment letters can be efficiently identified?

2.2 Prior Comment Letter Research

With the public availability of comment letters beginning in 2005, the literature is developing an understanding of the determinants of comment letters and their relation to financial reporting quality. At a macro level, the literature is also studying the effects of government monitoring of financial disclosures. Finally, other studies have been able to use information disclosed in comment letters to address research questions unrelated to comment letters themselves.

Early Comment Letter Studies

A number of academic studies consider comment letters in the context of financial reporting quality. To my knowledge, Schroeder et al. (1990) is the first study that reports on comment letters, describing the SEC's 1998 review of MD&A disclosure complexity. While this study does not specifically examine the impact of the comment letter process, it uses this targeted review setting to conduct textual analysis of annual report MD&As, presidents' letters, and footnotes. In addition to reporting the statistic that 95 percent of reviews resulted in a comment letter, they note that 122 of 345 firms receiving comment letters amended their filings, and half of these amendments involved expanding the MD&A. These results provide evidence that the SEC's reviews were effective in inducing firms to modify their disclosures.

Ertimur et al. (2006) is a more recent study of comment letters, conducted prior to the wide availability of comment letters issued in response to SOX reviews. Instead, they hand collect comment letter data based on IPO reviews, which is conducted by the same SEC Division of Corporation Finance staff, and is conceptually similar to reviews for already-public issuers. However, as their target firms are conducting an initial public offering, the subject firms and the type of issues identified will likely be much different from SOX reviews of public issuers, who may be expected to have more experience and have more developed financial reporting systems in place. Ertimur et al. (2006) describe the comment letter conversation duration and number of rounds, as well as categorize the issues discussed. Their research question is primarily concerned with the effect that management experience and corporate governance have on the cost of the comment letter process, and in particular they find that management experience is associated with shorter comment letters. Ertimur et al. (2006) also examine the effect of the comment letter process on the IPO firm information environment, but do not find significant evidence of comment letters affecting IPO underpricing or bid-ask spreads, though they do find that the number of comments and number of issues raised is associated with less market depth subsequent to the IPO.

Another early work to consider comment letters is Correia (2009). In her dissertation, Correia studies the link between political contributions from firms and executives, and SEC enforcement. Whether through executives exerting influence in the enforcement process or through a signaling effect, political contributions are predicted to indicate lower accounting quality as well as a lower probability of investigation. The empirical findings indicate that low accounting quality firms do target contributions to SEC-related Congressional committee members. By examining both comment letters and enforcement actions, Correia shows that politically connected firms are less likely to restate as a result of a comment letter, are less likely to be subject to enforcement actions, and pay lower monetary penalties.

Cassell et al. (2013) look at the main company-specific factors associated with receiving a comment letter. They validate that the SEC does indeed appear to conduct more frequent reviews of firms with factors specified by SOX Section 408, including past restatements, larger size, unusual price-to-earnings ratios, etc., and further note that letters are more frequent for firms with lower profits, weaker governance, higher complexity, and smaller audit firms. They also study the cost of remediation as proxied by the number of days and the number of rounds needed to complete the comment letter conversation. Restatements increase for smaller companies and for companies with smaller audit firms. Finally, Cassell et al. (2013) consider the effect of comment issue types on remediation costs and find that letters related to accounting issues such as classification and fair values take the longest to resolve.

Johnston et al. (2016) describe the range of resolutions to comment letters, and look at their effect on the information environment. In their sample, 17 percent of comment letters lead to some type of amendment, both major and minor. They also find changes in the information environment subsequent to comment letter resolution: analyst forecast accuracy improves, though absolute abnormal returns and trading volume around earnings announcements declines. It is not clear that these results support inferences of an improved information environment, as increasing ERCs are often viewed as a signal of better earnings quality (e.g., Chen, Cheng, and Lo 2013).

Comment Letter Topics

Early research into comment letter content (e.g., Ertimur et al. 2006; Johnston et al. 2016) relies upon the hand coding of comment letter comments into author-defined categories and subjects. Subsequent comment letter analysis generally relies on the Audit Analytics comment letter database, where issues are coded by the data provider. The Audit Analytics comment letter database facilitates topic analysis by coding comment letter issues into a standardized hierarchy of topics. Table 2.2 gives an overview of the Audit Analytics Issue Taxonomy, a three-level structure of issues from which Audit Analytics selects one or more items to indicate the subject of the comments in each letter. There are five high level categories of comment types, such as “Accounting Standards” and “Mergers & Acquisitions”. The second level comment issue categories allow for more specific identification of areas covered in the reviews, e.g. “Accounting Rule and Accounting Disclosure Type Issues”, and “Risk Factors Disclosure”. Brown, Tian, and Tucker (2015) identify risk factor-related comment letters when any second-level “Risk Factors Disclosure” classification items are identified for a comment letter. Within each of these second-level issues, there are currently more than 2,500 specific issue codes, which in some cases are granular down to the level of specific regulatory documents (e.g., “SEC Release No. 34-62934”), or may also broadly describe the whole letter (e.g., “Closing SEC letter associated with SEC commentary”). Table 2.3 shows the detailed third level of issues provided for just one second-level issue category, Accounting Rules and Accounting Disclosure Type Issues.

Comment Letter Information Content

Because investor processing costs for interpreting comment letters is high, Dechow et al. (2016) focus on the activities of firm insiders, who are best placed to recognize

Table 2.2: Audit Analytics' Comment Letter Issue Taxonomy

Comment Type (Level 1)	Comment Issue (Level 2)
Accounting Standards	• Accounting Rule and Accounting Disclosure Type Issues • EITF GAAP Standard Citations • FASB Accounting Standards Updates • FASB Concepts Statements • FIN (FASB Interpretation) guidance • FSP (FASB Staff Position) guidance • FTB (FASB Technical Bulletin) guidance • IAS (International Accounting Standards) • IFRS (International Financial Reporting Standards) • IFR Interpretations Committee • PCAOB Rules and Standards • SAB (Staff Accounting Bulletin) guidance • SFAS GAAP Standards • SIC (Standing Interpretations Committee) • SOP (Statement of Position) AICPA guidance
Mergers & Acquisitions	• Tender Offer Specific Comments
Registrations	• Registration Statement Specific Comments
Non-Standard and Other Disclosures	• Event Disclosure Matters (primarily 8K, or 6K items) • Federal Securities Statutes References • Legal Matters and Supreme Court Decisions • Whole Letter Description
Operational, Controls & Risk Assessments	• Disclosure and Internal Control Issues • MD&A Type Disclosure Issues • Risk Factors Disclosure
Securities Regulations	• Exchange Act Rules and Regulations • Investment Company Act of 1940 Rules and Regulations • Regulation AB • Regulation M-A References • Regulation S-K References • Regulation S-X References • SEC Releases • Securities Act Rules and Regulations

This table presents the top two (of three) levels of the Audit Analytics comment letter issue taxonomy. Each comment letter in the Audit Analytics database is coded with one or more of these issues. Each issue has sub-issues, and Table 2.3 lists the sub-issues associated with the Accounting Rule and Accounting Disclosure Type Issues category.

Table 2.3: Audit Analytics' Accounting Rule and Accounting Disclosure Type Issues

- Accounts receivable & cash reporting
- Acquisitions, mergers, and business combinations
- Asset retirement obligation (FAS 143)
- Asset sales, disposals, divestitures, reorganization
- Balance sheet classification of assets
- Capitalization of expenditures
- Cash flow statement classification errors
- Changes in accounting estimates
- Changes in accounting principles and interpretation
- Comprehensive income (Equity Section)
- Consolidation (FIN 46, variable interest, SIV, SPE & off-B/S)
- Consolidation, foreign currency/inflation issue
- Contingencies & Commit, legal, accounting
- Debt and/or equity classification
- Debt, quasi-debt, warrants & equity (BCF) security
- Deferred, stock-based and/or executive comp
- Deferred, stock-based options backdating only
- Deferred, stock-based SFAS 123 only (subcategory)
- Depreciation, depletion or amortization reporting
- Dividend and/or distribution
- EPS, ratio and classification of income statement
- Expense (payroll, SGA, other) recording
- Fair value measurement, estimates, use (incl. VSOE)
- Fin statement segment reporting (subcategory)
- Financial derivatives/hedging (FAS 133) acct
- Foreign (affiliate or subsidiary)
- Gain or loss recognition
- Intercompany accounting
- Inventory, vendor and/or cost of sales
- Investment in subs./affiliate
- Investments (SFAS 115) and cash and cash equivalents
- Lease, leasehold
- Liabilities, payables, and accrual estimate
- Loans receivable, valuation and allowances
- Loss reserves (LAEs, Reinsurance) disclosure
- Non-monetary exchange
- Pension and related Employee Plan
- Percentage of completion
- PPE fixed asset (value/diminution)
- PPE - Intangible assets and goodwill
- Research and Development
- Revenue recognition (incl deferred revenue)
- Subsidiary – US or foreign (subcategory)
- Tax expense/benefit/deferral/other
- Tax rate disclosure

This table presents presents the detailed (third level) issues relating to Accounting Standards: Accounting Rule and Accounting Disclosure Type Issues, from the Audit Analytics comment letter issue taxonomy.

important comment letters affecting their firms. They find that comment letters relating to revenue recognition are more likely to spur insider trading around the disclosure of such letters, and that insider trading is more pronounced for firms with high short interest, indicating the presence of sophisticated consumers of firm disclosures. Firms with greater levels of insider selling around the time of issuance had more negative returns post-disclosure period. The negative return drift may be due to investor inattention, and comment letters with above-median downloads have more negative post-disclosure returns.

Gietzmann and Isidro (2013a) examines the effect of comment letters on the firm information environment, by considering changes in institutional holdings in response to comment letters for a sample of multinational US-foreign cross listed firms who have foreign accounting issues raised by the SEC. This represents a joint test of the effect of government oversight of foreign accounting standards, and the effect of the comment letter process on institutional shareholders. They find that institutional investor portfolio rebalancing is associated with comment letters, indicating that these letters are informative signals. The effect is most pronounced for low-turnover institutions, which are also those hypothesized to be the more sophisticated consumers of firm disclosures. In addition to equity investors, debt investors are often sophisticated consumers of financial information, and Cunningham, Schmardebeck, and Wang (2016) examine the effect of comment letters on debt contracting. SEC reviews lead to higher loan costs, and comments relating to material errors (i.e., comments resulting in material restatements), management estimates, and asset valuations are associated the increase in loan costs.

In the specific setting of fair value-related comments, Bens, Cheng, and Neamtiu (2016) investigate whether SEC monitoring of fair value disclosures has an impact on uncertainty about the firms' fair value estimates, in particular whether the comment letter process reduces investor uncertainty about Level 2 and 3 asset values. Improved reporting quality may occur through two mechanisms: first, the SEC comment letter process itself may reveal more complete and accurate information, and second the presence of the SEC review may encourage other monitors, such as auditors, to provide better oversight. The findings of this study indicate that comment letters do indeed reduce investor uncertainty, as measured by reductions in the bid-ask spread and returns volatility following 10-K disclosure from the pre-comment letter to post-comment letter period.

Monitoring and the Effects of Regulatory Oversight

Naughton, Rogo, Sunder, and Zhang (2016) also examine SEC monitoring in the sample of US-foreign cross listed firms, to gauge the level of investor protections in a joint regulator oversight setting. Given the varying levels of legal and regulatory oversight in the issuers' home countries, they find that foreign firms are subject to less SEC monitoring than domestic issuers, and furthermore that the SEC's monitoring intensity is negatively associated with the strength of the issuers' home country regulatory infrastructure. Whether this is due to these firms having higher financial disclosure quality, or to the SEC prioritizing resources towards cross-listed firms from less stringent countries, the net effect is that the SEC moderates oversight activities to provide more consistent overall investor protection.

Comment Letters as Measures of Audit Quality

Another line of investigation builds upon the descriptive evidence in Cassell et al. (2013), that comment letters are associated with smaller auditors, indicating that comment letters may be a fruitful setting for investigating audit quality. This is a logical course of investigation: if SEC reviewers are able to identify disclosure deficiencies, inconsistencies, or other weaknesses based upon their reviews of public filings, then it seems reasonable to assume that the auditor's review of their client's financial statements should have been able to also identify the same issues. Granted, the firm or the auditor may have decided that the risk of receiving an SEC comment is outweighed by the cost of preemptively addressing all potential issues. In an unpublished working paper Lawrence, Lei, and Smith (2010) address find that the SEC generally issues comment letters with greater frequency for firms with characteristics identified in SOX Section 408, and that SEC reviews may prompt restatements, especially within firms that have weaker monitors, such as auditors.

Gietzmann and Pettinicchio (2013b) consider comment letters as an early warning signal of regulatory action, and examine the effects of comment letters on audit pricing. Their findings indicate that auditors adjust fees upwards both in the period during which the comment letter is received, which is not surprising considering auditors may be involved in their clients' responses, and adjust fees upwards in future periods. The authors find that this is consistent with auditors re-pricing risk following comment letter issuance, but these results may be difficult to interpret for future periods, as virtually all firms have now received publicly disclosed comment letters at some point. Hribar et al. (2014) validate the use of comment letters as an indicator of audit quality by showing that unexpected audit fees are related to comment letter issuance, as well as to the incidence of fraud and restatements. Baldwin, Hurtt, and MacGregor (2013) examines the auditor relationship and finds more frequent auditor changes when comment letter conversations take longer to resolve, concluding that the receipt of a comment letter can be interpreted as an indicator of lower audit quality.

Comment Letters and Governance

As disclosure quality is related to firm governance, comment letter also present an opportunity to examine interactions between firm governance, financial reporting, and government monitoring. Ettredge, Johnstone, Stone, and Wang (2011) examine comment letters related to Form 8-K, Item 4, specifically the failure to properly disclose circumstances surrounding auditor changes. This setting identifies firms that decide not to disclose bad news, as that bad news is subsequently revealed in the firm's response to the comment letter. Firms that fail to disclose bad news regarding auditor changes are associated with poor corporate governance and have lower external financing needs. In addition, when the news is bad (e.g., auditor resigned, disagreements, etc.), firms are less likely to comply with disclosure requirements, providing evidence that managers do strategically withhold bad news which may be subsequently revealed at the prompting of a comment letter.

In another study of compliance with disclosure regulations and the effects of monitoring, Robinson, Xue, and Yu (2011) consider another of the SEC's focused studies, whereby the SEC examined compliance with new compensation disclosure requirements enacted in 2006, by reviewing a random sample of annual proxy statements.They consider factors associated with noncompliance, and in particular if the noncompliance was associated with CEO compensation, proprietary costs, and media attention. They find that noncompliance was primarily related to higher levels of compensation and media scrutiny, but not other proprietary costs, as proxied by industry competitiveness. In this setting, the regulator's oversight and intervention in the disclosure deficiencies resulted in the correction of the deficiencies, but did not have an impact on executive compensation following the revised disclosures.

Comment Letters and Tax

Another set of studies utilizes comment letters to examine the effect of SEC monitoring on tax-related issues. Kubick, Lynch, Mayberry, and Omer (2016) look at the changes in tax avoidance behavior in the years surrounding tax-related comment letters. SEC monitoring appears to consider tax issues, because firms with more aggressive tax planning (lower GAAP and cash effective tax rates) are found to be more likely to receive comment letters with tax-related questions. Such firms subsequently report higher effective tax rates, consistent with managers adjusting assumptions in response to the SEC's questions. In a similar vein to Brown et al. (2015), Kubick et al. (2016) show spillover effects with firms who did not receive tax-related comments reducing their tax rates when competitors are questioned, potentially learnings from their competitors disclosures about the the threshold for regulatory scrutiny. In a related work, De Simone and Stomberg (2015) illustrate that firms with executives who possess prior tax expertise are better aware of these thresholds, as this expertise is associated with less frequent tax related comment letters and restatements, despite reporting lower GAAP effective tax rates.

Chen (2015) and Blouin, Krull, and Robinson (2014) use comment letters to examine the effect of regulatory oversight of the tax effects of foreign earnings. Chen (2015) show that tax-related comment letters encourage firms to disclose previously withheld information: the amount of foreign cash holdings. This is an important disclosure, as multinational firms are taxed on worldwide income, yet firms which designate foreign earnings as permanently reinvested offshore can avoid estimating deferred tax liabilities on the future repatriation of such profits. If investors are unaware of the amount of foreign cash, then SEC-prompted disclosures may be material to investors as it potentially reveals bad news about the value of cash, net of ultimate tax liabilities or other agency costs. After the SEC began asking firms about the amount of offshore cash holdings in 2011, Chen (2015) finds that firms that receive such a comment letter are 18 percent more likely to disclose foreign cash in the comment letter year, compared to firms that do not receive a comment letter, and the comment letter is the single largest predictor of foreign cash holdings disclosure.

Blouin et al. (2014) look at permanently reinvested earnings more broadly, finding that a large fraction of such earnings are held in financial assets and in tax haven jurisdictions. This is important because it indicates that a large fraction of such earnings may not be designated as permanently reinvested for operational reasons, but are instead so designated to affect the reported GAAP tax rate, and further that restrictions on the use of these assets result in agency costs. The SEC is more likely to issue comment letters to firms with higher levels of permanently reinvested earnings, that operate in tax havens, and that report lower effective tax rates. These results indicate that the SEC is concerned with improving disclosure of these value-relevant factors for which investors may have little insight absent the SEC-prompted disclosures.

Other Comment Letter Related Topics

Bozanic, Choudhary, and Merkley (2015) recognize that securities lawyers play a role in formulating financial disclosures and in dealing with regulators, and they study the effects of legal counsel on the comment letter resolution process. As may be expected, legal counsel is more likely to be involved in a comment letter response when more complex issues are raised by the SEC, and when the issues are not strictly accounting-related, as such issues are more likely the domain of the auditor and not outside counsel. Counsel is also more likely to be involved for firms that have faced class action lawsuits and have more complex annual reports. Finally, the presence of outside counsel is also associated with a more adversarial approach, as firms appear to involve such experts when they are withholding bad news, revealed through a greater price decline upon the filing of amendments associated with the comment letter conversation.

If disclosure requirements are unclear, firms may provide non-compliant disclosures because the ambiguity may leave them uncertain of how to provide acceptable disclosures. If enforcement priorities are unclear, firms may provide non-compliant disclosures if they believe the chances of enforcement are low. As the SEC is only required by SOX to review firms once every three years, comment letters can provide guidance to audit firms and to the firms who were not reviewed regarding both disclosure requirements and enforcement priorities. Such changes made by non-reviewed firms are termed spillover effects, and Brown et al. (2015) focus on how changes to risk factor disclosure between comment letter firms and no-letter firms are linked. They find that these effects are greater when the SEC comments on more salient peers: industry leaders, and close rivals. Firms also recognize patterns, via similar comments made to several industry peers, as evidence of greater enforcement priority. Consistent with firms learning about the SEC's disclosure requirements and enforcement priorities from these letters, firms making changes in response to peer firm comments have fewer comments of their own when they are reviewed.

To better understand the review process and the information it reveals to investors, Boone, Linthicum, and Poe (2013) examine the likelihood or receiving comments, and cost of resolution, in relation to the rules-based nature and the extent of management estimates required by specific accounting standards. They find that comments are more likely for rules-based standards, as deficiencies in such standards may be easier for a reviewer to identify. Comments are also more likely for standards involving estimates, as the SEC reviewer may request information about managers' estimates behind a disclosure. The length of time to respond to comments is not affected by rules based or complex standards, but it is positively related to management estimates. However, such estimates may be difficult for management to justify, or that explaining estimates is naturally more complicated, or the firm may provide incomplete responses in an attempt to avoid disclosure.

In more recent work utilizing comment letters in the IPO setting, Gupta and Israelsen (2015) examine the effect of the JOBS Act on IPO outcomes, and look at how comment letter content changes in the post-JOBS act period. They find that the SEC requests more "soft" information when firms reduce the disclosure of "hard" information, and that the SEC increases the length of their comment letters, providing evidence that the SEC exerts its influence to solicit more information for investors when firm disclosures are less informative.[1]

Cassell, Cunningham, and Lisic (2015) makes an initial effort to examine text properties of comment letters. By measuring the readability (average sentence and word length) of company responses to comment letters, they find that the complexity of the language in the letters is related to the cost of remediation, measured by the number of days it takes the SEC to respond to the company's initial response letter and the number of days it takes the SEC to close the filing review. Less complex response letters are also associated with a lower probability of a restatement stemming from the filing review. The authors conclude that simplicity in formulating responses to SEC comments can have a significant effect on regulators' reaction to the disclosure, however an alternative explanation is that complex issues and weak financial disclosure practices require more complex answers, making a causal inference difficult.

Finally, comment letters may also provide information unrelated to the comment letter process itself, but which is of interest to investors or researchers. Laurion,

[1]Hard information is defined as verifiable, and the authors give audit fees as an example. Soft information is non-verifiables, and the authors give a reduction in competitive advantage as an example.

Lawrence, and Ryans (2015) build a database of audit partner names, as audit partners are frequently copied by name in comment letter responses. Laurion et al. (2015) use this identification of audit partners to observe audit partner rotations, and examine the effect of partner rotation on audit quality using material restatements and changes in allowances as evidence of the effects of the new partner on firm disclosure.

To date, the comment letter literature has broadly investigated the determinants of receiving a comment letter, finding that they generally align with conditions specified in SOX, such as past financial statement restatements and high stock return volatility. A second theme is the association between comment letters and audit quality, via restatements or audit firm turnover. Finally, the comment literature illustrates the impact of comment letters on disclosure compliance and management estimates, using a variety of channels: fair value estimates, tax estimates, and compensation disclosures.

2.3 Textual Analysis in Accounting and Finance

As the comment literature develops, methods for classifying or interpreting the content of these letters may be valuable to future studies. To date, comment letters are primarily used as an indicator variable: the absence or presence of a comment letter is a proxy for disclosure or audit quality (e.g., Hribar et al. 2014). Since comment letters appear to be issued at a high rate when reviews are conducted, and since all firms are reviewed on a regular basis, the presence of a comment letter alone cannot provide a very precise signal. More granular analysis of comment letters can be achieved by identifying comment letters that are known to refer to a specific topic, such as revenue recognition, risk factors, or tax (e.g., Dechow et al. 2016; Brown et al. 2015; Blouin et al. 2014).

Issue categories are an effective way of identifying relevant comment letters to study, though to my knowledge, only Dechow et al. (2016) utilize specific topics as a proxy for comment letter importance, when they use revenue recognition comment letters as a proxy for more important reviews. Another way to examine the content of the comment letters is via various statistical text analysis techniques. Li (2010a) and Loughran and McDonald (2015) provide recent surveys of textual analysis in the accounting and finance literature. Relative to quantitative methods traditionally used in accounting and finance, textual analysis may be considered to be less precise. Textual analysis seeks to distill some statistical summary of the words or phrases used in a document into a quantitative signal. A wide variety of techniques are used based on the research question, and so an understanding of the methods and the types of research questions to which they are suited, as well as their potential shortcomings, is useful to understand research that deals with the textual elements of financial disclosures, the precise setting we face when examining comment letters.

Statistical text analysis has been used in accounting research as a response to the difficulty and cost of manual data collection for content analysis, which necessitates small sample sizes when such techniques are not used. A prominent method in the accounting and finance literature is dictionary based techniques, which use wordlists with pre-supposed meanings to identify the tone or topic of a text, without the need for manual coding. An early such example in the accounting literature is Bryan (1997), who look at the MD&A section of annual reports to assess their information content for predicting future performance. Forward looking discussions are identified by the presence of dictionary words associated with future actions, and the presence of such words in the MD&A are associated with future performance and investment activities. Longer-term associations are generally not significant.

Another way to examine variation in text is through complexity, as measured by document length and reading difficulty. This type of analysis is reminiscent of Schroeder et al. (1990), which reported on the SEC's review of MD&A complexity and the resulting comment letters. Li (2008) examines annual report complexity and its relation to firm performance and earnings persistence. The specific measure used is the Fog index, which is increasing in average number of words per sentence and the fraction of long words. The primary findings are that firms with lower earnings have more complex MD&As and firms with lower Fog scores have more persistent earnings.

To study a more narrow mechanism by which disclosure text may provide indications of accounting quality, Peterson (2012) focuses on the relation between revenue recognition discussion complexity and the probability of restating revenue. He finds that revenue recognition complexity increases restatement likelihood on an intentional and unintentional basis. While the relation to restatements is higher, complexity also appears to increase the threshold for enforcement and pricing consequences: complexity is associated with fewer AAERs, less negative event returns surrounding restatement announcements, and lower CEO turnover.

Complexity does not take into account content, beyond the possibility that certain types of content, for example bad news, may require more complex language to discuss compared to good news. A relatively simple method for content analysis is using dictionaries of words that are coded to signal the presence of some aspect of content. Words have been classified, for example, as: negative, positive, uncertain, litigious, and constraining. A simple technique for identifying the tone of a passage of text is to sum up the number of positive words and subtract the number of negative words, to give a net positive or negative tone score. Tetlock (2007) looks at media tone and stock market performance, using text from a daily Wall Street Journal column. Media pessimism predicts temporary declines in prices, and abnormal pessimism predicts higher trading volume. Kothari, Li, and Short (2009a) examine the tone of text from several sources, including management disclosures, analyst reports, and the news media. More positive tone is associated with lower cost of capital, price volatility, and analyst forecast dispersion, and negative tone is associated with higher levels of these measures.

Davis, Piger, and Sedor (2012) study the tone, classified as optimistic or pessimistic, of earnings announcements. Earnings announcements are one of the most important firm disclosures in terms of information content, and they mix quantitative earnings results with a more qualitative commentary that may communicate information about future performance. The study finds that there is an incremental contribution of earnings announcement tone such that it is associated with future firm performance and generates a market response. Related to earnings announcements, Larcker and Zakolyukina (2012) study earnings announcement conference call transcripts, and link “deceptive” financial reporting back to linguistic features of the conference call discussion. The authors label a conference call as being truthful or deceptive based upon subsequent outcomes (e.g., AAERs, restatements), and identify the presence of words associated with potential deceptiveness (e.g., anger, certainty, hesitations). Such word features are then used to predict deception outcomes out-of-sample, performing better than a random guess by 6-16 percent. This is a rate similar to or better than models based on financial and accounting variables.

Feldman, Govindaraj, Livnat, and Segal (2010) study the tone of 10-Q and 10-K reports, in particular focusing on the change in positive and negative MD&A tone compared to prior filings. They find that short window returns surrounding the filing date are significantly associated with the change in tone, after controlling for earnings surprises, but do not consider guidance, presumably as the MD&A tone incorporates the positive or negative guidance that they wish to study.

The overall impression these studies present is that textual analysis based on dictionary classifications borrowed from other domains can be effective, despite evidence that commonly used dictionaries can be misleading or ambiguous in the financial setting. For example, the word *decline* is classified as a negative word, but in the financial context, a company could report *declinining revenues*, a negative result, or, they could report *declining expenses* a positive result. Loughran and McDonald (2011) show that such word lists misclassify common words in financial documents. They find that almost 75 percent of 10-K words identified as negative are words typically not considered negative in financial contexts, and they generate a new word list, validated using returns, volume, class-action lawsuits, and material weaknesses.

Comment letters present a challenge to researchers studying the economic impact of their information content because they have an unstructured format and do not present consistent numerical statistics, such as earnings. This setting naturally lends itself to textual analysis techniques, in particular the concept of text classification, which attempts to determine the class of a document based upon the specific words or groups of words used in the text. The Naive Bayesian classification method is one of the most established methodologies used to classify texts, and is currently experiencing a renaissance in machine learning in a wide variety of fields, including accounting and finance (e.g., Lewis 1998; Loughran et al. 2015).

Text classification at its core refers to the identification of a passage of text as a belonging to a class. The classes into which a text is assigned can be arbitrarily defined. Table 2.4 gives some examples of classification schemes and studies incorporating the class definitions. Text classification has been used to study authorship (e.g., Mosteller and Wallace 1984), genre (e.g., Karlgren and Cutting 1994; Kessler, Numberg, and Schutze 1997), news category (e.g., Feldman and Dagan 1995; Dagan, Feldman, and Hirsh 1996), and the sentiment of movie reviews (e.g., Pang, Lee, and Vaithyanathan 2002). In the law literature, Talley and O'Kane (2012) identifies the properties of specific clauses within merger agreements.

The Naive Bayesian classification method is one of the most established methodologies used to analyze text (e.g., Lewis 1998; Loughran et al. 2015). Li (2010b) uses Naive Bayesian classification to automatically identify positive and negative sentences within MD&A sections of 10-K and 10-Q filings, to identify tone in a manner not dependent on the dictionary analysis. Li (2010b) finds that firms with more positive forward looking statements in their MD&As have better future performance, after controlling for other determinants, and also have better current performance, lower accruals, smaller size, lower market-to-book ratio, less return volatility, and lower MD&A complexity. A key finding in this work is that tone measures based on three commonly used dictionaries (Diction, General Inquirer, and

Table 2.4: Example Text Classification Schemes

Classification	Classes	Example Study
Authorship	{ *authored by Hamilton*, *authored by Jackson* }	Mosteller and Wallace (1984)
Text Genre	{ *legal*, *technical*, *nonfiction*, *fiction* }	Kessler, Numberg, and Schutze (1997)
News Subject	{ *oil industry*, *technology industry*, *auto industry*, ... }	McCallum and Nigam (1998)
Film Rating Opinion	{ *highly rated*, *neutral rated*, *low rated* }	Pang, Lee, and Vaithyanathan (2002)
Tone	{ *positive tone*, *negative tone* }	De Franco, Vasvari, Vyas, and Wittenberg-Moerman (2013)
Financial Disclosure Importance	{ *value relevant*, *not value relevant* }	This study

This table presents presents some examples of text classification schemes and example classes for which texts are associated, along with an example of a study implementing the classification scheme.

the Linguistic Inquiry and Word Count) are not effective in this setting.

De Franco, Vasvari, Vyas, and Wittenberg-Moerman (2013) use Naive Bayesian classification to identify sentence tone in sell-side debt analysts' discussions about debt-equity conflicts such as mergers and acquisitions, debt issuance, share repurchases, and dividend payments. As debt analysts routinely discuss these conflict events in their reports, the authors find that negative-tone discussions about conflict events are associated with increases in credit spreads and bond trading volume and predict higher bond offering yields for new issues. The informativeness of debt analysts' negative tone is greater when equity analyst reports have a more positive tone.

Huang, Zang, and Zheng (2014) uses Naive Bayesian classification of opinion in equity analyst reports to identify informative content beyond the simultaneously issued financial results, stock recommendations, and target prices. They find that investors react more strongly to negative than to positive text, indicating that analysts can significantly affect sentiment and highlight bad news. Analyst text is shown to have predictive value for future earnings growth, and it is more informative when the text is more confident and when it emphasizes non-financial topics.

In general, the accounting related studies that Naive Bayesian classification to identify tone or content categories, the authors begin by manually coding a fairly large sample of text passages to train the system to recognize the relevant categories. For example, Li (2010b) hand codes 30,000 MD&A passages for tone and content categories, De Franco et al. (2013) do the same for the tone of 5,933 debt analyst report passages, and Huang et al. (2014) for the tone of 10,000 equity analyst report sentences. Clearly this process is both costly and may be subject to researcher bias (Loughran et al. 2015). In addition, it is not immediately scalable to new settings as the researchers do directly provide a classification model that can be applied outside of their data set, instead researchers would have to replicate the process including hand coding a training set.

2.4 Financial Statement Restatements, Accounting, and Audit Quality

A recurring theme in research relating to accounting and audit quality is restatements. By their nature, material restatements, in particular of the financial statements, as opposed to other more minor amendments, reflect upon financial reporting quality and audit quality. Liu, Raghunandan, and Rama (2009) note that regulators and legislators have increased their attention to financial statement restatements in recent years, and that both the SEC and financial statement users consider restatements to be evidence of audit failure. In many ways, the concepts of accounting quality, earnings quality and audit quality are related and overlapping, and research tend to use similar measures to proxy for all of these concepts. In their survey of earnings quality, Dechow, Ge, and Schrand (2010) note that researchers measure earnings quality using earnings persistence, accruals, timeliness, loss avoidance, price response, restatements, AAERs, and class action lawsuits. Dechow, Ge, Larson, and Sloan (2011) identify both financial and nonfinancial measures that are predictive of restatements.

This study places an emphasis on material restatements, as the SEC monitoring of financial reports should, if the SEC can effectively identify errors in existing reports or flaws in firms' internal financial reporting systems, result in greater restatements when such errors are identified or later, when firms' improve their internal procedures. Hribar and Jenkins (2004) shows that restatements are important to investors, affecting a firm's cost of equity capital and decreasing expected future earnings. They find an economically large magnitude to the effect, with cost of equity increasing between 7 and 19 percent following the restatement. In a similar vein, Palmrose, Richardson, and Scholz (2004) examine the market reaction restatements and find an average abnormal return of about -9 percent. Restatements related to accounting issues including fraud, relating to a greater number of accounts, and those decreasing income cause a greater reaction.

Francis (2004) provides a survey of audit quality research, through the period of the enactment of SOX, which represents an important regulatory regime change, implementing limits on non-audit services, to improve audit quality. Kinney, Palmrose, and Scholz (2004), considers restatements from the perspective of audit quality, examining the impact of non-audit service fees on restatements, finding not a link between more non-audit fees and increased restatements, but instead that greater tax services fees result in lower restatements.

Accounting quality questions raised for one firm can have spillover effects to other firms in the same industry. Gleason, Jenkins, and Johnson (2008) finds that restatements affecting the value of the restating firm also affect the value of non-restating firms in the same industry, though this effect seems limited to more salient accounting issues, such as revenue restatements. Linking restatements to audit quality, they also find that the peer-effect is stronger for firms with the same auditor. Liu et al. (2009) illustrates that investors reveal their belief that restatements result from weak audits by being more likely to vote against reappointment of the firm's auditor after a restatement.

Chen et al. (2013) provide more recent evidence on the pricing effects of restatements and their implications of restatements for financial reporting quality. They find decreases in ERCs for up to three years following material restatement announcements, and for only one quarter with non-material restatements. Firms that take actions to improve credibility find improved ERCs, for example, by replacing the CEO, CFO, or external auditor.

While restatements are an obvious area of investigation relating to comment letters, internal control weaknesses may also provide insight into the the interactions between firm financial reporting practices, accounting quality, and regulatory oversight, as errors or inconsistencies and failure to comply with applicable accounting standards is evidence of inadequate internal controls. Internal control deficiencies are associated with information uncertainty and negative announcement returns. Doyle, Ge, and McVay (2007) make a value-relevant association between weaknesses estimated accruals that are not realized in cash flows. Internal control weakness disclosures are associated with abnormal returns and an increased cost of capital, leading to the conclusion that such disclosures indicate lower financial reporting credibility (e.g., Beneish, Billings, and Hodder 2008; Hammersley, Myers, and Shakespeare 2008; Ashbaugh-Skaife, Collins, Kinney, and LaFond 2009).

2.5 Investor Attention and Inattention to Financial Disclosures

Comment letters are more difficult to find and interpret than other commonly-read filings, which raises the possibility of investor inattention to this information source. The information contained in a complete comment letter conversation is distributed among several different EDGAR filings, and an investor needs to identify and read each related comment letter (Form UPLOAD) and company response (Form CORRESP), to observe the full scope of the conversation. The SEC's EDGAR website organizes comment letters chronologically according to filing date, the date that the document was processed by EDGAR, but not on the date the letter was disclosed, making it difficult for investors to identify timely comment letters.

Studies considering the market response to comment letters may therefore find it important to also consider the effects of potential investor inattention to these filings (e.g., Dechow et al. 2016). The flip side of inattention—attention—is a setting positively associated with security prices and market responses to information events. Merton (1987) develops an asset pricing model that indicates price levels will be associated with the fraction of investors who are aware of a security. On the other hand, investors are shown to have a limited ability to process all available information, constraining their activities to a subset of securities available to the (e.g., Kahneman 1973; Hirshleifer and Teoh 2003).

Stock prices appear to have a delayed response to earnings news (e.g., Bernard and Thomas 1989; Chan, Jegadeesh, and Lakonishok 1996). There are various explanations for this drift, including overconfidence (Daniel, Hirshleifer, and Subrahmanyam 1998), mean reversion (Barberis, Shleifer, and Vishny 1998), and underreaction due to processing limitations (Hong and Stein 1999). The only model that predicts investor inattention leading to greater drift is the underreaction explanation. There is evidence of underreaction to new information depending on both the salience of information (e.g., Chetty, Looney, and Kroft 2009) and investor inattention (e.g., DellaVigna and Pollet 2009), as well as the difficulty investors have processing information about related firms (Cohen and Frazzini 2008). The comment letter setting may experience underreaction due to processing costs and limited salience. Gietzmann et al. (2013a) find evidence of investor inattention to SEC comments on IFRS issues.

Hirshleifer, Lim, and Teoh (2009) examine the theory that limited investor attention causes market underreactions, by illustrating how reactions change when investors are presented with greater volumes of news. They find that price and volume reactions to earnings announcements is weaker, and post earnings announcement drift is stronger, when earnings announcements are more plentiful. In a related setting, DellaVigna et al. (2009) compare Friday earnings announcement responses to those occurring on other days. Consistent with expectations, Friday announcements have a lower immediate response and a greater delayed response. You and Zhang (2009) study the immediate and delayed market reaction to 10-K filings, and note that abnormal volume and return surrounding 10-K filings is associated with future profitability, indicating that useful information is recognized by the market, however, more complicated 10-K reports are associated with underreaction to the 10-K filing, as measured by document length. There is evidence of delayed responses to disclosures related to customer concentration (e.g., Patatoukas 2012) and footnote disclosures of resource valuations (e.g., Patatoukas, Sloan, and Zha 2015). Drake, Roulstone, and Thornock (2015) more directly measure attention to specific filings, looking at EDGAR search records to provide evidence that the market response to earnings surprise is increasing in EDGAR search volume.

Ljungqvist et al. (2014) examines a setting where arbitrageurs attempt to profit from their information collection activities surrounding short selling of potentially overvalued securities by raising attention through advertising their issuance of negative research reports. While limits to arbitrage can be high, making it difficult to profit from short sales, by advertising the potential overvaluation, such speculators can encourage current investors to sell, helping to correct the overvaluation by this alternative channel. Kovbasyuk and Pagano (2015) models a setting where multiple speculators identify the same mispriced security and publicize their information. As comment letters may be expected to reveal bad news on average, the activities of short sellers could be an important indication of investor attention to these disclosures (Dechow et al. 2016).

Taken together, there is evidence that investors do possess limited attention, resulting in delayed reactions to information which is more complex and less widely disseminated. Compared to other information releases, such as earnings announcements and periodic financial reports, comment letters appear prone to investor inattention. They are not released on an expected schedule, are more difficulty to identify in a timely manner, and are released as a set of separate documents, they are likely to be subject to investor inattention and as a result could be a setting where investors underreact to their information content.

Chapter 3

Textual Classification of SEC Comment Letters

This study seeks to identify important comment letters from their textual features, allowing readers of comment letters to potentially identify important comment letters before observing the market's response, which may be delayed if investors underreact to the letters' disclosure. Furthermore, the textual analysis that allows for the identification of important comment letters can also provide insights into the specific text features that are associated with importance, allowing for a greater understanding of the mechanisms by which the government monitoring process generating these letters reveals new information.

The particular monitoring process of interest is the mandatory periodic reviews of the annual financial statement, per Section 408 of SOX, and so the primary focus of this analysis is comment letters related to Form 10-K filings. Cassell et al. (2013) study determinants of receiving a comment letter and the costs of compliance, and Johnston et al. (2016) provide evidence that comment letters provide information in subsequent filings that improve the information environment. Bozanic, Dietrich, and Johnson (2014) find that firms make detectable changes to subsequent 10-Ks in response to comment letter issues, and Brown et al. 2015 find that firms make detectable changes to their risk-factor discussions when peers receive related comment letters. Dechow et al. (2016) provides evidence that there is information content in comment letters, observing abnormal insider trading around comment letter disclosure, but note a limited effect on stock returns. If comment letters are costly to process, then a delayed or limited market response is not surprising

(Hirshleifer et al. 2003).

I use realized abnormal returns following comment letter disclosure to classify documents in the training, identifying them as as *important* if post-disclosure abnormal returns are in the bottom quartile of the distribution. This approach eliminates the possibility of researcher coding bias, though it does increase the signal's noise, as abnormal returns will be driven by other information, especially as the return period increases. I focus on negative returns because comment letters result from a review that targets disclosure deficiencies and is intended to protect investors from fraud (SEC 2001; SOX 2002). If managers are more likely to withhold bad news (Kothari, Shu, and Wysocki 2009b), and if the SEC reviewers succeed in identifying disclosure deficiencies, then important comment letters will be more likely to result in a negative abnormal stock return when the information is revealed. The SEC has less incentive to protect against good news being withheld, and reviews finding compliant disclosure would either not generate a comment letter in the first place, or the identified issues would be minor, and therefore the disclosure of the resulting letter or changes in subsequent filings would not negatively affect returns. If text features in the comment letters associated with these negative returns are predictive of important comment letters for other firms, then firms with similar comment letter text will also experience negative stock returns following disclosure. The performance of the text to predict post-disclosure returns thus validates the model, and the first validation test is:

V1: Signaled comment letters are associated with negative post-disclosure returns.

Investor Inattention and EDGAR Views

A direct way to proxy for comment letter consumption is through the EDGAR log of document views.[1] A caveat to the use of this data is that EDGAR is not the only way for investors to access SEC filings, so I do not observe all occasions when a document is viewed. The EDGAR data itself is disseminated in two ways, through EDGAR's public web site and FTP file service, which encompass the traffic recorded by the log files used in this study.[2] EDGAR filings are also made available to data vendors via the Public Dissemination Service feed, which is a stream of all accepted filings (Drake et al. 2015). These feeds are the sources that populate commercial data services

[1] `http://www.sec.gov/data/edgar-log-file-data-set`

[2] `http://www.sec.gov`, `ftp://ftp.sec.gov`

such as Bloomberg, FactSet, and third party financial websites such as Morningstar Document Research. Therefore while the EDGAR logs represent a large volume of views, it is only a proxy for investor attention, as there is no way to capture all EDGAR filing views from all sources. It is also noteworthy that comment letters are not as widely available outside EDGAR as are other popular filings. Many corporate investor relations websites that claim to provide copies of all SEC filings often exclude comment letters (Dechow et al. 2016), and the most popular financial information sites, Yahoo Finance and Google Finance, do not provide access to comment letters through their firm-specific "SEC filings" pages. Dechow et al. (2016) use the EDGAR log files and find that comment letters are viewed at approximately one percent of the rate of views for the associated 10-K.

Given that investors appear to pay limited attention to comment letters, the information they contain may be incorporated in returns with a delay. Therefore, longer-term abnormal returns should provide an improved signal of comment letter importance, and the textual classification signal could be stronger when comment letters are known to have been viewed by investors. The second validation test is:

V2: The market response to signaled comment letters is greater when they are viewed.

3.1 Data

I collect firm fundamentals from Compustat, returns from CRSP, insider trades from Thompson Reuters Insider, and material restatements and internal control effectiveness reports from Audit Analytics. See Appendix A for definitions of all variables. I obtain copies of the daily EDGAR web logs from the SEC, for the period from June 2006 through January 2012. The log files are cleaned using a procedure similar to Drake et al. (2015).

I calculate cumulative abnormal returns from CRSP, for firms that trade on the NYSE, NASDAQ, or Amex exchanges, using a procedure similar to Campbell, Lo, and MacKinlay (1997). Specifically, cumulative abnormal returns are calculated using the market model: $CAR[a,b]_i = \prod_{t=a}^{b}(1 + AR_{it}) - 1$, where $CAR[a,b]_i$ is the cumulative abnormal return for firm i for day a through day b. AR_{it} is calculated as $AR_{it} = R_{it} - [\hat{\alpha}_i + \hat{\beta}_i R_{mt}]$, where AR_{it} is the abnormal return for firm i on day t, R_{mt} is the market return for day t using the S&P 500 index, and $\hat{\alpha}_i$ and $\hat{\beta}_i$ are estimated from the equation: $R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it}$, using a pre-event period from event day -300 to event day -46 trading days. I drop observations with less than 30 days of returns data in the estimation period, and observations without 90 days of post-event returns. Results are similar using size-adjusted returns.

I collect the full text of all SEC comment letters (Form UPLOAD) and company responses (Form CORRESP) directly from the ftp.sec.gov file transfer service, from June 2006 through January 2012, as this date range corresponds to the availability of EDGAR web logs. The daily EDGAR index files are utilized to determine each document's filing and disclosure dates. Filings may have different formats (PDF, HTML, and text), so I convert all to plain text. Comment letters and responses for the same CIK identifier, disclosed on the same day, are combined into a single conversation document.

Beginning with 55,688 separate conversations, I keep filings whose CIKs match to a firm in CRSP, the CRSP-Compustat Annual Fundamentals file, and Thomson Reuters Insider Trading database, 21,243 conversations. I keep conversations relating to Form 10-K filings, and those with sufficient returns data in CRSP to calculate abnormal returns for the 90 days post-comment letter disclosure, resulting in a final textual classification sample of 6,566 comment letter conversations for 3,527 unique firms. This sample is randomly divided into a training sample of 3,283 observations and a holdout sample of 3,283 observations.[3] I count the number of comment letters and responses in the conversation, count the number of questions in the comment letter, and identify if the comment letter relates to a revenue recognition topic, as prior research has shown that this is an important comment letter topic (e.g., Cassell et al. 2013; Dechow et al. 2016). Appendix B provides details on the preparation of the comment letter text for analysis.

Table 3.1A provides descriptive statistics for the textual classification sample. The mean market capitalization of firms in the sample is $6,021 million, which is somewhat larger than the mean Compustat population of $3,952 million over the same period, and is consistent with Cassell et al. (2013), who show that size is positively associated with comment letter receipt. The mean Book to Market ratio is 0.65, comparable to the Compustat population of 0.73 over the same period. Table 3.1B presents descriptive statistics for the sample of conversations known to be viewed more than median (2 times) over the three days post-disclosure, with 2,546 observations for 1,965 unique firms. The mean market value in this sample is $8,026 million, slightly larger

[3] A 50 percent holdout sample is used as it provides the lowest risk of inference errors (Schorfheide and Wolpin 2012).

Table 3.1: Textual Analysis Sample Descriptive Statistics

(A) All 10-K Comment Letters

	N	mean	sd	q10	q25	median	q75	q90
Market Capitalization	6,566	6,020	22,557	60	189	809	3,358	11,697
Book to Market	6,566	0.650	0.671	0.145	0.291	0.511	0.832	1.278
CAR[0,3]	6,566	-0.000	0.071	-0.059	-0.027	-0.003	0.021	0.055
CAR[0,90]	6,566	0.018	0.479	-0.372	-0.196	-0.036	0.127	0.390
Number of Questions	6,566	6.513	6.554	1.000	2.000	5.000	8.000	13.000
Conversation Items	6,566	4.912	2.514	3.000	3.000	4.000	6.000	8.000
Revenue Recognition	6,566	0.200	0.400	0.000	0.000	0.000	0.000	1.000
Insider Sales (% of shares out.)	6,566	0.052	0.421	0.000	0.000	0.000	0.000	0.062
EDGAR Views	6,566	2.164	2.605	0.000	1.000	2.000	3.000	4.000

(B) Above Median View 10-K Comment Letters

	N	mean	sd	q10	q25	median	q75	q90
Market Capitalization	2,546	8,026	25,680	68	226	1,020	4,617	16,787
Book to Market	2,546	0.664	0.687	0.144	0.296	0.523	0.858	1.340
CAR[0,3]	2,546	-0.002	0.076	-0.058	-0.028	-0.004	0.019	0.050
CAR[0,90]	2,546	-0.020	0.391	-0.360	-0.197	-0.051	0.096	0.295
Number of Questions	2,546	6.896	6.637	1.000	3.000	5.000	9.000	14.000
Conversation Items	2,546	4.944	2.421	3.000	3.000	5.000	6.000	8.000
Revenue Recognition	2,546	0.165	0.372	0.000	0.000	0.000	0.000	1.000
Insider Sales (% of shares out.)	2,546	0.052	0.456	0.000	0.000	0.000	0.000	0.050
EDGAR Views	2,546	3.992	3.323	3.000	3.000	3.000	4.000	5.000

This table presents descriptive statistics for all comment letter firms used in the textual classification sample in Panel a and the subset of firms with above median EDGAR views (> 2) in Panel b. Refer to Appendix A for variable definitions.

Table 3.2: Univariate Correlations between Selected Variables

		1	2	3	4	5	6	7
1	Naive Bayes signal	1.00	0.26	0.11	-0.01	0.02	0.04	-0.01
2	Number of questions	0.23	1.00	0.17	-0.03	0.03	0.12	0.03
3	Revenue recognition	0.11	0.22	1.00	0.04	-0.04	-0.08	-0.05
4	Insider sales rank	-0.01	-0.04	0.03	1.00	0.03	-0.14	-0.02
5	Market capitalization	-0.05	-0.08	-0.10	0.22	1.00	-0.08	0.10
6	Book to market	0.01	0.12	-0.11	-0.17	-0.28	1.00	-0.00
7	EDGAR requests	-0.01	0.08	-0.08	-0.03	0.09	0.03	1.00

Upper triangle is Pearson correlations, lower triangle is Spearman. *Market capitalization* ($ millions) is from Compustat at the most recent fiscal year end prior to the comment letter disclosure date ($CSHO \times PRCCF$) . *Book to market* is book value of equity at the most recent fiscal year end ($SEQ/$ market capitalization). *Number of questions* is the number of questions asked by the SEC in the first comment letter of a conversation. *Revenue recognition* is a dummy variable equal to 1 if the first comment letter of a conversation has at least one revenue recognition related question. *Insider sales rank* is 1 if insider sales as a percent of shares outstanding between disclosure date -15 days and +15 days is 0, and is set to 2 to 5 for firms with insider sales in the first to fourth quartile of non-zero insider sales. *Insider sales (percent of shares outstanding)* is the percentage of shares outstanding (Compustat CSHO at most recent fiscal year end) sold by insiders between disclosure date -15 days and +15 days.

than the full sample.

For all firms with comment letter conversations, $CAR[0,3]$ is negligible (0.000), while $CAR[0,90]$ is 0.018. The mean positive return for all firms can be attributed to some small-firm outliers. Excluding firms with market capitalization of less than $25 million reduces the mean $CAR[0,90]$ to 0.005 ($p > .35$), all other results are unaffected by excluding these firms. Firms where the comment letters are downloaded more than 2 times have a mean $CAR[0,3]$ of -0.002, while $CAR[0,90]$ is -0.020. This provides preliminary indications that comment letters that were read soon after disclosure appear to disclose bad news on average. Investors may become aware of comment letters that contain bad news, or bad news released through some other channel may cause investors to find and download concurrently released comment letters. Earnings announcements and filings of 10-Ks and 10-Qs are evenly distributed throughout the event window for both groups of firms, and as a result such such announcements should not bias the results.

The mean number of questions in the initial comment letter is 6.513 for all 10-K comment letters, and 6.896 for comment letters viewed more than 2 times. The number of items in a conversation (SEC comment letters and company responses) is nearly identical at 4.912 for all comment letters and 4.944 for comment letters viewed more than 2 times. The fraction of all 10-K comment letters mentioning revenue recognition issues is 0.200 for all conversations, and 0.165 for comment letters viewed more than 2 times. Insider sales as a percentage of shares outstanding sold by officers and directors in the window from disclosure date -15 days to +15 days is a mean of 0.052 percent for all 10-K comment letters, and 0.052 percent for comment letters viewed more than 2 times. In untabulated tests, size is the main factor associated with greater numbers of EDGAR views.

3.2 Naive Bayesian Classification

In general terms, the Naive Bayesian classification procedure estimates the class of a document based on the frequencies of words or short phrases, collectively referred to as *features*, present in the document. Classes may be arbitrarily defined, for example: authorship, subject matter, or in this setting, importance. To implement Naive Bayesian classification, a model is trained by calculating the relative frequencies of each feature appearing in the training documents for each class. When a new document is examined, the feature frequencies are calculated and the document is assigned the class with the most-similar feature distribution.

Formally: let d be a document in a set $D = \{d_1, ..., d_k\}$ consisting of k documents. Let $F = \{f_1, ..., f_m\}$ be the set of m possible *features* that can appear in D. Let $n_i(d)$ be the number of times feature f_i appears in document d. Then each document will have a vector representation $\mathbf{d} = (n_1(d), ..., n_m(d))$.

The naive Bayes classifier assigns a document to a class c^* from among n classes $(c_1, ..., c_n)$, where $c^* = \arg\max_c P(c|d)$. Consider Bayes' rule:

$$P(c|d) = P(c) \times \frac{P(d|c)}{P(d)} \quad ,$$

then under the assumption that the f_is are conditionally independent given the document's class, the probability that a document belongs to class c is:

$$P(c|d) = P(c) \times \frac{\prod_{i=1}^{m} P(f_i|c)^{n_i(d)}}{P(d)} \quad . \tag{3.1}$$

I prepare the text for analysis by converting all characters to lowercase, and removing all punctuation and numbers. The document set is converted into a term document matrix, using either single words as the feature set (unigram), or single words and consecutive 2-word combinations (unigram + bigram).[4] The term document matrix has one row for each document vector. Finally, I remove any features that appear in fewer than 5 percent of the documents, which makes the computations less costly, and generally consist of items such as web site addresses, companies' and individuals' names, and hence don't have a consistent information value for the classification. The total feature set is 2,549 words in the unigram feature set and 4,472 in the unigram + bigram feature set.

The probabilities in Equation 3.1 are calculated from the sample: $P(c)$ is the prior probability, or the relative frequency of class c in the sample, in this case, bottom quartile returns occur with frequency 0.25; $P(f_i|c)$ is the conditional probability, the relative frequency of f_i among all features in the sample; $P(d)$ is the probability of the predictor—a document—and is the same for every observation and so can be dropped without affecting the maximization. $P(c|d)$ then is the posterior probability, the probability the document belongs to a class, given its feature set. I randomly

[4] E.g., the text "internal controls" appearing in a document would be represented by two features ("internal", "controls") in a unigram representation of the document, one feature ("internal controls") in a bigram representation, and three features ("internal", "controls", "internal controls") in a unigram + bigram representation.

select 50 percent of the comment letter sample as a training sample, which is a set of documents of known class to to calculate the probabilities $P(c)$ and $P(f_i|c)$ in Equation 3.1.

It is possible that a feature never appears in any document in a given class. This would result in a posterior probability of zero, and so a method of compensating is "add one smoothing", where one is added to the count of each feature in calculating the frequency numerator, and m is added to the denominator. Secondly, the multiplication of many small probabilities can lead to floating point overflow errors, which is corrected by instead adding the logarithms of each probability. Limiting our analysis to two classes: c_I and c_U for *important* and *unimportant* respectively, the maximization problem simplifies to:

$$\log(P(c_I|d)) = \log(P(c_I)) + \sum_{i=1}^{m} \log(P(f_i|c_I)) \times n_i(d)$$
$$\log(P(c_U|d)) = \log(P(c_U)) + \sum_{i=1}^{m} \log(P(f_i|c_U)) \times n_i(d) \quad ,$$

where

$$P(c_j) = \frac{|c_j|}{|D|} \quad \text{and}$$
$$P(f_i|c_j) = \frac{\left(\sum_{d \in c_j} n_i(d)\right) + 1}{\left(\sum_{d \in D} n_i(d)\right) + k} \quad .$$

A document is assigned to class c_I if $\log(P(c_I|d)) > \log(P(c_U|d))$, but class c_U otherwise. Hereinafter, I refer to documents classified as *important* by the Naive Bayesian algorithm as having a *Signal* value of TRUE, or simply "signaled", but otherwise documents classified as *unimportant* have a *Signal* value of FALSE.

To validate the classification, signaled comment letters should be associated with bottom-quartile post disclosure returns. I first check the precision that signaled comment letters have bottom-quartile abnormal returns in the three and 90-days post disclosure, and I test the underreaction setting by conditioning the precision performance on comment letters with above-median EDGAR downloads.

I test the statistical significance of abnormal returns associated with the signal, both with and without controls for the number of questions, the presence of revenue recognition comments, and the level of insider sales. This leads to the following OLS regression model:

$$\begin{aligned} \text{CAR}_i = \beta_0 + \beta_1 I(\text{Signal})_i + \beta_2 \log(\text{Num. Questions})_i \\ + \beta_3 I(\text{Revenue Recognition})_i + \beta_4 \text{Insider Sales Rank}_i + \varepsilon_{i,t} \quad , \end{aligned} \tag{3.2}$$

where *CAR* is either the three-day (*CAR[0,3]*) or 90-day (*CAR[0,90]*) cumulative abnormal return. *Number of Questions*, *Revenue Recognition*, and *Insider Sales Rank* are included to observe if the signal has power to explain returns in addition to other possible indicators of important comment letters (e.g., Cassell et al. 2013; Dechow et al. 2016). Refer to Appendix A for variable definitions.

Classification Performance

Table 3.3 reports the effectiveness of the Naive Bayes classification model for identifying important comment letters, presenting the results given varied parameter choices. This table gives the precision of the signal to identify comment letters with subsequent bottom-quartile abnormal returns, as a first validation test of the classification model. Results are listed for the full sample (*All*) and for the sample known to have been viewed on EDGAR (*Views* > 2) in the three days post-disclosure. Stronger results for the *Views* > 2 sample provide evidence supporting potential inattention. Table 3.4 provides empirical results for the determinants of EDGAR views, illustrating that the primary determinant of greater comment letter views is firm size.

The *Signal* is *CAR[0,3] (CAR[0,90])* when training documents are classified as important if cumulative abnormal return are in the bottom quartile from day 0 after disclosure through day +3 (+90). *Frequency* identifies whether the Naive Bayes classifier uses the *frequency* count of each feature, or *presence*, which assigns a value of 1 if a feature appears at least once. *Documents* refers to the number of conversations in the combined training and holdout sample (50 percent of the documents are used for training, and 50 percent for testing the classifier effectiveness). *Precision* is the ability of the classification to correctly predict the importance of a comment letter, as realized by the relevant CAR signal. The baseline precision is approximately 25 percent for full sample, because I base the signal on bottom-quartile returns, but the exact frequency in the training sample varies somewhat as the observations are randomly selected but the bottom quartile threshold value is fixed. The increase i7n precision column (*Inc. Prec.*) presents the percent improvement in the rate at which the model signals bottom quartile firms over the rate at which bottom quartile

Table 3.3: Naive Bayes Classification Performance

	Model Paramaters				Model Performance	
	Signal	Sample	Frequency	Documents	Precision (%)	Inc. Prec. (%)
1	CAR[0,3]	All	frequency	6,566	24.93	0.55
2	CAR[0,3]	Views > 2	frequency	2,546	28.27	8.08
3	CAR[0,90]	All	frequency	6,566	28.82	10.66
4	CAR[0,90]	Views > 2	frequency	2,546	32.47	40.11
5	CAR[0,3]	All	presence	6,566	26.06	5.09
6	CAR[0,3]	Views > 2	presence	2,546	31.57	20.68
7	CAR[0,90]	All	presence	6,566	30.13	15.70
8	CAR[0,90]	Views > 2	presence	2,546	26.70	15.23

This table presents the effectiveness of the Naive Bayes classifier where the training documents are a random sample of 50 percent of the conversations, selected from the entire sample period. The feature set used is all unigrams + bigrams (all single words as well as all consecutive two word sequences) that appear in more than 5 percent or more of the sample documents. *Signal* refers to the measure used to identify important comment letters in the training sample (50 percent of documents): *CAR[0,3]* signals an important comment letter if the cumulative abnormal return is in the bottom quartile of returns from disclosure day 0 to disclosure day +3, and *CAR[0,90]* signals an important comment letter if the cumulative abnormal return is in the bottom quartile of returns from disclosure day 0 to disclosure day +90. Classification testing is run on *All* comment letter conversations, or on only those that are known to have been viewed on EDGAR more than the median number of times in the three days after disclosure (*Views* > 2). *Frequency* refers to whether the classifier uses the *frequency* or the count of the number of times each feature appears in the document, or *presence*, which equals 1 if the feature is present at least once in the document. *Documents* is the number of conversations in the combined training and testing sample (50 percent of the documents are used for training, and 50 percent for testing the classifier effectiveness). *Precision* refers to the fraction of comment letter conversations classified as important in the test sample that did in fact have bottom quartile CAR per the relevant signal. The increase in precision *Inc. Prec.* is the percentage increase in the fraction of comment letters identified as important the fraction occurring in the test sample, and represents the ability of the Naive Bayes classifier to identify important comment letters versus random chance.

Table 3.4: Determinants of EDGAR views

	(1) Log EDGAR Views (OLS)	(2) I(EDGAR Views > 2) (Logit)
Intecept	-0.096	-14.914^{***}
	(-0.393)	(-9.051)
Log(Market Capitalization)	0.039^{***}	0.148^{***}
	(5.802)	(4.704)
I(Dividend)	0.062^{**}	0.262^{**}
	(2.448)	(2.030)
I(Acquisition)	-0.107^{***}	-0.647^{***}
	(-3.313)	(-3.937)
ROA	-0.087	-0.693^{*}
	(-1.043)	(-1.805)
Sales Growth	-0.030	0.042
	(-0.697)	(0.212)
Accruals	0.189	1.020
	(1.305)	(1.519)
Special Items	0.070	0.338
	(0.300)	(0.297)
Num. Business Segments	0.003	0.021
	(0.449)	(0.657)
Num. Geographic Segments	-0.003	0.001
	(-0.658)	(0.038)
I(Secondary Offering)	-0.056	-0.360
	(-1.078)	(-1.441)
Age	0.000	0.008
	(0.031)	(1.205)
Book-to-Market	-0.001	-0.020
	(-0.061)	(-0.202)
Adj. R^2	0.329	
Num. obs.	2544	2544

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

This table presents firm-related determinants of comment letter views on EDGAR.

abnormal returns appear in the holdout sample, e.g. if important documents were identified at a rate of 27.5 percent when the baseline is 25 percent, the increase in precision is 10 percent ($(27.5 - 25)/25 * 100$ percent).

The results reported in Table 3.3 support the model validation tests, V1 and V2. The models provide predictive power to signal comment letters in the holdout sample that have bottom-quartile abnormal returns following disclosure. Considering the 90-day CARs as the signal, the ability to identify important comment letters is between 10.66 percent and 40.11 percent greater than random chance, supporting V1. The improvement in power is significantly stronger using the 90-day CAR signal, as opposed to the three-day CAR signal, indicating underreaction to comment letter disclosures. The three-day CAR signal appears to provide little ability to identify important comment letters (0.55 to 5.09 percent increase in precision) in the all comment letter sample, though the precision improves to 8.08 to 20.68 percent when the comment letters have above median views. For the 90-day CAR signal, precision improves from 10.66 to 15.70 percent in the all comment letter sample to 15.23 to 40.11 percent in the above median view sample. This ability to more precisely identify important comment letters when they have been viewed supports V2 and provides evidence that inattention affects short term returns.

A benefit of the Naive Bayesian classification procedure is that the model reveals the features that appear with greatest frequency in each class—allowing researchers to gain insight into specific features driving the classification. Table 3.5 provides a list of the features with the greatest frequency differential between important and unimportant comment letters. For example, the feature with the greatest ratio of frequency in important letters to frequency in unimportant letters is "continue monitor", which has a frequency of 0.08 in important comment letters but a frequency of only 0.02 in unimportant comment letters. As an example of how this term may be used in an important comment letter, consider the following excerpt from a company correspondence in the sample:

> "...We have explored different borrowing alternatives with Key Bank, the lender under that facility, and other parties, but to date determined that the terms of these alternatives were not acceptable. We *continue* to *monitor* whether credit facilities may be available to us on acceptable terms. We may also have to pursue various other strategies to secure any necessary additional financing, which may include, without limitation, public or private offerings of debt or equity securities..."

Table 3.5: Terms with Greatest Frequency Differential Between Signaled Important and Unimportant Comment Letters

	Feature	Freq. Important	Freq. Unimportant	Ratio
1	continue monitor	0.08	0.02	4.78
2	quantitatively	0.10	0.02	4.27
3	straightline	0.13	0.03	3.94
4	severity	0.24	0.07	3.29
5	income continuing	0.19	0.06	2.94
6	rental	0.49	0.17	2.93
7	loan portfolio	0.37	0.13	2.84
8	accounting guidance	0.21	0.07	2.81
9	recoveries	0.19	0.07	2.78
10	brand	0.32	0.11	2.78
11	allowance loan	0.74	0.27	2.75
12	pension	0.67	0.25	2.72
13	commodity	0.31	0.11	2.70
14	real estate	1.77	0.66	2.68
15	estate	1.92	0.72	2.67
16	revised disclosures	0.13	0.05	2.66
17	leased	0.22	0.08	2.65
18	publicly traded	0.11	0.04	2.62
19	historical experience	0.15	0.06	2.61
20	senior management	0.23	0.09	2.61
21	payout	0.75	0.29	2.57
22	revising	0.13	0.05	2.55
23	credit quality	0.25	0.10	2.54
24	note consolidated	0.13	0.05	2.53
25	real	1.97	0.78	2.52
26	effective tax	0.39	0.15	2.51
27	safety	0.56	0.23	2.47
28	prior period	0.18	0.07	2.47
29	revenues expenses	0.06	0.03	2.41
30	monitor	0.37	0.16	2.37

This table presents the training sample features with the greatest difference in frequencies among documents signaled as important and unimportant based on having bottom-quartile 90-day post-disclosure abnormal returns. For example, feature (1), *continue monitor*, appears with a frequency of 0.08 per conversation in important documents, but with a frequency of only 0.02 in unimportant documents, thus it appears 4.78 times more frequently in important than in unimportant documents.

This conversation provides evidence that management has liquidity concerns, and reveals consideration of a secondary equity offering. The three- and 90-day CAR for this firm after this comment letter conversation was disclosed was -3.1 percent and -35.3 percent respectively.

Inspecting important comment letters with features identified in the Table 3.5 list such as *senior management* and *payout* may indicate that broad issues such as governance plays a role in some important comment letters. Features such as *loan portfolio*, *recoveries*, *severity*, *allowance loan*, and *credit quality* indicate that financing and distress related issues may be important. These are also terms associated with management estimates, and thus examination of these issues could reflect both on the potential for restatements, as estimates are revisited, and on internal controls, which ensure reliable financial reporting and compliance with disclosure regulations.

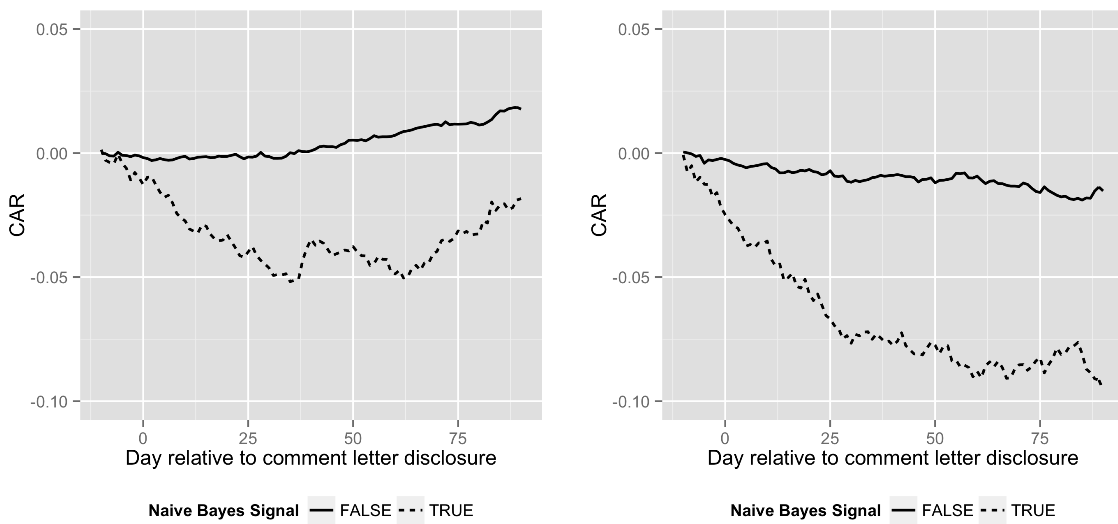
The following empirical tests are limited to the holdout sample, and the estimated signal for important comment letters is Model 3 in Table 3.3, the 90-day CAR classification model with the lowest increase in precision (+10.66 percent). The following results should therefore be downward-biased if other model parameter selections result in a greater discriminatory power to identify important comment letters.

3.3 Signaled Comment Letters and Abnormal Returns

Figure 3.1A illustrates the mean CAR from comment letter disclosure date -10 days to +90 days, for holdout sample comment letter conversations, partitioned by the signal. 90 days after disclosure, firms whose comment letters are not signaled have a mean CAR of +1.77 percent, and firms with signaled comment letters have a mean CAR of -1.84 percent, providing support for the classification model. Figure 3.1B illustrates mean CAR over the same period for firms with above median views from the EDGAR web site. Firms with above median EDGAR views whose comment letters are not signaled have a mean CAR of -1.52 percent at disclosure date +90 days, and firms with signaled comment letters have a mean CAR of -9.54 percent, providing evidence that the classification is more powerful for comment letters known to have been read by investors. In addition, the lower returns for signaled comment in this setting indicates that it is not solely the investor views of the comment letters

Figure 3.1: Comment Letter Disclosure Cumulative Abnormal Returns

(A) Holdout Sample (B) Holdout Sample Views > 2



This figure illustrates cumulative abnormal returns from ten days prior to 90 days after disclosure of holdout sample comment letters, partitioned on the Naive Bayes signal of importance. Panel A illustrates the results for all firms, and Panel B illustrates the results for firms whose comment letters were observed to be viewed on the EDGAR web site more that twice in the three days following disclosure. Refer to Appendix A for variable definitions.

Table 3.6: Signaled Comment Letters and Abnormal Returns

	CAR[0,3]				CAR[0,90]			
	All	All	Views > 2	Views > 2	All	All	Views > 2	Views > 2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
I(Signal)	0.008	0.007	−0.013**	−0.012*	0.012	0.018	−0.058*	−0.059**
	(1.228)	(1.021)	(−2.103)	(−1.918)	(0.343)	(0.493)	(−1.936)	(−1.963)
Num. questions		0.0003		0.00003		−0.002*		0.001
		(1.514)		(0.115)		(−1.775)		(0.309)
Revenue recognition		−0.005		−0.009*		0.040		−0.020
		(−1.510)		(−1.846)		(1.461)		(−0.780)
Insider sales rank		−0.001		0.001		−0.022***		−0.016**
		(−0.908)		(0.755)		(−4.137)		(−2.243)
Constant	0.001	0.001	−0.002	−0.003	0.011	0.053***	−0.007	0.020
	(0.390)	(0.318)	(−0.791)	(−0.527)	(1.297)	(3.071)	(−0.547)	(0.721)
Observations	3,283	3,283	1,273	1,273	3,283	3,283	1,273	1,273
Adjusted R^2	0.0004	0.001	0.001	0.001	−0.0003	0.003	0.001	0.001

*p<0.1; **p<0.05; ***p<0.01

This table presents regression results for Equation 3.2, using all observations in the holdout sample in Columns (1), (2), (5), and (6), and the subset of observations with above median EDGAR views (> 2) in Columns (3), (4), (7), and (8). Standard errors are robust. Columns (1) to (4) utilize three-day CAR as the dependent variable, and Columns (5) to (8) utilize 90-day CAR as the dependent variable. Refer to A for variable definitions.

that cause the price decline, but that the signal is effective at identifying firms with lower returns.

Table 3.6 examines the statistical significance of abnormal return differences associated with the signal, utilizing Equation 3.2. I regress the signal on short term (three-day) and long term (90-day) CAR, for holdout sample firms. Columns (1) to (4) consider the ability of signal to predict three-day abnormal returns. There is no statistical significance for the signal to predict returns in Columns (1) and (2), where all comment letters are used. In Column (3) I test the set of observations where the the comment letters were viewed, and the coefficient on signal is -0.013 percent ($p < 0.05$), when no additional comment letter characteristics are included as controls. Column (4) reports a similar coefficient of -0.012 ($p < 0.1$) when controls for other features related to comment letter importance are included (e.g., Cassell et al. 2013; Dechow et al. 2016). See Appendix A for variable definitions. The results of Columns (1) to (4) imply a -1.2 to -1.3 percent abnormal return in the three-days post-comment letter disclosure for signaled comment letters, but only when the comment letters are viewed.

Columns (5) to (8) regress the signal on 90-day abnormal returns. When the comment letters were not viewed, in Columns (5) and (6), the coefficient on the signal is insignificant. When the comment letters were viewed, in Columns (7) and (8), the coefficients are negative and significant at -0.058 ($p < 0.1$) when no controls are included and -0.059 ($p < 0.05$) when controls are added. The results of Columns (5) to (8) imply a -5.8 to -5.9 percent abnormal return in the 90-days post-comment letter disclosure for signaled comment letters, but only when the comment letters were viewed. Together these results indicate that when investors are known to have viewed the comment letters, the signal predicts negative returns over both the three- and 90-day period following disclosure, jointly supporting the model validation and potential inattention.

3.4 Robustness Analyses

To provide evidence that the naive Bayes classification technique provides power to identify important comment letters in time-series out of sample settings, I test the robustness of the technique using documents from the first half of the sample, by comment letter disclosure date, as the training sample, and the remaining out of sample comment letters as the holdout sample. Table 3.7 illustrates that the increase in precision for identifying comment letters versus random chance is generally comparable to the results from the random holdout sample reported in Table 3.3. Although two of the models provide no additional identification precision, the remaining six models provide an increase in precision for identifying important comment letters of between +8.45 percent and +61.15 percent.

I also investigate whether insider sales surrounding comment letter disclosure can be used to signal importance, as an alternative to market returns, for the Naive Bayes model (e.g., Dechow et al. 2016). In untabultated results, I find that the classification model is ineffective using this specification, insofar as signaled comment letters have no greater levels of insider trading than other comment letters. While market returns may be expected to give an unbiased response to new information, executive behavior may not be unbiased. Some executives may decide to sell stock surrounding the release of a comment letter that they deem important, though other executives may consider this a violation of insider trading norms. If important comment letters generate insider trades for some observations but not for others, than the Naive Bayes classification algorithm would have difficulty distinguishing the text features of the important comment letters.

Table 3.7: Naive Bayes Classification Performance for Time Based Training Sample

	Model Paramaters				Model Performance	
	Signal	Sample	Frequency	Documents	Precision (%)	Inc. Prec. (%)
1	CAR[0,3]	All	frequency	6,566	24.42	10.02
2	CAR[0,3]	Views > 2	frequency	2,546	18.37	-5.00
3	CAR[0,90]	All	frequency	6,566	20.12	-2.73
4	CAR[0,90]	Views > 2	frequency	2,546	24.00	12.18
5	CAR[0,3]	All	presence	6,566	31.75	43.05
6	CAR[0,3]	Views > 2	presence	2,546	20.97	8.45
7	CAR[0,90]	All	presence	6,566	33.33	61.15
8	CAR[0,90]	Views > 2	presence	2,546	24.24	13.31

This table presents the effectiveness of the Naive Bayes classifier, where the training documents are the first 50 percent selected by date disclosed. The feature set used is all unigrams + bigrams (all single words as well as all consecutive two word sequences) that appear in more than 5 percent or more of the sample documents. *Signal* refers to the measure used to identify important comment letters in the training sample (50 percent of documents): *CAR[0,3]* signals an important comment letter if the cumulative abnormal return is in the bottom quartile of returns from disclosure day 0 to disclosure day +3, and *CAR[0,90]* signals an important comment letter if the cumulative abnormal return is in the bottom quartile of returns from disclosure day 0 to disclosure day +90. Classification testing is run on *All* comment letter conversations, or on only those that are known to have been viewed on EDGAR more than the median number of times in the three days after disclosure (*Views* > 2). *Frequency* refers to whether the classifier uses the *frequency* or the count of the number of times each feature appears in the document, or *presence*, which equals 1 if the feature is present at least once in the document. *Documents* is the number of conversations in the combined training and testing sample (50 percent of the documents are used for training, and 50 percent for testing the classifier effectiveness). *Precision* refers to the fraction of comment letter conversations classified as important in the test sample that did in fact have bottom quartile CAR per the relevant signal. The increase in precision *Inc. Prec.* is the percentage increase in the fraction of comment letters identified as important the fraction occurring in the test sample, and represents the ability of the Naive Bayes classifier to identify important comment letters versus random chance.

Chapter 4

Effects of Signaled Comment Letters

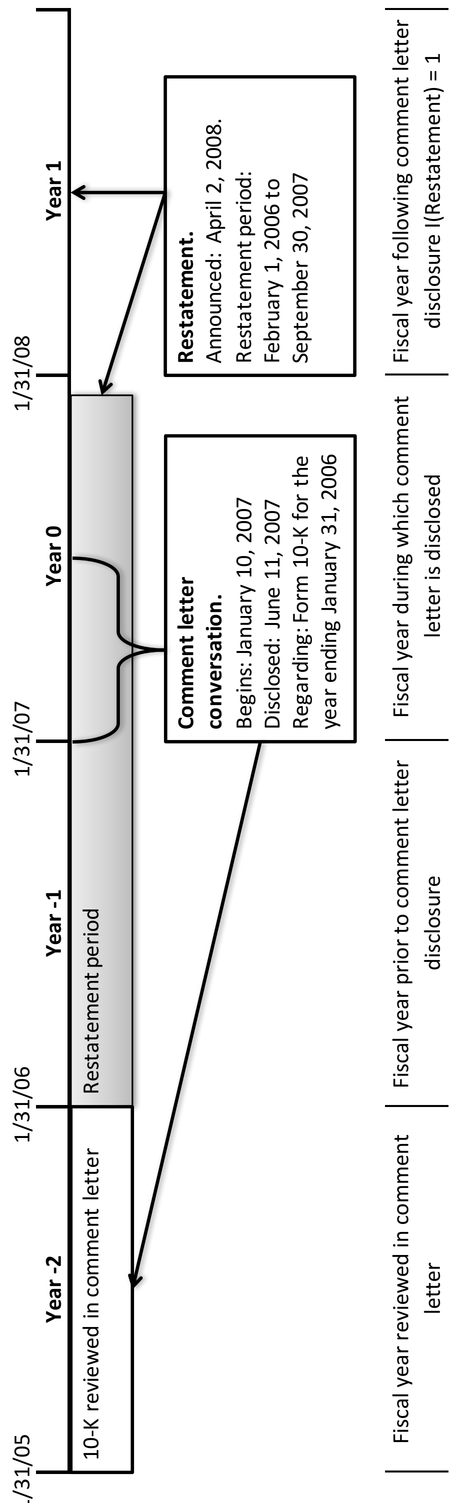
4.1 Hypothesis Development

Comment Letters and Financial Reporting Quality

SEC reviews conducted in accordance with SOX Section 408 and the SEC's Full Disclosure Program aim to protect investors from fraud and misrepresentation, and to ensure that disclosures comply with relevant laws and regulations (SEC 2001; SOX 2002). If some managers strategically avoid disclosing bad news, and such undisclosed information is not reflected in market prices (e.g., Grossman and Stiglitz 1980; Bloomfield 2002), then efforts by the SEC to improve disclosures through the review process should reveal information when the related correspondence is disclosed, in amendments or periodic disclosures while the review process is underway, or in subsequent periods.

I study the three years surrounding the disclosure of 10-K comment letters, considering performance and events in the year before, during, and after, the disclosure of the comment letter. Figure 4.1 provides an illustrative example of the timing of these events for Dillards, Inc. Dillards has a comment letter disclosed during Year 0, and the comment letter is discussing a Form 10-K disclosed during Year -1. Dillards has a restatement announced during Year 1, though this restatement covers time periods covering both Year -1 and Year 0. The restatement is not directly prompted by the comment letter, though it may have resulted from additional

Figure 4.1: Illustrative Timeline for Dillards, Inc.



management attention or auditor scrutiny as a result of the comment letter process.

If the comment letter process either reveals that a firm had no significant disclosure deficiencies, or if the comments resulted in disclosure improvements with no bad news being revealed, then earnings should not be affected by the review process, and the stock market response could be positive, consistent with prior literature regarding disclosure quality and performance (e.g., Lang and Lundholm 1993; Francis, LaFond, Olsson, and Schipper 2005; Francis, Nanda, and Olsson 2008). On the other hand, more important comment letters could result in the release of negative information that management was withholding (e.g. Kothari et al. 2009b), and earnings could decline as a result of a comment letter if management estimates are revised more skeptically evaluated by auditors in subsequent periods. The first hypothesis follows (in alternative form):

H1: Signaled comment letters are associated with lower earnings and earnings persistence.

Important comment letters may impact financial reporting and audit processes. Auditors are often included in the comment letter correspondence (Laurion et al. 2015), and the auditor may modify their assessment of audit risk, identify areas of financial reporting weakness, and recognize internal control weaknesses as a result of issues raised by the SEC. Management investigations made to provide responses to SEC questions could lead to changes in accounting assumptions and policies, uncovering errors resulting in material restatements. Material restatements reflect financial reporting quality and have an effect on returns (e.g., Hribar et al. 2004; Kinney et al. 2004; Palmrose et al. 2004; Gleason et al. 2008; Liu et al. 2009; Dechow et al. 2011; Francis 2011; Chen et al. 2013). The second hypothesis follows (in alternative form):

H2: Signaled comment letters are positively associated with material restatements.

While comment letters may identify actual errors or material misstatements requiring a restatement, this same process may reveal failures of internal controls over financial reporting. If the SEC correctly identifies material disclosure requirements with which the issuer has not complied, then this is evidence that the issuer does not have adequate financial reporting capabilities and controls. Internal control weaknesses are associated with information uncertainty and negative announcement returns (e.g., Doyle et al. 2007; Beneish et al. 2008; Hammersley et al. 2008; Ashbaugh-Skaife et al. 2009). The third hypothesis follows (in alternative form):

H3: Signaled comment letters are positively associated with internal control weaknesses.

4.2 Descriptive Statistics

Table 3.1A provides descriptive statistics for the textual classification sample. The mean market capitalization of firms in the sample is \$6,021 million, which is somewhat larger than the mean Compustat population of \$3,952 million over the same period, and is consistent with Cassell et al. (2013), who show that size is positively associated with comment letter receipt. The mean Book to Market ratio is 0.65, comparable to the Compustat population of 0.73 over the same period. Table 3.1B presents descriptive statistics for the sample of conversations known to be viewed more than median (2 times) over the three days post-disclosure, with 2,546 observations for 1,965 unique firms. The mean market value in this sample is \$8,026 million, slightly larger than the full sample.

For all firms with comment letter conversations, $CAR[0,3]$ is negligible (0.000), while $CAR[0,90]$ is 0.018. The mean positive return for all firms can be attributed to some small-firm outliers. Excluding firms with market capitalization of less than \$25 million reduces the mean $CAR[0,90]$ to 0.005 ($p > .35$), all other results are unaffected by excluding these firms. Firms where the comment letters are downloaded more than 2 times have a mean $CAR[0,3]$ of -0.002, while $CAR[0,90]$ is -0.020. This provides preliminary indications that comment letters that were read soon after disclosure appear to disclose bad news on average. Investors may become aware of comment letters that contain bad news, or bad news released through some other channel may cause investors to find and download concurrently released comment letters. Earnings announcements and filings of 10-Ks and 10-Qs are evenly distributed throughout the event window for both groups of firms, and as a result such such announcements should not bias the results.

The mean number of questions in the initial comment letter is 6.513 for all 10-K comment letters, and 6.896 for comment letters viewed more than 2 times. The number of items in a conversation (SEC comment letters and company responses) is nearly identical at 4.912 for all comment letters and 4.944 for comment letters viewed more than 2 times. The fraction of all 10-K comment letters mentioning revenue

Table 4.1: Earnings, Restatement, and Internal Control Sample Descriptive Statistics

	N	mean	sd	q10	q25	median	q75	q90
Naive Bayes Signal	2,544	0.094	0.291	0.000	0.000	0.000	0.000	0.000
CAR[0,3]	2,544	0.001	0.082	-0.055	-0.027	-0.003	0.020	0.056
CAR[0,90]	2,544	0.016	0.494	-0.362	-0.198	-0.038	0.123	0.373
Earnings	2,544	0.001	0.182	-0.123	0.001	0.032	0.077	0.121
I(IC Weakness)	2,544	0.036	0.187	0.000	0.000	0.000	0.000	0.000
I(Restatement)	2,544	0.073	0.260	0.000	0.000	0.000	0.000	0.000
EDGAR Views	2,544	2.256	3.503	0.000	1.000	2.000	3.000	4.000
Market Capitalization	2,544	7,907	28,661	61	209	981	3,970	14,722
ΔReceivables	2,544	-0.002	0.043	-0.040	-0.015	-0.000	0.012	0.034
ΔInventory	2,544	0.000	0.027	-0.019	-0.003	0.000	0.004	0.021
Soft Assets	2,544	0.591	0.261	0.195	0.393	0.622	0.812	0.930
Leverage	2,544	3.067	5.336	0.215	0.580	1.559	3.560	8.820
Book to Market	2,544	0.666	0.637	0.163	0.305	0.511	0.827	1.302
I(Dividend)	2,544	0.463	0.499	0.000	0.000	0.000	1.000	1.000
I(Acquisition)	2,544	0.123	0.328	0.000	0.000	0.000	0.000	1.000
ΔEarnings	2,544	-0.000	0.149	-0.085	-0.022	0.000	0.022	0.085
Sales Growth	2,544	0.085	0.260	-0.168	-0.035	0.067	0.179	0.344
Accruals	2,544	-0.019	0.083	-0.100	-0.048	-0.008	0.018	0.055
Special Items	2,544	-0.014	0.052	-0.032	-0.009	-0.000	0.000	0.002
Business Segments	2,544	2.281	1.736	1.000	1.000	1.000	3.000	5.000
Geographic Segments	2,544	2.697	2.588	1.000	1.000	2.000	4.000	6.000
I(Secondary Offering)	2,544	0.058	0.233	0.000	0.000	0.000	0.000	0.000
Age	2,544	18.131	8.998	6.000	11.000	17.000	27.000	31.000
I(Big4)	2,544	0.789	0.408	0.000	1.000	1.000	1.000	1.000

This table presents descriptive statistics for all comment letter firms the holdout sample with sufficient data for tests of earnings persistence, the incidence of restatements, and the incidence of internal control weaknesses. *Signal* indicates that the comment letter was identified as important by the Naive Bayesian classification. Refer to A for variable definitions.

recognition issues is 0.200 for all conversations, and 0.165 for comment letters viewed more than 2 times. Insider sales as a percentage of shares outstanding sold by officers and directors in the window from disclosure date -15 days to +15 days is a mean of 0.052 percent for all 10-K comment letters, and 0.052 percent for comment letters viewed more than 2 times. In untabulated tests, size is the main factor associated with greater numbers of EDGAR views.

To study financial performance and reporting quality in the years adjacent to comment letter issuance, I use comment letters in the textual classification holdout sample that have the required Compustat control variables for two years before and one year after comment letter disclosure, resulting in a sample of 2,544 conversations for 1,801 unique firms. Table 4.1 provides descriptive statistics for these firms, which have a mean market capitalization of \$7,908 million, slightly larger than the all comment letter sample of \$6,021 million and slightly smaller than the above-median EDGAR view sample of \$8,026 million.

4.3 Earnings and Earnings Persistence

To study the effect of signaled comment letters on financial performance, I test H1 by examining the relationship between signaled comment letters, earnings, and earnings persistence. To study the relation between earnings and signaled comment letters, I examine the following logit regression model:

$$
\begin{aligned}
I(\text{Signal})_{i,o} = {} & \beta_0 + \beta_1\text{Earnings}_{i,t} + \beta_2\text{Accruals}_{i,t-1} \\
& + \beta_3 I(\text{Dividend})_{i,t-1} + \beta_4\text{Special Items}_{i,t-1} \\
& + \beta_5\text{Num. Bus. Segments}_{i,t-1} + \beta_6\text{Num. Geo. Segments}_{i,t-1} \\
& + \beta_7 I(\text{Secondary Offering})_{i,t-1} + \beta_8 I(\text{Acquisition})_{i,t-1} \\
& + \beta_9\text{Age}_{i,t} + \beta_{10}\text{Book to Market}_{i,t-1} \\
& + \beta_{11}\log(\text{Market Capitalization})_{i,t-1} + \varepsilon_{i,t} \quad . \qquad (4.1)
\end{aligned}
$$

To study the relation between signaled comment letters and earnings persistence, I examine the following OLS regression model:

$$\begin{aligned}\text{Earnings}_{i,t} = {} & \beta_0 + \beta_1 I(\text{Signal})_{i,0} + \beta_2 \text{Earnings}_{i,t-1} \\ & + \beta_3 I(\text{Signal})_{i,0} * \text{Earnings}_{i,t-1} \\ & + \beta_4 \text{Accruals}_{i,t-1} + \beta_5 I(\text{Dividend})_{i,t-1} + \beta_6 \text{Special Items}_{i,t-1} \\ & + \beta_7 \text{Num. Bus. Segments}_{i,t-1} + \beta_8 \text{Num. Geo. Segments}_{i,t-1} \\ & + \beta_9 I(\text{Secondary Offering})_{i,t-1} + \beta_{10} I(\text{Acquisition})_{i,t-1} \\ & + \beta_{11} \text{Age}_{i,t} + \beta_{12} \text{Book to Market}_{i,t-1} \\ & + \beta_{13} \log(\text{Market Capitalization})_{i,t-1} + \varepsilon_{i,t} \quad . \end{aligned} \tag{4.2}$$

I include fixed effects for year and Fama-French 49 industry membership. The fiscal year in which the comment letter is disclosed is defined as $t = 0$. These models are estimated for $t = -1$, the year before the comment letter is disclosed, $t = 0$, the year of disclosure, and $t = 1$, the year following disclosure. Firm-comment letter observations, i, are from the Naive Bayesian holdout sample with available control variables. $Signal_{i,0}$ is equal to 1 if the Naive Bayes classification model indicated importance, but 0 otherwise, and can only be evaluated at $t = 0$. The measure of earnings is return on assets (Compustat $ibadj_{i,t}/at_{i,t}$). See Appendix A for all other variable definitions. Control variables have been shown in prior literature to affect earnings persistence (e.g., Li 2008), and are defined in Appendix A. The coefficient of interest is β_3, the interaction term between *Signal* and the prior years' earnings. If *Signal* is associated with lower earnings persistence, then β_3 will be negative.

Figure 4.2A illustrates the level of earnings for firms in the holdout sample, partitioned by the signal of comment letter importance. Firms receiving important comment letters have significantly lower—on average, negative—earnings in the year prior to the year the comment letter was disclosed ($t = -1$), compared to firms without signaled letters. Year $t - 1$ is the fiscal year that the SEC reviews for the comment letter disclosed in year $t = 0$, indicating that firms with lower profits are more likely to generate signaled comment letters. Earnings tend to increase but remain negative in year $t = 0$ and $t = 1$. Table 4.2 reports on the difference in means for the key analysis and control variables in year $t = 0$, conditioned on the signal. Firms with lower *Earnings*, higher incidences of *Restatement*, and higher incidences of internal control (*Weakness*) are more likely to have signaled comment letters. Signaled firms also tend to have larger *Market Capitalization* ($p < 0.1$), a greater proportion of *Soft Assets* ($p < 0.05$), greater *Leverage* ($p < 0.05$), a greater *Book to Market* ratio ($p < 0.05$), a higher rate of secondary equity offerings (*Secondary*

Table 4.2: Characteristics of Signaled Comment Letter Firms

	N.B. Signal=1	N.B. Signal=0	Difference	
CAR[0,3]	0.010	0.000	0.009	
CAR[0,90]	0.011	0.016	-0.005	
Earnings	-0.035	0.005	-0.040	***
I(IC Weakness)	0.076	0.032	0.044	**
I(Restatement)	0.130	0.067	0.063	***
EDGAR Views	2.008	2.281	-0.273	**
Market Capitalization	11,824	7,503	4,320	*
ΔReceivables	-0.001	-0.002	0.001	
ΔInventory	0.001	0.000	0.000	
Soft Assets	0.627	0.587	0.040	**
Leverage	3.928	2.978	0.951	**
Book to Market	0.779	0.654	0.125	**
I(Dividend)	0.395	0.471	-0.076	**
I(Acquisition)	0.134	0.121	0.013	
ΔEarnings	0.020	-0.002	0.023	
Sales Growth	0.110	0.082	0.027	
Accruals	-0.018	-0.019	0.001	
Special Items	-0.021	-0.013	-0.008	*
Business Segments	2.168	2.292	-0.124	
Geographic Segments	2.479	2.719	-0.240	
I(Secondary Offering)	0.105	0.053	0.052	**
Age	17.546	18.191	-0.645	
I(Big4)	0.765	0.791	-0.027	

This table compares differences in means of key variables for holdout sample firms with comment letters, conditioned on the Naive Bayesian classification signaling an important comment letter. Variables are measured at the end of the fiscal year in which the comment letter is disclosed ($t = 0$). N = 238 observations where the Naive Bayesian Signal is 1 and N = 2,306 observations where it is 0. Refer to A for variable definitions.

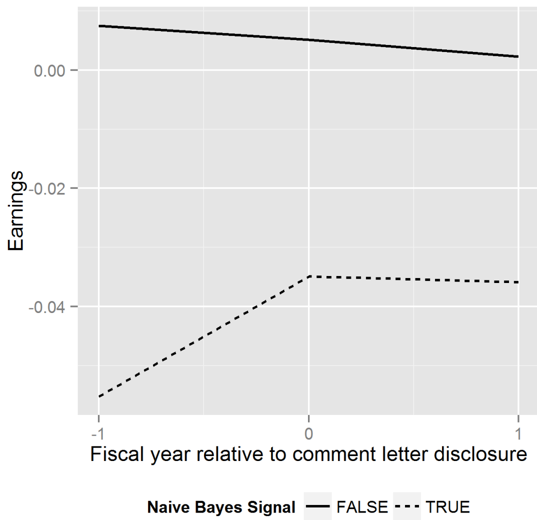
Offerings; $p < 0.05$), but lower *EDGAR Views* ($p < 0.05$), lower incidence of *Dividend* payments ($p < 0.05$), and *Special Items* ($p < 0.1$). Other characteristics are similar.

Table 4.3 models Equation 4.1 to study the relation between firms' *Earnings* and *Signal*. Columns (1) to (3) examine profitability in the year before, during and after the comment letter conversation, respectively. *Earnings* only predict *Signal* if they are low in the year prior to comment letter issuance (Column (1) coefficient on *Earnings* of -1.228 ($p < 0.01$)). Signaled comment letters do not appear to be associated with significantly different earnings in the year the comment letter is issued ($t = 0$) or the following year ($t = 1$). The marginal effect of a 1 percent decline in return on assets is a 3 percent increase in having a comment letter identified as important. While neither the SEC's stated policies nor Section 408 of SOX target firms with low earnings or losses, this result builds on Cassell et al. (2013), who note that loss firms are more likely to receive a comment letter, as this result indicates that firms with lower earnings are more likely to receive important comment letters. It does not appear that signaled comment letters help to predict lower future earnings, controlling for other determinants of profitability, the level of earnings may not be a mechanism for signaled comment letters to affect returns.

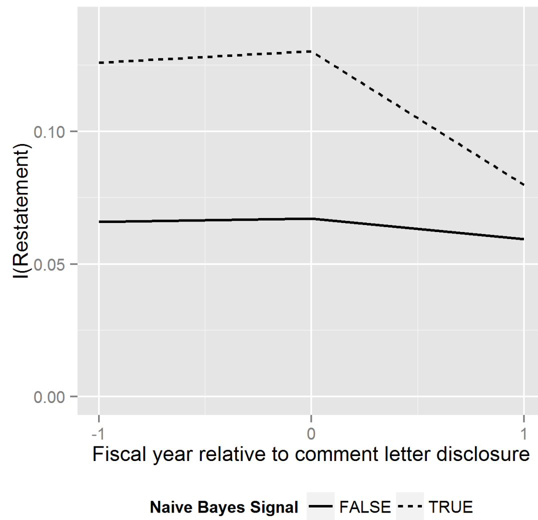
I study the relation between signaled comment letters and earnings persistence in Table 4.4, implementing Equation 4.2, including year and industry fixed effects. The coefficient on the interaction term, I(*Signal*) * $Earnings_{t-1}$, captures the change in persistence for firms receiving important comment letters. Columns (1) to (3) examine earnings persistence in the year before, during and after the comment letter conversation, respectively, for profit firm-years. The coefficient on I(*Signal*) * $Earnings_{t-1}$ in Column (1) of -0.493 ($p < 0.01$) indicates that for profit firms with signaled comment letters, earnings persistence declines in the year prior to the comment letter review. The interaction coefficient is also negative in Column (3) at -0.334 ($p < 0.01$), indicating that profit firms with signaled comment letters have lower earnings persistence in the year following the review. This finding could have a valuation impact, as information disclosed in signaled comment letters may reveal uncertainty about future earnings for profit firms. Columns (4) to (6) analyze loss firms. The interaction term in Column (4) of 0.655 ($p < 0.01$) relates to the year prior to the comment letter ($t = -1$), as firms with higher loss persistence were more likely to receive a signaled comment letter. In the year of the comment letter conversation, losses were less persistent, with the coefficient on the interaction term being -0.173 ($p < 0.05$). In the year following the comment letter conversation, reported in Column (6), the effect of signal on persistence is insignificant. Overall these results support

Figure 4.2: Earnings, Restatements, and Internal Control Weaknesses for Fiscal Years Surrounding Comment Letter Disclosure

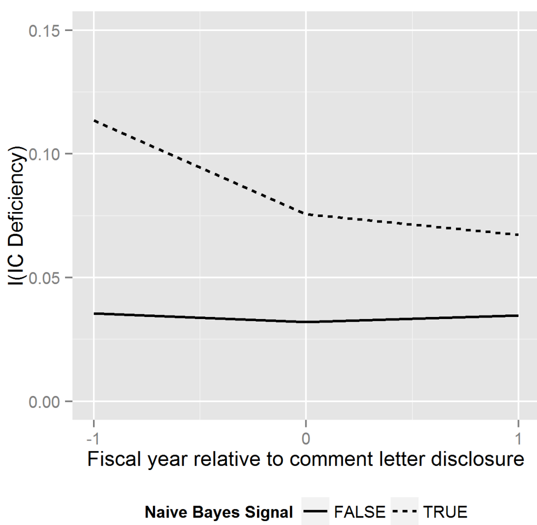
(A) Earnings, Partitoned by *Signal*



(B) Incidence of Restatments, Partitioned by *Signal*



(C) Incidence of Internal Control Weaknesses, Partitioned by *Signal*



This figure illustrates the differences in financial performance and reporting quality variables for holdout sample firms in the year before ($t = -1$), the year of ($t = 0$), and the year after ($t = 1$) comment letter disclosure, partitioned by the signal of importance. Panel A illustrates the difference in earnings for signaled comment letter firms. Panel B illustrates the difference in the rate of material restatements for signaled comment letter firms. Panel C illustrates the difference in internal control weaknesses for signaled comment letter firms. Refer to Appendix A for variable definitions.

Table 4.3: Signaled Comment Letters and Earnings

	I(Signal)$_0$		
	t=-1	t=0	t=1
	(1)	(2)	(3)
Intecept	−34.080	−34.212	−34.199
	(−0.008)	(−0.008)	(−0.008)
Earnings$_t$	−1.228***	0.085	−0.323
	(−2.734)	(0.172)	(−0.727)
Accruals$_t$	−0.041	0.463	0.472
	(−0.047)	(0.473)	(0.485)
I(Dividend)$_t$	−0.321*	−0.273	−0.376**
	(−1.673)	(−1.414)	(−1.987)
Special Items$_t$	1.221	−2.636*	−0.361
	(0.839)	(−1.842)	(−0.244)
Business Segments$_t$	0.027	0.024	0.035
	(0.579)	(0.508)	(0.748)
Geographic Segments$_t$	−0.025	−0.031	−0.034
	(−0.699)	(−0.885)	(−0.995)
I(Secondary Offering)$_t$	−0.251	0.597**	0.171
	(−0.772)	(2.054)	(0.538)
I(Acquisition)$_t$	0.011	0.183	−0.192
	(0.048)	(0.837)	(−0.830)
Age$_t$	0.004	0.004	0.002
	(0.379)	(0.436)	(0.193)
Book to Market$_{t-1}$	0.348***	0.313**	0.389***
	(2.761)	(2.488)	(3.298)
Log(Market Capitalization)$_{t-1}$	0.069	0.047	0.077*
	(1.526)	(1.029)	(1.708)
Observations	2,544	2,544	2,544
Pseudo R^2	0.075	0.074	0.074

*p<0.1; **p<0.05; ***p<0.01

This table presents results of the Equation 4.1 logit regression of *Earnings* on *Signal*, the Naive Bayesian signal of comment letter importance, for holdout sample firms, including industry and year fixed effects. Year $t = -1$ is the fiscal year prior to comment letter disclosure, and is the year under review by the SEC, year $t = 0$ is the year of disclosure, and $t = 1$ is the year following. Refer to A for variable definitions.

Table 4.4: Signaled Comment Letters and Earnings Persistence

	$Earnings_t$ (Profit Firms)			$Earnings_t$ (Loss Firms)		
	t=-1	t=0	t=1	t=-1	t=0	t=1
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	−0.028	−0.012	0.012	−0.027	−0.077	0.038
	(−0.277)	(−0.143)	(0.123)	(−0.356)	(−0.338)	(0.156)
$I(Signal)_0$	0.026**	−0.009	0.010	0.064**	−0.005	0.038
	(2.329)	(−1.066)	(1.041)	(2.335)	(−0.157)	(1.074)
$Earnings_{t-1}$	0.727***	0.595***	0.765***	0.573***	0.656***	0.641***
	(15.701)	(15.969)	(17.586)	(14.231)	(11.809)	(12.028)
$I(Signal)_0$ * $Earnings_{t-1}$	−0.493***	−0.019	−0.334***	0.655***	−0.173**	0.144
	(−3.875)	(−0.174)	(−2.807)	(8.387)	(−2.182)	(1.449)
$Accruals_{t-1}$	−0.147***	0.010	0.014	−0.278***	−0.185**	−0.170*
	(−4.462)	(0.348)	(0.443)	(−3.569)	(−2.223)	(−1.760)
$I(Dividend)_{t-1}$	0.015***	0.011***	0.009*	0.014	0.016	−0.014
	(2.773)	(2.629)	(1.960)	(0.603)	(0.562)	(−0.440)
Special $Items_{t-1}$	−0.305**	−0.666***	−0.937***	−0.755***	−0.750***	−0.775***
	(−2.484)	(−6.996)	(−8.068)	(−7.175)	(−6.291)	(−5.974)
Business $Segments_{t-1}$	0.002	−0.0005	−0.0003	0.007	−0.003	−0.0003
	(1.377)	(−0.445)	(−0.235)	(1.232)	(−0.466)	(−0.046)
Geographic $Segments_{t-1}$	−0.002	0.0002	0.001	0.001	0.008*	0.001
	(−1.612)	(0.288)	(1.131)	(0.198)	(1.833)	(0.194)
I(Secondary $Offering)_{t-1}$	−0.049***	−0.047***	0.012	−0.041*	−0.085***	−0.078**
	(−4.250)	(−4.599)	(0.974)	(−1.763)	(−2.865)	(−2.525)
$I(Acquisition)_{t-1}$	0.001	−0.005	−0.008	−0.019	0.004	−0.035
	(0.132)	(−0.975)	(−1.413)	(−0.708)	(0.135)	(−1.066)
Age_{t-1}	0.0003	−0.0001	−0.00004	0.0004	0.001	0.002
	(1.176)	(−0.303)	(−0.156)	(0.418)	(1.028)	(1.472)
Book to $Market_{t-1}$	−0.040***	−0.028***	−0.018***	−0.022*	−0.013	−0.009
	(−6.554)	(−6.074)	(−3.878)	(−1.867)	(−0.888)	(−0.636)
Log(Market $Cap.)_{t-1}$	0.002*	0.004***	0.004***	−0.003	0.006	0.017**
	(1.779)	(4.267)	(3.304)	(−0.565)	(0.911)	(2.448)
Observations	1,871	1,899	1,925	673	645	619
Adjusted R^2	0.241	0.283	0.260	0.508	0.433	0.434

*p<0.1; **p<0.05; ***p<0.01

This table presents results of the Equation 4.2 OLS regression of I(*Signal*), $Earnings_{t-1}$, and I(*Signal*) * $Earnings_{t-1}$ on $Earnings_t$, for holdout sample firms, including industry and year fixed effects. Standard errors are robust. Profit firms are shown in Columns (1) to (3) and loss firms are shown in Columns (4) to (6). Year $t = -1$ is the fiscal year prior to comment letter disclosure, and is the year under review by the SEC, year $t = 0$ is the year of disclosure, and $t = 1$ is the year following. Refer to A for variable definitions.

H1, specifically that receiving a signaled comment letter is associated with a lower persistence of profits in the following year, a result that may explain some of the negative abnormal returns associated with signaled comment letters.

4.4 Restatements

To study the association between signaled comment letters and higher rates of material restatements, I test H2 by examining the following logit regression model:

$$\begin{aligned} I(\text{Restatement})_{i,t} = {} & \beta_0 + \beta_1 I(\text{Signal})_{i,0} + \beta_2 I(\text{Restatement})_{i,t-1} \\ & + \beta_3 \text{Accruals}_{i,t} + \beta_4 I(\Delta\text{Receivables})_{i,t} + \beta_5 \Delta\text{Inventory}_{i,t} \\ & + \beta_6 \text{Soft Assets}_{i,t} + \beta_7 \text{Leverage}_{i,t} \\ & + \beta_8 I(\text{Secondary Offering})_{i,t} + \beta_9 \Delta\text{Earnings})_{i,t} \\ & + \beta_{10} \text{Big4}_{i,t} + \beta_{11} \text{Age}_{i,t} + \beta_{12} \text{Book to Market}_{i,t-1} \\ & + \beta_{13} \log(\text{Market Capitalization})_{i,t-1} + \varepsilon_{i,t} \quad . \end{aligned} \tag{4.3}$$

I include fixed effects for year and Fama-French 49 industry membership. As with Equation 4.1, $t = 0$ is the fiscal year in which the firm receives a comment letter, and this model is estimated for t =-1, 0 , and 1. $Restatement_{i,t}$ is an indicator variable equal to 1 if Audit Analytics reports a material restatement announced during year t, but 0 otherwise. See Appendix A for all other variable definitions. Control variables have been shown in prior literature to predict restatements (e.g., Dechow et al. 2011), and are defined in Appendix A. The coefficient of interest is β_1 which will be positive if firms with signaled comment letters are more likely to materially restate their financials in year t.

To study the effects of important comment letters on restatements, Table 4.5 gives the results of the regression model specified in Equation 4.3. Columns (1) to (3) used the signal and lagged restatements as the only control, including industry and year fixed effects. In Column (1), the coefficient on I(*Signal*) of 0.770 ($p < 0.01$) indicates that past restatements are positively associated with receipt of a signaled comment letter, consistent with the SEC targeting firms with material restatements, as required by SOX Section 408. The magnitude of this effect is similar to that of Column (2) where the coefficient on *Signal* of 0.745 ($p < 0.01$) indicates that important comment letters are also associated with increases in material restatements during the year of the SEC review. Column (3) indicates a lower, but still positive impact of signaled

Table 4.5: Signaled Comment Letters and Restatements

	I(Restatement)$_t$					
	t=-1	t=0	t=1	t=-1	t=0	t=1
	(1)	(2)	(3)	(4)	(5)	(6)
I(Signal)$_0$	0.770***	0.745***	0.354*	0.740***	0.692***	0.382*
	(3.467)	(3.376)	(1.332)	(3.300)	(3.054)	(1.414)
I(Restatement)$_{t-1}$	0.084	0.567**	0.491*	0.080	0.523**	0.425
	(0.322)	(2.242)	(1.781)	(0.304)	(2.051)	(1.525)
Accruals$_t$				0.617	−1.226	−0.574
				(0.603)	(−1.234)	(−0.505)
ΔReceivables$_t$				−2.200	3.170*	−0.148
				(−1.269)	(1.947)	(−0.090)
ΔInventory$_t$				−0.312	−0.769	1.927
				(−0.131)	(−0.261)	(0.764)
Soft Assets$_t$				0.404	0.409	−0.173
				(0.911)	(0.948)	(−0.370)
Leverage$_t$				−0.004	0.010	0.008
				(−0.199)	(0.607)	(0.497)
I(Secondary Offering)$_t$				0.408	0.570*	0.566
				(1.221)	(1.864)	(1.592)
ΔEarnings$_t$				−0.683	0.608	0.251
				(−1.215)	(1.225)	(0.464)
I(Big4)$_t$				0.280	0.117	0.895***
				(1.201)	(0.532)	(3.353)
Age$_t$				−0.001	0.007	−0.005
				(−0.074)	(0.678)	(−0.436)
Book to Market$_{t-1}$				0.182	−0.070	−0.047
				(1.143)	(−0.433)	(−0.309)
Log(Market Cap.)$_{t-1}$				−0.075	−0.183***	−0.204***
				(−1.399)	(−3.294)	(−3.385)
Intecept	3.740	−33.004	4.071	3.900	−31.747	5.329
	(0.001)	(−0.007)	(0.001)	(0.001)	(−0.007)	(0.001)
Observations	2,544	2,544	2,544	2,544	2,544	2,544
Pseudo R^2	0.061	0.070	0.069	0.069	0.092	0.091

*p<0.1; **p<0.05; ***p<0.01

This table presents results of the Equation 4.3 OLS regression of I(*Signal*) on I(*Restatement*)$_t$, for holdout sample firms, including industry and year fixed effects. Standard errors are robust, and significance is one-tailed for I(*Signal*) as it is expected to increase restatements, and two-tailed for all other variables. Columns (1) to (3) include only the signal and lagged restatements as predictor variables, and Columns (4) to (6) include controls shown in prior literature to affect restatements. Year $t = -1$ is the fiscal year prior to comment letter disclosure, and is the year under review by the SEC, year $t = 0$ is the year of disclosure, and $t = 1$ is the year following. Refer to A for variable definitions.

comment letters on restatements in the year following disclosure, with a coefficient on *Signal* of 0.354 ($p < 0.1$, one-tailed). Including controls in Columns (4) to (6), results are similar. In Column (6) the coefficient on *Signal* of 0.382 ($p < 0.1$, one-tailed, as I predict an increase in restatements) indicates a 47 percent increase in the odds of a restatement, a result that is not diminished by including controls shown in prior research to explain restatements. These results support H2. While the association between comment letters and past and current restatements has already been shown (e.g., Cassell et al. 2013; Dechow et al. 2016), the finding that signaled comment letters may be able to identify future restatements indicates that the review process identifies undisclosed financial reporting deficiencies. Prior research has demonstrated an effect of restatements on returns, so this association may also be a source of negative announcement returns for signaled comment letters (e.g., Hribar et al. 2004).

4.5 Internal Control Weaknesses

To study the association between signaled comment letters and increased internal control weaknesses, I test H3 by examining the following logit regression model:

$$\begin{aligned} I(\text{Weakness})_{i,t} = {} & \beta_0 + \beta_1 I(\text{Signal})_{i,0} + \beta_2 I(\text{Weakness})_{i,t-1} \\ & + \beta_3 \log(\text{Market Capitalization})_{i,t-1} + \beta_4 \text{SalesGrowth}_{i,t} \\ & + \beta_5 \text{Inventory}_{i,t} + \beta_6 \text{Accruals}_{i,t} + \beta_7 \text{Leverage}_{i,t} \\ & + \beta_8 \Delta\text{Receivables}_{i,t} + \beta_9 \Delta\text{Inventory}_{i,t} + \beta_{10} \text{Soft Assets}_{i,t} \\ & + \beta_{11} I(\text{Secondary Offering})_{i,t} + \beta_{12} \Delta\text{Earnings}_{i,t} \\ & + \beta_{13} \text{Big4}_{i,t-1} + \beta_{14} \text{Age} + \beta_{15} \text{Book to Market} + \varepsilon_{i,t} \quad . \end{aligned} \tag{4.4}$$

I include fixed effects for year and Fama-French 49 industry membership. $Weakness_{i,t}$ is an indicator variable equal to 1 if Audit Analytics reports that internal controls were ineffective during year t, but 0 otherwise. This model is estimated for t =-1, 0 , and 1. Control variables have been shown in prior literature to predict restatements (e.g., Ogneva, Subramanyam, and Raghunandan 2007), and are defined in Appendix A. The coefficient of interest is β_1 which will be positive if firms with signaled comment letters are more likely to report an internal control Weakness in year t.

To study the effects of important comment letters on restatements, I implement the regression model specified in Equation 4.4 and report the results in Table 4.6.

Table 4.6: Signaled Comment Letters and Internal Control Weaknesses

	I(Weakness)$_t$					
	t=-1	t=0	t=1	t=-1	t=0	t=1
	(1)	(2)	(3)	(4)	(5)	(6)
I(Signal)$_0$	1.123***	0.388	0.551**	1.170***	0.280	0.377
	(4.222)	(1.171)	(1.716)	(4.192)	(0.801)	(1.076)
I(Weakness)$_{t-1}$	2.460***	3.212***	2.750***	2.361***	3.141***	2.407***
	(9.975)	(11.137)	(9.440)	(9.163)	(10.152)	(7.673)
Sales Growth$_t$				−0.471	−0.402	−0.049
				(−1.161)	(−0.858)	(−0.119)
Inventory$_t$				0.382	−0.592	−1.191
				(0.265)	(−0.364)	(−0.849)
Accruals$_t$				−1.324	0.080	−2.479*
				(−1.051)	(0.058)	(−1.861)
Leverage$_t$				0.034	0.059***	0.053***
				(1.410)	(3.090)	(2.905)
ΔReceivables$_t$				−3.502	−3.939	6.436***
				(−1.566)	(−1.577)	(2.842)
ΔInventory$_t$				−2.495	4.548	−0.826
				(−0.969)	(1.007)	(−0.267)
Soft Assets$_t$				−0.044	0.335	0.149
				(−0.071)	(0.525)	(0.238)
I(Secondary Offering)$_t$				−0.091	−0.274	0.583
				(−0.210)	(−0.572)	(1.292)
ΔEarnings$_t$				−1.960***	0.696	0.534
				(−2.814)	(0.998)	(0.882)
I(Big4)$_t$				−0.574**	0.117	−0.466
				(−2.001)	(0.357)	(−1.528)
Age$_t$				−0.009	−0.027	−0.014
				(−0.574)	(−1.633)	(−0.915)
Book to Market$_{t-1}$				−0.035	0.302	0.098
				(−0.168)	(1.381)	(0.535)
Log(Market Cap.)$_{t-1}$				−0.206**	−0.215**	−0.290***
				(−2.471)	(−2.393)	(−3.181)
Intecept	20.349	−40.405	−36.443	23.050	−37.805	−32.847
	(0.002)	(−0.003)	(−0.003)	(0.002)	(−0.003)	(−0.003)
Observations	2,544	2,544	2,544	2,544	2,544	2,544
Pseudo R^2	0.220	0.259	0.175	0.274	0.303	0.258

*p<0.1; **p<0.05; ***p<0.01

This table presents results of the Equation 4.4 OLS regression of I(*Signal*) on I(*Weakness*)$_t$, for holdout sample firms, including industry and year fixed effects. Standard errors are robust, and significance is one-tailed for I(*Signal*) as it is expected to increase weaknesses, and two-tailed for all other variables. Columns (1) to (3) include only the signal and lagged internal controls weakness as predictor variables, and Columns (4) to (6) include controls shown in prior literature to affect internal controls weakness. Year $t = -1$ is the fiscal year prior to comment letter disclosure, and is the year under review by the SEC, year $t = 0$ is the year of disclosure, and $t = 1$ is the year following. Refer to A for variable definitions.

Columns (1) to (3) used the signal, with lagged internal control weaknesses as the only control, including industry and year fixed effects. In Column (1), the coefficient on *Signal* of 1.123 ($p < 0.01$) indicates that past weaknesses are positively associated with receipt of a signaled comment letter. Column (2) reports no significant increase in internal control weaknesses due to the signaled comment letter, likely because any weaknesses identified in the comment letter reivew will not be disclosed until the following annual report in time $t = 1$. The coefficient on *Signal* in Column (3) is 0.551 ($p < 0.05$, one tailed, as I predict an increase in weaknesses) indicates an increase in weaknesses reported in the year following receipt of a signaled comment letter, representing an increase in the odds of reporting a material weakness of 74 percent, controlling for past internal control weakness. Columns (4) to (6) include additional control variables shown in prior literature to be associated with internal control weaknesses. Signaled comment letters in are associated with weaknesses reported in year $t - 1$, with a coefficient of 1.170 ($p < 0.01$). Column (5) reports no significant increase in weakness in the year of the signaled comment letter disclosure, similar to Column (2). Column (6) reports that the signal no longer has a significant effect on weaknesses reported in the year following, indicating that the increase in internal control weaknesses reported in the following year can be explained by the control variables. While internal control weaknesses have been shown to have an effect on returns (e.g., Hammersley et al. 2008), the limited association between signaled comment letters and internal control weaknesses indicates that even if signaled comment letters help to reveal internal control weaknesses to management and auditors, remedial steps can be taken to resolve the weaknesses prior to the next audit report.

Chapter 5

Conclusion

This study uses Naive Bayesian text classification to signal important SEC comment letters, using negative stock returns following disclosure as the measure of importance. The resulting signal is used on a holdout sample of comment letters, to demonstrate that text analysis is effective (up to 40 percent more precise than chance) at identifying comment letters associated with negative abnormal returns. I study the effects of signaled comment letters on returns, and find some evidence of underreaction to comment letters, as the signal is only predictive of abnormal returns for comment letters that were known to be viewed on EDGAR in the days immediately after disclosure. For firms with above-median comment letter views, abnormal returns following signaled disclosure is significantly more negative 90 days after disclosure (-5.8 percent) than three days after disclosure (-1.3 percent). I study the effect of signaled comment letters on earnings and earnings persistence, noting lower persistence of profits in the year before and the year following signaled comment letters. I study the effect of signaled comment letters on material restatements, finding higher levels of material restatements both in the year before and the year after signaled comment letters. Signaled comment letters are related to internal control weaknesses the year prior to the SEC review, however future weaknesses do not appear to be explained by signaled comment letters.

The implications of this study have broad applicability to the discussion of the role of government monitoring of financial disclosures, and by association, the auditors that review financial disclosures. Some comment letters are reactive, resulting from reviews conducted as a result of prior restatements or other factors identified by SOX and the SEC as triggers for more frequent reviews. Other comment letters appear to be inconsequential, at least from a valuation perspective, as they deal with complex disclosure regulations, are infrequently read by investors, and have little effect on stock prices. On the other hand, some comment letters do appear to be consequential, and this study shows that the text of comment letters can be used to identify firms with undisclosed performance and financial disclosure deficiencies, supporting their use as a source of information about firms' financial reporting and audit quality. Comment letters also appear to change managers' and auditors' behavior, resulting in future changes to reported earnings and higher restatements in the year following a review. SEC reviews appear to ask questions and prompt disclosures that are not otherwise highlighted by securities analysts, a category of stakeholder who ostensibly review corporate disclosures, but who have little incentive or power to ask similar questions or request additional disclosures. Overall, these results suggest that there is value to the role of the government in reviewing financial disclosures, and that the benefits of this disclosure are in many cases material to investors.

References

Ashbaugh-Skaife, H., D. W. Collins, W. R. Kinney, and R. LaFond (2009). "The effect of SOX internal control deficiencies on firm risk and cost of equity". *Journal of Accounting Research* 47.1, 1–43.

Baldwin, J., D. Hurtt, and J. MacGregor (2013). "The Relationship between SEC Comment Letters and Subsequent Auditor Change". *Working Paper, Baylor University*.

Barberis, N., A. Shleifer, and R. Vishny (1998). "A model of investor sentiment". *Journal of Financial Economics* 49.3, 307–343.

Beneish, M. D., M. B. Billings, and L. D. Hodder (2008). "Internal control weaknesses and information uncertainty". *The Accounting Review* 83.3, 665–703.

Bens, D. A., M. Cheng, and M. Neamtiu (2016). "The Impact of SEC Disclosure Monitoring on the Uncertainty of Fair Value Estimates". *The Accounting Review* 91.2, 349–375.

Bernard, V. and J. Thomas (1989). "Post-earnings-announcement drift: delayed price response or risk premium?" *Journal of Accounting Research* 27, 1–36.

Bloomfield, R. J. (2002). "The "Incomplete Revelation Hypothesis" and Financial Reporting". *Accounting Horizons* 16.3, 233–243.

Blouin, J., L. Krull, and L. Robinson (2014). "The Location, Composition, and Investment Implications of Permanently Reinvested Earnings". *Working Paper, University of Oregon*.

Boone, J. P., C. L. Linthicum, and A. Poe (2013). "Characteristics of accounting standards and SEC review comments". *Accounting Horizons* 27.4, 711–736.

Bozanic, Z., P. Choudhary, and K. J. Merkley (2015). "Securities Law Expertise and Corporate Disclosure". *Working Paper, Georgetown University*.

Bozanic, Z., J. R. Dietrich, and B. Johnson (2014). "The SEC Comment Letter Process and Firm Disclosure". *Working Paper, The Ohio State University*.

Brown, S. V., X. Tian, and J. W. Tucker (2015). "The Spillover Effect of SEC Comment Letters on Qualitative Corporate Disclosure: Evidence from the Risk Factor Disclosure". *Working Paper, Arizona State University*.

Bryan, S. H. (1997). "Incremental information content of required disclosures contained in management discussion and analysis". *Accounting Review*, 285–301.

Campbell, J. Y., A. W.-C. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Cassell, C. A., L. M. Cunningham, and L. L. Lisic (2015). "The Consequences of Writing Not So Readable Responses to SEC Comment Letters". *Working Paper, University of Arkansas*.

Cassell, C. A., L. M. Dreher, and L. A. Myers (2013). "Reviewing the SEC's Review Process: 10-K Comment Letters and the Cost of Remediation". *The Accounting Review* 88.6, 1875–1908.

CFA Institute (2014). *Candidate Body of Knowledge*. URL: http://www.cfainstitute.org/programs/cfaprogram/courseofstudy/Pages/cbok.aspx.

Chan, L. K., N. Jegadeesh, and J. Lakonishok (1996). "Momentum strategies". *Journal of Finance* 51.5, 1681–1713.

Chen, N. (2015). "Repatriation Taxes and the Value of Cash Holdings". *Working Paper, University of California at Irvine*.

Chen, X., Q. Cheng, and A. K. Lo (2013). "Is the decline in the information content of earnings following restatements short-lived?" *The Accounting Review* 89.1, 177–207.

Chetty, R., A. Looney, and K. Kroft (2009). "Salience and Taxation: Theory and Evidence". *American Economic Review* 99.4, 1145–1177.

Cohen, L. and A. Frazzini (2008). "Economic links and predictable returns". *Journal of Finance* 63.4, 1977–2011.

Correia, M. M. (2009). *Political connections, SEC enforcement and accounting quality*. PhD Thesis, Stanford University.

Cunningham, L. M., R. Schmardebeck, and W. Wang (2016). "Examining the Usefulness of the SEC Filing Review Process: Evidence from Debt Contracting". *Working Paper, University of Tennessee*.

Dagan, I., R. Feldman, and H. Hirsh (1996). "Keyword-based browsing and analysis of large document sets". *Proceedings of the Fifth Annual Symposium on Document Analysis and Information Retrieval–SDAIR, Las Vegas, Nevada*. Citeseer, 191–208.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). "Investor psychology and security market under-and overreactions". *Journal of Finance* 53.6, 1839–1885.

Davis, A. K., J. M. Piger, and L. M. Sedor (2012). "Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language". *Contemporary Accounting Research* 29.3, 845–868.

De Franco, G., F. P. Vasvari, D. Vyas, and R. Wittenberg-Moerman (2013). "Debt analysts' views of debt-equity conflicts of interest". *The Accounting Review* 89.2, 571–604.

De Simone, L. and B. Stomberg (2015). "How Internal Tax and Legal Expertise Affect Corporate Income Taxes". *Working Paper, Stanford University.*

Dechow, P. M., W. Ge, C. R. Larson, and R. G. Sloan (2011). "Predicting Material Accounting Misstatements". *Contemporary Accounting Research* 28.1, 17–82.

Dechow, P. M., A. Lawrence, and J. Ryans (2016). "SEC Comment Letters and Insider Sales". *The Accounting Review* 91.2, 401–439.

Dechow, P., W. Ge, and C. Schrand (2010). "Understanding Earnings Quality: A Review of the Proxies, Their Determinants and Their Consequences". *Journal of Accounting and Economics* 50.2, 344–401.

DellaVigna, S. and J. M. Pollet (2009). "Investor inattention and Friday earnings announcements". *Journal of Finance* 64.2, 709–749.

Deloitte & Touche LLP (2014). *SEC Comment Letters – Including Industry Insights: A Recap of Recent Trends.* URL: `http://deloitte.wsj.com/riskandcompliance/2014/12/05/sec-comment-letters-a-recap-of-recent-trends/`.

Doyle, J. T., W. Ge, and S. McVay (2007). "Accruals quality and internal control over financial reporting". *The Accounting Review* 82.5, 1141–1170.

Drake, M. S., D. T. Roulstone, and J. R. Thornock (2015). "The Determinants and Consequences of Information Acquisition via EDGAR". *Contemporary Accounting Research* 32.3, 1128–1161. ISSN: 1911-3846.

Ertimur, Y. and M. E. Nondorf (2006). "IPO Firms and the SEC Comment Letter Process". *Working Paper, Duke University.*

Ettredge, M., K. Johnstone, M. Stone, and Q. Wang (2011). "The effects of firm size, corporate governance quality, and bad news on disclosure compliance". *Review of Accounting Studies* 16.4, 866–889.

Feldman, R. and I. Dagan (1995). "Knowledge Discovery in Textual Databases KDT". *Proceedings of the First International Conference on Knowledge Discovery and Data Mining.* Vol. 95, 112–117.

Feldman, R., S. Govindaraj, J. Livnat, and B. Segal (2010). "Management's tone change, post earnings announcement drift and accruals". *Review of Accounting Studies* 15.4, 915–953.

Francis, J., D. Nanda, and P. Olsson (2008). "Voluntary Disclosure, Earnings Quality, and Cost of Capital". *Journal of Accounting Research* 46.1, 53–99.

Francis, J., R. LaFond, P. Olsson, and K. Schipper (2005). "The market pricing of accruals quality". *Journal of Accounting and Economics* 39.2, 295–327.

Francis, J. R. (2004). "What do we know about audit quality?" *The British Accounting Review* 36.4, 345–368.

Francis, J. R. (2011). "A framework for understanding and researching audit quality". *Auditing: A Journal of Practice & Theory* 30.2, 125–152.

Gietzmann, M. B. and H. Isidro (2013a). "Institutional Investors' Reaction to SEC Concerns about IFRS and US GAAP Reporting". *Journal of Business Finance & Accounting* 40.7-8, 796–841.

Gietzmann, M. B. and A. K. Pettinicchio (2013b). "External Auditor Reassessment of Client Business Risk Following the Issuance of a Comment Letter by the SEC". *European Accounting Review* 23.1, 57–85.

Gilbert, D. (2014). "Regulators ask Exxon Why No Writedowns?" *The Wall Street Journal*. URL: `http://blogs.wsj.com/corporate-intelligence/2014/02/03/regulators-ask-exxon-why-no-writedowns/`.

Gleason, C. A., N. T. Jenkins, and W. B. Johnson (2008). "The Contagion Effects of Accounting Restatements". *The Accounting Review* 83.1, 83–110.

Greenlight Capital (2011). *GAAP-uccino*. URL: `http://online.wsj.com/public/resources/documents/EinhornGMCRpresentation_Oct2011_VIC.pdf`.

Grossman, S. and J. Stiglitz (1980). "On the Impossibility of Informationally Efficient Markets". *American Economic Review* 70.3, 393–408.

Gupta, S. and R. D. Israelsen (2015). "Hard and Soft Information: Firm Disclosure, SEC Letters, and the JOBS Act". *Working Paper, Indiana University*.

Hammersley, J. S., L. A. Myers, and C. Shakespeare (2008). "Market reactions to the disclosure of internal control weaknesses and to the characteristics of those weaknesses under Section 302 of the Sarbanes Oxley Act of 2002". *Review of Accounting Studies* 13.1, 141–165.

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009). "Driven to distraction: Extraneous events and underreaction to earnings news". *Journal of Finance* 64.5, 2289–2325.

Hirshleifer, D. and S. H. Teoh (2003). "Limited attention, information disclosure, and financial reporting". *Journal of Accounting and Economics* 36.1, 337–386.

Hong, H. and J. C. Stein (1999). "A unified theory of underreaction, momentum trading, and overreaction in asset markets". *Journal of Finance* 54.6, 2143–2184.

Hribar, P. and N. T. Jenkins (2004). "The effect of accounting restatements on earnings revisions and the estimated cost of capital". *Review of Accounting Studies* 9.2-3, 337–356.

Hribar, P., T. Kravet, and R. Wilson (2014). "A new measure of accounting quality". English. *Review of Accounting Studies* 19.1, 506–538. ISSN: 1380-6653.

Huang, A. H., A. Y. Zang, and R. Zheng (2014). "Evidence on the Information Content of Text in Analyst Reports". *The Accounting Review* 89.6, 2151–2180.

Johnston, R. and R. Petacchi (2016). "Regulatory Oversight of Financial Reporting: Securities and Exchange Commission Comment Letters". *Contemporary Accounting Research* Forthcoming.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Karlgren, J. and D. Cutting (1994). "Recognizing text genres with simple metrics using discriminant analysis". *Proceedings of the 15th conference on Computational Linguistics*. Vol. 2. Association for Computational Linguistics, 1071–1075.

Kessler, B., G. Numberg, and H. Schutze (1997). "Automatic detection of text genre". *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 32–38.

Kinney, W. R., Z.-V. Palmrose, and S. Scholz (2004). "Auditor Independence, Non-Audit Services, and Restatements: Was the US Government Right?" *Journal of Accounting Research* 42.3, 561–588.

Kothari, S., X. Li, and J. E. Short (2009a). "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.

Kothari, S., S. Shu, and P. Wysocki (2009b). "Do managers withhold bad news?" *Journal of Accounting Research* 47.1, 241–276.

Kovbasyuk, S. and M. Pagano (2015). "Advertising arbitrage". *Working Paper, University of Naples Federico II.*

Kubick, T. R., D. P. Lynch, M. A. Mayberry, and T. C. Omer (2016). "The Effects of Regulatory Scrutiny on Tax Avoidance: An Examination of SEC Comment Letters". *The Accounting Review* Forthcoming.

Lang, M. and R. Lundholm (1993). "Cross-sectional determinants of analyst ratings of corporate disclosures". *Journal of Accounting Research*, 246–271.

Larcker, D. F. and A. A. Zakolyukina (2012). “Detecting deceptive discussions in conference calls”. *Journal of Accounting Research* 50.2, 495–540.

Laurion, H., A. Lawrence, and J. Ryans (2015). “U.S. Audit Partner Rotation”. *Working Paper, University of California at Berkeley*.

Lawrence, J., G. Lei, and D. Smith (2010). “SEC Comment Letters and Financial Statement Restatements”. *Working Paper, University of Nebraska at Lincoln*.

Lewis, D. D. (1998). “Naive (Bayes) at forty: The independence assumption in information retrieval”. *Machine learning: ECML-98*. Springer, 4–15.

Li, F. (2008). “Annual report readability, current earnings, and earnings persistence”. *Journal of Accounting and Economics* 45.2, 221–247.

Li, F. (2010a). “Textual Analysis of Corporate Disclosures: A Survey of the Literature”. English. *Journal of Accounting Literature* 29, 143–165.

Li, F. (2010b). “The Information Content of Forward-Looking Statements in Corporate Filings. A Naive Bayesian Machine Learning Approach”. *Journal of Accounting Research* 48.5, 1049–1102.

Liu, L.-L., K. Raghunandan, and D. Rama (2009). “Financial restatements and shareholder ratifications of the auditor”. *Auditing: A Journal of Practice & Theory* 28.1, 225–240.

Ljungqvist, A. and W. Qian (2014). “How Constraining Are Limits to Arbitrage? Evidence from a Recent Financial Innovation”. *NBER Working Paper 19834*.

Loughran, T. and B. McDonald (2011). “When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks”. *Journal of Finance* 66.1, 35–65.

Loughran, T. and B. McDonald (2015). “Textual Analysis in Accounting and Finance: A Survey”. *Working Paper, University of Notre Dame*.

McCallum, A. and K. Nigam (1998). “A Comparison of Event Models for Naive Bayes Text Classification”. *AAAI-98 Workshop on Learning for Text Categorization*. Vol. 752, 41–48.

Merton, R. C. (1987). “A simple model of capital market equilibrium with incomplete information”. *Journal of Finance* 42.3, 483–510.

Mosteller, F. and D. L. Wallace (1984). *Applied Bayesian and Classical Inference*. Springer.

Naughton, J. P., R. Rogo, J. Sunder, and R. Zhang (2016). “SEC Monitoring of Foreign Firms’ Disclosures”. *Working Paper, University of British Columbia*.

Ogneva, M., K. R. Subramanyam, and K. Raghunandan (2007). “Internal control weakness and cost of equity: Evidence from SOX Section 404 disclosures”. *The Accounting Review* 82.5, 1255–1297.

Palmrose, Z.-V., V. J. Richardson, and S. Scholz (2004). "Determinants of market reactions to restatement announcements". *Journal of Accounting and Economics* 37.1, 59–89.

Pang, B., L. Lee, and S. Vaithyanathan (2002). "Thumbs up?: Sentiment classification using machine learning techniques". *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*. Vol. 10. Association for Computational Linguistics, 79–86.

Patatoukas, P. N. (2012). "Customer-Base Concentration: Implications for Firm Performance and Capital Markets". *The Accounting Review* 87.2, 363–392.

Patatoukas, P. N., R. G. Sloan, and J. Zha (2015). "On the Pricing of Mandatory DCF Disclosures: Evidence from Oil and Gas Royalty Trusts." *The Accounting Review* 90.6, 2449–2482.

Pershing Square (2013). *Who wants to be a Millionaire?* URL: http://factsabout-herbalife.com/wp-content/uploads/2013/01/Who-wants-to-be-a-Millionaire.pdf.

Peterson, K. (2012). "Accounting complexity, misreporting, and the consequences of misreporting". *Review of Accounting Studies* 17.1, 72–95.

Prescience Point (2013). *A Rock in Peril.* URL: http://www.presciencepoint.com/uncategorized/boulder-brands-inc-bdbd-feb-26-2013/.

Revsine, L., D. W. Collins, B. Johnson, and F. Mittelstaedt (2011). *Financial Reporting and Analysis*. Fifth. McGraw-Hill Irwin.

Robinson, J. R., Y. Xue, and Y. Yu (2011). "Determinants of disclosure noncompliance and the effect of the SEC review: Evidence from the 2006 mandated compensation disclosure regulations". *The Accounting Review* 86.4, 1415–1444.

Sandler, L. (2013). "Muddy Waters Secret China Weapon is on SEC Website". *Bloomberg News*. URL: http://www.bloomberg.com/news/2013-02-19/muddy-waters-secret-china-weapon-is-on-sec-website.html.

Schorfheide, F. and K. I. Wolpin (2012). "On the use of holdout samples for model selection". *American Economic Review* 102.3, 477–481.

Schroeder, N. and C. Gibson (1990). "Readability of Management's Discussion and Analysis". *Accounting Horizons* 4.4, 78–87.

SEC (2001). *Comment Letter Follow-Up (Audit 326).* U.S. Securities and Exchange Commission. URL: www.sec.gov/about/oig/audit/326fin.pdf.

SEC (2015). *FY 2016 Congressional Budget Justification, FY 2016 Annual Performance Plan, and FY 2014 Annual Performance Report.* U.S. Securities and

Exchange Commission. URL: `http : / / www . sec . gov / about / reports / secfy16congbudgjust.shtml`.

SOX (2002). *The Sarbanes-Oxley Act of 2002. Public Law 107-204 [H.R. 3763]*. U.S. House of Representatives.

Talley, E. and D. O'Kane (2012). "The Measure of a MAC: A Machine-Learning Protocol for Analyzing Force Majeure Clauses in MA Agreements". *Journal of Institutional and Theoretical Economics* 168.1, 181–201.

Tetlock, P. C. (2007). "Giving content to investor sentiment: The role of media in the stock market". *Journal of Finance* 62.3, 1139–1168.

You, H. and X.-J. Zhang (2009). "Financial reporting complexity and investor underreaction to 10-K information". *Review of Accounting Studies* 14.4, 559–586.

Appendix A

Variable Definitions

Variable	Definition
Accruals	Operating earnings - cash flow from operations, normalized by total assets (Compustat $(oiadp - oancf)/at$.
I(Acquisition)	Indicator variable if the firm made a material acquisition (greater than 5 percent of assets) during the fiscal year (Compustat 1 if $acq/at > 0.05$ but 0 otherwise).
Age	Number of years the firm has appeared in the Compustat annual file.
I(Big4)	Indicator variable if the firm has a Big-4 auditor.
Book to Market	Book value of equity divided by market value of equity (Compustat $seq/(csho * prcc_f)$), winsorized at the one percent level.
Business Segments	Number of business segments (Compustat segment file *stype*="BUSSEG").
CAR[0,3]	Three day cumulative abnormal return from the close prior to comment letter disclosure date through the close three trading days after the disclosure date. Calculation details are described in Section 3.1.
CAR[0,90]	90 day cumulative abnormal return from the close prior to comment letter disclosure date through the close 90 trading days after the disclosure date. Calculation details are described in Section 3.1.

Continued.

Variable	Definition
Conversation Items	Number of total letters (Form UPLOAD) and company responses (Form CORRESP) in the comment letter conversation.
I(Dividend)	Indicator variable if the firm paid a dividend during the fiscal year (Compustat 1 if $dvc > 0$ but 0 otherwise).
Earnings	Income before extraordinary items - adjusted for common stock equivalents normalized by total assets, winsorized at the one percent level (Compustat $ibadj/at$).
ΔEarnings	Earnings_t - Earnings_{t-1}.
EDGAR Views	Number of document downloads of the first comment letter (Form UPLOAD) in a conversation (SEC EDGAR web log files).
Geographic Segments	Number of geographic segments (Compustat segment file $stype$="GEOSEG").
I(Weakness)	Indicator variable if an internal control Weakness is reported at the fiscal year end (Audit Analytics). 1 if NOTEFF_ACC_RULE=1 or NOTEFF_FIN_FRAUD=1 or NOTEFF_OTHER=1 or NOTEFFERRORS=1.
Insider Sales	Insider sales as a percentage of shares outstanding. Sum of the number of shares (SHARES) sold from disclosure date - 15 days to disclosure date +15 days for officers and directors having ROLECODE of CEO, D, O, H, DO, OD, VC, OB, OP, OT, CB, AV, CFO, CI, CO, CT, EVP, OX, P, S, SVP, VP (Thompson Reuters Insider Trading), divided by shares outstanding at the prior year end (Compustat *csho*) * 100.
Insider Sales Rank	Equals 1 if Insider Sales is 0, and is set to 2 to 5 for firms with Insider Sales in the first to fourth quartile of non-zero insider sales.
Inventory	Inventory as a fraction of total assets, winsorized at the one percent level (Compustat $invt/at$)
ΔInventory	Change in inventories as a fraction of total assets, winsorized at the one percent level (Compustat $invt_t/at_t - invt_{t-1}/at_{t-1}$)

Continued.

Variable	Definition
Leverage	Debt to equity (Compustat $(dltt + lt)/seq$).
Market Capitalization	Market capitalization of common equity ($ millions) (Compustat $csho * prcc_f$).
Number of Questions	Number of itemized questions asked by the SEC in the first comment letter of the conversation. The methodology for determining the number of questions is described in Appendix B
ΔReceivables	Change in receivables as a fraction of total assets, winsorized at the one percent level (Compustat $rect_t/at_t - rect_{t-1}/at_{t-1}$)
I(Restatement)	Indicator variable if a material restatement was announced during the fiscal year (Audit Analytics).
I(Revenue Recognition)	Indicator variable if revenue recognition questions are asked by the SEC in the first comment letter of the conversation. The methodology for determining if a revenue recognition question is present is described in Appendix B
Sales Growth	Sales growth, winsorized at the one percent level (Compustat $(sale_t - sale_{t-1})/sale_{t-1}$)
I(Secondary Offering)	Indicator variable if the firm had a material issuance of equity during the fiscal year (Compustat 1 if $sstk/at > 0.1$ but 0 otherwise).
I(Signal)	Indicator variable if the Naive Bayesian classification algorithm identifies a comment letter conversation as *important*, based on the methodology discussed in Section 3.2. The classification settings are: Unigram+Bigram feature set, term frequency, and bottom quartile of CAR[0,90] by year as the signal of importance for the training comment letters.
Soft Assets	Fraction of assets that are neither cash nor property, plant, and equipment, winsorized at the one percent level (Compustat $(at - ppent - che)/at$.
Special Items	Special items as a fraction of total assets winsorized at the one percent level (Compustat spi/at).

Appendix B

Comment Letter Preparation

1. Remove common english "stop words", i.e. frequent words that are ineffective in distinguishing important from unimportant documents:

   a, about, above, after, again, against, all, am, an, and, any, are, as, at, be, because, been, before, being, below, between, both, but, by, cannot, could, couldn't, did, do, does, doing, down, during, each, few, for, from, further, had, has, have, having, he, her,here, hers, herself, him, himself, his, how, i, if, in, into, is, it, its, itself, me, more, most, my, myself, no, nor, not, of, off, on, once, only, or, other, ought, our, ours, ourselves, out, over, own, same, she, should, so, some, such, than, that, the, their, theirs, them, themselves, then, there, these, they, they've, this, those, through, to, too, under, until, up, very, was, we, were, what, when, where, which, while, who, whom, why, with, would, you, your, yours, yourself, yourselves

2. Determine if document is related to a Form 10-K: Text between the string `"Re:"` and `"Dear "` contains the string `"Form 10-K "`

3. Count the number of questions in the first comment letter:

   Identify paragraphs that begin with the regular expression

```
"( |\n|\t)([1-9][.]|[1-9][)]|[1-9][0-9][.]|[1-9][0-9][)])
(Please|We|It|Pursuant|Refer|In|To|Revise|Tell|You|
On|The|Discuss|For|Although|Further|If|Describe)"
```

This extracts a list of questions, as well as the number at the beginning of each question (e.g., {"3", "3. Please revise your discussion of..."}). The number of items in the list is compared to the extracted number of the final question, and if there is a disagreement, the smaller number is selected. I manually check 100 documents and find that this method identifies the number of comments exactly correctly in 90% of documents, and the total number of questions identified is 96% accurate.

4. Identify revenue recognition related comment: True if text between `"Dear "` and the end of the document satisfies the regular expression

```
"([Rr]evenue [Rr]ecognition)|([Rr]ecognize [Rr]evenue)|
(ASC 605)|(SAB 101)|(SAB 104)|(EITF 99-19)|(FAS 48)|
(EITF 01-9)|(FAS 45)|(SOP 97-2)|(SOP 98-9)|(EITF 00-21)|
(EITF 08-1)|(EITF 08-2)|(EITF 08-9)|(EITF 01-3)|(EITF 00-24)|
(EITF 95-1)"
```