THREE ESSAYS ON UNORTHODOX AUDIT EVIDENCE

By

Kyunghee Yoon

A dissertation submitted to the

Graduate School-Newark

Rutgers, The State University of New Jersey

in partial fulfillment of requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

Written under the direction of

Dr. Miklos A. Vasarhelyi

and approved by

____

Dr. Miklos A. Vasarhelyi

____

Dr. Alexander Kogan

____

Dr. Carolyn B. Levine

____

Dr. Theodore Mock

____

Dr. Kevin Moffitt

____

Newark, New Jersey

May 2016

ABSTRACT OF THE DISSERTATION

Three Essays on Unorthodox Audit Evidence

By Kyunghee Yoon

Dissertation Director:

Professor Miklos A. Vasarhelyi

A recent survey of audit practitioners (Trompeter and Wright 2010) indicates that most large audit firms use business risk audit (BRA) approaches that require a compressive understanding of the client's industry and strategy and highlight the importance of gathering audit evidence from a wide variety of sources. However, the significance of nonfinancial information as primary audit evidence and the ways to measure and weight nontraditional or nonfinancial audit evidence remain uncertain. In response to these issues, the goal of this dissertation is to examine if nontraditional audit evidence can deliver relevant and reliable information to auditors and to suggest ways to measure and weight it.

The first essay investigates the value of the optimistic tone of management qualitative disclosures in 10-K and 8-K filings, including press releases, on initial audit fees and changes in audit fees, based on prior studies about the significance of the optimistic tone of management qualitative disclosures in predicting a firm's future performance and identifying management fraudulent behaviors. The empirical results show that the optimistic tone of qualitative information from 10-K and 8-K reports is negatively associated with successive audit fees. In addition, the association between the optimistic tone and audit fees is changed for firms that receive going concern opinions, indicating that auditors respond differently to the optimism of management when they audit high-risk clients.

The second essay describes that substantive analytical procedures (SAPs) have the potential to provide a high level of assurance for revenue accounts and can be especially useful in examining revenue accounts since nontraditional audit evidence is often needed as an independent benchmark to verify revenue accounts and the population of underlying transactions tends to be extremely large. However, audit firms tend to focus more on tests of details, such as audit sampling in substantive tests of details, in recent years in order to avoid possible negative outcomes from PCAOB inspections caused by moderate or weak SAPs. An examination of the existing literature suggests that audit sampling in substantive tests of details and SAPs are often complementary, even if SAPs do not offer high assurance. In some situations, either one could be more effective. Therefore, this paper argues that the auditor must consciously examine the factors affecting the effectiveness and efficiency of substantive tests before selecting the appropriate procedures to improve audit quality.

The third essay examines how weather variables play an important role in improving the effectiveness of SAPs for revenue accounts. Prior studies in economics, marketing, and finance show the influence of weather on sales. Specifically, unfavorable weather conditions are likely to hinder customers' store visits, thereby decreasing sales. Thus, the models proposed in this study are tested by using daily and weekly aggregated sales revenue accounts from a multi-location retail firm with homogeneous operations in the US. Since the influence of weather on sales varies depending on seasons and regions, appropriate ways to integrate the weather variables in the proposed models are suggested. The empirical results indicate that weather variables have less value in forecasting store-level sales accounts than selected peer stores sharing similar macroeconomic characteristics but provide incremental values in improving error detection.

ACKNOWLEDGEMENTS

"If we knew what it was we are doing, it would not be called research would it?" - Albert Einstein

I would never have been able to finish my dissertation without the guidance of my committee members, assistance from friends, and support from my family.

I would like to express my deepest gratitude to my advisor, Dr. Miklos A. Vasarhelyi. He brought me to academia and encouraged me to explore on my own. When my step faltered, he often told me the quote above. Without his trust and consistent support, I could not overcome the many crisis situations during my PhD studies. I would like to thank to Dr. Alexander Kogan, who continuously found time to discuss issues with me and patiently guided me. I am deeply grateful to him for helping me to sort out the technical details of my work. I am also very grateful to Dr. Carolyn B. Levine. She patiently supported my development and exploration of research ideas. In particular, she led me to develop the technical skills in my early PhD years, which are a significant basis for my dissertation.

I would like to thank Dr. Theodore Mock and Dr. Kevin Moffitt for serving on the committee and supporting me in improving my dissertation with their insightful comments. I appreciate Lucas Hoodguin for his thoughtful comments as an audit practitioner. Also, I deeply appreciate Dr. Michel Alles, who always read my work and gave me valuable comments. I am grateful to Dr. Ann Medinets, Dr. John Peter Krahel, Dr. Paul Byrnes, and Kate Thomas for their help with the manuscript.

My dissertation would not have been possible without support from the large accounting firm and the chance to work and talk with audit practitioners. Especially, I would like to express my gratitude to Tim Pearce for his kind support.

I would like to acknowledge my gratitude to Dr. Dan Palmon for his consistent support as department chair. I would like to specially thank to my friends and colleagues at Rutgers Business Schools, Dr. Victoria Chiu, Dr. Qi Liu, Dr. Hussein Issa, and Dr. Khrystyna Bochkay et al.

Most importantly, none of this would have been possible without the love and patience of my family. To Dr. Dimitri Lazo, my former professor, I would like to express my warmest appreciation for his help. Particularly, for my husband, Ju Myung Song and my mom, Jong Ok Kim, words cannot describe how grateful I am for their support and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

Even though auditing standards (e.g., AICPA 2012a) require that auditors use nonfinancial information as audit evidence, whether nonfinancial information is utilized as primary audit evidence remains uncertain (e.g., Brazel et al. 2014). Accordingly, in this dissertation, unorthodox or nontraditional audit evidence is defined as nonfinancial information such as analytical evidence relating to nonfinancial performance indicators (Eilifsen et al. 2001). Traditional audit evidence, on the other hand, is defined as financial measures.

My research interest is in the role of nontraditional audit evidence on audit procedures. Especially, by analyzing various forms of nonfinancial information, such as management qualitative disclosures and weather variables and by studying different stages of audit procedures, such as risk assessments in pre-engagement stage and substantive tests, the significance of nontraditional audit evidence is examined.

This chapter introduces the motivation and method of this thesis and provides a literature review of the related concept of nontraditional audit evidence. The chapter two examines whether the tone of qualitative management disclosures plays a role in initial audit fee decisions, the chapter three studies the significance of substantive analytical procedures developed from nonfinancial information for auditing revenue accounts, and the chapter four explores the form of nonfinancial information, weather variables, as audit evidence to improve substantive analytical procedures for revenue accounts. The last chapter concludes the dissertation by summarizing the findings, discussing the limitations, and pointing out future research areas.

AUDIT EVIDENCE

Audit requires the collection and evaluation of audit evidence. Audit evidence is defined as “all of the information used by auditors in arriving at the conclusions on which audit opinion is based” (AICPA 2006a). Audit evidence is obtained generally through interviewing entity personnel, inspecting accounting records, the auditor’s knowledge, and testing documents (Louwers et al. 2013). Collected information should result in sufficient, appropriate audit evidence (AICPA 2006a). The sufficiency of audit evidence is related to the auditor’s judgment regarding whether the information is enough to verify management assertions, and appropriate audit evidence should be reliable and relevant.

The forms of audit evidence classified by auditing standards (AICPA 2006a) as follows:

1) Inspection of tangible assets

Inspection of tangible assets is the examination of the existence and quantity of the firm’s tangible assets. Physical examination is valuable for verifying the occurrence of production operations or acquiring goods (Whittington and Pany 2001). Even though the physical examination of tangible assets provides highly pervasive audit evidence, it would not offer sufficient evidence in order to verify rights and obligations, or support the valuation of these assets (AICPA 2006a).

2) Inquiry

Inquiry results in audit evidence obtained by interviewing people within or outside of the entity who have knowledge about transactions and operations. Although inquiry is widely used as primary audit evidence, management inquiry is generally considered as less reliable audit evidence to verify internal control or material misstatement at the assertion level, compared to information obtained from outsiders.

3) Inspection of records or documents

Inspection of records or documents offers audit evidence to confirm authorization, the existence of assets, and the effectiveness of internal controls. The reliability of this type of audit evidence largely depends on whether it is internal or external, and if it is internal, then the level of internal controls (Louwers et al. 2013). In addition, the way to obtain audit evidence (direct or indirect), the type of records (original documents or photocopies), and the form of documents (paper, electronic, or other medium) affect the evaluation of this type of audit evidence (AICPA 2006a).

4) Confirmation from a third party

Conformation is "the process of collecting a representation or of existing condition directly from third party" (AICPA 2006a). Even though confirmation is often used to address the existence assertion for accounts receivable and is considered as the best source of evidence because it is come from external entities, compared to internal documentation, such as sales invoices or bill of landing, the effectiveness of confirmation largely depends on confirmation response rates (Johnson et al. 1981).

5) Analytical procedures

Analytical procedures are the process to evaluate financial information by examining plausible relationships among both financial and nonfinancial data. They can be conducted by simple scanning, which requires the auditor's professional judgment to identify unusual fluctuations or items within account balance, transitions, subsidiary leger, reconciliation and other detailed reports (AICPA 2006a). Alternatively, analytical models, such as ratio tests or regression analysis, can be used to set the expectations for an account or class of transactions. Comparing to other audit evidence, such as recalculation, confirmation, and inspection of documents, analytical procedures are often considered to be “soft” audit evidence (Louwers et al. 2013). On the other hand, some studies argue that the level of accuracy of expectations generated by analytical models determines whether analytical procedures are “soft” or “hard” evidence (e.g., Loebbecke and Steinbart 1987).

6) Recalculation and reperformance

Recalculation, the process of checking the mathematical accuracy of documents or records, and reperformance, the auditor’s independent execution of procedures or controls, are often facilitated by information technology (AICPA 2006a).

BUSINESS RISK AUDIT (BRA)

Business risk audit (BRA) methodologies have been adopted since the 1990s by large audit firms (Bell et al. 2008). Prior to the adoption of BRA, approaches transaction based audit (TBA) approaches dominate audit practice. The TBA, a bottom-up approach to auditing, begins with transactions or events and moves up to the financial statements (Bell et al. 1997). Under the TBA approaches, the auditor concentrates on risks related to the account balances, on the class of transactions, and on the client’s accounting systems (Bell et al. 1997) and relies mainly on tests of details.

Since many important business risks have relationships to audit concerns (Eilifsen et al. 2001), the BRA approach considers the dynamic environments of the client and uses this understanding as the basis for inspecting financial statement risk (Robson et al. 2007). Therefore, under the top-down BRA approach, the auditor needs an understanding of the client's business risks, as measured by the client's business strategies and business models (Knechel 2007).

Utilizing the BRA might be beneficial to the auditors because: 1) the audit environment has become more complicated (e.g., a client's business model has been changed or a client's senior management is likely to involve financial statement fraud) (Peecher et al. 2007); 2) it might reduce audit costs (Knechel 2007); and 3) it enhances the recognition of management fraud (e.g., external sources provides an independent benchmark to evaluate management assertions (Messier et al. 2013a)).

Especially, the BRA approach influences the forms and volume of audit evidence since it encompasses various forms of audit evidence and requires complex and comprehensive risk assessments and the reduction of audit efforts in accord with such assessed risks (Bell et al. 2008). For instance, under the TBA approach confirmation from the third party is widely used audit evidence, but under the BRA approach analytical procedures developed from the auditor's understanding of the client and its environment are significant audit evidence (Bell et al. 2005). Along these lines, Bell et al. (2005) define the concept of evidentiary triangulation as incorporating evidence from multiple sources. Specifically, the auditor collects audit evidence from three sources, such as business strategy, management information intermediaries, and management business representations. Business strategies include the firm's stated strategies, its process, and economic events. Management business representations are related to audit evidence from sources like accounting journals, ledgers, financial statements, MD&A, and press releases. Management information intermediaries cover the firm's internal controls and information systems (Peecher et al. 2007).

A number of studies examine the significance of synthesizing audit evidence from multiple sources, such as assessing fraud risks (Trotman and Wright 2012), auditing complex estimates (e.g., fair value estimate or impairments) (Griffith et al. 2015a; Griffith et al. 2015b), and evaluating material misstatements (Schultz et al. 2010 ; Kopp and O'Donnell 2005). On the other hand, O'Donnell and Schultz (2005) indicate that favorable comprehensive evaluations regarding a client's business risk assessments lead to an overly positive assessment of detailed account-level information.

MOTIVATIONS, RESEARCH QUESTIONS AND RESEARCH METHODS

Prior studies have examined the value of nonfinancial or nontraditional measures, such as the number of employees, the size of stores, or customer satisfaction rates, as independent benchmarks for evaluating financial statements (Bell et al. 2005; Ittner and Larcker 1998; Knechel 2007). In addition, auditing standards (e.g., AICPA 2012a) highlight the use of nonfinancial information. Along these lines, some studies examine the significance of nontraditional information as audit evidence but find that auditors are not likely to rely on nonfinancial information as primary audit evidence (Brazel et al. 2014; Cohen et al. 2000) since auditors struggle to use it during their judgment processes (Cohen et al. 2000; Trotman and Wright 2012). Luft (2009) challenges the use of nonfinancial information due to difficulties with precise measurement and appropriate weighting. Accordingly, it remains uncertain whether nontraditional audit evidence is primarily used for audit decisions and delivers value in improving audit quality, and how to measure it precisely and weight it appropriately. This dissertation is an attempt to respond these issues.

Previous literature suggests that the tone of management qualitative disclosures is one of the indicators that can help auditors to understand a firm's potential future performance and management fraudulent behaviors (Li 2006; Rogers et al. 2011), which are known to be factors that auditors consider when they assess risks in the pre-engagement stage. (Krishnan and Krishnan 1997; Morgan and Stocken 1998). Numerical information regarding business risks are highly correlated with each other, but management qualitative disclosure might provide independent, forward looking information or risk factors not captured by historically based numerical information (Li 2006). Generally, the optimistic tone of management qualitative disclosures is related to the firm's positive future performance. On the other hand, since "overly optimistic" management disclosures are considered as red flags (AICPA 2002), the discrepancy between management's perceived risks and auditor's perceived risks might discourage the auditor from relying on the optimistic tone of management disclosures. Since going-concern judgments are likely to be consistent with the engagement partner's evaluations (Wilks 2002), a successor auditor's going-concern opinion regarding whether a client can continue operations in the next twelve month might suggest very high business risk to the successor auditor in the pre-engagement stage. Accordingly, as an indicator of the auditor's perceived risk, a successor auditor's going concern opinion is used. Based on this, chapter two explores whether there is association between optimistic management qualitative disclosures and audit fees in initial engagements and whether the negative association between optimistic management qualitative disclosures and initial audit fees is weaker for firms that receive going concern opinions.

To explore the influence of publicly available management disclosures on initial audit fee decisions, this dissertation looks at 696 initial audit fee decisions from 2010 to 2013. Management qualitative disclosures from 10K and 8K filings from one year prior to the dismissal/resignation date to one day prior to the date when the predecessor auditor is dismissed or resigned are collected from SEC EDGAR. To capture the level of optimism in these disclosures, positive and negative words are counted, based on Financial Sentiment Dictionaries (Loughran and McDonald 2011), and optimism is measure by the difference between positive words and negative words divided by the sum of the positive and negative words. The measured tone of optimism is the included in the traditional audit fee model suggested by Hay et al. (2006).

Chapter three responds to the concerns of the Public Company Accounting Oversight Board (PCAOB) regarding the quality of substantive analytical procedures (SAPs). Recently, the PCAOB highlighted the problem of insufficient SAPs for large value income statement accounts, such as the revenue account (e.g., PCAOB 2014) due to its concern that if SAPs cannot provide a high level of assurance (the difference between the auditor's expected value and the account balance is smaller than the performance materiality) then SAPs, in fact, provide no assurance (PCAOB 2011). Prior studies regarding substantive tests show that SAPs detect risky areas even if they offer only moderate or weak assurance, resulting in more tests of details in these areas (e.g., Knechel 1988) and utilizing nonfinancial information in analytical procedures is beneficial to identify financial statement fraud (Brazel et al. 2009; Brazel et al. 2014). Nonetheless, in response to the PCAOB's notices, the auditor is likely to avoid utilizing SAPs and employs tests of details, such as audit sampling in substantive tests of details

(Christensen et al. 2015) instead because it is difficult to set precise expectations for revenue accounts (Glover et al. 2015) and because it is easier to document to audit sampling than SAPs (Trompeter and Wright 2010). Similarly, Christensen et al. (2015) suggest that the costs and benefits of using sampling in substantive tests of details instead of SAPs should be examined for accounts like revenue. This essay attempts to respond to the uncertainty about the significance of SAPs for identifying misstatements by examining the cost and benefits to the auditor from utilizing sampling without SAPs developed with nontraditional information and whether SAPs can offer a high level of assurance for revenue accounts.

First, by exploring prior studies dealing with SAPs and sampling that indicates related risks, this essay illustrates that cases where SAPs are more effective than audit sampling and *vice versa* to describe costs and benefits of audit sampling as a substantive test without conducting SAPs developed with nontraditional information. Second, this essay conducts a meta-analysis of prior studies with the outcome of SAPs for revenue accounts to examine whether SAPs could be persuasive audit evidence providing high levels of assurance.

The final essay in chapter four suggests relevant and reliable audit evidence for revenue accounts. Studies in economics, marketing, and finance indicate that weather influences sales in certain industries, such as retailers. In particular, these studies show that unfavorable weather conditions may negatively affect retail sales. In addition, unlike other external information, such as gross domestic product (GDP), weather variables are updated often, so the auditor can use this data during the audit. Accordingly, weather variables can be relevant audit evidence for retailers. Since weather variables are not affected by management, they could actually be more reliable audit evidence than internal sources. Accordingly, weather variables have meaningful potential as audit evidence to verify a retailer's revenue accounts. This study examines the correlation between weather variables and the sales revenue account and the incremental value that weather information offers to enhance the performance of SAPs.

In order to test these issues, this essay examines store-level sales revenue for a publicly-held retailer operating in a large number of locations around the US. Since this firm tries to provide homogeneous service across all stores, this setting is useful to measure the influence of weather on sales. Weather variables are measured different ways: heating degree days and cooing degree days, and the apparent temperatures developed from the heat index and the wind chill index. As a control variable, the average sales amount of peer stores is used (Allen et al. 1999). Peer stores are selected for each store based on location and annually updated macroeconomic indicators that are highly correlated with sales. It is assumed that the selected peer stores share not only similar macroeconomic characteristics related to sales but also firm-wide reputations and industry competition. Two types of statistical models are developed: multivariate regression models and times series models. Each statistical model is developed with and without weather variables and a variable generated by peer stores. Since the relationship between weather and sales is nonlinear and modified depending on regions and seasons, different approaches are tested, such as transformation of weather variables (e.g., centering data) and utilizing different statistical approaches like polynomial regression models and stepwise regression models. As in prior studies, the expectations set by these models are measured by the accuracy of expectations and the rate of false positives and false negatives.

The contribution of this dissertation is to investigate the impact of nontraditional audit evidence on the audit decisions and the audit quality. Even though some audit evidence, such as management qualitative disclosures, is commonly used in audit practice during risk assessments in the pre-engagement stage, academia rarely examines the influence of such factors on audit fee decisions. On the other hand, although some audit evidence such as SAPs developed with nonfinancial or nontraditional information are recommended to identify misstatements in academia, audit firms are likely to avoid using it because of the PCAOB's inspections. Furthermore, although weather information has great potential to be reliable and relevant audit evidence for a retailer's revenue accounts, it is unknown whether it is used as audit evidence in audit practice and is rarely examined in academia. Therefore, the contribution of dissertation consist of filling the gaps these discrepancies created between academic findings and audit practice and suggesting new types of audit evidence and approaches to measure and evaluate it.

The remainder of this dissertation is as follows: The three essays are contained in chapters two, three and four. The last chapter summarizes the findings, discusses the limitations of this work, and points out possible areas for future research.

Chapter 2 Initial Audit Fees and the Tone of Management Qualitative Disclosures

INTRODUCTION

This study investigates the effects of textual information found in 10-K and 8-K filings on audit fees for initial engagements, and whether these effects vary with auditors' perceived risk levels. This is a direct extension of the audit pricing model. Whereas prior studies often measure clients' business risks by using financial proxies (e.g., return on asset (ROA), leverage, earnings, and the current ratio), this study focuses on indicators from textual information. In particular, this study examines two main issues: 1) The extent that successor auditors use management qualitative disclosures when they engage new clients; and 2) Whether the level of a client's business risks alters the fee premium associated with management qualitative disclosures.

Previous literature provides evidence that clients with high-risk businesses are likely to pay high audit fees (O'Keefe et al. 1994; Lyon and Maher 2005; Venkataraman et al. 2008). In spite of the growing body of evidence regarding the determinants of audit fees and the role of textual information in the market, few studies have attempted to combine these research streams.

A major motivation for this idea is found in audit practice. Audit firms often conduct background checks by evaluating a variety of sources, such as news articles and microeconomic variables (Louwers et al. 2013). In addition, in the last decade, audit firms have started to utilize a business risk approach that requires a comprehensive knowledge of a client's industry and strategy (Trompeter and Wright 2010).

This study proposes that management's qualitative disclosures provide incremental evidence to assess a firm's future performance. A growing volume of accounting and finance research uses textual analysis to understand the extent that qualitative information is related to stock price, analysts' behavior, and future earnings (Li 2006; Lehavy et al. 2011; Merkley 2013). The results suggest that management's textual disclosures provide additional evidence to help predict a firm's future performance and optimistic (pessimistic) disclosures tend to have a relationship to positive (negative) future outcomes. Accordingly, this study hypothesizes that firms with pessimistic management disclosures might be assessed higher audit risk premiums.

On the other hand, since auditing standards (AICPA 2002) and Roger et al. (2011) suggest that overly optimistic management disclosure behavior is a possible fraud indicator, in certain cases the auditors might not consider optimistic management qualitative disclosures as an indicator of low audit risks premiums. In addition, whether auditors recognize this inconsistency between auditors' perceived risks and management assertion is related to auditor professional skepticism (Feng and Li 2014). Consequently, this study tests whether the association between audit fees and qualitative disclosures are modified by the level of auditor's perceived client business risks. As an indicator of auditors' perceived client business risks, going-concern judgment of successor auditors is used since going-concern judgments are likely to be consistent with the engagement partner's evaluations (Wilks 2002).

Unlike most other external stakeholders, auditors have access to a firm's internal information, including strategies, decisions, and disaggregated financial information. Consequently, their dependence on qualitative information might be limited, especially in ongoing engagements with established clients. To understand the influence of external qualitative information, this study tests audit fee decisions for initial engagements since successor auditors might not possess sufficient internal client information in the pre-engagement stage.

Evaluating the influence of qualitative management disclosures on audit fee decisions is based on enhanced versions of the audit fee model that add quantified textual information to the financial variables in the original models (Hay et al. 2006). The language from various textual disclosures is quantified as the difference between positive words and negative words divided by the sum of positive words and negative words in an effort to understand the auditor's assessment of the client's risk.

Results are generally consistent with the hypothesis of this study. The tone of qualitative disclosures tends to be influential in explaining the level of initial audit fees after auditor replacement, indicating that successor auditors make efforts to overcome information asymmetry in new engagement settings.

However, this association between initial audit fees and optimistic management qualitative disclosures is differentially modified depending on whether the textual information comes from the 10-K filings or the 8-K filings of high-risk clients. Auditors of risky clients are likely to reduce audit fee premiums on optimistic 10-K filings, but to increase audit fee premiums on optimistic 8-K filings. These results may indicate that auditors of high-risk clients analyze the tone of qualitative disclosures in each filing for different purposes. Auditors of high-risk clients use the tone of 8-K filings to evaluate client business risks, but use the tone of 10-K filings to detect possible fraudulent management behavior. These differences might be related to the characteristics of each filing. Because of strong regulatory requirements, management is likely to disclose more credible qualitative information in 10-K filings to avoid possible litigation, thereby keeping a certain level of pessimistic tone in these filings (Davis and Tama-Sweet 2012). Accordingly, auditors of high-risk clients may consider overly optimistic 10-K reports as "red flags", and reflect this concern in higher audit fees. On the other hand, 8-K filings are considered to be more flexible and strategic than 10-K filings. For example, market participants are likely to respond more to press releases on 8-K filings than to the Management Discussion and Analysis (MD&A) in 10-K or 10-Q filings (Levi 2008). Similarly, auditors might weight on the tone of 8-K filings to evaluate client business risks for high-risk clients.

The findings in this paper relate to other research on the implications of qualitative information on a firm's risk evaluation. Despite extensive literature on the usefulness of qualitative information in predicting future firm performance, little is known about whether external auditors, who also assess their clients' business risks, perceive these qualitative sources as legitimate risk indicators.

In addition, this paper contributes to prior research with regard to audit fee decisions and risk evaluations. First, it provides additional variables related to developing a theory of audit fees. In the finance literature, some scholars argue that textual variables can offer additional unique and independent variables to existing models with quantitative proxies (Li 2006). Similarly, Hay et al. (2006) argue that numerical variables which have suggested in prior studies are likely to be inconsistent with the audit fee model because it has inherent issues, such as inadequate control variable proxies and omitted variables. The approach presented here provides new insights into the conditions under which audit firms may consider quantitative factors when developing their risk management strategies.

Second, understanding auditing decisions is important in generating realistic theories of auditor behavior (Gibbins and Newton 1994). Although audit firms tend to operate with a comprehensive audit approach (Trompeter and Wright 2010), prior research largely develops audit fee models based exclusively on limited financial factors and models, and few studies include qualitative factors. This essay shows that auditors' risk management strategies are not limited to numerical information.

The remainder of this essay is divided into six sections: the next section summarizes the literature, the third part describes the research design and sample selection, the fourth section explains the results, the fifth illustrates a robustness test, and the sixth contains additional analysis. Finally, the section seven provides conclusions and limitations.

RELATED LITERATURE AND HYPOTHESIS DEVELOPMENT

Theory

Many studies have explored the fundamental audit pricing model first proposed by Simunic (1980). The model states:

$$E(C) = cq + E(d)E(\theta),$$

Where $E(C)$ is expected total cost to the auditor; c is the per unit factor cost of external auditor resources (including both explicit and implicit costs), q is the quantity of resources used by the auditor during audit procedures, $E(d)$ is the expected present value of future losses that may arise from a particular periodic audit, and $E(\theta)$ is the likelihood that the auditor will have to pay for those losses.

As Simunic (1980) suggests, audit fees are primarily comprised of two factors: 1) Effort (cq); and 2) Expected loss (E(d)E(θ)). These two factors are negatively correlated, so expected losses tend to decrease as audit production or effort increases. Certain aspects, such as client size and complexity of operations, can also be linked to audit fees because those factors cause auditors to use more efforts in performing audit procedures. Finally, expected future losses contain costs that may occur from actual or threatened litigation, as well as damaged reputation (Seetharaman et al. 2002).

Client Business Risk and Audit Fee

The audit risk model in SAS No.107 (AICPA 2006b), used in the planning stage, provides a framework for evaluating the risks of issuing unqualified opinions on financial statements that are materially misstated. Business risks are generally considered to have two components: 1) Client business risk, which is related to the client's continued existence; and 2) Auditor business risk, which is the risk of possible litigation and other potential costs related to audit failures (Colbert et al. 1996; AICPA 2006b; Ethridge et al. 2011). High business risk prompts auditors to change fee premiums in an effort to cover potential future losses and/or allow for greater audit effort (Simunic 1980; Pratt and Stice 1994; Seetharaman et al. 2002).

Generally, the elements of the audit risk model (i.e. inherent risk, control risk, and detection risk) are changed by business risks (Brumfield et al. 1983). When client business risk is high, financial statements are more susceptible to material misstatement. In such cases, a client who does not have sufficient resources to provide reliable reporting might be more likely to manipulate financial reports to hide poor performance (Stanley 2011), and auditors might face increased higher litigation and audit failure risks

(*E(d)E(θ)*). To overcome these heightened risks, auditors are motivated to increase audit effort (*cq*) so as to attain a tolerable level of audit risk, which necessitates a higher audit fee premium (*E(C)*). Prior literature offers empirical evidence to support the argument that client business risk is linked to variances in audit fee levels, and to measure such risks, financial proxies such as financial condition and/or stock price variability are often adopted (Stice 1991; Choi et al. 2004). Nevertheless, the use of these proxies to measure client business risk may paint an incomplete picture. According to SAS No. 109 AU Section 314 (AICPA 2006c),

> "Business risk is broader than the risk of material misstatement of the financial statements, although it includes the latter. For example, a new entrant to the marketplace with the competitive advantage of brand recognition and economies of scale may represent a business risk to a manufacturer's ability to garner as much shelf space at retailers and compete on price."

To avoid audit failure and reduce risk, audit firms are progressively adopting a more comprehensive business risk audit approach (Trompeter and Wright 2010). This entails gaining an understanding of client business strategies and industry characteristics (Bell et al. 2008). Similarly, Krishnan et al. (2012) and Chen et al. (2012) find an inverse correlation between management's earnings forecasts and audit fees. They report that optimistic earnings forecasts are related to lower audit fees and lower probability of auditor resignation. Bentley et al. (2013) find that client business strategy (e.g., product differentiation or competitive cost) is associated with financial statement irregularities and higher audit fees.

Auditor Changes and Information Asymmetry

There are two ways for firms to change auditors: auditor-initiated resignation and client-initiated dismissal. Client-initiated change is largely related to audit fees (Ettredge et al. 2007; Griffin and Lont 2010). However, the causes of auditor resignation are often unclear. Prior studies identify several factors related to auditor resignation, including client corporate governance structure (Lee et al. 2004), auditor industry specialization (Cenker and Nagy 2008), litigation risks (Johnstone and Bedard 2004; Zhan Shu 2000), and deteriorating financial condition and enhanced litigation risk (Krishnan and Krishnan 1997).

Auditors and clients often conclude engagement negotiation, including audit fees, before starting audit procedures. Once fees are decided, they usually remain unchanged for the fiscal year (Hackenbrack and Hogan 2005). When audit fees are negotiated with clients, auditors do not know the exact risk levels. Accordingly, in the pre-engagement stage, external auditors investigate new clients and examine changes in the circumstances of existing clients by collecting relevant information about the firm (Bell et al. 2002). Since their audit fee decisions, especially the first year of audit, might incorporate their lack of prior knowledge, they might collect a wider variety of information. This paper focuses on new engagements in order to examine the influence of qualitative information from management more clearly.

Qualitative Information and Client Business Risk

In the evolving information age, accessing textual information incurs progressively lower cost due to technological advances that facilitate this process. Consequently, researchers have begun to consider how textual information plays a role in marketplace decisions. For instance, Kothari et al. (2009) analyze the content of various types of textual information, such as news and analyst reports, as well as footnotes and MD&A in 10-K forms. They find that when content analysis yields favorable (unfavorable) information, the firm's risk, as measured by cost of capital, stock return volatility, and analyst forecast dispersion, declines (rises). These results imply that market participants, including creditors, analysts, and shareholders, recognize textual information and consider it to be relevant.

Previous studies addressing qualitative information provide new factors that cannot be perceived from existing variables (Levine and Smith 2011). News from management or journalists can provide additional evidence relative to major concerns, such as the risk of material misstatements (Larcker and Zakolyukina 2012) and aggressive financial reporting behavior (Patelli and Pedrini 2013). Most studies in this domain examine whether textual information has the power to predict future stock returns and/or earnings.

Whereas financial statements are subject to generally accepted accounting principles (GAAP), textual information in 10-K reports is not. Moreover, under the Private Securities Litigation Reform Act of 1995 (PSLRA), which establishes a legislative safe harbor protection from liability for voluntary disclosures of financial projections and other forward-looking information, management has more freedom to formulate its voluntary disclosures. However, the SEC and federal courts have treated forecasts and general expressions of optimism as actionable under federal securities law (Palmiter 2008). Accordingly, management may have strategic intentions when placing information in annual reports, so analyzing this information could prove useful for understanding the implications of these disclosures. Textual information in financial statements might be more suitable for estimating future stock price because, unlike numerical data in financial statements, textual information (e.g., MD&A) often discusses customer demands or market competition (Li 2010). In this sense, textual information is considered more forward-looking and therefore more valuable for predicting future business performance. Li (2006) shows that the risk sentiment (i.e. risk, risks, and risky) expressed in annual reports can predict future returns in a cross-sectional setting. More specifically, during the year following the annual report filing date, entities with large increases in risk sentiment experience significantly more negative returns than organizations with small increases in risk sentiment.

The SEC requires that firms submit 8-K filings no more than four days after a material event such as acquisition of assets, bankruptcy, or changes in management or the board of directors. Item 2.02 of 8-K filings, Results of Operations and Financial Conditions, covers public announcement or press releases regarding material events resulting from a firm's operations or financial conditions. Ma (2012) finds that firms' stock returns react to the language of 8-K filings. Particularly, because Item 2.02 relates to voluntarily disclosures and is often disclosed in narrative form, some studies examine whether the linguistic tone of press releases is informative to shareholders. For instance, by measuring approximately 20,000 earnings press releases issued from 1998 to 2003, Davis et al. (2012) examine whether managers employ positive (negative) linguistic styles when issuing earnings press releases regarding expectations of future firm performance. The authors find a significant positive (negative) association between level of optimistic (pessimistic) tone in earnings press releases and future ROA. This leads to the following hypothesis:

H1: There is negative association between optimistic linguistic tone in management disclosures and audit fees in an initial engagement, *ceteris paribus*.

Qualitative Disclosures, Management Strategic Decisions, and Auditors' Going Concern Opinions

Auditors must evaluate whether clients are able to continue their business for one year beyond the financial statement date. This relates to their evaluation of client business risks[1]. When auditors consider a client as high-risk, does this judgment affect the association between the tone of qualitative disclosures and audit fee decisions?

Optimistic management disclosures do not always relate positively to a firm's future performance. Even though management faces litigation risks from stakeholders for overly optimistic disclosures, it is possible for management to manipulate financial statements for its own benefits. In this context, Rogers et al. (2011) suggested that overly optimistic management disclosure behavior is a potential fraud indicator. If management is optimistic in its filings, but the auditor considers the client to be risky, how is this inconsistency reflected in the audit fee decision? This inconsistency between management assertions and audit evidence is related to the auditor's professional

[1] Going concern opinions are issued after audit procedures are finished, but audit fee negotiations are concluded before audit procedures are performed. To link going-concern opinion and client business risks reflected in audit fee decisions requires a strong assumption that auditors fully anticipate their clients' high business risks before they assess the clients' internal information. Nevertheless, this assumption is also used in audit fee models. In addition, Wilks (2002) found that going-concern judgments tend to be consistent with the engagement partner's evaluations. Thus, audit fee models are generally developed using financial indicators from the fiscal year in which the auditors accept their clients, so the audit fee models may not account for possible errors in the auditors' evaluations. This issue will be discussed in the robustness test later in this paper.

skepticism, which refers to the auditor's attitude not to accept management assertions without validation. Professional skepticism plays an important role in audit practice because of its relevance to audit failures. Consequently, an auditor may consider management disclosures to be less credible if the auditor believes that there is a high possibility that the client may go bankrupt, but management anticipates positive future performance.

In summary, when auditors negotiate audit fees, they may rely differentially on management's qualitative disclosures depending on whether they perceive the client to be high-risk or low-risk. To identify risk factors, they might rely more on information from qualitative disclosures for high-risk clients than for low-risk clients. However, it is an empirical question whether they weight on the information from the qualitative disclosures for high-risk clients in the opposite direction from the tone of management in disclosures because they might think the tone is less credible. Consequently, this leads to the following hypothesis:

H2: The negative association between the optimistic tone in management qualitative disclosures and audit fees will be weaker for firms receiving going-concern opinions, *ceteris paribus*.

RESEARCH DESIGN

Sample Selection

Table 2.1 describes the sample selection procedure. The procedure begins with firms that changed auditors for the fiscal years 2010 to 2013 in Audit Analytics (n = 3,913). Only 1,030 of these firms have the needed financial information within Compustat and audit fee information within Audit Analytics. 137 of the 1,030 retained firms are foreign firms, and another 28 firms do not include business segment information. Firms in financial industries (SIC codes 6000-6999), as well as those missing either 10-K filings or audit opinion information are excluded as well. Among the remaining 715 observations, 19 firms (about 2.7%) filed no 8-Ks during the targeted period and were removed. Thus, the final sample consists of 696 observations.

Table 2.1 Sample selection

Firms changing auditors for the fiscal years 2010 - 2013 in AuditAnalytics	3,913
Less: Financial variable data or audit fee data missing in Compustat and AuditAnalytics	-2,883
Less: Foreign firms	-137
Less: Variable data missing in segment data	-28
Less: Financial industry	-30
Less: 10-K filings missing in EDGAR (e.g., S-1)	-42
Less: Audit opinion data missing in AuditAnalytics	-78
Less: No 8-K filings for the targeted period	-19
Total	696

This model examines whether textual information is an *ex ante* indicator for audit fees for new engagements. As Figure 2.1 presents, the disclosures filed in the EDGAR database for the period between a year plus one day before the dismissal/resignation date and one day before the dismissal/resignation date are downloaded. HTML tag information is deleted, and all remaining textual information in the 10-K and 8-K documents is used for analysis as in Li (2006). 8-K filings often contain exhibits. For example, Table 2.2 shows, firms generally announce the incidence of press release in 8-K filings and then the contents of the press release are disclosed separately in exhibits. Accordingly, in this paper not only 8-K filings but also related exhibits are analyzed. Basic statistics with regard to 8-K filings are described in Panel D of Table 2.3 and the average number of 8-K reports is 12.20.

Table 2.2 Example of 8-K filings

Panel A: Example of 8-K filings

Item 2.02 Results of Operations and Financial Condition

On November 6, 2012, Akorn, Inc. issued a press release announcing financial results for the quarter ended September 30, 2012. A copy of the press release is furnished as Exhibit 99.1 to this report.

The information in this report, including the exhibit hereto shall not be deemed to be “filed” for purposes of Section 18 of the Securities Exchange Act of 1934 (the “Exchange Act”) or otherwise subject to the liabilities of that section, nor shall it be deemed to be incorporated by reference in any filing under the Securities Act of 1933 or the Exchange Act, except as shall be expressly set forth by specific reference in such a filing.

Panel B: Example of Exhibits with 8-K filings

Akorn Reports First Quarter 2013 Financial Results

- Reports Record Revenue of $73.9 million and Adjusted EPS of $0.13 -

LAKE FOREST, Ill.--(BUSINESS WIRE)--May 7, 2013--Akorn, Inc. (NASDAQ: AKRX), a niche generic pharmaceutical company, today reported financial results for its first quarter ended March 31, 2013.

Raj Rai, Chief Executive Officer, commented, "We are pleased with our first quarter results although we were behind in the launches of certain products that were approved late last year due to capacity constraints with our contract manufacturing partners as well as market challenges. We expect resolution sometime in the second half of this year. We are also excited about the establishment of our new R&D center in Vernon Hills, Illinois. The new center, with its added capacities and capabilities, will make it possible for us to file 35 to 40 ANDAs with the USFDA from our US and India facilities starting next year, which is the cornerstone of our long term growth strategy."

First Quarter 2013 Highlights

- Achieved record consolidated revenue of $73.9 million, up 43% over the prior year quarter.
- Received FDA approval for 2 ANDAs, Naphazoline Hydrochloride 0.025% with Pheniramine Maleate 0.3% and Clindamycin Phosphate Injection in 5% Dextrose premix in three strengths, with a combined IMS addressable market size of $90 million.
- Filed 4 ANDAs and completed the development on an additional 2 ANDAs with a combined annual IMS market size of approximately $640 million.
- Completed modernization and the first phase of capacity expansion at the Company’s Somerset, New Jersey ophthalmic manufacturing plant.
- Opened a new, 19,000 square foot research and development center in Vernon Hills, Illinois; designed to accommodate 35 to 40 ANDA filings per year and expand into the development of specialty formulations such as carbapenems, hormones and oncolytics.

Financial Results for the Quarter Ended March 31, 2013

Consolidated revenue for the first quarter of 2013 was $73.9 million, up 43% over the prior year quarter’s consolidated revenue of $51.7 million. The increase in consolidated revenue was driven by the sale of new products launched late in 2012, organic growth of established products and products re-launched in prior periods, and a full quarter’s sales generated by Akorn India. Consolidated gross margin for the first quarter of 2013 was 53.0% compared to 59.8% in the comparable prior year period. The decrease in gross margin was primarily the result of lower margins from Akorn India, which began operations upon completion of the Kilitch acquisition on February 28, 2012, as well as the impact of various new products launched late in 2012 which generate lower gross margins as a result of being either partnered or manufactured through third parties, and also as a result of a shift in product mix on established products.

Selling, general and administrative expenses were $12.3 million in the first quarter of 2013 compared to $10.3 million in the first quarter of 2012, with a large part of the increase related to increasing our sales infrastructure to support a growing product portfolio. R&D expenses were $6.0 million in the first quarter of 2013, an increase of $3.1 million over the prior year quarter and consistent with 2013 guidance. Increased 2013 R&D spending is the result of three factors: the Generic Drug User Fee Act (“GDUFA”) fees associated with the projected 25 abbreviated new drug application (“ANDA”) filings for 2013; the cost of bio-equivalence (“BE”) studies associated with high-value products; and the increased internal R&D costs due to the build out and staffing of the new R&D facility.

Non-GAAP adjusted net income for the first quarter of 2013 was $14.4 million, or $0.13 per diluted share, compared to non-GAAP adjusted net income of $10.6 million, or $0.10 per diluted share, in the prior year quarter.

After collecting all available textual information from EDGAR, targeted words are counted using Python's Natural Language Tool Kit.

Figure 2.1 Timeline of textual information

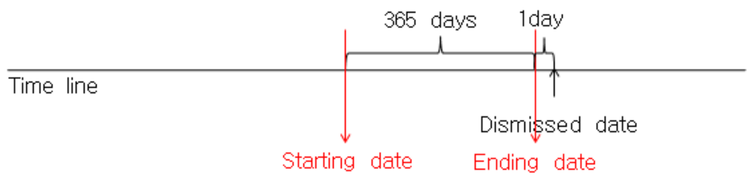


Table 2.3 Sample Composition

Panel A: Sample Composition by Industry

	Distribution in targeted sample	
	#	%
100-1999: Agricultural, Mining and Construction	66	9.48%
2000-3999: Manufacturing	366	52.59%
4000-4999: Transportation and Utilities	52	7.47%
5000-5999: Wholesale and Retail	62	8.91%
7000-8999 : Services	149	21.41%
9999 : Nonclassifiable Establishments	1	0.00%
Total	696	100.00%

Panel B: Sample Composition by Fiscal Year

Year	#	%
2010	193	27.73%
2011	157	22.56%
2012	156	22.41%
2013	190	27.30%
Total	696	100.00%

Panel C: Direction of Auditor Switch

	#	%
Big 4 to Big 4	113	16.24%
Non-Big 4 to Big 4	70	10.06%
Big 4 to non- Big 4	86	12.36%
Non-Big 4 to non-Big 4	427	61.35%
Total	696	100.00%

Panel D: Descriptive Statistics of the Number of 8-K filings

Variable	Mean	Q1	Med.	Q3
Number of 8-K filings	12.20	6.00	10.00	15.50

Table 2.4 Correlation among targeted variables and the dependent variables in audit feel models

	LogAuditFee	*Optimism10k*	*Optimism8K*	*ROAearnings*	*Size*	*Invrec*	*NumSeg*	*Foreign*	*Merge*	*Special*	*Leverage*	*Loss*	*BTM*	*Growth*	*CurrentRatio*	*Big4*	*IW*	*GC*
Optimism10k	-0.23																	
Optimism8K	-0.06	0.25																
ROAearnings	0.20	0.02	0.02															
Size	0.84	-0.16	0.01	0.32														
Invrec	-0.03	0.06	0.13	-0.11	-0.11													
NumSeg	0.40	-0.04	0.01	0.07	0.42	0.01												
Foreign	0.44	-0.09	0.08	0.08	0.40	0.13	0.23											
Merge	0.33	-0.10	-0.03	0.05	0.30	-0.08	0.17	0.18										
Special	0.25	-0.16	-0.13	0.01	0.15	0.00	0.08	0.18	0.35									
Leverage	-0.06	0.01	-0.02	-0.04	-0.16	-0.01	-0.04	-0.07	-0.03	0.01								
Loss	-0.30	-0.12	-0.23	-0.11	-0.45	-0.07	-0.26	-0.22	-0.14	0.08	0.09							
BTM	0.08	-0.09	-0.04	0.05	0.16	-0.09	0.03	0.05	0.01	-0.02	-0.06	-0.06						
Growth	-0.06	-0.02	-0.03	-0.01	-0.08	0.03	0.05	-0.05	-0.02	-0.04	-0.02	0.06	-0.01					
CurrentRatio	-0.06	0.07	0.03	0.07	0.05	-0.10	-0.05	0.01	-0.06	-0.11	-0.09	-0.13	0.07	0.01				
Big4	0.61	-0.15	-0.02	0.07	0.53	-0.04	0.21	0.31	0.25	0.21	-0.03	-0.22	0.01	0.04	0.01			
IW	0.21	-0.13	-0.01	0.03	0.18	-0.04	0.05	0.07	0.08	0.05	-0.01	-0.08	0.03	-0.02	0.02	0.19		
GC	-0.45	-0.01	-0.12	-0.19	-0.58	0.00	-0.21	-0.26	-0.16	0.04	0.16	0.41	-0.13	0.10	-0.21	-0.31	-0.05	
Resignation	-0.11	0.04	0.11	0.04	-0.10	0.09	-0.03	-0.09	-0.09	-0.11	0.00	-0.04	0.01	-0.02	-0.04	-0.24	-0.01	0.02

This table reports Pearson correlations for the targeted variables and the dependent variable. The sample consists of changing auditors for fiscal year between 2010 and 2013. Variable definitions are provided in Table A.1 of APPENDIX A.

Proxy of Textual Information

In analyzing and quantifying textual information, this study adopts a rule-based (dictionary) approach instead of a statistical method. Although there are some benefits associated with the statistical approach[2], there is a multi-faceted rationale for favoring a rule-based paradigm. First, as described above, prior studies indicate related characteristics of textual information on a client's business risks and litigation risks (i.e., optimism). Hence, exploring characteristics of risky firms' textual information might not be required. Second, an appropriate dictionary that has been applied in other studies (Bonini et al. 2011; Rogers et al. 2011; Vadnais 2012) is available. To measure optimism, this study uses the Financial Sentiment Dictionaries[3] from Loughran and McDonald (Loughran and McDonald 2011). This dictionary is developed for textual analysis of financial areas and offers several lists of words that allow for financial sentiment analysis. For example, elements in the dictionary include negative, positive, uncertain, litigious, strong modal and weak modal words. This analysis is particularly interested in the use of negative and positive words to measure the tone of management qualitative disclosures. For convenience, a sample of actual dictionary words is presented in Table 2.5. The following formula (1) is incorporated (Henry 2006; Rogers et al. 2011).

$$Optimism = \frac{(the\ number\ of\ positive\ words - the\ number\ of\ negative\ words)}{(the\ number\ of\ positive\ words + the\ number\ of\ negative\ words)} \quad (1)$$

---

[2] Li (2011) describes the benefits of a statistical approach and the possible situations in which a dictionary approach is acceptable. Based on his frame to determine an appropriate approach, a dictionary approach is considered to be appropriate in this paper.

[3] Available at http://www.nd.edu/~mcdonald/Word_Lists.html)

The current study considers the following set of individual variables from textual information:

Optimism10k = the tone of optimism calculated by the different between the number of positive words and the number of negative words divided by total number of positive and negative words in a firm's 10-K from the last fiscal year the replaced auditor was engaged;

Optimism8k = the tone of optimism calculated by the different between the number of positive words and the number of negative words divided by total number of positive and negative words in 8-K coverage from the year prior to dismissed date to a day prior to the date of dismissal.

Table 2.5 Sample words of the dictionary

Category	Examples
Positive Words	able, boost, breakthroughs, brilliant, charitable, enjoyed, popular, rewards…
Negative Words	assaults, assertions, bad, bail, bailout, balk, damage, cut, dispute, risky, slow, unlawful..

Audit Fee Model

To analyze the relationship between audit fees and quantitative factors, existing quantitative factors from recent studies (e.g., Francis and Wang 2005; Krishnan et al. 2005; Ghosh and Pawlewicz 2009; Choi et al. 2010; Stanley 2011) are controlled.

$$\begin{aligned} LogAuditFee_n = {} & B_0 + B_1 Optimism_{n-1} + B_2 RoaEarnings_n + B_3 Size_n + B_4 InvRec_n + B_5 NumSeg_n \\ & + B_6 Foreign_n + B_7 Merg_n + B_8 Special_n + B_9 Leverage_n + B_{10} CurrentRatio_n + B_{11} Loss_n \\ & + B_{12} BTM_n + B_{13} Growth_n + B_{14} Big4_n + B_{15} Resignation_n + B_{16} GC_n + B_{17} IW_n \end{aligned}$$

(2)

Where:

LogAuditFee = natural log of initial audit fee (Audit Analytics);

Optimism = proxies of textual information in 10-K or 8-K filings (described above);

ROAEarnings= earnings, calculated as operating income after depreciation (OIADP) divided by total asset (AT);

Size = natural log of total assets (AT);

InvRec = inventory (INVT) plus accounts receivable (RECT) divided by total assets (AT);

NumSeg = the number of business segments;

Foreign = 1 if the firm has foreign operations (TXFO), 0 otherwise;

Merg = 1 if the firm reported the item related to acquisition and merger (AQP), 0 otherwise;

Special = 1 if the firm reported special items (SPI), 0 otherwise;

Leverage = the difference between total liabilities (LT) and current liabilities (LCT) divided by total assets (AT);

CurrentRatio = current assets (ACT) divided by current liabilities;

Loss = 1 if a firm's net income (NI) < 0, 0 otherwise;

BTM = the difference between total assets (AT) and total liabilities (LT) divided by market value of common equity (PRCC_F × CSHO);

Growth = the percentage of change in sales (SALE) from period n-1 to period n;

Big4 = 1 if a successor auditor is one of the Big 4, 0 otherwise;

Resignation = 1 if a predecessor auditor initiated auditor resignation, 0 otherwise;

GC = 1 if a successor auditor issues a going-concern opinion, 0 otherwise;

IW = 1 if a successor auditor indicates internal control weakness, 0 otherwise.

The targeted independent variables are indicators drawn from textual information (*Optimism10k* and *Optimism8k*). Each selected independent variable arises from previous studies demonstrating that auditors seek to resolve uncertainty about client risk by adjusting audit fees.

Firm size is one of the most important attributes of audit fee decisions because firm size affects audit efforts (O'Keefe et al. 1994). This study uses the natural log of total assets (*AT*) to proxy for firm size because the dollar amount of total assets would be disproportionately large.

Indicators related to client business risks are also controlled to examine the incremental values of the targeted variables. ROA (*ROAearnings*), leverage (*Leverage*), and current ratio (*CurrentRatio*) are proxies for risk and are therefore controlled (Stanley 2011). ROA is commonly considered to measure a firm's business performance. In addition, leverage and current ratio are indicators of the client's business failure risk. Leverage can capture long-term financial liquidity, and the current ratio captures short-term liquidity.

As a proxy for client complexity, organizational complexity (*NumSeg* and industry) and geographical complexity (*Foreign*) are controlled (Hay 2010). Inherent risk, the probability of material misstatement before considering the effectiveness of internal control, is often proxied by the amount of a firm's inventories and accounts receivable (*InvRec)* or additional audit procedures (*Special* and *Merg*). Accordingly, these variables are controlled.

Some variables related to a firm's performance are also controlled. The ratio of a firm's book value to market value (*BTM)* is a proxy for the firm's market value and is therefore controlled. Profitable firms generally pay lower audit fees, so loss (*Loss*) is included as a control variable. As an indicator of a firm's maturity, which may be related to systematic risks, sales growth (*Growth*) is controlled.

Factors related to the auditor and the auditor's decisions are controlled as well. A binary variable for Big 4 audit firms (*Big4*) is added because these firms are often considered to provide higher audit quality, and consequently, higher audit fee premiums (Francis and Wang 2008). Auditor-initiated resignations signal risky clients, so firms whose predecessor auditors resigned are likely to be charged high audit fees (*Resignation*) by the successor auditor. An internal control weakness opinion (*IW*) is related to control risks, and a going-concern opinion (*GC*) is related to business risk, so these factors are also controlled.

RESULTS

Descriptive Statistics

Table 2.6 provides descriptive statistics for the dependent variable, the proxy of textual information, and control variables. The average of *LogAuditFee* is 12.42 (the average of actual audit fee is \$596,004.44). The tone of both of 10-K filings and 8-K filings is negatively skewed, although the tone of 8-K reports is slightly less pessimistic than the tone of 10-K reports; the mean of *Optimism10K* and *Optimism8K* are -0.33 and -0.31 respectively. When the tone of optimism is treated as a dichotomous variable (if the tone of optimism > 1 then 1, otherwise 0), that negativity becomes more apparent. Only eight firms out of 696 (1.15%) have an optimistic tone in their 10-K filings. On the other hand, 91 firms in the full sample (13.07%) use an optimistic tone in their 8-K reports.

The descriptive statistics show that the sample contains various sizes of firms. For instance, only 25 percent of sample is the client of big4, and the ratio of firms receiving going-concern opinion is 26 percent. Larger firms are likely to hire big audit firms, and the ratio of firms receiving going-concern opinions varies depending on the firm size. In this line, Carson et al. (2013) showed that the larger firms (market cap > \$500 million) are less likely to receive going-concern opinions (0.33%), whereas the smaller firms (market cap < $75 million) are more likely to receive going-concern opinions (36.70 %).

Table 2.4 presents univariate correlations among individual variables from textual information and the dependent variable. Generally, between 10-K and 8-K reports there is a relatively moderate correlation (0.25), indicating that each variable presents a different aspect of client risk, but the relationship is still statistically significant ($p < 0.01$). In addition, as anticipated, the tone of optimism in qualitative disclosures is negatively correlated with audit fees (correlation with *Optimism10K* is -0.23 and with *Optimism8k* is -0.06), but only the tone of optimism in 10-K filings is statistically significant ($p < 0.01$). Textual variables are not highly correlated with financial indicators of client business risk. *Optimism10k* and *Optimism8k* are correlated with *ROAearnings* (0.02 and 0.02 respectively), with *Leverage* (0.01 and -0.02 respectively), and with *CurrentRatio* (0.07 and 0.03 respectively). Only the correlation between *Optimism10K* and *CurrentRatio* are marginally significant ($p < 0.1$).

Table 2.6 Descriptive statistics of the targeted variable and independent variables

Variable	Mean	Std.	Lower Quartile	Median	Upper Quartile
Audit Fee ($)	596,004.44	1,211,063.99	1,01,826.00	228,527.50	605,760.50
LogAuditFee	12.42	1.30	11.53	12.34	13.31
Optimism10k	-0.33	0.13	-0.42	-0.34	-0.25
Optimism8k	-0.31	0.29	-0.50	-0.41	-0.19
ROAearnings	-1.44	11.69	-0.31	-0.02	0.07
Size	3.85	2.60	2.27	4.01	5.48
Invrec	0.28	0.22	0.09	0.25	0.44
NumSeg	1.67	1.09	1.00	1.00	2.00
Foreign	0.29	0.45	0.00	0.00	1.00
Merge	0.18	0.39	0.00	0.00	0.00
Special	0.63	0.48	0.00	1.00	1.00
Leverage	0.32	1.25	0.01	0.09	0.32
Loss	0.60	0.49	0.00	1.00	1.00
BTM	0.48	6.57	0.10	0.42	0.85
Growth	1.18	11.60	-0.07	0.09	0.38
CurrentRatio	2.72	4.43	0.87	1.66	3.08
Big4	0.26	0.44	0.00	0.00	1.00
IW	0.05	0.22	0.00	0.00	0.00
GC	0.26	0.44	0.00	0.00	1.00
Resignation	0.25	0.43	0.00	0.00	0.50

Test of H1

Hypothesis 1 explores the relationship between the level of optimism in management disclosures and audit fees. Table 2.7 reports the estimation and t-statistics results clustered by firm and year (Petersen 2009). Panel A contains three regressions with proxies for the level of optimism found in management qualitative disclosures. Column 1 of Panel A shows the results of the audit fee model with *Optimism10K*. The model explains initial audit fees relatively sufficiently (R^2 = 0.81) with some variables (*Size*, *Invec*, *NumSeg*, *Foreign*, *CurrentRatio*, *Loss*, *BTM*, and *Big4*) proving significant at $p < 0.01$. Interestingly, unlike the tone of optimism in 10K filings (*Optimism10k*) some business risk indicators do not play an important role in explaining audit fees[4]. The coefficient for *Optimism10k* (-0.45) is negative which is consistent with Hypothesis 1 and indicates that firms using an optimistic tone in their 10-K reports receive lower audit fee premiums.

Columns 2 and 3 of Table 2.7 the multivariate regression results for the audit fee model with *Optimism8K* only and with both *Optimism10k* and *Optimsim8k* respectively. As predicted, the estimated coefficient of *Optimism8K* (-0.19) is negative and statistically significant. A comparison of Models 1 and 2 shows that *Optimism8k* has more explanatory power for initial audit fee decisions than *Optimism10k*. This might be related to the characteristics of 10-K reports, since they are more uniformly pessimistic and less variable than 8-K filings. Finally, Column 3 of Panel A describes the results of the audit fee model with *Opitimism10k* and *Optimism8k*. The negative coefficients for two variables are statistically significant (β = -0.38, *t*-value = -2.19 for the tone in 10-K reports and β = -0.17, *t*-value = -5.43 for the tone in 8-K reports). Accordingly, Hypothesis 1 is supported.

[4] In the audit fee model without textual variables, these variables are still not statistically significant at the $p < 0.1$ level. This indicates that correlations between textual variables and these variables do not meaningfully affect the results.

Table 2.7 Regression of optimism in qualitative disclosures on initial audit fees for H1

		Model1		Model2		Model3	
		(1)		(2)		(3)	
	Pred	Est.	t-stat	Est.	t-stat	Eat.	t-stat
Optimism10k	H(-)	-0.45	-2.60**			-0.38	-2.19**
Optimism8k	H(-)			-0.19	-6.66***	-0.17	-5.43***
ROAearnings	(-)	-0.00	-1.22	-0.00	-1.27	-0.00	-1.21
Size	(+)	0.37	21.28***	0.38	21.83***	0.37	21.27***
Invrec	(+)	0.28	1.67*	0.29	1.74*	0.28	1.71*
NumSeg	(+)	0.06	4.63***	0.06	5.15***	0.06	4.56***
Foreign	(+)	0.24	3.42***	0.25	3.79***	0.25	3.59***
Merge	(+)	0.08	0.93	0.08	1.04	0.08	0.98
Special	(+)	0.12	1.53	0.12	1.55	0.11	1.46
Leverage	(+)	0.06	1.09	0.06	1.16	0.06	1.13
CurrentRatio	(-)	-0.02	-2.10**	-0.02	-2.19**	-0.02	-2.24**
Loss	(+)	0.21	2.08**	0.21	2.15**	0.19	1.99**
BTM	(-)	-0.01	-4.26***	-0.01	-4.80***	-0.01	-4.66***
Growth	(-)	0.00	-1.45	0.00	-1.39	0.00	-1.53
Big4	(+)	0.53	16.05***	0.53	15.44***	0.53	15.06***
Resignation	(+)	0.06	0.86	0.08	1.10	0.07	1.00
GC	(+)	0.02	0.26	0.02	0.26	0.02	0.19
IW	(+)	0.22	1.27	0.25	1.31	0.22	1.25
Intercept		9.97	63.01***	10.04	67.34***	10.00	69.40***
Firm/ Year		Included		Included		Included	
R^2		0.81		0.81		0.81	
# observation		696		696		696	

*, **, ***denote the significance levels of 0.1, 0.05, and 0.01 (two –tailed), respectively, using t statistics adjusted for firm and year clustering (Petersen 2009).

Test of H2

Hypothesis 2 looks at the effect of going-concern opinions on the relationship between the level of optimism in management disclosures and audit fees. To test whether the auditors' assessment of high business risks modifies the audit fee premium for firms disclosing negative qualitative information, the interaction terms *Optimism10k* × *GC*, *Optimism8k* × *GC*, and both *Optimism10* × *GC* and *Optimism8k* × *GC* are included separately in the audit fee model. Table 2.8 contains estimates from equation (1) using the full sample. Model 1 shows a significant positive coefficient for *Optimism10k* × *GC*

(β = 1.04, *t*-value = 3.14), along with a significant negative coefficient for *Optimism10k* (β = -0.68, *t*-value = -4.93). These results indicate that auditors reflect their perceived business risks (*GC*) on analyzing the tone of 10-K filings. Specifically, when auditors perceive high client risks, as management discloses more optimistic points of view in their10-K reports, auditors consider them as risker clients.

On the other hand, Column 2 of Table 2.8 shows a significant negative coefficient for *Optimism8k* × *GC* (β = -0.12, *t*-value = 2.16), along with a significant negative coefficient for *Optimism8k* (β = -0.16, *t*-value = -2.61). These results indicate that auditors weight the tone of 8-K reports when they audit high-risk clients. Consistent with these results, Column 3 of Panel C, which contains two compounding effects, presents a large, significant, and negative coefficient for *Optimism10k* × *GC* (β = 1.09, *t*-value = 2.65), but a positive significant coefficient for *Optimism8k* × *GC* (β = 0.22, *t*-value = -4.67).

What factors contribute to these differences between the two filings? The answer may relate to the characteristics of each filing. The answer may relate to the characteristics of each filing. Even though some parts of 10-K filings (e.g., MD&A) and of 8-K filings (e.g., Item 2.02 – Results of Operations and Financial Condition) are voluntary disclosures, the qualitative information in 10-K reports is likely to be more pessimistic and less likely to change from period to period. Brown and Tucker (2011) find that firms experiencing *large* economic change are more likely to modify MD&A disclosures. On the other hand, Davis and Tama-Sweet (2012) show that the tone of MD&A in 10-K and 10-Q filings tends to be more pessimistic than that of press releases.

Specifically, they argue that since the market is more likely to respond to press releases than to 10-K or 10-Q filings, management uses press releases more strategically.

Accordingly, when the tone of 10-K filings is unusually optimistic even though auditors consider the client as high-risk, auditors treat the optimistic tone of 10-K reports as a red flag indicating potential management fraud. On the other hand, auditors use 8-K filings to understand underlying reason why a client contains business risks (e.g., bankruptcy filings or information from press releases), and rely more heavily on these filings for high-risk clients.

These results imply that successor auditors on a new engagement consider qualitative information from 10-K and 8-K reports in deciding audit fees. The auditors place greater weight on 8-K reports, but respond to inconsistencies between the tone in 10-K filings and their judgment regarding client business risks with audit fees. Thus, auditors understand the characteristics of each filing and set the different standards to evaluate the tone of management.

Table 2.8 Regression of optimism in qualitative disclosures on initial audit fees for H2

		Model1		Model2		Model3	
		(1)		(2)		(3)	
	Pred	Est.	t-stat	Est.	t-stat	Est.	t-stat
GC×Optimism10k	H(+/-)	1.04	3.14***			1.09	2.65***
GC×Optimism8k	H(+/-)			-0.12	-2.61***	-0.22	-4.67***
Opitmism10k	(-)	-0.68	-4.93***			-0.65	-4.44***
Optimism8k	(-)			-0.16	-5.77***	-0.10	-2.39**
ROAearnings	(-)	-0.00	-1.22	-0.00	-1.26	-0.00	-1.20
Size	(+)	0.38	23.33***	0.38	21.88***	0.37	23.50***
Invrec	(+)	0.29	1.74*	0.29	1.71*	0.29	1.75*
NumSeg	(+)	0.06	4.51***	0.06	5.19***	0.06	4.43***
Foreign	(+)	0.24	3.39***	0.25	3.81***	0.24	3.57***
Merge	(+)	0.08	0.92	0.08	1.02	0.08	0.92
Special	(+)	0.11	1.41	0.12	1.59	0.11	1.42
Leverage	(+)	0.06	1.14	0.06	1.18	0.06	1.21
CurrentRatio	(-)	-0.02	-2.09**	-0.02	-2.16**	-0.02	-2.19**
Loss	(+)	0.21	2.11**	0.21	2.15**	0.19	2.02**
BTM	(-)	-0.01	-3.94***	-0.01	-4.72***	-0.01	-4.38***
Growth	(-)	0.00	-1.48	0.00	-1.38	0.00	-1.53
Big4	(+)	0.51	16.25***	0.53	15.65***	0.52	15.69***
Resignation	(+)	0.08	1.09	0.08	1.08	0.09	1.22
GC	(+)	0.38	3.78***	-0.02	-0.24	0.30	2.59**
IW	(+)	0.21	1.36	0.25	1.32	0.21	1.35
Intercept		9.89	58.12***	10.01	66.18***	9.90	64.49***
Firm/ Year		Included		Included		Included	
R^2		0.81		0.81		0.81	
# observation		696		696		696	

*, **, ***denote the significance levels of 0.1, 0.05, and 0.01 (two –tailed), respectively, using t statistics adjusted for firm and year clustering (Petersen 2009).

ADDITIONAL ANALYSIS

Changes in Audit Fees

In this paper, the main audit fee model examines the association between initial audit fee decisions and the tone of qualitative disclosures. As an additional analysis, this section looks at how audit fees changed from predecessor auditors to successor auditors. This test can also provide some insight into whether qualitative disclosures play a role in changing audit fees.

Table 2.9 shows estimation results for the audit fee change model, which is developed from the audit fee model (equation 2) with variables measuring the first difference between the last fiscal year with the predecessor auditor and the first year with the successor auditor ($\Delta LogAuditFee_{t,t-1} = LogAuditFee_{t}$ - $LogAuditFee_{t-1}$). To control for the effect of unpredicted audit fees in the prior fiscal year, the residuals from the audit fee model (equation 2) in the last year with the predecessor auditor are included in the model (Francis and Wang 2005; Stanley 2011).

Results show that changes in the tone of 10-K reports do not have explanatory power for changes in audit fees (β = -0.31, *t*-value = -0.68). As discussed above, the qualitative information of 10-K filings are less likely to be modified from one year to the next. Consequently, the changes in the optimistic tone of 10-K filings might not adequately capture changes in client business risks. However, the coefficient for changes in the tone of 8-K filings ($\Delta Optimism8k$) in column 2 is negative (-0.21) and statistically significant at p <0.01, which indicates that changes in the optimistic tone in 8-K reports have a negative association with changes in audit fees. The results support a meaningful association between audit fees and qualitative management disclosures in 8-K reports.

Table 2.9 Regression of changes in optimism in qualitative disclosures on changed audit fees for additional analysis

		Model1		Model2	
		(1)		(2)	
	Pred	Est.	t-stat	Est.	t-stat
$\Delta Optimism10k_{t-2,t-1}$	H(-)	-0.31	-0.68		
$\Delta Optimism8k_{t-2,t-1}$	H(-)			-0.21	-3.32***
$\Delta ROAearnings_{t-1,t}$	(-)	-0.20	-2.38**	-0.19	-2.21**
$\Delta Size_{t-1,t}$	(+)	0.34	6.82***	0.35	9.09***
$\Delta Invrec_{t-1,t}$	(+)	0.02	0.12	0.03	0.13
$\Delta NumSeg_{t-1,t}$	(+)	-0.01	-0.31	-0.01	-0.29
$\Delta Foreign_{t-1,t}$	(+)	0.04	0.65	0.04	0.73
$\Delta Merge_{t-1,t}$	(+)	0.03	0.76	0.05	1.41
$\Delta Special_{t-1,t}$	(+)	0.08	1.42	0.07	1.35
$\Delta Leverage_{t-1,t}$	(+)	0.28	1.50	0.29	1.75*
$\Delta CurrentRatio_{t-1,t}$	(-)	0.00	15.88***	0.00	10.33***
$\Delta Loss_{t-1,t}$	(+)	0.12	1.84*	0.13	2.36**
$\Delta BTM_{t-1,t}$	(-)	0.00	-0.56	0.00	-0.55
$\Delta Growth_{t-1,t}$	(-)	0.00	0.25	0.00	0.25
$\Delta Big4_{t-1,t}$	(+)	0.38	13.13***	0.38	14.69***
Resignation	(+)	0.05	0.66	0.06	0.71
$\Delta GC_{t-1,t}$	(+)	-0.11	-1.86*	-0.12	-2.42**
$\Delta IW_{t-1,t}$	(+)	0.04	0.28	0.02	0.21
$Unexpected_{t-1}$	(-)	-0.45	-20.84***	-0.45	-19.17***
Intercept		0.28	1.86*	0.25	2.04**
Firm/ Year		Included		Included	
R^2		0.54		0.55	
# observation		418		418	

*, **, ***denote the significance levels of 0.1, 0.05, and 0.01 (two –tailed), respectively using t statistics adjusted for firm and year clustering (Petersen 2009).

Big4 and Non-big4

In this section I estimate the audit fee model separately depending on the size of successor auditors in Table 2.10. Results show that different variables play an important role in explaining initial audit fees for Big 4 and non-Big 4 auditors, indicating that these firms have differing fee strategies. The coefficients on *Size*, *Foreign*, and *BTM* are significant in both models, but the coefficients on *ROAearnings*, *Growth*, *Resignation*, and *IW* are significant in the audit fee models for Big 4 auditors, but not for non-Big 4 auditors. By contrast, coefficients for *NumSeg*, *CurrentRatio*, and *Loss* are significant in for non-Big 4 auditors, but not for Big 4 auditors. Interestingly, the coefficients on *Optimism10k* and *Optimism8k* are insignificant in the audit fee model for Big 4 successor auditors, but the coefficient on *Optimism8k* is negative and highly significant (p-value < 0.01). The results indicate that Big 4 auditors do not consider the tone of qualitative disclosures for audit fee decisions, whereas non-Big 4 auditors do.

Table 2.11 presents two estimation results for equation (2) with *GC* × *Optimism10k* and *GC* × *Optimism8k*, depending on the size of successor auditors. For Big 4 auditors (Column 1) the coefficient for *GC* × *Optimism10k* is positive and significant (β = 4.42, t-value = 2.65), but the coefficient for *GC* × *Optimism8k* is negative and significant (β= -5.11, t-value = -7.08). Nevertheless, the coefficients for *Optimism10k* and *Optimism8k* are insignificant in the audit fee model for Big 4 auditors. These results suggest that the association between audit fees and the tone of qualitative disclosures are only obvious when Big 4 auditors consider the clients to be risky. Column 2 illustrates the estimation results for the audit fee model for non-Big 4 auditors. The coefficient for *GC* × *Optimism10k* is insignificant, but *GC* × *Optimism8k* is negative and significant (β= -0.17, t-value = -3.38). Also, the coefficients for *Optimism10k* and *Optimism8k* are negative and significant (β = -0.61, t-value = -2.17and β = -0.61, t-value = -2.43 respectively). These results show that non-Big 4 auditors reflect the tone of qualitative disclosures in their audit fee decisions. The association between audit fees and the tone of optimism in 8-K filings is greater for high-risk clients, but the association between audit fees and the tone of optimism in 10-K filings is not modified by the level of risk.

In summary, Big 4 and non-Big 4 auditors reflect the tone of qualitative management disclosures differently in their audit fee decisions. The association between audit fees for Big 4 auditors and the optimistic tone of qualitative information is significant only for high-risk clients. On the other hand, non-Big 4 auditors' fee decisions reflect the tone of qualitative management disclosures, but the association between the tone of optimism on 10-K filings and audit fees is not reduced for risky clients.

These results could be explained by the Big 4 auditors' portfolio management where the large audit firms tend to avoid risker clients (Johnstone and Bedard 2004). Compared to non-Big 4 firms, the Big 4 auditors are more likely to use higher standards when screening potential clients, and evaluate various components more rigorously to avoid potential losses (Rama and Read 2006; Ettredge et al. 2007). Alternatively, these results might be caused by sampling bias in each group of auditors (Big 4 versus non-Big 4) since specific types of clients are likely to be audited by Big 4 firms. For instance, according to the Student's t test between Big 4 clients and non-Big 4 clients, Size is significantly different (difference = -3.14, p-value < 0.01), as is ROAearnings (difference = -1.92, p-value < 0.01). However, the means of Optimism8k are not different between subgroups, although Big 4 clients' Optimism10k are more pessimistic than of non-Big 4 (difference = 0.04, p-value <0.01). Consistent with previous studies, Big 4 auditors are likely to work with larger and less risker clients than non-Big 4 auditors. Accordingly, Big 4 auditor might value qualitative disclosures only when they work for clients with high business risks. Generally, these results support Hypothesis 2.

Table 2.10 Regression of optimism in qualitative disclosures on initial audit fees for the additional analysis

		Big 4		Non-Big 4	
		(1)		(2)	
	Pred	Est.	t-stat	Est.	t-stat
Optimism10k	H(-)	-0.06	-0.22	-0.30	-1.27
Optimism8k	H(-)	-0.07	-0.56	-0.19	-4.59***
ROAearnings	(-)	-0.45	-2.46**	-0.00	-1.17
Size	(+)	0.32	8.47***	0.38	19.50***
Invrec	(+)	-0.08	-0.25	0.32	1.54
NumSeg	(+)	0.04	0.93	0.08	7.79***
Foreign	(+)	0.34	9.02***	0.23	2.98***
Merge	(+)	0.07	0.81	0.01	0.16
Special	(+)	0.18	1.10	0.10	1.58
Leverage	(+)	0.21	1.49	0.06	1.20
CurrentRatio	(-)	-0.02	-1.39	-0.02	-2.13**
Loss	(+)	0.07	0.99	0.14	1.98*
BTM	(-)	-0.09	-2.29**	-0.01	-4.26***
Growth	(-)	-0.01	-2.65***	0.00	0.57
Resignation	(+)	0.26	3.18***	0.03	0.29
GC	(+)	0.44	1.50	0.04	0.45
IW	(+)	0.53	10.61***	-0.27	-1.55
Intercept		11.42	43.68***	9.91	58.24***
Firm/ Year		Included		Included	
R^2		0.76		0.71	
# observation		183		513	

*, **, ***denote the significance levels of 0.1, 0.05, and 0.01 (two –tailed), respectively using t statistics adjusted for firm and year clustering (Petersen 2009).

Table 2.11 Regression of optimism in qualitative disclosures on initial audit fees for the additional analysis

		Big4		Non-Big4	
		(1)		(2)	
	Pred	Est.	t-stat	Est.	t-stat
GC×Optimism10k	H(+/-)	4.42	2.65***	0.96	1.53
GC×Optimism8k	H(+/-)	-5.11	-7.08***	-0.17	-3.38***
Opitmism10k	(-)	-0.12	-0.44	-0.61	-2.17**
Optimism8k	(-)	-0.04	-0.34	-0.12	-2.43**
ROAearnings	(-)	-0.49	-2.50**	-0.00	-1.17
Size	(+)	0.31	11.40***	0.39	21.64***
Invrec	(+)	-0.13	-0.37	0.33	1.53
NumSeg	(+)	0.04	0.89	0.08	6.83***
Foreign	(+)	0.35	8.58***	0.22	2.99**
Merge	(+)	0.11	1.44	0.02	0.23
Special	(+)	0.15	0.80	0.10	1.45
Leverage	(+)	0.18	1.18	0.06	1.28
CurrentRatio	(-)	-0.02	-1.46	-0.02	-2.10**
Loss	(+)	0.05	0.85	0.14	2.03**
BTM	(-)	-0.08	-2.15**	-0.01	-3.44***
Growth	(-)	-0.01	-3.13***	0.00	0.72
Resignation	(+)	0.25	3.31***	0.04	0.52
GC	(+)	0.03	0.04	0.30	1.40
IW	(+)	0.53	14.73***	-0.26	-1.46
Intercept		11.50	48.37***	10.01	66.18***
Industry/ Year		Included		Included	
R^2		0.77		0.72	
# observation		183		513	

*, **, ***denote the significance levels of 0.1, 0.05, and 0.01 (two –tailed), respectively using t statistics adjusted for firm and year clustering (Petersen 2009).

ROBUSTNESS TEST

Successor auditors often do not have the necessary current financial information when negotiating audit fees (Hackenbrack et al. 2014). Nevertheless, the majority of audit fee studies, including studies regarding initial engagements use the current financial information (e.g., Huang et al. 2009). The main reason that audit fees studies are performed with contemporaneous financial variables is because researchers assume that auditors decide audit fees based on reasonable expectations regarding client risks. Nonetheless, initial engagements are generally considered as the point when audit fees change because firms often change auditors to reduce audit fees. Accordingly, examining audit fees with financial variables of the first fiscal year with successor auditors might not be appropriate. To overcome this concern, the audit fee model can also be developed with indicators from prior fiscal years as a robustness test. This is performed by also using Equation (2), but using lagged values for all indicator variables. Table 2.12 illustrates the estimation of results for equation (2) with these lagged values. Overall, lagged indicators are found to have lower explanatory powers in audit fee decisions ($R^2 = 0.79$). This result supports the current trend to use contemporaneous information in audit fee research.

However, a few lagged indicators may play significant roles in understanding audit fee decisions, such as a predecessor auditor's internal control weakness opinion, $IW_{n\text{-}1}$. Consequently, it would be useful to examine the value of lagged indicators in future research. Interestingly, the negative coefficient for *Optimism8k* is statistically significant (β = -0.25, *t*-value = -4.78), but the negative coefficient for *Opitimism10k* is not significant (β = -0.25, *t*-value = -1.15). To explore the non-significant coefficient for *Opitimism10k*, the correlation matrix between *Opitimism10k* and contemporaneous variables is compared to the matrix between *Opitimism10k* and lagged variables, but results show no large differences. Nevertheless, it is possible that *Opitimism10k* is more affected by lagged indicators than by contemporaneous indicators, thereby reducing the orthogonal explanatory powers of audit fee decisions. Since *Opitimism10k* is formulated from all possible textual information in 10-Ks, some parts of textual 10K filings are already included in the quantitative information of 10-Ks (e.g., footnotes).

In summary, the robustness test with lagged indicators suggests that the tone of 8-K filings plays an important role in deciding audit fees when auditors cannot reasonably predict a client's financial information. These results partially support the Hypothesis 1.

Table 2.12 Regression of optimism in qualitative disclosures on initial audit fees for a robustness Test

	Pred	Est.	t-stat
Optimism10k	H(-)	-0.25	-1.15
Optimism8k	H(-)	-0.25	-4.78***
ROAearnings	(-)	-0.01	-2.99***
Size	(+)	0.43	32.10***
Invrec	(+)	0.21	1.20
NumSeg	(+)	0.06	3.25***
Foreign	(+)	0.20	2.01**
Merge	(+)	0.06	0.57
Special	(+)	0.12	3.45***
Leverage	(+)	0.10	3.36***
CurrentRatio	(-)	0.00	-3.17***
Loss	(+)	0.09	1.23
BTM	(-)	-0.03	-1.43
Growth	(-)	0.00	-1.06
Resignation	(+)	0.00	-0.02
GC	(+)	-0.02	-0.25
IW	(+)	0.51	4.17***
Intercept		9.99	38.06
Firm/ Year		Included	
R^2		0.79	
# observation		594	

*, **, ***denote the significance levels of 0.1, 0.05, and 0.01 (two –tailed), respectively using t statistics adjusted for firm and year clustering (Petersen 2009).

CONCLUSIONS, LIMITATIONS, AND FUTURE RESEARCH

Although audit firms commonly consider various sources of information regarding a client in the pre-engagement stage, there is little research to support this audit practice. This study examines whether an auditor's judgment of audit risks is affected by certain client characteristics reflected in qualitative factors. By analyzing textual information from 10-K filings and 8-K filings, including press releases, this study examines whether textual information can explain potential engagement risks as measured by audit fee premiums.

The audit fee model can be enhanced by including quantified textual information. Successor auditors tend to have less access to clients' internal information than predecessor auditors, so they make greater efforts to analyze available information in order to overcome a lack of information. Prior literature indicates that textual information delivers orthogonal evidence, which quantitative attributes might not provide, to predict a firm's future performance (Li 2006). In addition, in certain cases qualitative management disclosures could be the indicators of management fraudulent behaviors (Rogers et al. 2011). Accordingly, auditors might depend on textual data to collect information that cannot be provided by financial attributes alone.

The empirical results of this paper indicate that qualitative information plays an important role in explaining audit fees. The difference between the number of positive words and the number of negative words scaled by the sum of negative and positive words in 10-K filings and 8-K filings is negatively associated with audit fees for initial engagement. By adding the compounding effect between optimistic tone in management disclosures and the auditor's going-concern opinion, the effect of the level of client business risks perceived by successor auditors on the association between audit fee decisions and the tone of qualitative disclosures can be assessed. The results indicate that high perceived business risks lead to a stronger association between audit fees and the tone of 8-K filings. By contrast, high perceived business risks weaken the association between audit fees and the tone of 10-K filings because auditors consider 10-K filings of high-risk clients to be less credible. This implies that auditors depend on qualitative information differently, depending on the client's risk level.

There are some limitations in this research. First, it is based on a simple word counting method to quantify qualitative factors. More sophisticated approaches to this evaluation would likely provide more precise results. Second, events like bankruptcies and earnings news releases are disclosed in 8-K filings, so the proxy for these sources might not capture the unique tone of these filings. Future research could distinguish and measure these information sources separately.

Based on the results, some interesting further research questions arise. This study focused on new engagements, but whether auditors use qualitative information for client retention decisions is an interesting topic to explore. In addition, this study uses only qualitative information that is publicly available. Evaluating if data is available the impact of internal qualitative sources such as minutes of board meetings or emails on audit decisions would be of value. Finally, it would also be interesting to consider whether auditors weigh qualitative information or quantitative attributes more heavily when these factors conflict.

Chapter 3 Auditing the Revenue Account: Does Audit Sampling Replaced By Substantive Analytical Procedures Improve Audit Quality?

INTRODUCTION

Substantive analytical procedures (SAPs) are utilized as audit evidence regarding assertions related to account balance or class of transactions (PCAOB 2010). The auditor develops expectations by employing plausible relationships among financial and nonfinancial information based on his/her knowledge and understanding of the client's industry and business, and then examines whether the expectation of the value of the account is materially different from the reported value.

Previous literature generally finds that analytical procedures are effective to identify financial misstatements (e.g., Knechel 1988), and a commonly used audit approach for the revenue account is a combination of SAPs indicating high risk areas and tests of details on designated risky areas because often it is hard to set precise expectations using SAPs (Glover et al. 2015). Nevertheless, the Public Company Accounting Oversight Board (PCAOB)'s inspections have continually expressed concerns with the performance of SAPs (Messier et al. 2013b), and it only considers SAPs as substantive evidence if the difference between the auditor's expectation and the client's recorded account is within materiality[5] (performance materiality or tolerable misstatement) (PCAOB 2011). If that threshold is exceeded, then SAPs are deemed to

[5] The PCAOB did not clarify whether it is performance materiality or overall materiality. According to audit standards, performance materiality is used to identify misstatements in particular classes of transactions or account balances (AICPA 2012c). Accordingly, performance materiality should be the threshold to identify misstatement in revenue accounts. Nevertheless, revenue accounts are often used as a benchmark of overall materiality and are large. Accordingly, it seems that Glover et al. (2015) consider as the rule to be overall materiality for revenue accounts.

deliver no evidence, and the inspector equates an absence of substantive details testing with a lack of assurance (Glover et al. 2015).

On the other hand, the high assurance provided by SAPs is often difficult for large income statement accounts like revenue (Glover et al. 2015). Therefore, some auditors have modified their approach for revenue testing, employing sampling rather than substantive analytical procedures (Christensen et al. 2015). In responding to current audit practices regarding SAPs, Glover et al. (2015) argue that low and moderate SAPs could provide benefits to improve audit quality if they are combined with other forms of audit evidence. Also, Christensen et al. (2015) suggest that further research is needed to examine the costs and benefits of using sampling instead of SAPs for accounts like revenue.

In response to current issues related to SAPs for revenue accounts, this essay examines two issues: 1) the cost and benefits of sampling in substantive tests of details instead of SAPs for revenue accounts; and 2) whether SAPs can offer a high level of assurance for revenue accounts. These questions are addressed by exploring possible risks and benefits of each substantive test suggested by prior studies and by conducting a meta-analysis based on the outcomes of prior studies of SAPs for revenue accounts.

A few older studies shown that the combination of audit sampling and SAPs can reduce sample size and detection risks (Kinney 1979; Knechel 1988). Nonetheless, these studies did not extensively explore whether the value of SAPs are consistent even if conditions that might affect their effectiveness are modified. Based on exploring prior studies, this essay argues that both SAPs and audit sampling should be deliberately applied as substantive tests, since the effectiveness and the efficiency of each audit procedure is often influenced by such factors as the characteristics of error patterns, the number of total line items, and internal control weakness. SAPs could be more effective and efficient when the population is very large and client business risk components remain a consideration. In addition, even if audit sampling is utilized with moderate or weak SAPs, the SAPs could still identify areas containing deviations from the auditor's expectation, thereby reducing audit efforts. In addition, unusual moderate or weak SAPs may offer evidence of financial statement fraud.

This essay contributes to the audit literature, audit practice, and regulation by showing the value of SAPs. Prior studies have indicated conditions that could possibly affect the effectiveness and the efficiency of each substantive test, but few studies summarize the influential components on substantive tests. This essay addresses the question of whether moderate or weak SAPs combined with sampling could improve audit quality. Auditors might consider this idea when they choose appropriate substantive tests for certain clients. Finally, based on the argument this paper makes, regulator might need to consider possible unintended effects on audit quality caused by their regulations.

The remainder of this paper is divided into four sections: the next section summarizes SAPs and audit sampling including its criticism, the third section illustrates its research method, the fourth section describes the evaluation of SAPs in terms of audit effectiveness and efficiency, and the last section provides conclusions and limitations.

CURRENT ISSUES

Auditing standards indicate that audit evidence includes all information, whether obtained from audit procedures or other sources, used by the auditor in arriving at the conclusions upon which the auditor's opinion is based (AICPA 2006a). Substantive procedures are designed to obtain direct evidence about dollar amounts in account balances. (Louwers et al. 2013). Some substantive procedures must be performed in all audits, and substantive tests commonly consist of substantive analytical procedures (SAPs) and tests of detail.

Substantive Analytical Procedures

Analytical procedures are a test of the reasonableness of reported financial statement items. They are required to be conducted during planning and review, and often as the part of tests of detail or substantive analytical procedures. A principle underlying the use of analytical procedures is that reasonable relationships among data may be expected. Generally, as described in Figure 1, analytical procedures contain four steps:

1) Developing an expectation,

2) Deciding a tolerable difference (e.g., performance materiality or the desired level of assurance),

3) Comparing the difference between the expectation the auditor develops and the reported amount, and investigating substantial differences, and

4) Assessing explanations (e.g., from management or staff) and related evidence.

In the planning stage, analytical procedures are used to support the auditor in planning and extent of other auditing procedures. For instance, preliminary analytical procedures are used to determine the necessity and extent of further audit procedures. They are employed in the review stage as a global review of the financial information. As a substantive test, they are used to collect audit evidence about particular assertions related to account balances or classes of transactions. SAPs could be applied as substantive tests or corroborated with other test of details, such as audit sampling (AICPA 2012b). Stated differently, SAPs could offer audit evidence itself as a source on the reasonableness of an account balance and designate risky areas for further audit effort.

The effectiveness and efficiency of SAPs are decided by the 1) financial assertion being examined, 2) plausibility and the certainty of the relationship, 3) accessibility and reliability of data employed to set the expectation, and 4) accuracy of the expectation (AICPA 2012a). Thus, an expectation is effective and efficient in cases where the targeted account is generally predictable and has related reliable factors that could explain changes in account effectively. Figure 3.1 illustrates the framework of SAPs adapted from Messier et al. (2012) and the process for determining effective and efficient SAPs partly adapted from audit standard (AICPA 2012a).

SAPs offer assurance that partly depends on the accuracy of the expectation derived. In this line, Loebbecke and Steinbart (1987) distinguish between SAPs and preliminary analytical procedures by measuring the level of accuracy and false positives or negatives resulting from analytical models. Accordingly, the procedures to set expectations as substantive tests are more intricate than the procedures during planning and review. For instance, SAPs usually address a single account, unlike analytical procedures used during the planning or the review phase, which deal with either a single account or the financial statements as a whole.

Scholars have been aware of the benefit of analytical procedures in the audit process for a long time. Biggs and Wild (1985) find that over 40 percent of material errors were at first detected by analytical procedures. Wallace (1982) states that “if something is in error or if some key aspect of operations or the environment affect recorded accounting numbers differently than expected, such information is value of the auditors. Similarly, when the balance appears reasonable at some level of precision, a contribution is made to assessing the overall reasonableness of financial statements.”

Because the expectations in SAPs are often derived not only from the account balance in the prior year, but also the auditor’s knowledge and understanding of economic conditions, SAPs are considered important for business risk audit approaches (Bell et al. 1997; Eilifsen et al. 2001), which are increasingly popular (Trompeter and Wright 2010). Business risk audit approaches involve understanding the client’s business environment and the role of substantial transactions in terms of the client’s environment. SAPs are an effective way to integrate those perspectives by adding reliable sources. An auditor's expectations of financial accounts are derived from a comprehensive knowledge combining “the field of auditing, system theory, and business strategy” (Bell et al. 1997). In a similar line, Bell et al. (2005) argue that “evidentiary triangulation”, the combination of evidence from multiple sources, is particularity useful in improving audit quality when the auditor is concerned about international management fraud.

Figure 3.1 Overview of substantive analytical procedures

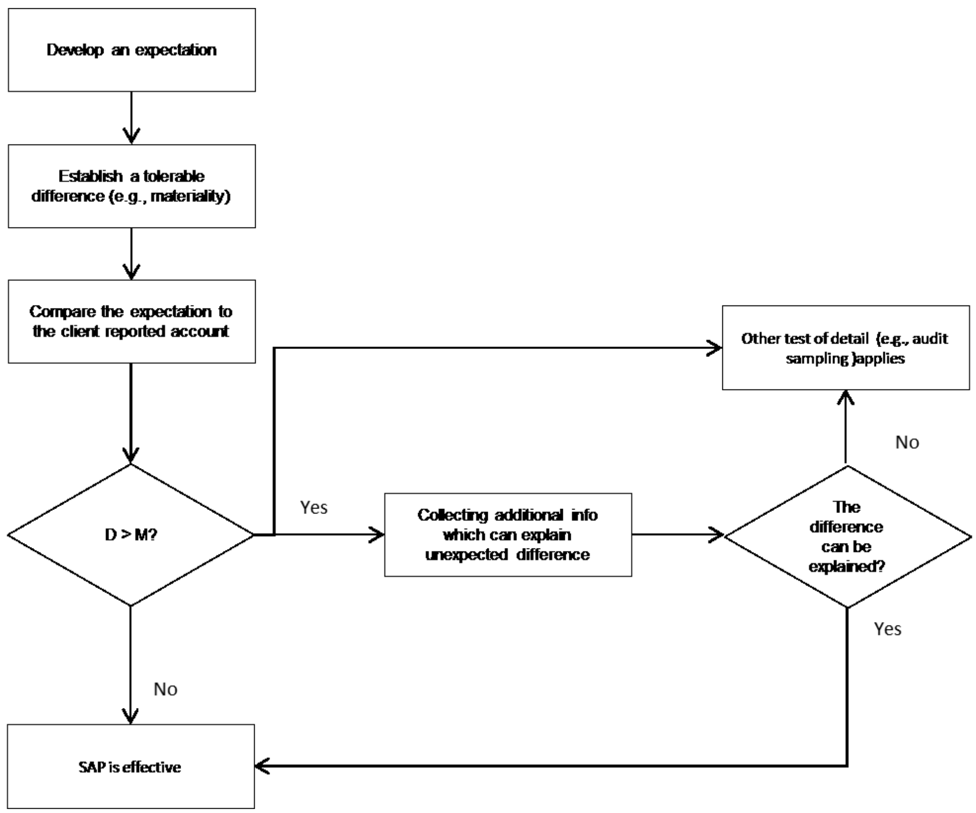


Illustration based on Messier et al. (2012) and AICPA (AICPA 2012a)

Where,
D indicates the difference between the expectation and the client's reported account;
M indicates that the tolerable difference.

Criticism of Substantive Analytical Procedures

SAP effectiveness is frequently influenced by the firm's operating environment. For instance, auditing standards indicate that relationships are more predictable in stable environments than in dynamic environments and with income statement accounts rather than with balance sheet accounts, but they are less predictable with transactions subject to management discretion (AICPA 2012a).

Some studies empirically show that economic stability and constant sales patterns affect the accuracy of expectations and the capacity for error detection (Chen and Leitch 1998, 1999).

In addition, the application of SAPs in an audit is directly affected by the level of client risk. Blocher and Willingham (1985) insist that SAPs offer a negative rather than a positive assurance, and should not be used when risk or materiality is high. Trompeter and Wright (2010) report that, in audit practice, there is a strong link between internal controls and dependence on SAPs. In high risk areas or when a client's internal controls are weak, the auditor is less likely to use SAPs, relying instead on tests of details (Trompeter and Wright 2010).

The value of SAP is susceptible to the specific environment. Analytical procedures might be valuable to identify risky areas that the auditor needs to explore in further tests of detail, but not as substantive audit evidence providing a high level of assurance. Lowers et al. (2013) argue that even though analytical procedures are valuable in terms of cost efficiency generally, auditors perceive analytical procedures as "soft" evidence, as opposed to "hard" evidence, such as recalculation, confirmation, and inspection of documents.

Audit Sampling in Substantive Tests of Details

Auditors have discretion in terms of the amount of audit evidence that must be collected. The auditor examines the entire population of items in an account, items having certain characteristics, and/or a certain number of items that are less than 100 percent of the total. According to audit standards (AICPA 2012d), examining the entire populations is useful when 1) the population consists of a small number of large value items; 2) the risk is very high; 3) other approaches for selecting items for examination do not offer "sufficient appropriate audit evidence"; or 4) testing full the population could be automated effectively. On the other hand, in certain cases, the auditors might select specific items, such as key or high-risk items (e.g., suspicious or unusual items), or items that exceed a certain amount, to validate a large part of the entire account value. Auditors may also select particular items to acquire knowledge about the nature of the client's business or transactions. Those two approaches are not referred as to audit sampling and are not used to project to the entire population.

Audit sampling is "the application of an audit procedure to less than 100 percent of the items within an account balance or class of transactions for the purpose of evaluating some characteristic of the balance or class" (AICPA 2012b). Audit sampling is used in two fundamental ways. It can be used, first, as a test of internal controls (attributes sampling) to decide whether these controls are performing effectively in preventing or detecting and correcting misstatements. Second, audit sampling can be used as a substantive procedure (variables sampling) to decide whether an account balance or class of transactions is properly recorded. In this study, variables sampling is explored.

Figure 3.2 illustrates the steps of the audit sampling process and factors that the auditor should consider during the procedures. This model is adapted from (Elder et al. 2013). The major steps of audit sampling are as follows:

1) The planning phase: The auditor determines the audit objective, defines the error or the misstatement, and decides on the sample size based on applied sampling method (statistical/nonstatistical).

2) The performing phase: The auditor selects sample item and performs audit sampling. If any misstatement is recognized, then the auditor conducts further investigations.

3) The evaluation phase: The auditor formulates the projected misstatement and upper limit on the misstatement, and concludes the acceptability of the population.

Substantive tests of detail are considered to provide potentially high levels of audit assurance. An auditor has the vital tasks of drawing conclusions regarding the precision of the client's reported transactions, but in many cases, the auditor cannot inspect the full population, necessitating sampling to gather audit evidence regarding transactions or account balances (Rittenberg et al. 2010).

However, not all audit procedures are appropriate for sampling. For example, inquiries, observations, and SAPs would be difficult to perform by sampling (Louwers et al. 2013). While audit sampling is not required as a substantive test of detail, sampling is normally employed for any balance or class of transactions when target testing cannot effectively conduct a test of detail in practice (Christensen et al. 2015).

Criticism of Audit Sampling

Since audit sampling as tests of details does not directly assure the total population, it is significant to control risks related to audit sampling. Well-designed sampling can allow the auditor to make valid inferences about the total population, but limiting circumstances can reduce the value of sampling. If the selected observations represent the population well (sampling risks), and human errors in the application of sampling procedures are well controlled (nonsampling risks) (Guy 1981), then the outcome of audit sampling could provide appropriate audit evidence.

Teitlebaum and Robinson (1975) make the point that the auditor needs to consider factors like the nature of the population, including error distributions. Specially, they argue that the auditor is required to deal with three simultaneous difficulties: highly shewed population, rare occurrence of error, and limited sample sizes necessitated by time and cost constraints. In a similar line, Smieliauskas (1986) notes that the outcome of statistical sample approaches is affected by the book value distribution, the quantity of strata, the nature of stratification, and the materiality level used. Therefore, the auditor should project the distribution of book values and errors sufficiently and then choose the sampling method. Nevertheless, the auditor, particularly in the first year of audit, is often not aware of the required information needed to achieve effective audit sampling.

In summary, the intrinsic limitations of audit sampling include the following. First, sampling risks cannot be eliminated. There is no perfect approach by which an auditor can project the populations by using a subset of the population. Second, because of the characteristics of financial populations (e.g., positively skewed distribution), it is often difficult to apply a statistical sampling approach. Third, if the auditor strictly considers sampling assumptions, the number of samples is likely to be large. Accordingly, the efficiency of audit sampling has been brought into question (Kachelmeier and Messier Jr 1990). In addition, Bell et al.(1997) argue that expectations at the entity level developed from details of transactions samples might have limited value since the sampling approach does not reflect an understanding of the client's competitors and business risks.

Figure 3.2 Step of audit sampling

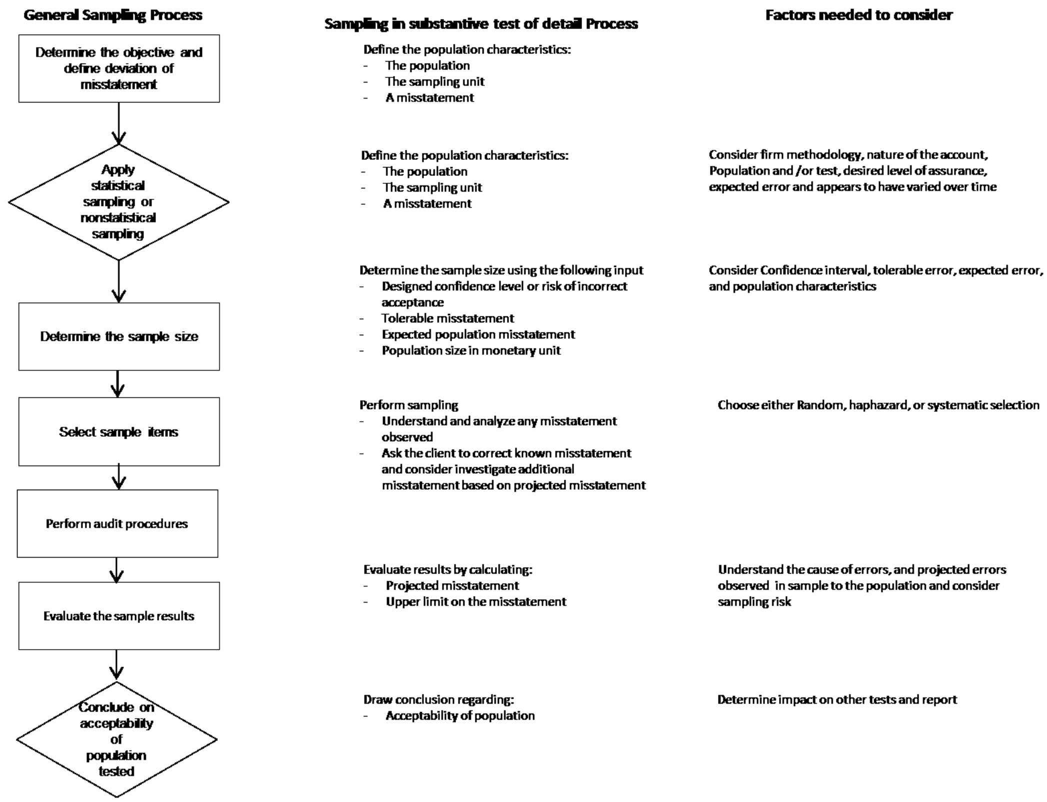


Illustration based on Elder et al. (2013)

Substantive Analytical Procedures and Audit Sampling for Testing Revenue Account

Since the level of assurance delivered by SAPs is related to the accuracy of expectations derived from them, it is significant to establish an appropriate accuracy threshold based on materiality and the level of assurance desired from the procedures (AICPA 2012a). Because of the ambiguity of the term "level of assurance desired from the procedure", there have been long discussions to clarity the level of assurance that SAPs should provide (Glover et al. 2015). If the client's control risk is not high, a traditional approach to obtaining substantive evidence is to integrate the evidence obtained from SAPs with the substantive evidence obtained through detail testing (Glover et al. 2015). Discussions regarding the level of assurance desired from SAPs has determined that SAPs provide a different level of assurance (i.e. moderate or low assurance) (AICPA 2012a). That is, if the evidence from SAPs is not sufficiently persuasive, then an auditor should consider corroborating it with other audit procedures.

Nevertheless, the PCAOB has raised the issue of the insufficient performance of analytical procedures as substantive tests (Messeir et al. 2013b; PCAOB 2008). with particular concern in the case of revenue accounts (PCAOB 2014). The PCAOB find that auditors unsuccessfully examined differences that exceeded the materiality threshold (PCAOB 2011). Based on interviews with audit firm partners, Glover et al. (2015) report that they understand the inspectors' point regarding the insufficient SAPs, but that SAPs can deliver substantive audit evidence even if the difference between the expectation derived from the SAPs and the reported account value does not exceed materiality. Christensen et al. (2015) also reported that, in response to these increased concerns, some firms encourage tests of detail for the revenue account. Trompeter and Wright (2010) anticipate in advance that auditors would consider detail tests as relatively justifiable audit evidence to evade possible issues with the PCAOB. In this line, Christensen et al. (2015) indicate that future research is needed to examine the cost and benefits of replacing audit sampling with SAPs. In response to this issue, this essay explores the costs and benefits of audit sampling and SAPs.

On the other hand, Glover et al. (2015) argue that reduced usage of SAPs due to the PCAOB's concerns might actually reduce the audit quality. In addition, they maintain that it is very difficult to achieve a high level of assurance for large value income statements accounts like revenue, and that SAPs providing even a low or medium level of assurance could still deliver meaningful audit evidence when corroborated by other forms of audit evidence. However, since auditors are still less likely to use regression models than ratio tests (Trompeter and Wright 2010), it is uncertain whether the problem is that SAPs inherently cannot offer a high level of audit assurance for revenue accounts or that the auditor's failure to conduct sufficient SAPs are largely due to their overly simple SAP approaches.

RESEARCH METHOD

To understand the costs and benefits of audit sampling in place of substantive analytical procedures, it is necessary to examine the value of substantive analytical procedures and audit sampling in substantive tests of details to identify the probability of misstatements. By exploring prior studies on audit efficiency and effectiveness, the costs and benefits of each substantive test are addressed.

In addition, as discussed above, since a major concern of the PCAOB is that auditors fail to establish precise expectations, the outcomes of prior studies in the SAP domain is examined. First, the literature was searched in Google Scholar by using key words "substantive analytical procedures" and "audit analytical procedures[6]" without time limitation. Behavioral studies are not included[7]. To evaluate the performance of SAPs, meta-analysis is conducted. Meta-analysis, referring to "the analysis of analyses", is the statistical analysis of findings from studies in order to combine the results (Glass 1976). Generally, prior studies evaluate the effectiveness of SAPs by addressing: 1) How accurate the expectations derived from SAPs are; and 2) How well SAPs detect misstated

[6] Many studies did not indicate whether the study is for SAPs or preliminary analytical procedures.
[7] Messier et al. (2013b) summarize behavior studies in the analytical procedures domain.

revenue. A survey of the literature will help to evaluate these issues. While it is unnecessary to assume that the measurements of audit effectiveness used in various studies of SAPs are comparable, these might provide overall insights of the performance of audit procedures to identify misstatement.

The outcomes of Mean Absolute Percentage Error (MAPE) were used to evaluate the accuracy of expectations derived from SAPs and formulated as follows. Absolute Percentage Error (APE) of each item is calculated by an absolute difference between a predicted value and an actual value divided by the actual value. The average of APEs is called as MAPE. Fundamentally, it represents the absolute percentage of the actual value that is not represented in the predicted value. Consequently, smaller MAPE is more desirable. However, information auditors' materiality threshold is not publicly available. In addition, the detection of misstatement considers not only materiality, but also other qualitative information (e.g., economic condition). As a result, it is difficult to generalize materiality thresholds. Since the overall materiality is generally 0.2-2 percent of a revenue account (Eilifsen and Messier 2015), this paper considers materiality as two percent of the revenue account.

The results of false negative (Type 1 errors) and false positive (Type 2 errors) related to the revenue account were evaluated to assess how well SAPs detect misstated revenue. False negatives and false positives are related to misclassification of misstated items or the correct items with SAPs. Many prior studies test various settings of experiments using various sizes of seeded errors and alpha and/or the distributions of errors. Depending on the settings, the outcomes of error detections vary. Accordingly, because of a wide variety of experimental settings, this paper does not include the statistical analysis of false negatives/positives from prior studies. Not all studies on this topic could be included for meta-analysis because of a lack of necessary information (e.g., standard deviation of absolute percentage errors[8]). In addition, some studies that do not include the revenue account in the test were excluded (e.g., Kogan et al. 2014). To evaluate the consensus of MAPEs among prior studies, seven studies[9] are selected as shown in Table 3.1. In summary this essay explores whether audit sampling outperforms SAPs for revenue accounts.

[8] Many studies do not show the standard deviation of absolute percentage errors (APEs) to formulate a meta-analysis for mean absolute percentage errors. A few studies show the required values only for subsamples. In this case, the standard deviations of APEs are based on the APEs of the subsamples.
[9] Two studies are the minimum number of studies to conduct meta-analysis (Valentine et al. 2010).

Table 3.1 Summary of performance of SAP in prior studies

	Research	Journal	#of firms	Best model[10]	Added variables
(1)	Kinney (1978)*	TAR	6	ARIMA	
(2)	Lev (1980)*	JAR	573	Regression	macroeconomic index
(3)	Kinney & Salamon (Kinney and Salamon 1982)	JAR	200 simulated data	Regression with Statistical Techniques for Analytical Review (STAR)	
(4)	Loebbecke and Steinbart (1987)*	AJPT	38	Sub martingale	
(5)	Kinney (1987)	AJPT	1	Pattern analysis of cross sectional changes in several ratios	
(6)	Knechel (1988)	TAR	2	Regression with Statistical Techniques for Analytical Review (STAR)	
(7)	Wheeler and Pany (1990)	TAR	5	X-11	
(8)	Lorek et al. (1992)*	AJPT	78	ARIMA	
(9)	Dzeng (1994)*	AJPT	1	VAR	related lagged account and macroeconomic index
(10)	Chen and Leitch (1998)	AJPT	5	Stepwise regression	related account and macroeconomic index
(11)	Chen and Leitch [11](1999)	AJPT	90	Stepwise regression	related account and macroeconomic index
(12)	Allen et al. (1999)*	AJPT	1	Regression	accounts of peer store, nonfinancial information, and related account
(13)	Leich and Chen (2003)	AJPT	3	Structural Equation Model	related account
(14)	Hoitash et al. (2006)*	AJPT	5,747 quarterly observations	Time series model	accounts of peer firms and related account
(15)	Vandervelde et al. (2008)	AJPT	4	Stepwise regression with related accounts	related account
(16)	Kogan et al. (2014)	AJPT	1	Regression	related account

TAR: The Accounting Review
JAR: Journal of Accounting Research
AJPT: Auditing: A Journal of Practice & Theory
* indicates the paper used for meta-analysis.

EVALUATION

Audit efficiency

The cost of SAPs is mainly related to computational efforts. On the other hand, the cost of audit sampling as a substantive test consists of a fixed cost when conducting a

[10] Depending on the measurement, the best model is often different. The models selected as the best model in this study is based on the MAPE values.
[11] This study doesn't provide a specific MAPE for the revenue account.

sampling application and a variable cost when connected to the number of transactions tested (Kinney 1979). Obviously, audit sampling is more efficient than testing 100 percent of a population, but it might not be more efficient than SAPs. Specifically, the application of SAPs could reduce audit hours in less risky areas (Biggs et al. 1989).

As a common audit approach for revenue accounts, when an SAP detects an area of discrepancy, the auditor then uses a test of detail in this area (e.g., Statistical Technique for Analytical Review (STAR)[12] rule). Consequently, if audit sampling is conducted without an SAP, the number of observations the auditor needs to select and examine would be increased, thereby increasing audit cost and lowering audit efficiency. Along this line, Kinney (1979) and Knechel (1988) pointed out that the number of samples used without SAP is larger than the number used with SAPs.

Audit effectiveness

How Do SAPs Improve Audit Effectiveness for Revenue Accounts?

Unlike audit sampling, SAPs fundamentally examine the total population (Knechel 1986). Accordingly, a precise SAP might provide higher level of assurance than audit sampling. In addition, in certain audit environments, an SAP could improve audit effectiveness. Knechel (2007) argue that

> "Alternative audit approaches were needed in an environment where millions of transactions occurred in a short period of time and were processed at the speed of technology without leaving a paper trail to be observed at the convenience of the auditor. Two traditional sources of audit evidence could have helped with these complexities – testing internal controls and analytical tests of overall results – but their increased use led to further questions about the conventional wisdom of auditing." (p. 390)

[12] Prior studies suggest the STAR investigation rule by investigating any month for which upper confidence limit exceeds the desired thresholds (Kinney 1979).

This argument has been frequently applied to the revenue cycle. A number of transactions in certain industries, such as retail or financial services, occur in a timely fashion. If the auditor uses only a small number of transactions, it is difficult to conclude that audit sampling sufficiently represents such a large population. As indicated previously, if the auditor attempts to satisfy statistical sampling assumptions for this type of population, then a large number of observations must be examined, which may be impossible to do if conducted manually. In a case where the auditor does not need to confirm transactions by conducting manual audit procedures (e.g., sending a confirmation email and checking a response from the customer) and SAPs could confirm the existence of transactions by using plausible relationships with related information, then the auditor might not be required to gather sufficient and appropriate audit evidence by employing tests of detail such as audit sampling in substantive test of details.

By showing the accuracy of analyst forecast for revenue, Glover et al.(2015) insist that it is fundamentally difficult to predict revenue accounts precisely. Unlike analysts, the auditor can access a client's internal information, so the expectations for revenue generated by the auditor might be more precise than the expectation formulated by analysts. Nevertheless, if the auditor heavily relies on internal information provided by management, the auditor might fail to set the appropriate expectation based on professional skepticism (Griffith et al. 2015b). Consequently, to the auditor should take account of various sources to prevent possible biases. As discussed previously, unlike SAPs, audit sampling has limited ability to consider these sources.

How Accurately Do SAPs Predict Revenue?

To examine the second research question, meta-analysis is conducted. As a preliminary assessment, the degree of heterogeneity among the findings of selected studies is ascertained by formulating the I^2 statistic. Since a non-zero I^2 is founded and each study examines the effectiveness of SAPs by using the different level of disaggregated data and a different scope of information, random effect models that restrict maximum likelihood estimators are applied to integrate outcomes.

Table 3.2 Results of meta-analysis

Estimate	Lower bound	Upper bound	Std. error	p-value
0.054	0.022	0.085	0.016	<0.001

Figure 3.3 Forest plot of meta-analysis for MAPE

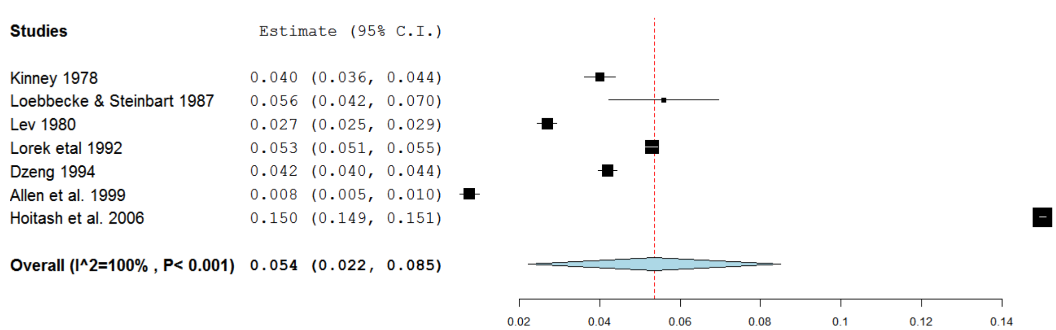


Figure 3.3 illustrates a forest plot, a graph for displaying MAPE values and associated confidence intervals from prior studies. Each square represents the outcome of

each study. When outcomes are compared in a meta-analysis, each result is not given equal weighting (e.g., studies with wide confidence intervals tend to be given a lower weighting than those with narrow confidence intervals). In the forest plot the size of square indicates the weight; a larger square indicates a higher weighting and *vice versa*. The center line of diamond indicates the summary estimate (i.e. 0.054) and lateral tips of diamond represents the associated confidence intervals.

As the forest plot shows, the accuracy levels of SAPs are generally not within the materiality range. However, by adding external indicators (e.g., macroeconomic indicator) and contemporaneous relevant accounts (e.g., account receivable), and by using more disaggregated data (e.g., store-level), the predictive accuracy of SAPs is likely to be increased. Since Hoitash et al. (2006) examine a large number of firms, MAPE values vary and are higher than for other studies. Therefore, based on prior studies, more rigorously developed SAPs might provide a higher level of assurance. However, these results might not sufficiently present the effectiveness of SAPs.

Even though prior studies indicate that the auditor should set expectations with the absence of knowledge of unaudited amounts because it can cause the auditors to be biased toward management assertions (Pike et al. 2013), the recent trend of empirical studies on SAPs tends to use contemporary unaudited accounts (e.g., account receivable or cost of goods sold) in models to set expectations (Hoitash et al. 2006; Chen and Leitch 1998, 1999; Vandervelde et al. 2008). These studies assume that unaudited accounts do not contain material errors and then seed errors based on common fraud schemes, so their approaches are appropriate. The benefit to the auditor from adding contemporaneous relevant accounts (e.g., account receivable or cost of goods sold) in statistical models is gaining an understanding of the client's business environment, such as market conditions in that year, which revenue amounts in previous years could not offer. In addition, "uncoordinated errors" can be captured by those models. For example, management may have added fake revenue transactions, thereby increasing revenue and account receivable, but does not formulate the cost of goods sold and inventory amounts related to the faked revenue transactions. In that case, adding accounts receivable in the model to predict the revenue account could identify misstated revenue. On the other hand, models containing relevant contemporaneous accounts could not detect such coordinated errors. For instance, if management has manipulated all possible related accounts, then there is no way to capture misstatements by adding relevant contemporaneous accounts.

Accordingly, it seems obvious that the experimental settings of academic literature testing the effectiveness of SAPs could not reflect genuine audit environments. In addition, the outcomes of many prior studies are not based on real data, but are based on simulated data generated by using quarterly financial statement accounts. Moreover, in practice the auditor commonly uses simple ratio tests as well as a wide variety of sources of information (Trompeter and Wright 2010). Consequently, it is possible there is a performance gap for SAPs between the outcomes of academic studies and the results of audit practice. However, those studies suggest that, if the auditor sets expectations based on rigorous SAPs, it is possible for the SAPs to be effective.

Even if the size of errors resulting from SAPs exceeds the threshold of materiality (i.e. SAPs provide moderate or weak assurance), could SAPs deliver values to improve audit quality? Because it is impossible to identify whether the variance between the auditor's analytical expectations and the client's unaudited is caused by genuine errors, it is difficult to reach a conclusive answer. Stated differently, it is possible that the expectations from rigorously developed SAPs could not be closed enough to the reported revenue account. In this case, the auditor might consider the variations as “red flags” and conduct tests of detail. For example, since a firm’s sales pattern is generally well structured, its SAPs commonly provide a high level of assurance. Suppose that, in this audit period, the SAPs of revenue offer only a moderate or low assurance, indicating that existing models cannot provide clear expectations of sales. In this case, moderate or low SAPs could be considered as a signal of financial statement fraud. Audit standards note that if the outcomes derived from SAP show unusual or unexpected account fluctuations, the auditor should study the possibility of financial statement fraud (AICPA 2012a). Thus, moderate or weak SAPs could provide evidence. Accordingly, it may be rash to conclude that moderate or weak SAPs do not deliver any assurance for revenue account.

Does Audit Sampling outperform Substantive Analytical Procedures?

Both SAPs and audit sampling trade effectiveness for efficiency. Accordingly, since SAPs and audit sampling include different types of intrinsic risks, it is difficult to determine which procedure outperforms the other. The significant issue when gauging performance is how to handle those intrinsic risks. Table 3.3 illustrates possible risks related to each substantive test.

SAPs contain two types of risks: false positive and false negative. If the auditor classifies correct transactions or account balances as misstated, that is false positive. False negative arises if the auditor judges misstated transactions or account balances as correct.

Audit sampling carries two risks: sampling risk and nonsampling risk. Sampling risks are connected to both false positives and false negatives. False positives occur if the auditor concludes that a fairly stated population (i.e. class of transactions or account balance) is misstated. False negatives refer to situations in which the misstated amount of the population exceeds the tolerable level, but the auditor concludes that the population is fairly stated (Louwers et al. 2013).

Table 3.3 Risks of SAPs and audit sampling

Risk	Description of risks	SAP	Audit Sampling
False Positive	Risk of incorrect acceptance	✓	✓
False Negative	Risk of incorrect rejection	✓	✓
Nonsampling risks	Risk of an auditor's incorrect judgment or execution		✓

Risks related to false positives and negatives are related to the performance of SAPs and audit sampling. The performance of SAPs is largely related to employing an appropriate statistical model that can properly reflect the trend of sales and the relationships with independent variables, and adding reliable variable(s) that can explain the sales account and other factors (stable environment and so on). Some factors are given, but risks associated with determining proper statistical models can be handled by the auditor. Finding a trustworthy variable that has predictive value for sales account might not be possible because, in some cases, the auditor fails to find any reliable predictor highly correlated with sales.

The performance of audit sampling is associated to sampling approaches (statistical or nonstatistical), sample size, and sample selection (random, systematic, or haphazard). Some significant factors and costs/benefits related to this decision are illustrated in Table 3.4. It is important to note that nonsampling risks stems from the auditor. Unlike sampling risk, nonsampling risk depends on an audit firm's quality control (Puttick et al. 2008). Accordingly, utilizing sampling effectively requires careful approaches for each step of procedures. Neter and Loebbecke (1975) conclude that all sampling approaches do not perform equally well in every case, but error rates in populations, the effectiveness of sample selections, and evaluation methods affect sampling risks. In addition, Caster et al.(2000) list factors affecting the effectiveness of audit sampling other than sampling and nonsampling risks: "auditing and firm standard, decision tool/aid, an auditor's knowledge, experience, motivation, ability, ethics, decision heuristics, and environmental factors."

In summary, risks arising during audit procedures cannot be completely eliminated and are generally unknown in advance, but by applying appropriate approaches, the auditor might reduce these risks. Accordingly, the comparison of effectiveness between SAPs and audit sampling lies in how effectively the auditor adjusts risk by utilizing appropriate substantive tests and applying tests correctly.

In certain circumstances, audit sampling might outperform SAPs, and in other cases, SAPs might outperform audit sampling. However, it is important to note that in many instances, SAPs and audit sampling are complementary. For example, even though SAPs offer a high level of assurance, it is possible that evenly distributed small errors are less likely to be captured by an SAP. On the other hand, if the small percentage of errors is widely distributed in the transactions audit sampling performs better than SAPs[13]. Accordingly, as traditional audit approaches suggest, the combination of SAPs and audit sampling could provide sufficient and appropriate audit evidence by itself.

As Figure 3.1 shows, SAPs are often conducted first, and in case SAPs find unexpected, unusual account functions, then other approaches are considered. It shows that the auditor might judge SAPs plus audit sampling to be less effective than substantive tests in advance based on circumstance, but cannot forecast perfectly whether SAPs would be strong in advance. Even though the expectation of sales account value is difficult to determine precisely, SAPs could be beneficial under certain circumstances. In summary, Table 3.5 presents the conditions affecting the use of SAPs.

[13] Sampling approaches should be carefully decided as well based on error tainting magnitude, the percentage of errors in the recorded amount of a sample item. For instance, Hoogduin et al. (2015) argue that if error tainting magnitude is small (i.e., 50 percent or less) and items are randomly ordered then Monetary Unit Sampling (MUS) with systematic selection performs better than MUS randomized systematic selection.

Table 3.4 Factors the auditor need to consider for audit sampling procedures

Panel A: Cost and benefits related to nonstatistical approache

Cost	Benefits
Audit judgments to determine sample items are required.	An additional software is not required
No objective way to control and measure sampling risk	Auditor's prior expectations regarding errors in account can be reflected.
	Less time to plan, select, evaluate the sample

(Adapted from Rittenberg et al. (2010))

Panel B: Factors affecting to decide statistical sampling approaches

Factors	Monetary Unit Sampling	Classical Variable Sampling
Overstatements are more concerned	✓	
It is difficult to estimate standard deviation	✓	
No or few misstatements are anticipated	✓	
A low level of variation with line item values		✓

(Based on Lowers et al. (2013))

Panel C: Factors affecting sample size decisions

Factor	Conditions leading to smaller sample size	conditions leading larger sample size	Related factor for substantive sample planning
Inherent risk	Low assessed level of inherent risk.	High assessed level of inherent risk	Allowable risk of incorrect acceptance.
Internal control weakness	Low assessed level of control risk.	High assessed level of control risk.	Allowable risk of incorrect acceptance
Reliance on other substantive test	Low assessment of risk associated with other relevant substantive procedures.	High assessment of risk associated with other relevant substantive procedures.	Allowable risk of incorrect acceptance
Measure of tolerable misstatement for a specific account.	Larger measure of tolerable misstatement.	Smaller measure of tolerable misstatement.	Tolerable misstatement
Expected size and frequency of misstatements.	Smaller misstatements or lower frequency	Larger misstatements or higher frequency.	Assessment of population characteristics.
Number of items in the population.	Virtually no effect on sample size unless population is very small.		
Choice between statistical and nonstatistical sampling	Ordinarily, sample sizes are comparable.		

(Adapted from AICPA 1983 AU 350)

Table 3.5 Conditions related to effective and efficient SAPs

Question	SAP applies
A client's internal control risks are high.	No
The population is extremely large.	Yes
Sales are highly related to business risk factors	Yes
The trend of sales is structured	Yes
It is expected that a large size of errors is randomly distributed	Yes

SUMMARY AND DISCUSSION

Even though prior studies have explored the value of using sophisticated SAPs, PCAOB inspectors have increased concern regarding SAPs , possibly because, in practice, auditors are still less likely to use even relatively simple statistical models like regression (Trompeter and Wright 2010), thereby creating deficient predictions of the account. If the auditor conscientiously conducts SAPs, these concerns can be avoided. Furthermore, skipping SAPs to avoid possible negative inspection outcome might not be the intent of the PCAOB.

By exploring prior studies, this essay attempts to respond to this current issue and to indicate the factors that might impair audit quality. If internal controls are reliable, SAPs might be more effective than audit sampling in case where the number of line items are extremely large, the external business conditions can be considered to the examine revenue cycle, the sales trend is structured, and/or the large errors are randomly distributed. On the other hand, certain audit sampling approaches are more effective than SAPs when small errors are evenly distributed in large areas, internal control weakness is a concern, and/or materiality/risk is high. If internal controls are reliable, then even a moderate or weak SAP could still reduce the scope of audit sampling. In addition, in some cases, the level of assurance of the SAPs (i.e. high/moderate/low) could provide audit evidence in itself.

SAPs play an important role in business risk audits and have been known to improve audit effectiveness and efficiency.

By showing the accuracy of analyst forecast for revenue, Glover et al.(2015) insist that it is fundamentally difficult to predict revenue accounts precisely. Unlike analysts, the auditor can access a client's internal information, so the expectations for revenue generated by the auditor might be more precise than the expectation formulated by analysts. Nevertheless, if the auditor heavily relies on internal information provided by management, the auditor might fail to set the appropriate expectation based on professional skepticism (Griffith et al. 2015b). Consequently, to the auditor should take account of various sources to prevent possible biases. As discussed previously, unlike SAPs, audit sampling has limited ability to consider these sources.

In practice, it is important to apply SAPs and audit sampling carefully based on factors that might influence the effectiveness and the efficiency of substantive tests. This study extends the audit literature in terms of sampling and analytical procedures. Prior studies commonly examine the significance of each substantive test during the audit or suggest how to enhance the test. In addition by responding to the concerns of the PCAOB, this study illustrates the unintended consequences of regulations. Therefore, this paper highlights that the inspector and the auditor need to reconsider the value of SAPs in auditing revenue account. Nevertheless, this study only examines prior studies.

Accordingly, as future research, it could be valuable to examine empirically how SAPs and audit sampling could improve audit quality in certain circumstances suggested in this paper.

Chapter 4 Weather Variables as Audit Evidence

INTRODUCTION

This paper examines two questions related to the relevance of weather indicators in substantive analytical models: (1) Are weather variables correlated with a sales? (2) Do weather variables contain incremental information to enhance analytic expectations?

The purpose of this study is to advance substantive analytical procedures by adding weather indicators as an additional, external information source. The effect of weather on business activities and related human behavior has been explored in various business domains: marketing, economics, and even finance. Retail firms are especially sensitive because unfavorable weather discourages store visits, thereby reducing sales. Accordingly, weather variables could be relevant audit evidence for revenue accounts in the retail industry.

Analytical procedures are required at the planning and review phases of the audit and are recommended in substantive testing (AICPA 2012a). Auditors examine a client's historical data and other relevant information during these procedures to allow them to consider the reasonableness of financial results, thereby offering a more comprehensive view. The literature indicates that substantive analytical procedures have the power to discover misstatements and irregularities, thereby improving the effectiveness of audit procedures (Biggs et al. 1989; Knechel 1988; Messier et al. 2013b).

In response to increased demand for audit effectiveness and efficiency, researchers have tried for decades to determine which information can enhance analytical power. Their efforts include the use of nonfinancial information, in accordance with AICPA recommendations (AICPA 2012a). Prior studies have highlighted that the auditor needs to integrate information from a variety of sources in order to enhance audit quality (e.g.,Griffith et al. 2015b). Trompeter and Wright (2010) argue that advanced technology has caused noticeable changes in audit approaches and analytical procedures. Auditors are likely to utilize a variety of nonfinancial information to identify fraudulent financial statements. For instance, analyst reports and internet searches are often used as indicators for analytical procedures.

Nevertheless, previous studies suggest that auditors' analytic expectations are not directly affected by nonfinancial information, but are used as supplementary components (Cohen et al. 2000; Brazel et al. 2012). Luft (2009) points out that challenges to the use of nonfinancial information mainly relate to difficulties in precise measurement and appropriate weighting. Consequently, the importance and proper employment of nonfinancial information in analytical procedures remain unclear.

Auditors can benefit from utilizing weather indicators as audit evidence. Weather data is timely and easily available during the audit procedures. Weather information is updated regularly and is location-specific. This makes it suitable for sophisticated time and location models. Furthermore, the existing literature has been limited to utilizing nonfinancial information from internal sources, such as the number of employees or the size of locations. This data might not improve the effectiveness of analytical models because these types of internal client sources are susceptible to fraudulent management manipulation. Thus, financial "red flags" might be obscured or missing in a client's accounts. On the other hand, weather indicators, a form of external nonfinancial information, can be used as independent benchmarks because they are not affected by internal account errors and data restrictions.

By using daily store-level sales data for a retailer operating in multiple U.S. locations, this essay examines whether adding weather variables to various statistical models enhances these substantive analytical procedures for revenue accounts. Empirical results show that weather indicators such as temperatures, wind speed, and humidity explain around 90 percent of weekly sales. This analysis shows that the sales patterns of the studied firm correlate with the weather over the relevant time period. Accordingly, the sales patterns might be related to weather variables, although it is highly probable that other variables also affect the sales trend. In addition, a variable derived from peer stores sharing similar macroeconomic characteristics have more powerful values in forecasting store sales than weather variables.

This study is an extension of Kogan et al. (2014) and Allen et al. (1999) in that it examines data disaggregated by time and location, but it differs in two respects. First, it adds weather indicators to enhance the accuracy and the precision of analytical predictions. Even though weather can affect revenues, prior studies have rarely looked at its significance for business activity. Furthermore, weather indicators can provide more reliable and timely information for examining revenue accounts than traditional audit evidence. Accordingly, utilizing weather indicators has great potential to improve the effectiveness of substantive analytical procedures. This study can offer useful insights into audit practice and academia in terms of the significance of weather variables as audit evidence and the approaches to utilizing it in audit procedures. In addition, this study uses data comprising a large number of stores in the U.S., providing the necessary heterogeneity to improve expectations. Allen et al. (1999) do not find the enhancement of analytical procedures by utilizing store-level disaggregated data and explain that, due to the homogeneous nature of operation across thirty stores at a given firm, they cannot find substantial difference between the disaggregate and aggregate approaches. Nevertheless, even though individual stores may offer similar services and products, local customer characteristics and economic conditions for each store create a variance in sales that would be different from stores in other areas. This paper contributes to reassessing the value of disaggregated data by location.

This paper is organized as follows. The first section discusses the related academic literature and articulates the hypotheses that this study tests. The second section discusses research methodology, and the third section presents the major findings. Finally, the last section summarizes and discusses the results, and also suggests future research opportunities.

LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Weather and Sales in Retail

The economic influence of weather has long been studied. Over thirty years ago, Engle et al. (1986) found the relationship between temperature and electricity sales. By analyzing 70 years of data, Lazo et al.(2011) show that overall around three percent of U.S. annual GDP is affected by weather variability. Tol (2000) documents the influence of weather on the consumption of energy and the tourism industry. Even though certain sectors such as retail, fashion manufacturing, service, and transportation are considered weather-sensitive, weather may affect the outcome of business activities differently depending on the industry (Larsen 2006). For instance, precipitations might hinder customers' visit to stores, but tourism could benefit from gloomy weather.

The literature shows that retail is one of the most weather-sensitive sectors. Marketing and economic studies have examined the association between weather and retail sales (Murray et al. 2010; Larsen 2006; Dutton 2002). For example, Starr-McCluer (2000) examines various types of retailers (e.g., sellers of durable/nondurable goods) and concludes that overall temperatures have a relatively strong explanatory power for retail firms' sales. She argues that unfavorable weather conditions, such as cold temperatures or precipitation, can interrupt customers' store visits, reducing sales. Outcomes for retailers in general, and especially for durable goods outlets, are affected by unfavorable weather conditions. Nevertheless, the influence of lagged unfavorable weather is likely to be offset in every quarter.

Management often discusses risks related to weather in MD&A. The word "weather" appears in 50,704 10-K filings from 1994 to 2015. For instance, the 2010 10-K report of Friendly's Restaurants explains that "results for the year were negatively impacted" in part by "unusually cool weather in the northeast, especially in the summer months". Nike Inc. states in its 2015 10-K report that "weather events…impacted two of the three major holiday periods of the 2014/2015 ski season and adversely affected the ski industry in general."

Substantive Analytical Procedures and Nonfinancial Information

A fundamental assumption underlying the use of analytical procedures is that there are reasonable relationships in financial data and nonfinancial data. Accordingly, auditors develop expectations based on their examination of a client's industry and business. They then compare their prediction to the client's actual account balance. Auditors conduct further investigations when they recognize a material difference between their expectation and the client's reported amount. Types of analytical procedures used by auditors can be classified as planning, substantive, or review. This study concentrates on analytical procedures as substantive tests. Even though analytical procedures are only required during planning and review phases, the desire to improve audit effectiveness and efficiency leads auditors to adopt analytical procedures during the substantive testing phase as well (Tabor and Willis 1985; Wright and Ashton 1989).

Audit standards suggest that nonfinancial information should be considered when performing analytical procedures (AICPA 2012a). Similarly, the literature suggests that nonfinancial information can offer supplementary evidence to overcome existing auditing procedures. For example, if management manipulates earnings to beat forecasts, then the earnings would be similar to or higher than the forecasted earnings. Management can also be motivated to be consistent with industry trends and/or its budget (Beneish 1997). It is possible to miss such fraudulent behavior using only financial information provided solely by management. In this vein, the PCAOB indicates that "red flags" can be missed when management manipulates accounts to reach expected trends (PCAOB 2004). Studies have shown that the auditor needs to collect and examine various sources of information (Bell et al. 1997; Bell et al. 2005; Griffith et al. 2015b).

On the other hand, the effects of nonfinancial information on analytical procedures are still in question. Cohen et al. (2000) find that auditors are more likely to depend more on financial trends than on nonfinancial trends when deciding the audit scope for analytical procedures. Also, Brazel et al. (2014) show that auditors react to inconsistency between financial factors and nonfinancial factors, but few use nonfinancial information itself in analytical procedures.

Some studies link nonfinancial information and audit procedures, but nonfinancial information is often limited to client-provided sources. Allen et al.(1999) use the number of pounds of product serviced and the number of working days in their model, and Brazel et al. (2014) conduct behavioral experiments with senior auditors to examine the impact of nonfinancial information, such as the number of employees and production spaces. However, there are limitations to the use of nonfinancial information in the research setting because they do not see a dramatic variance over the period. For instance, the production space that Brazel et al. (2014) use might not change on a weekly or monthly basis, but remains almost fixed over the entire period. These measures might be useful to understand the size of operations or business strategy, but they do not provide sophisticated information to examine operations on a timely basis. On the other hand, the nonfinancial information that Allen et al. (1999) use in their model vary over the period, but like internal financial information, this can be modified by management.

Weather indicators as Audit Evidence

Since big data provides a large amount of information and tamper-resistant data, Yoon et al. (2015) argue that big data can be complementary audit evidence. Big data can be defined in various ways, but generally it is explained by four dimensions: volume, variety, velocity and veracity (Buhl et al. 2013). Big data is described as a large amount of data containing various types of rapidly changing information, and frequently analyzed by sophisticated techniques, such as data mining or pattern recognition.

In this paper, weather variables are considered as big data evidence since weather components are updated on a timely and locational basis, thereby generating a large amount of data. Even though weather is quantitative information, the influence of weather on business outcomes is often nonlinear (Starr-McCluer 2000).

Adding weather indicators to analytical models can enhance error recognition because these indicators are not affected by internal accounting errors or irregularities. In addition, weather indicators are contemporaneously available during the audit procedures. Weather information is continuously updated, so auditors can access it to examine transactions and accounts. For example, economic indicators, such as GDP, could be useful to understand financial outcomes (Lev 1980), but it is difficult for it to be timely, relevant information since the annual GDP is revealed only after considerable time has passed.

Accessibility of information is an additional benefit that auditor might get from utilizing weather indicators. Since external auditors depend heavily on clients' internal information, they might find it difficult to obtain information that the clients do not want to share. The relationship with a client and the manner of client communication play important roles in the acquisition of audit evidence (Bennett and Hatfield 2012). By using external big data, external auditors can access information free from the restrictions imposed by the client, reducing dependence on client relationships and providing an opportunity to enhance the quality and sufficiency of audit evidence. Accordingly, this essay tests the following hypothesis:

> **H:** Firm-wide revenue expectations developed from models with weather variables and financial information yield more accurate and precise predictions than firm-wide revenue expectations derived from models with only financial information.

RESEARCH METHOD

Data

The data employed in this research is obtained from a world-wide accounting firm. The targeted firm is a publicly held multi-location retail firm with a homogeneous worldwide operation, but only observations from the U.S. are used. The firm sells durable goods and tries to offer consistent services and products across stores, although store sizes differ. Twenty-four monthly observations are provided for 1,901 operating locations[14] from fiscal year 2011 to fiscal year 2012. Over the targeted fiscal years, the external auditor expressed an unqualified opinion on annual reports and did not express concerns regarding the effectiveness of the firm's internal controls. Even though daily and weekly balances were not audited separately, the auditors did not find signals of misstatements in the revenue account. Accordingly, this study assumes that these balances do not contain material errors.

Revenue is tested in this study because it can be related to external nonfinancial information and changes dynamically over the period. Other studies examining the usefulness of nonfinancial information test revenue accounts as well (Ittner and Larcker 1998; Cohen et al. 2000). Previous studies develop models by using relevant accounts, such as cost of goods sold, but in this study, other accounts are not included since these accounts are not as frequently updated. The dataset includes store-specific daily sales revenue and the address of each store.

[14] The total number of operating units is more than the number of stores used in this study. Some stores are dropped due to a lack of data. Accordingly, it does not mean that the total number of operating units of this firm is 1,901.

The weather data is obtained from Wunderground.com, a source used by many previous studies (e.g., Iyer et al. 2010).

Disaggregated Data

Audit standards emphasize that "when expectations are developed at a more detailed level, it is more likely that the analytical procedure will more effectively address the assessed risk of misstatement to which it is directed." (AICPA 2012a). Previous literature indicates that higher frequency data (disaggregated or micro-level data) can deliver better performance for analytical procedures (Dzeng 1994; Knechel 1988; Wheeler and Pany 1990), but there has been no consensus as to the preferable level of aggregation. Kogan et al. (2014) compare more disaggregate weekly or daily data, and show that there is no optimal level of aggregation. They argue that the ideal level of aggregation depends on the distribution of potential errors. Correspondingly, this paper generates weekly and daily disaggregated data.

Auditing Standard No. 5 (PCAOB 2007) also indicates that "when a company has multiple locations or business units, the auditor should identify significant accounts and disclosures and their relevant assertions based on the consolidated financial statements." In addition, the PCAOB has raised concerns regarding the auditor's failure to examine location-specific revenue accounts (PCAOB 2014). Nevertheless, studying the optimal level of aggregation by location has been difficult due to a lack of appropriate data.

Allen et al. (1999) study a firm operating offering similar services in thirty locations and argue that, by analyzing the information across the locations, they can provide a useful foundation to measure the power provided at the store-level in their model. In addition, because external auditors often conduct analytical procedures targeted toward a certain line of service or product (Hirst and Koonce 1996), this setting can provide useful insight reflected by the practice.

Allen et al. (1999) employ data disaggregated both monthly and by location from a multi-location service firm and present mixed results regarding the value of disaggregated data by location. They find that the summation of individual location account balance expectations is inferior to expectations derived from aggregate models unless the individual location models contain peer location observations of the account balance. They infer that there is not enough heterogeneity, which might provide unique aspects over thirty stores in their case, so the local-level data does not outperform the firm-level aggregated data. Nevertheless, the data from 1,901 stores across the U.S. used in this study could provide sufficient heterogeneity generated by various levels of local economy and customer characteristics.

Control Variables

To develop expectations at the store level, this study uses contemporaneous sales accounts of peer stores. Allen et al. (1999) use all twenty-nine available locations as peer stores, but in this case, the number of locations is too large. To choose the most relevant peer stores, this study assumes that stores sharing similar economic characteristics have similar sales patterns, thereby having more explanatory powers to understand the targeted store's sales account. First, macroeconomic factors related to the sales account, which are location-based and updated annually, are selected. Powerlytics provides geographically specific macroeconomic indicators from the Internal Revenue Service (IRS) tax returns and data from the U.S. Census Bureau and the Department of Labor. To match annually updated macroeconomic indicators with daily store sales, the average amount of store sales for a year is used. One year lagged macroeconomic indicators are matched because it is assumed that auditors would not be able to access contemporaneous macroeconomic indicators during the audit period. By running a correlation matrix among location-level lagged macroeconomic indicators and store sales, highly correlated macroeconomic indicators are selected[15]. Because those economic indicators are highly correlated with each other, Principal Component Analysis with Varimax Rotation is used to formulate independent factors. At the end three factors are generated and peer store are selected by ranking these three factors. Each factor should be within the integer (the total number of stores/7) of each peer. Each year, each store has a certain number of peers (Hoitash et al. 2006)[16]. The indicator from peer stores is calculated by the following formula:

$$P_t = \frac{\sum_1^N p_{i,t}}{N} \tag{1}$$

Let P_t be the indicator variable reflecting the average amount of sales in peer stores at t where *N* is the number of peer stores, and $p_{i,t}$ is daily/weekly sales of each peer store (*i*=1…N.)

Evaluation of Weather Indicators

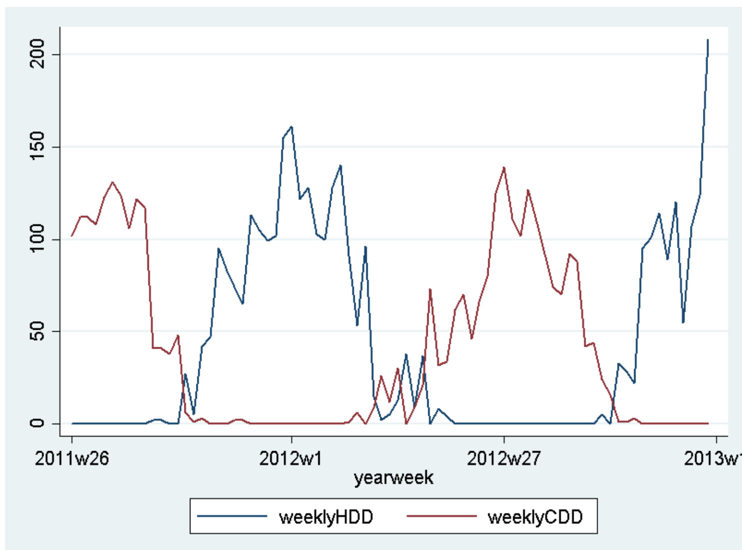
To evaluate unusual weather conditions that might affect sales, "cooling degree days"(*CDD*) and "heating degree days" (*HDD*) are often utilized (Larsen 2006; Starr-McCluer 2000). In cases when the average daily temperature exceeds 65° Fahrenheit, the difference between 65° and the average daily temperature is considered as "cooling degree days" since air conditioning might be used to reduce the temperature. Similarly,

[15] Home mortgage interest and points, unemployment compensation, average real estate taxes, schedule SE total self-employment income, mortgage, losses (rent and royalties, business income or loss, depreciation expense or depletion) rental, the number that filed as single, the count of returns with Schedule F attached are selected as highly correlated lagged macroeconomic indicators to store sale (r > = 0.2).

[16] Hoitash et al.(2006) illustrate the way to select peers in detail in TABLE 2 in their paper. Detailed illustration regarding the peer store is available in the Appendix.

"heating degree days" is calculated in the same way as "cooling degree days" if the average daily temperature is below 65° Fahrenheit (Larsen 2006). The average daily temperature is measured by the maximum daily temperature plus the minimum daily temperature divided by two. Figure 4.1 shows that heating degree days and cooling degree days behave in opposite directions.

Figure 4.1 Heating degree days and cooling degree days



Apparent Temperature (*AT)* is determined by the combination of Heat Index (*HI*) and Wind Chill Index (*WCI*). Heat index and wind chill index are used to evaluate unfavorable weather conditions (Feinberg and Genethliou 2005). These indices are developed and used by the National Oceanic and Atmospheric Administration (NOAA)'s National Weather Service (NWS). When the temperature at a particular grid point falls to 50° F or less, wind chill will be used to calculate the apparent temperature. When the temperature at a grid point rises above 80° F, the heat index will be used. Between 51°

and 80° F, the apparent temperature will be the ambient air temperature, so the average temperature is used for the apparent temperature.

Heat index[17] presents the level of discomfort caused by the combined effects of air temperature and humidity. Heat Index (*HI*) (Rothfusz and Headquarters 1990) is calculated as follows:

$$HI = -42.379 + 2.04901523 \times \mathrm{T} + 10.14333127 \times \mathrm{RH} - 0.22475541 \times \mathrm{T} \times \mathrm{RH} - 0.00683783 \times \mathrm{T} \times \mathrm{T} - 0.05481717 \times \mathrm{RH} \times \mathrm{RH} + 0.00122874 \times \mathrm{T} \times \mathrm{T} \times \mathrm{RH} + 0.00085282 \times \mathrm{T} \times \mathrm{RH} \times \mathrm{RH} - 0.00000199 \times \mathrm{T} \times \mathrm{T} \times \mathrm{RH} \times \mathrm{RH}$$

Where:

T = Air temperature, °F
RH = Relative humidity in percent

Wind chill index is estimated by the air temperature and the wind speed, and useful to determine dangers from winter winds and freezing temperatures. Wind Chill Index (*WC*I) (Williamson 2003) is formulated as follows:

$$WCI = 35.74 + 0.6125 \times T - 35.75 \times V^{0.16} + 0.4275 \times V^{0.16}$$

Where:

V = Wind velocity, mph
T = Air temperature, °F

Charts in Appendix C present the threshold of danger weather conditions based on heat index and wind chill index. The weekly weather indices are the average of daily indices. Because the impact of weather on human behaviors is nonlinear, apparent

[17] Heat index is modified in certain weather conditions. See http://www.srh.noaa.gov/ama/?n=heatindex for more information. Appendix C presents a detailed formula for *HI*.

temperature is transformed by squaring apparent temperatures, titled as Squared Apparent Temperature (*SAT*).

Statistical Models

Previous literature shows that more statistically sophisticated models perform better than simpler models (Chen and Leitch 1999; Kinney Jr 1978; Kinney and Salamon 1982; Lorek et al. 1992; Stringer 1975). Recently, more sophisticated statistical methods have been suggested, such as artificial neural networks (Koskivaara 2004), a three-stage least square (Leitch and Chen 2003), and automated equilibrium correction modeling (Omura and Willett 2006). Kogan et al. (2014) compare the widest range of statistical models, such as simple linear regression, Simultaneous Equation Model, Vector Autoregressive Model (VAR) and GARCH. They find that VAR models and linear regression models tend to perform better than others in their study.

In this study, due to the characteristics of the data that provide only a single account (sales), very few of these models can be utilized. Accordingly, only the time series model with/without exogenous variables and multivariate regression models are tested in this study. In time series models, in order to decide the optimal lag length, Partial Autocorrelation Function is analyzed. Based on this, AR (1) models[18] for in the weekly data are tested.

Let Y_t be a weekly firm-level account balance series under audit, where t is the current week. A weekly firm-level account balance is determined by the total amount of store account balances, where j is the number of stores and y is a store-level account balance. To test the hypothesis, multivariate regression model with the peer store

[18] This model is for the weekly data, so for the daily data AR(7) is used.

indicator and weather indicators and AR (1) with the peer store indicator and weather indicators are tested as follows:

1. Multivariate regression model

$$\widehat{Y_t} = \sum_{i=1}^{j}(\beta_0 + \beta_1 W_{i,t}) \quad (1)$$
$$\widehat{Y_t} = \sum_{i=1}^{j}(\beta_0 + \beta_1 P_{i,t} + \beta_2 W_{i,t}) \quad (2)$$

2. Time series model with/without an exogenous variable

$$\widehat{Y_t} = \sum_{i=1}^{j}(\beta_0 + \beta_1 y_{i,t-1}) \quad (3)$$
$$\widehat{Y_t} = \sum_{i=1}^{j}(\beta_0 + \beta_1 y_{i,t-1} + \beta_2 W_{i,t}) \quad (4)$$
$$\widehat{Y_t} = \sum_{i=1}^{j}(\beta_0 + \beta_1 y_{i,t-1} + \beta_8 P_{i,t}) \quad (5)$$
$$\widehat{Y_t} = \sum_{i=1}^{j}(\beta_0 + \beta_1 y_{i,t-1} + \beta_2 P_{i,t} + \beta_3 W_{i,t}) \quad (6)$$

Let Y_t be a weekly firm-level account balance series under audit, where t is week (*j* = 52 weeks) and where *P* is an average account balance of that store's peer stores.

Model Evaluation

This work employs two approaches to efficiency and effectiveness measures for analytical procedures to compare the statistical models provided above. Prior studies of analytical procedures commonly measure the performance of analytical models by evaluating how well a model predicts the account balance (predictive accuracy) and detects errors (a false positive, Type I error and a false negative, Type II error). Often, these criteria do not consistently present the most powerful model (Dezeng 1994). However, it would be meaningful to examine how weather indicators play a role in improving such criteria.

Predictive accuracy

Many previous studies have examined the expectations derived from models with regard to their predictive accuracy, so the models can be compared on their ability to detect errors. This is especially true for the Mean Absolute Percentage Error (MAPE), which is calculated as the absolute value of the difference between the predicted value and the actual account balance. It is desirable to have a smaller MAPE. As in previous studies, this study adopts the out-of-sample forecasting approach to test the accuracy of the forecasting models. A training set to develop statistical models and a testing set to assess the accuracy of a prediction are separated depending on the strategies utilized. This will be illustrated below in a later section. A MAPE value is calculated as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \widehat{Y}_t}{Y_t}\right|$$

Where, Y_t = Actual firm reported sales
$\widehat{Y}_t$ = Forecast sales
n = 52 (365 for daily data)

There are no clear guidelines to achieve a certain level of a MAPE. However, the commonly used materiality threshold of a revenue account is 0.2-2 percent (Eilifsen and Messier 2015). This desired accuracy might be too high, especially with the limited scope of information used in this study. Some studies often end up with relatively high MAPEs. For instance, the best MAPE value formulated by Kogan et al. (2014) is 0.3919. Even though that study is not comparable to the current research because they use the different information scopes, it is important to achieve a high level of accuracy to resolve current issues related to the performance of analytical procedures in audit practice (e.g., PCAOB 2014). Accordingly, the targeted MAPE value in this study is the level of performance materiality (e.g., 2 percent).

False positive and false negative error counts

A false positive occurs when an analytical model points out a material misstatement in a true account balance. This is related to audit efficiency. On the other hand, a false negative arises when the model fails to identify a materially misstated account balance, which is connected to audit effectiveness.

Since it is assumed that there are no errors in the given data, a false positive is tested by using the given data. When a model detects errors in the given data, these are considered as a false positive. To test a false negative, diverse levels of simulated errors are seeded into the given data. This approach is often called "seeding errors" and is useful to control the setting of an experiment. This is especially true because the power of error detection is largely dependent on the amount of errors. Consequently, various sizes of errors are seeded: 0.5 percent, 1 percent, 2 percent, and 4 percent of the sales revenue account.

It is important to note that both errors size and error distribution affect the performance of error detection. For instance, the best scenario suggested by Kogan et al. (2014) is that the total number of errors is seeded into one observation. It is highly probable that the model will capture this error. On the other hand, the worst scenario would be if the total number of errors is evenly distributed into many observations. In this case, each selected observation would contain a small amount of error. Because this study examines daily and weekly sales revenue values, a dual test for the error seeding experiment is used: 1) the total amount of weekly basis errors is seeded into one day (the best scenario), and 2) the total amount of weekly basis errors is evenly distributed into operating days in that week (the worst scenario). For example, if the weekly sales amount is $700,000, and every day the firm earns the same amount, and 1 percent of sales amount is seeded (error = \$7,000), then in the best scenario, one of the days in that week is randomly selected and errors are seeded in that day's sales. For example, the revenue for Tuesday is changed from \$100,000 to \$107,000. In the worst scenario, the revenues from all seven days in the week are changed from 100,000 to \$101,000.

To detect errors, it is necessary to employ the combination of a predicted account balance and a prediction interval. The size of the prediction interval impacts the effectiveness of analytical tests and is determined by α. If α is high, this will result in a low detection rate, and vice versa. As has been done in prior studies, the value of α is 0.05 (Kogan et al. 2014). Therefore, the total number of experiments is the number of statistical models × four different sizes of seeded errors × two strategies to seed errors.

RESULTS

Store level analysis

The first analysis attempts to discern whether contemporaneous weather indicators for an individual location are associated with customer purchasing behavior and with store-level revenues. The association between weather and sales has often been considered to be nonlinear, so it is important to understand how the association is modified over a period of time and across regions. To understand the explanatory power of weather variables, this study examines a multivariate regression analysis by using weekly store-level sales and weekly weather variables. As a control variable, the average weekly sales of peer stores of an individual store are added. Table 3.1 illustrates increased R^2 by adding weather indicators and the P value of weather indicators. When two weather indicators (i.e. heating degree days and cooling degree days) are included, the P value of F test is examined.

As Table 4.1 illustrates, when the average sales amounts for peer stores are controlled, heating degree days increase the explanatory power of weekly store level revenue by 2.5 percent. Cooling degree days and the combination of heating degree days and cooling degree days increase the explanatory power by 0.1 percent and 2.7 percent respectively. Apparent temperature and squared apparent temperature increases the explanatory power of weekly store level revenues by 1.5 percent and 0.9 percent respectively. Those are lower percentages than those found by Starr-McCluer (2000), who indicates that 10 percent of total monthly revenues for retailers in the U.S. are explained by heating/cooling degree days. The biggest reason for the different results is the control variables used. Starr-McCluer (2000) adds industry general variables as control variables, such as changes in real labor income. Models in this study instead include a store-specific control variable, peer stores, which might contain factors related to macro-economic conditions, industry, and the general popularity of the firm in the market. Without peer stores, weather variables explain generally 7 to 10 percent of sales.

Table 4.1 Increased R^2 by adding weather variables- fiscal year 2011

	HDD		**CDD**		**HDD&CDD**		**AT**		**SAT**	
	Incr.R^2	Sign.	Incr.R^2	Sign.	Incr.R^2	Joint Sign.	Incr.R^2	Sign.	Incr.R^2	Sign.
Sales	0.025	0.000	0.001	0.000	0.027	0.000	0.015	0.000	0.009	0.000
Δ Sales	0.0001	0.000	0.007	0.000	0.009	0.000	0.002	0.000	0.003	0.000
	ΔHDD		**ΔCDD**		**ΔHDD &ΔCDD**		**ΔAT**		**ΔSAT**	
Sales	0.000	0.810	0.001	0.000	0.001	0.000	0.0000	0.308	0.002	0.000
Δ Sales	0.0002	0.000	0.007	0.000	0.008	0.000	0.0001	0.010	0.026	0.000

Panels A, B, C, and D of Table 4.2 illustrate whether the association between weather components and store sales revenues change on a seasonal basis. The association between heating degree days and store revenues is stronger in spring and winter than in summer and autumn (4 percent and 7.7 percent respectively), but the association between cooling degree days and store revenues is not robust in summer. Similarly, percentage changes in squared apparent temperature increases R^2 for store sales and percentage changes of store sales meaningfully in winter (7.5 percent and 12.3 percent respectively). This suggests that the sales of this firm are particularly sensitive to unfavorable weather conditions (cold weather) in winter and spring seasons, but only moderately sensitive to unfavorable weather conditions (hot weather) in summer.

Figure 4.2 demonstrates the seasonal patterns of weather's impact on sales. Overall results are consistent with the multivariate regression analysis examined above. The correlation between daily sales revenues and daily weather indicators is computed specifically for an individual date. For example, the correlation coefficient for March 29, 2011 is estimated by using observations of store-level sales and weather variables on this date. Then by using daily correlation coefficients, the monthly average correlation coefficients are formulated and illustrated in Figure 4.2, which shows that weather does not always have meaningful explanatory power for store-level sales. Accordingly, it is important to consider seasonality for forecasting revenues.

Figure 4.3 illustrates the regional impact of weather on sales. First, the correlation between daily store-level sales and daily weather components is estimated. The average of correlation coefficients in individual states is illustrated by the map. Heating degree days are negatively correlated with the sales of stores located in the New England, Plains, and northern Far West regions, but barely correlated with sales of stores located in the Southeast, especially Florida. Cooling degree days are negatively correlated in southern Far West regions and Florida. Interestingly, cooling degree days are positively correlated with some northern regions. In this case, customers in these regions are more likely to respond positively to warm weather even though temperature is mild (higher than 65° Fahrenheit).

In summary, some analyses are conducted to show how weather and contemporary weather information has explanatory power for store sales revenue. Compared to the control variable, the mean sales of peer stores, weather variables play a less vital role in explaining store-level sales. Nevertheless, unfavorable weather conditions in the winter and spring are meaningful in understanding store sales. In addition, the effect of weather on sales varies according to region. Therefore, it is significant to modify models to forecast store-level sales revenues.

Table 4.2 Increased R^2 by adding weather variables - the seasons of fiscal year 2011

Panel A: Fiscal Year 2011- Spring (March, April, and May)

	HDD		CDD		HDD&CDD		AT		SAT	
	Incr.R^2	Sign.	Incr.R^2	Sign.	Incr.R^2	Joint Sign.	Incr.R^2	Sign.	Incr.R^2	Sign.
Sales	0.041	0.000	0.003	0.000	0.048	0.000	0.023	0.000	0.031	0.000
Δ Sales	0.001	0.000	0.006	0.000	0.006	0.000	0.002	0.000	0.003	0.000
	ΔHDD		**ΔCDD**		**ΔHDD &ΔCDD**		**ΔAT**		**ΔSAT**	
Sales	0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.048	0.003	0.000
Δ Sales	0.004	0.000	0.009	0.000	0.013	0.000	0.003	0.000	0.075	0.000

Panel B: Fiscal Year 2011-Summer (June, July, and August)

	HDD		CDD		HDD&CDD		AT		SAT	
	Incr.R^2	Sign.	Incr.R^2	Sign.	Incr.R^2	Joint Sign.	Incr.R^2	Sign.	Incr.R^2	Sign.
Sales	0.006	0.000	0.013	0.000	0.014	0.000	0.013	0.000	0.012	0.000
Δ Sales	0.002	0.000	0.000	0.001	0.004	0.000	0.000	0.009	0.000	0.003
	ΔHDD		**ΔCDD**		**ΔHDD &ΔCDD**		**ΔAT**		**ΔSAT**	
Sales	0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.001
Δ Sales	0.000	0.024	0.000	0.101	0.000	0.021	0.009	0.000	0.010	0.000

Panel C: Fiscal Year 2011-Fall (September, October, and November)

	HDD		CDD		HDD&CDD		AT		SAT	
	Incr.R^2	Sign.	Incr.R^2	Sign.	Incr.R^2	Joint Sign.	Incr.R^2	Sign.	Incr.R^2	Sign.
Sales	0.002	0.000	0.002	0.000	0.003	0.000	0.003	0.000	0.003	0.000
Δ Sales	0.000	0.053	0.000	0.607	0.000	0.118	0.000	0.021	0.000	0.041
	ΔHDD		**ΔCDD**		**ΔHDD &ΔCDD**		**ΔAT**		**ΔSAT**	
Sales	0.000	0.604	0.001	0.000	0.001	0.000	0.000	0.117	0.001	0.000
Δ Sales	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.001	0.000

Panel D: Fiscal Year 2011-Winter (December, January, and February)

	HDD		CDD		HDD&CDD		AT		SAT	
	Incr.R^2	Sign.	Incr.R^2	Sign.	Incr.R^2	Joint Sign.	Incr.R^2	Sign.	Incr.R^2	Sign.
Sales	0.077	0.000	0.044	0.000	0.085	0.000	0.080	0.000	0.080	0.000
Δ Sales	0.008	0.000	0.000	0.006	0.009	0.000	0.006	0.000	0.008	0.000
	ΔHDD		**ΔCDD**		**ΔHDD &ΔCDD**		**ΔAT**		**ΔSAT**	
Sales	0.002	0.000	0.003	0.000	0.006	0.000	0.000	0.464	0.004	0.000
Δ Sales	0.000	0.516	0.009	0.000	0.009	0.000	0.000	0.131	0.123	0.000

Figure 4.2 The impact of weather on sales by month

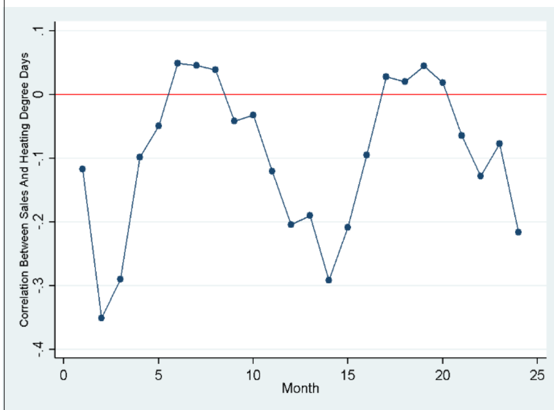


Figure 4.2.A Correlation between sales and HDD

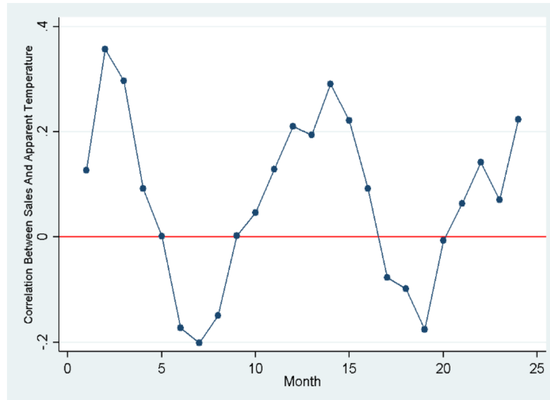


Figure 4.2.C Correlation between sales and AT

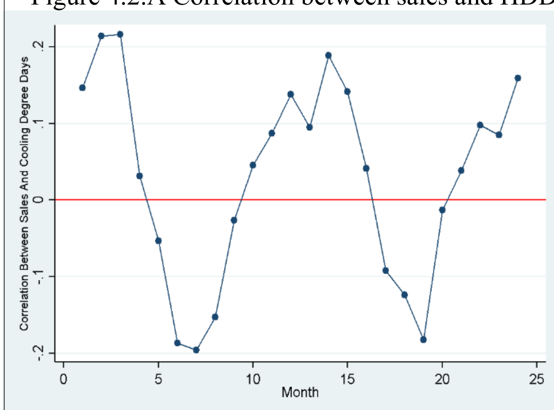


Figure 4.2.B Correlation between sales and CDD

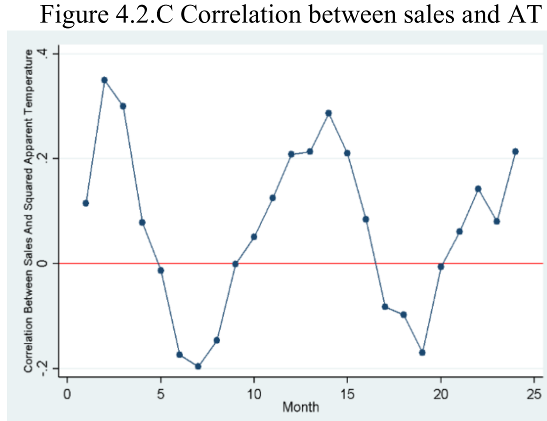


Figure 4.2.D Correlation between sales and SAT

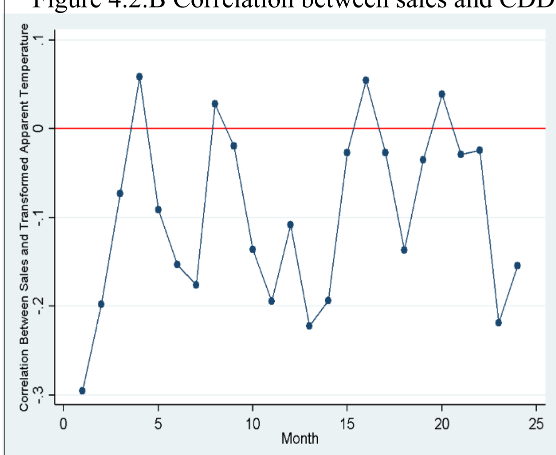
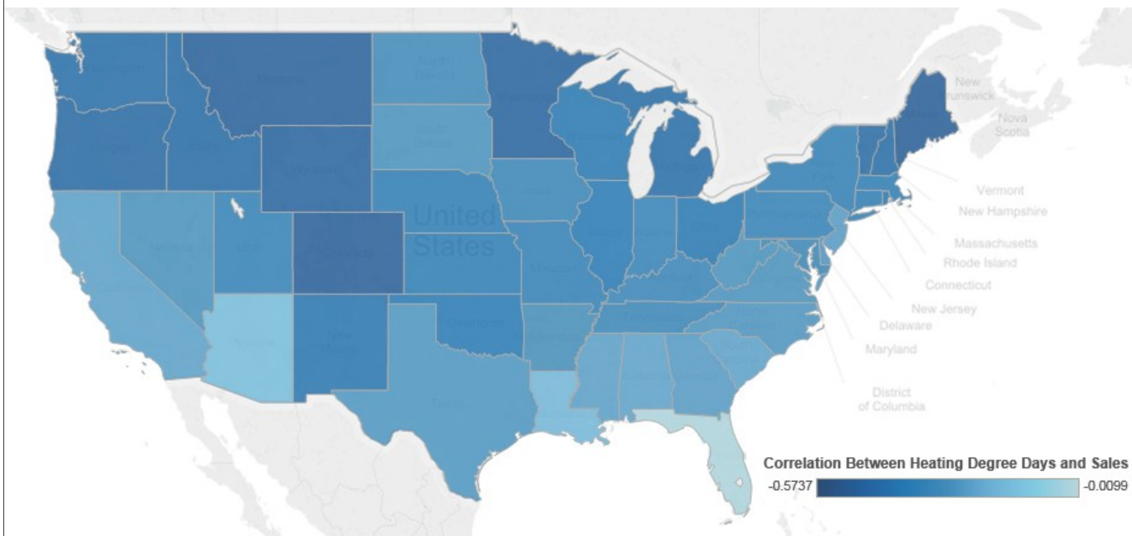


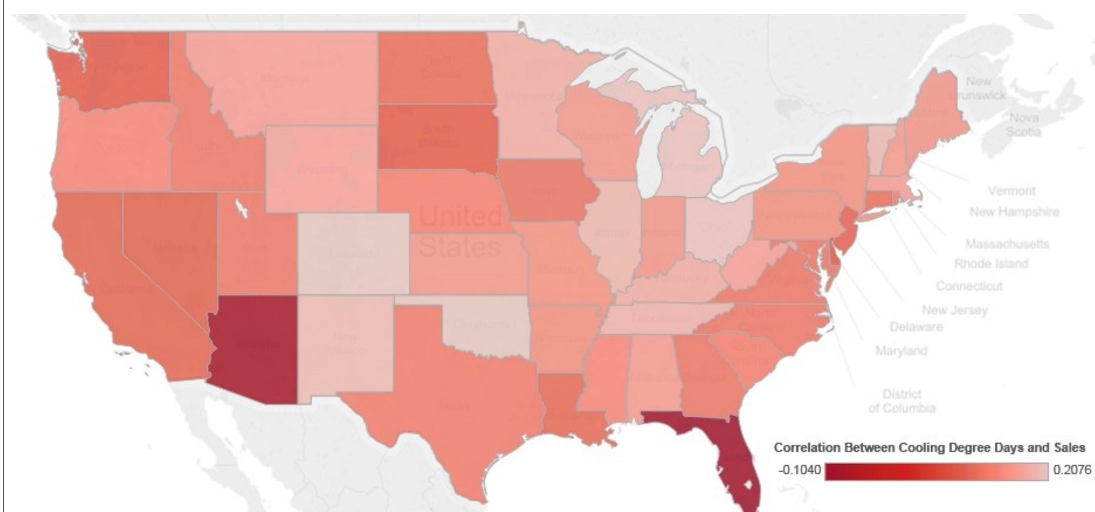
Figure 4.2.E Correlation between sales and CAT

Figure 4.3 The impact of weather on sales by region

Panel A: Analysis of heating degree days by state



Panel B: Analysis of cooling degree days by state



Results – Prediction Accuracy of Fixed models

Panel A of Table 4.3 summarizes the outcomes for the comparison of MAPE values derived from weekly heating degree days, cooling degree days, apparent temperature, and squared apparent temperature. The MAPE value computed from heating degree days and cooling degree days is 0.076 and those from apparent temperature and squared apparent temperature are 0.084 and 0.089 respectively. Without any other control variables, over 90 percent of weekly firm-level sales can be forecasted by aggregated weather indicators. Nevertheless, it might be too optimistic to conclude the value of weather components on sales forecast based on this result.

As noted in Panel A of Table 4.3, a MAPE value estimated by the time series model containing lagged sales is 0.063, indicating strong seasonal patterns in weekly sales. By using lagged sales, the auditor can forecast the following weekly sales with 94 percent accuracy. The statistical model involving lagged sales and weather variables, including heating degree days and cooling degree days, apparent temperature, and squared apparent temperature develops a marginally improved MAPE, 0.061, 0.061, and 0.062 respectively. Adding weather indicators in the model with lagged sales does not radically improve the forecast accuracy of weekly sales. Table 4.4 reports Wilcoxon Signed Rank Test ($\mathrm{H0}: MAPE_a = MAPE_b$). There is no significant difference among MAPE values resulting from models with weather variables, but there is a significant difference between a model with the average sales of peer stores and models with weather indicators. However, the standard deviations of MAPEs from models with cooling degree days and heating degree days, apparent temperatures, and squared apparent temperatures are slightly smaller than those resulting from the model with lagged sales (0.036, 0.042, 0.042, and 0.047 respectively). It follows that weather variables might reduce forecast errors.

Based on these results, the seasonality of sales seems to be reasonably apparent. This seasonality resembles the seasonality of weather, providing significant predictive value for forecasting sales. For instance, Wilcoxon Signed Rank Test between the model with heating degree days and cooling degree days (Model (1)) and the model with lagged weekly sales (Model (10)) is not rejected at the 0.5 significance level.

Adding peer stores improves the accuracy of the sales forecast. The MAPE value derived from the model having the average weekly sales of peer stores is 0.029. MAPEs calculated by peers and weather indicators are not reasonably higher than the MAPE resulting from the average sales of peers (both heating degree days and cooling degree days, apparent temperature, and squared apparent temperature are 0.029, 0.028, and 0.028 respectively). Weather variables do not offer incremental explanatory power that peer stores cannot provide. Or, perhaps, the current approach does not effectively capture the influence of weather on sales.

Contemporaneous weather indicators have explanatory power for weekly sales accounts. However, adding a weather indicator is less likely to improve the effectiveness of analytical procedures to forecast the sales account when other variables (lagged sales and/or the sales of peer stores) are considered. Nevertheless, the approaches applied in this section might not adequately represent the influence of weather on sales. To overcome this issue, two additional sets of experiments are performed.

Table 4.3 Results of predictive accuracy- by using weekly store level data

Model #	Included variables	# of obs.	MAPE	Std.	Min APE	Max APE
Panel A: Fixed Model						
(1)	HDD & CDD	52	0.076	0.051	0.001	0.196
(2)	AT	52	0.084	0.056	0.000	0.193
(3)	SAT	52	0.089	0.058	0.000	0.204
(4)	CAT	52	0.078	0.047	0.001	0.158
(5)	Peer	52	0.022	0.019	0.000	0.083
(6)	Peer & HDD & CDD	52	0.026	0.020	0.001	0.074
(7)	Peer & AT	52	0.024	0.016	0.000	0.061
(8)	Peer & SAT	52	0.023	0.015	0.002	0.056
(9)	Peer & CAT	52	0.026	0.017	0.001	0.059
(10)	$Sales_{n-1}$	52	0.063	0.047	0.001	0.202
(11)	$Sales_{n-1}$ & HDD & CDD	52	0.061	0.036	0.008	0.151
(12)	$Sales_{n-1}$ & AT	52	0.061	0.042	0.003	0.165
(13)	$Sales_{n-1}$ & SAT	52	0.062	0.044	0.004	0.180
(14)	$Sales_{n-1}$ & CAT	52	0.058	0.037	0.001	0.148
(15)	$Sales_{n-1}$ & Peer	52	0.030	0.025	0.001	0.108
(16)	$Sales_{n-1}$ & Peer & HDD & CDD	52	0.029	0.020	0.000	0.079
(17)	$Sales_{n-1}$ & Peer & AT	52	0.028	0.022	0.001	0.090
(18)	$Sales_{n-1}$ & Peer & SAT	52	0.028	0.023	0.001	0.097
(19)	$Sales_{n-1}$ & Peer & CAT	52	0.028	0.021	0.000	0.085
Panel B: Stepwise Model						
(1)	HDD & CDD	52	0.059	0.038	0.006	0.177
(2)	AT	52	0.058	0.040	0.009	0.173
(3)	SAT	52	0.058	0.041	0.009	0.175
(4)	CAT	52	0.060	0.041	0.009	0.171
(5)	Peer	52	0.044	0.032	0.000	0.126
(6)	Peer & HDD & CDD	52	0.044	0.029	0.002	0.127
(7)	Peer & AT	52	0.043	0.030	0.003	0.125
(8)	Peer & SAT	52	0.044	0.030	0.003	0.126
(9)	Peer & CAT	52	0.045	0.030	0.003	0.122
(10)	$Sales_{n-1}$	52	0.058	0.045	0.003	0.177
(11)	$Sales_{n-1}$ & HDD & CDD	52	0.058	0.044	0.002	0.177
(12)	$Sales_{n-1}$ & AT	52	0.058	0.044	0.002	0.178
(13)	$Sales_{n-1}$ & SAT	52	0.058	0.044	0.002	0.178
(14)	$Sales_{n-1}$ & CAT	52	0.060	0.046	0.003	0.179
(15)	$Sales_{n-1}$ & Peer	52	0.055	0.042	0.001	0.166
(16)	$Sales_{n-1}$ & Peer & HDD & CDD	52	0.055	0.041	0.001	0.167
(17)	$Sales_{n-1}$ & Peer & AT	52	0.055	0.041	0.001	0.167
(18)	$Sales_{n-1}$ & Peer & SAT	52	0.055	0.041	0.002	0.167
(19)	$Sales_{n-1}$ & Peer & CAT	52	0.056	0.042	0.000	0.168

Table 4.4 Pairwise comparison - fixed model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(2)	-1.45																	
(3)	-1.79	-1.95																
(4)	-0.44	1.06	1.61															
(5)	5.72*	5.68*	5.93*	6.10*														
(6)	6.27*	5.60*	5.56*	6.03*	-0.97													
(7)	5.95*	6.22*	6.24*	5.76*	-0.59	0.46												
(8)	5.84*	6.17*	6.24*	5.83*	-0.33	0.26	1.49											
(9)	6.02*	5.45*	5.56*	6.23*	-1.67	0.21	-0.66	-0.67										
(10)	1.50	2.33*	2.75*	2.29*	-5.79*	-4.35*	-4.96*	-5.25*	-4.79*									
(11)	2.15*	2.61*	2.98*	2.63*	-5.81*	-5.54*	-5.45*	-5.52*	-5.71*	0.42								
(12)	1.80	2.80*	3.21*	2.27*	-5.67*	-4.50*	-5.49*	-5.69*	-4.90*	0.31	-0.41							
(13)	1.73	2.65*	3.09*	2.17*	-5.74*	-4.39*	-5.38*	-5.60*	-4.77*	0.29	-0.38	-0.07						
(14)	2.30*	2.90*	3.23*	2.94*	-5.56*	-4.88*	-5.07*	-5.23*	-5.16*	0.81	1.38	0.94	0.87					
(15)	4.37*	4.95*	5.18*	4.94*	-2.38*	-0.64	-1.29	-1.54	-0.66	6.20*	5.86*	6.05*	6.12*	5.53*				
(16)	5.15*	4.96*	5.10*	5.43*	-1.89	-0.64	-1.50	-1.61	-0.68	5.56*	6.27*	5.56*	5.53*	5.99*	0.50			
(17)	4.76*	5.29*	5.57*	5.18*	-1.43	-0.23	-0.88	-1.19	-0.30	6.17*	6.13*	6.26*	6.27*	5.76*	2.26*	0.66		
(18)	4.72*	5.22*	5.51*	5.12*	-1.48	-0.22	-0.87	-1.19	-0.31	6.24*	6.06*	6.22*	6.27*	5.73*	2.81*	0.74	0.40	
(19)	5.02	4.98	5.10*	5.31*	-1.68	-0.26	-1.09	-1.19	-0.35	5.83*	6.20*	5.62*	5.57*	6.20*	0.50	0.73	-0.72	-0.79

(1) HDD & CDD
(2) AT
(3) SAT
(4) CAT
(5) Peer
(6) Peer & HDD & CDD
(7) Peer & AT
(8) Peer & SAT
(9) Peer & CAT
(10) Salesn-1
(11) Salesn-1 & HDD & CDD
(12) Salesn-1 & AT
(13) Salesn-1 & SAT
(14) Salesn-1 & CAT
(15) Salesn-1 & Peer
(16) Salesn-1 & Peer & HDD & CDD
(17) Salesn-1 & Peer & AT
(18) Salesn-1 & Peer & SAT
(19) Salesn-1 & Peer & CAT

The Wilcoxon signed rank test which is a nonparametric test. H0 is the averages of two variables are equal. * indicates significance at the 5 % level

Table 4.5 Pairwise comparison - stepwise model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(2)	0.50																		
(3)	-0.24	-1.34																	
(4)	-0.56	-2.25*	-1.80																
(5)	5.66*	5.50*	5.50*	6.07*															
(6)	6.19*	5.29*	5.21*	5.46*	-0.46														
(7)	5.89*	6.18*	6.10*	6.03*	0.61	0.42													
(8)	6.00*	6.25*	6.22*	6.16*	0.13	-0.19	-0.52												
(9)	5.82*	5.86*	5.77*	6.16*	-0.40	-0.13	-1.39	-0.81											
(10)	0.31	0.05	0.08	0.31	-1.83	-1.63	-1.84	-1.78	-1.63										
(11)	0.33	0.17	0.25	0.40	-1.80	-1.54	-1.88	-1.78	-1.59	-0.08									
(12)	0.39	0.05	0.18	0.32	-1.89	-1.60	-1.86	-1.77	-1.62	0.35	-0.37								
(13)	0.35	0.07	0.15	0.33	-1.93	-1.57	-1.83	-1.74	-1.67	0.43	0.05	0.11							
(14)	0.03	-0.24	-0.16	0.02	-2.08*	-1.87	-2.06	-1.98*	-1.86	-0.10	-0.14	-0.67	-0.54						
(15)	0.67	0.45	0.49	0.69	-1.50	-1.32	-1.53	-1.48	-1.35	4.44*	3.45*	3.19*	3.28*	4.10*					
(16)	0.72	0.47	0.53	0.77	-1.58	-1.28	-1.54	-1.48	-1.35	3.43*	3.51*	3.75*	3.80*	3.64*	-0.18				
(17)	0.70	0.46	0.56	0.69	-1.58	-1.29	-1.55	-1.48	-1.43	3.59*	3.95*	4.03*	3.85*	3.57*	0.02	0.19			
(18)	0.72	0.49	0.55	0.69	-1.59	-1.31	-1.57	-1.52	-1.42	3.20*	3.77*	4.02*	4.11*	3.43*	-0.05	0.16	0.15		
(19)	0.64	0.30	0.42	0.51	-1.65	-1.41	-1.64	-1.61	-1.50	3.45*	3.63*	3.70*	3.82*	4.48*	0.20	0.46	-0.17	-0.01	0.99

(1)	HDD & CDD	(11)	Salesn-1 & HDD & CDD
(2)	AT	(12)	Salesn-1 & AT
(3)	SAT	(13)	Salesn-1 & SAT
(4)	CAT	(14)	Salesn-1 & CAT
(5)	Peer	(15)	Salesn-1 & Peer
(6)	Peer & HDD & CDD	(16)	Salesn-1 & Peer & HDD & CDD
(7)	Peer & AT	(17)	Salesn-1 & Peer & AT
(8)	Peer & SAT	(18)	Salesn-1 & Peer & SAT
(9)	Peer & CAT	(19)	Salesn-1 & Peer & CAT
(10)	Salesn-1		

The Wilcoxon signed rank test which is a nonparametric test. H0 is the averages of two variables are equal. * indicates significance at the 5 % level

Transformed Weather Variables

As Figure 4.2 shows, the correlation between weather and sales changes from period to period, creating a nonlinear relationship. Prior literature suggests a square transformation of weather indicators. Nevertheless, as noted in Table 4.3, squared apparent temperature (SAT) does not significantly improve the predictive accuracy. Therefore, two additional approaches are tested.

First, a polynomial regression analysis is examined. By conducting an F-test, the suggested model is the fourth degree of polynomial ($y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$) if the other variables such as peer stores and lagged sales are not considered. Nevertheless, the suggested model is modified by adding other independent variables, so suggested partial polynomial regression models are also empirically tested to find the best model.

Centering data is one approach to transform data, and is formulated as the difference between the median of data points and a data point. Centering apparent temperature could more efficiently explain sales than apparent temperatures since it might reduce the wide variation of correlation between weather and sales. By squaring the centered data, the absolute difference from the median apparent temperature is a set for models.

As Figure 4.2.E illustrates, by transforming apparent temperature, the correlation coefficients derived from apparent temperatures and sales are generally positive values. Centering apparent temperature is also tested for the polynomial regression analysis. Again, the fourth degree of polynomial is suggested by F-test. However, few partial polynomial regression models with transformed apparent temperature are tested since the optimal degree of polynomial is modified by adding control variables.

Two different approaches are tested and described in Table 4.6. While some models are moderately different, a low degree of polynomial is generally preferred, thereby suggesting to test squared centering apparent temperature (CAT) as an additional weather indicator. The transformed weather variable is calculated as follows:

$$Centering\ Apparent\ Temperature_{i,t} = \left\{median\ (Apparent\ Temperature) - Apparent\ Temperature_{i,t}\right\}^2$$

Table 4.6 Prediction accuracy - polynomial regression and transformed apparent temperature

Model #	Included variables	# of obs.	MAPE	Std.	Min APE	Max APE
Panel A: Polynomial regression						
(20)	AT & SAT & AT^3 & AT^4	52	0.077	0.049	0.001	0.180
(22)	AT & SAT	52	0.077	0.049	0.002	0.170
(23)	AT & AT^3	52	0.076	0.050	0.001	0.175
(24)	AT & AT^4	52	0.075	0.051	0.001	0.185
(25)	Peer & AT & SAT & AT^3 & AT^4	52	0.077	0.049	0.001	0.180
(26)	Peer & AT & SAT	52	0.077	0.049	0.002	0.170
(27)	Peer & AT & AT^3	52	0.076	0.050	0.001	0.175
(28)	Peer & AT & AT^4	52	0.075	0.051	0.001	0.185
(29)	$Sales_{n-1}$ & AT & SAT & AT^3 & AT^4	52	0.063	0.040	0.003	0.175
(30)	$Sales_{n-1}$ & AT & SAT	52	0.059	0.036	0.001	0.161
(31)	$Sales_{n-1}$ & AT & AT^3	52	0.059	0.036	0.000	0.159
(32)	$Sales_{n-1}$ & AT & AT^4	52	0.060	0.036	0.003	0.152
(33)	$Sales_{n-1}$ & Peer & SAT & AT^3 & AT^4	52	0.031	0.023	0.001	0.112
(34)	$Sales_{n-1}$ & Peer & AT & SAT	52	0.028	0.020	0.001	0.079
(35)	$Sales_{n-1}$ & Peer & AT & AT^3	52	0.028	0.020	0.001	0.080
(36)	$Sales_{n-1}$ & Peer & AT & AT^4	52	0.029	0.020	0.001	0.079
Panel B: Polynomial regression with transformed apparent temperature						
(37)	CAT^2	52	0.078	0.047	0.001	0.158
(38)	CAT & CAT^2 & CAT^3 & CAT^4	52	0.076	0.048	0.002	0.168
(39)	CAT & CAT^2	52	0.076	0.048	0.002	0.181
(40)	Peer & CAT^2	52	0.026	0.017	0.001	0.059
(41)	Peer & CAT & CAT^2 & CAT^3 & CAT^4	52	0.027	0.018	0.001	0.065
(42)	Peer & CAT & CAT^2	52	0.027	0.020	0.000	0.072
(43)	$Sales_{n-1}$ & CAT^2	52	0.058	0.037	0.001	0.148
(44)	$Sales_{n-1}$ & CAT & CAT^2 & CAT^3 & CAT^4	52	0.059	0.036	0.001	0.161
(45)	$Sales_{n-1}$ & CAT & CAT^2	52	0.063	0.040	0.003	0.175
(46)	$Sales_{n-1}$ & Peer & CAT^2	52	0.028	0.021	0.000	0.085
(47)	$Sales_{n-1}$ & Peer & CAT & CAT^2 & CAT^3 & CAT^4	52	0.028	0.021	0.000	0.088
(48)	$Sales_{n-1}$ & Peer & CAT & CAT^2	52	0.031	0.023	0.000	0.112

Predictive Accuracy – Stepwise Model

As previously noted, the influence of weather on sales varies by region and season. To find a suitable model for each region and period, stepwise regression is analyzed. A store-basis analysis is preferred to achieve a region-specific model. However, since only a limited number of observations are given (four observations per a store for a month), alternative options [19] are considered. One of these is a state and county basis disaggregation. Nevertheless, some states (e.g., New York) are too large to capture regional characteristics. In addition, one or a few stores may be located in one county, so the model would still have issues related to a lack of observations. Managers are assigned to a certain number of closely located stores, and this information is also provided. Grouping stores by assigned managers could be a way to provide a sufficient enough number of observations to execute statistical models and also contains regional characteristics that might affect the association between weather and sales.

To reflect seasonality, sophisticated time series models such as seasonal ARIMA models with exogenous variables were considered, but to combine regional and temporal specific components, this study uses stepwise regression models based on each individual group and month. This means that every month's weather indicators are separately selected by the stepwise regression model for an individual group of stores. Unlike the fixed model procedures, the base period is limited to fiscal year 2011. For instance, the stepwise regression model based on 2011 fiscal year observations of the five stores assigned to manager A selects heating degree days in December, but cooling degree days in July. By using selected variables and formulated coefficients with contemporary weather/peer store components, 2012 fiscal year observations of these stores are predicted.

---

[19] Because the daily basis data increases the number of observations, it provides an opportunity to compare explanatory values. The store basis and state basis are analyzed and illustrated in Appendix C. The store basis data generates generally higher MAPE than others, but the differences between store basis and the basis of the assigned manager are modest.

Results – Stepwise Regression Models

Panel B of Table 4.5 illustrates the MAPE values derived from stepwise regression models. Variables are presented differently as base variables, but depending on a group determined by the assigned manager and month, selected variables vary. Time series models (9) to (15) are generated based on a related stepwise regression model. Model (9), having base variables as lagged sales, heating degree days, and cooling degree days, contains variables selected by Model (1) resulting from predictor variables including heating degree days and cooling degree days. For example, Model (1) selects heating degree days for group A in December. Model (9) derives from lagged sales and heating degree days for group A in December. Panel C shows the percentage of selected variables.

Stepwise regression models outperform the fixed models overall. MAPE values resulting from heating degree days and cooling degree days, apparent temperature, and squared apparent temperature are 0.059, 0.058, and 0.058 respectively. Unexpectedly, some models result in worse predictive accuracy than the fixed models. This becomes especially apparent when models contain the sales of peer stores. A MAPE value derived from the stepwise regression model with peer sales is 0.044, which is larger than 0.022resulting from the regression model.

Table 4.7 shows how often each variable is selected by the stepwise regression variable selection process. Generally, only five to ten percent of the total number of models include weather variables. The sales of peer stores are most often selected (32-34 percent of all models). When both weather variables and the sales of peer stores are considered as the base variables, the percent of weather variables selected is reduced to four to seven percent of the total number of models. These results indicate that there is a correlation between weather variables and the sales of peer stores, and generally the sales of peer stores have higher predictive values in explaining the sales of a targeted store.

It is important to understand in which regions and during which periods weather variables can add value in forecasting. Simply adding weather components might cause overfitting problems, reducing predictive accuracy.

Table 4.7 Percentage of selected variables by stepwise regression model

Variable	**Base Variables**						
	HDD &CDD	AT	SAT	Peer	Peer & HDD & CDD	Peer & AT	Peer & SAT
Model #	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HDD	6.96%	11.24%	10.93%	34.61%	4.49%		
CDD	5.93%				4.24%		7.36%
SAT						7.48%	
Peer					32.96 %	33.30%	33.21%

Results – Disaggregated Data

In this section, I test the accuracy of predictions by using daily store sales. Because daily sales patterns might have higher fluctuations than weekly sales patterns, it is expected that the prediction accuracy would be worse. Excepting two outliers (Christmas and Thanksgiving), MAPE values are described in Table 4.8. As expected, MAPE values of daily sales are higher than those of weekly sales. Weather indicators are likely to lose their explanatory values for sales forecasts. A MAPE value derived from heating degree days, cooling degree days, apparent temperature, squared apparent temperature, and centering apparent temperature is 0.119, 0.123, 0.126, and 0.127 respectively. Impressively, the sales of peer stores kept a similar level of predictive values. The MAPE value derived from the model with the sales of peer stores is 0.027.

Even if stepwise model selection procedures increase MAPE values, the increased amount is smaller than that of the weekly sales. A MAPE value resulting from weather indicators, heating degree days, cooling degree days, apparent temperature, squared apparent temperature and centering apparent temperature, is 0.100, 0.101, 0.100, and 0.100.

Table 4.8 Predictive accuracy- by using daily store level

Model #	Included variables	# of obs.	MAPE	Std.	Min APE	Max APE
Panel A: Fixed Model						
(1)	HDD & CDD	363	0.119	0.114	0.001	0.704
(2)	AT	363	0.123	0.116	0.000	0.785
(3)	SAT	363	0.126	0.118	0.001	0.856
(4)	CAT	363	0.127	0.121	0.002	0.828
(5)	Peer	363	0.027	0.026	0.000	0.230
(6)	Peer & HDD & CDD	363	0.028	0.027	0.000	0.159
(7)	Peer & AT	363	0.027	0.026	0.000	0.165
(8)	Peer & SAT	363	0.027	0.026	0.000	0.188
(9)	Peer & CAT	363	0.027	0.026	0.000	0.176
(10)	$Sales_{n-1} \cdots Sales_{n-7}$	363	0.096	0.112	0.001	0.940
(11)	$Sales_{n-1} \cdots Sales_{n-7}$ & HDD & CDD	363	0.096	0.110	0.000	0.931
(12)	$Sales_{n-1} \cdots Sales_{n-7}$ & AT	363	0.096	0.111	0.000	0.918
(13)	$Sales_{n-1} \cdots Sales_{n-7}$ & SAT	363	0.096	0.111	0.000	0.919
(14)	$Sales_{n-1} \cdots Sales_{n-7}$ & CAT	363	0.096	0.111	0.000	0.953
(15)	$Sales_{n-1} \cdots Sales_{n-7}$ & Peer	363	0.033	0.040	0.000	0.356
(16)	$Sales_{n-1} \cdots Sales_{n-7}$ & Peer & HDD & CDD	363	0.032	0.038	0.000	0.344
(17)	$Sales_{n-1} \cdots Sales_{n-7}$ & Peer & AT	363	0.032	0.038	0.000	0.337
(18)	$Sales_{n-1} \cdots Sales_{n-7}$ & Peer & SAT	363	0.031	0.038	0.000	0.336
(19)	$Sales_{n-1} \cdots Sales_{n-7}$ & Peer & CAT	363	0.033	0.040	0.000	0.361
Panel A: Stepwise Model						
(1)	HDD & CDD	363	0.100	0.108	0.000	0.811
(2)	AT	363	0.101	0.108	0.000	0.813
(3)	SAT	363	0.101	0.108	0.000	0.812
(4)	CAT	363	0.100	0.108	0.000	0.816
(5)	Peer	363	0.044	0.049	0.000	0.299
(6)	Peer & HDD & CDD	363	0.044	0.049	0.000	0.292
(7)	Peer & AT	363	0.045	0.049	0.000	0.296
(8)	Peer & SAT	363	0.044	0.049	0.000	0.295
(9)	Peer & CAT	363	0.044	0.049	0.000	0.300
(10)	$Sales_{n-1} \cdots Sales_{n-7}$	363	0.099	0.130	0.000	1.038
(11)	$Sales_{n-1} \cdots Sales_{n-7}$ & HDD & CDD	363	0.098	0.120	0.001	1.035
(12)	$Sales_{n-1} \cdots Sales_{n-7}$ & AT	363	0.099	0.120	0.001	1.035
(13)	$Sales_{n-1} \cdots Sales_{n-7}$ & SAT	363	0.099	0.120	0.000	1.034
(14)	$Sales_{n-1} \cdots Sales_{n-7}$ & CAT	363	0.098	0.120	0.001	1.037
(15)	$Sales_{n-1} \cdots Sales_{n-7}$ & Peer	363	0.070	0.079	0.000	0.611
(16)	$Sales_{n-1} \cdots Sales_{n-7}$ & Peer & HDD & CDD	363	0.071	0.079	0.000	0.610
(17)	$Sales_{n-1} \cdots Sales_{n-7}$ & Peer & AT	363	0.071	0.079	0.000	0.614
(18)	$Sales_{n-1} \cdots Sales_{n-7}$ & Peer & SAT	363	0.071	0.079	0.000	0.613
(19)	$Sales_{n-1} \cdots Sales_{n-7}$ & Peer & & CAT	363	0.071	0.079	0.000	0.617

Results – Error Detection Performance

This section illustrates the performance of error detections. By using weekly store level data, I can find the number of observations that each model identifies as errors and determine whether those are actual misstated observations or not. Table 4.9 shows that false positives occur in the range between 5.9 percent and 8.6 percent. Interestingly, weather indicators are likely to outperform the sales of peer stores. The model incorporates heating and cooling degree days, apparent temperature, squared apparent temperature, and centering apparent temperature flags correct observations as errors 5.9, 5.5, and 5.2 percent of the total observations respectively.

On the other hand, most models are less likely to capture genuine errors in either the worst or best scenarios described earlier. In the worst case scenario, where weekly errors are evenly distributed among operating days, about 90 percent of errors are not captured by the models. In this case, where four percent of weekly sales are evenly distributed into operating days, centering apparent temperature marginally outperforms the sales of peer stores in flagging 88.3 percent and 88.5 percent of errors. In the best case scenario, especially where four percent of weekly sales are distributed into one of daily sales, 69 percent of cases are not captured by the model with peer stores. The model with weather indicators, heating degree days, cooling degree days, apparent temperature, squared apparent temperature, and centering apparent temperature, is likely to miss 79.3, 79.5, 79.6, and 79.5 percent of errors, respectively.

Since the results of error detections are largely skewed toward low false positive rates, modifying α adjusts the proportion between false positive and false negative s. α is therefore modified from 0.1 to 0.5. Among the outcomes resulting from modified α, the performance of error detections at $\alpha = 0.2$ is summarized in Table 4.10. Again, models with weather variables are less likely to classify correct accounts as errors than models with average sales of peer stores. For instance, false positives resulting from the model with the squared apparent temperature is 38 percent, but false positives derived from the model with the average sales of peer stores is 44 percent. On the other hand, the average sales of peer stores play an important role in reducing false negative. In the best scenario where four percent of weekly sales are seeded into one day's sales in that week, a false negative rate resulting from the model with the average peer stores sales is 29.9 percent. In this case, false negative rates derived from the model with weather variables range from 41.3 percent to 42.9 percent. However, the best model in terms of a false negative rate is the one with the combination of the weather indicator (heating degree days and cooling degree days), lagged sales, and the average sales of peer stores (26.3 percent).

Models with weather indicators are likely to have lower false positives than models with other variables, at the cost of higher false negatives. Models containing lagged sales, the variable from the average sales of peer stores, and weather indicators are likely to have the lowest false negative rate among models, indicating the incremental values of weather variables in reducing false negative rates.

Table 4.9 Error detection- alpha = 0.05

		False Positive	False Negative							
Model #	Included variables		Worst Scenario				Best Scenario			
			0.5%	1%	2%	4%	0.5%	1%	2%	4%
(1)	HDD & CDD	0.059	0.938	0.934	0.925	0.901	0.937	0.928	0.900	0.793
(2)	AT	0.055	0.942	0.938	0.929	0.907	0.938	0.929	0.901	0.795
(3)	SAT	0.052	0.944	0.941	0.932	0.911	0.938	0.929	0.901	0.796
(4)	CAT	0.085	0.912	0.910	0.903	0.883	0.938	0.928	0.900	0.795
(5)	Peer	0.086	0.912	0.909	0.903	0.885	0.930	0.916	0.867	0.691
(6)	Peer & HDD & CDD	0.086	0.911	0.909	0.901	0.879	0.930	0.915	0.866	0.681
(7)	Peer & AT	0.086	0.912	0.909	0.903	0.883	0.930	0.916	0.866	0.684
(8)	Peer & SAT	0.087	0.911	0.908	0.902	0.882	0.930	0.916	0.866	0.684
(9)	Peer & CAT	0.045	0.952	0.949	0.942	0.923	0.930	0.916	0.867	0.689
(10)	$Sales_{n-1}$	0.066	0.932	0.929	0.922	0.904	0.941	0.934	0.910	0.797
(11)	$Sales_{n-1}$ & HDD & CDD	0.064	0.933	0.929	0.921	0.897	0.941	0.933	0.908	0.793
(12)	$Sales_{n-1}$ & AT	0.063	0.934	0.931	0.924	0.902	0.942	0.934	0.909	0.795
(13)	$Sales_{n-1}$ & SAT	0.063	0.934	0.931	0.924	0.902	0.942	0.934	0.909	0.796
(14)	$Sales_{n-1}$ & CAT	0.065	0.933	0.930	0.922	0.90	0.941	0.933	0.909	0.795
(15)	$Sales_{n-1}$ & Peer	0.078	0.920	0.918	0.910	0.888	0.932	0.919	0.872	0.676
(16)	$Sales_{n-1}$ & Peer & HDD & CDD	0.078	0.920	0.918	0.910	0.888	0.931	0.918	0.869	0.669
(17)	$Sales_{n-1}$ & Peer & AT	0.078	0.921	0.919	0.912	0.890	0.931	0.918	0.870	0.671
(18)	$Sales_{n-1}$ & Peer & SAT	0.079	0.920	0.918	0.911	0.890	0.931	0.918	0.870	0.671
(19)	$Sales_{n-1}$ & Peer & CAT	0.077	0.922	0.919	0.913	0.892	0.932	0.92	0.873	0.675

Table 4.10 Error detection - alpha = 0.2

		False Positive	False Negative							
Model #	Included variables		Worst Scenario				Best Scenario			
			0.5%	1%	2%	4%	0.5%	1%	2%	4%
(1)	HDD & CDD	0.422	0.575	0.572	0.562	0.536	0.683	0.667	0.604	0.413
(2)	AT	0.391	0.606	0.603	0.598	0.576	0.690	0.673	0.609	0.425
(3)	SAT	0.385	0.613	0.611	0.605	0.585	0.690	0.674	0.611	0.429
(4)	CAT	0.420	0.577	0.575	0.569	0.549	0.676	0.661	0.598	0.422
(5)	Peer	0.440	0.558	0.556	0.551	0.529	0.615	0.595	0.515	0.299
(6)	Peer & HDD & CDD	0.449	0.551	0.549	0.544	0.521	0.617	0.596	0.514	0.287
(7)	Peer & AT	0.442	0.557	0.556	0.550	0.529	0.618	0.598	0.516	0.290
(8)	Peer & SAT	0.444	0.556	0.555	0.548	0.526	0.618	0.597	0.516	0.291
(9)	Peer & CAT	0.444	0.556	0.554	0.548	0.525	0.616	0.597	0.516	0.296
(10)	$Sales_{n-1}$	0.408	0.590	0.587	0.579	0.556	0.715	0.699	0.620	0.377
(11)	$Sales_{n-1}$ & HDD & CDD	0.417	0.583	0.580	0.573	0.550	0.712	0.696	0.617	0.373
(12)	$Sales_{n-1}$ & AT	0.409	0.591	0.589	0.583	0.562	0.714	0.699	0.621	0.378
(13)	$Sales_{n-1}$ & SAT	0.409	0.590	0.589	0.583	0.560	0.714	0.699	0.621	0.379
(14)	$Sales_{n-1}$ & CAT	0.417	0.582	0.579	0.571	0.547	0.713	0.697	0.617	0.373
(15)	$Sales_{n-1}$ & Peer	0.422	0.577	0.576	0.569	0.543	0.631	0.611	0.517	0.269
(16)	$Sales_{n-1}$ & Peer & HDD & CDD	0.430	0.570	0.569	0.563	0.538	0.628	0.608	0.512	0.263
(17)	$Sales_{n-1}$ & Peer & AT	0.425	0.574	0.573	0.567	0.543	0.629	0.609	0.514	0.265
(18)	$Sales_{n-1}$ & Peer & SAT	0.427	0.572	0.572	0.567	0.542	0.629	0.609	0.513	0.266
(19)	$Sales_{n-1}$ & Peer & CAT	0.429	0.571	0.570	0.563	0.538	0.632	0.611	0.516	0.268

Discussion of severe weather

Not only are weather variables related to temperature, but they are also related to unfavorable weather conditions, such as tornadoes, precipitation, and storms. Although I conduct an extensive analysis with additional weather variables, sufficient results could not be found. There are several possible reasons.

First, severe weather events occur rarely and in limited regions. For example, by using the Saffir-Simpson Hurricane Wind Scale (SSHWS) and the Fujita scale, it is possible to examine how severe thunder and winter storms affect firm level sales. But any major or minor influence on sales is found. Since this firm owns a number of stores in the U.S., when a limited number of operating units are affected by severe weather events the firm is still too large to see the influence of the events on firm level sales. Perhaps the auditor may be interested in those events when they examine individual store level sales, but not a huge multi-locational firm. In this case study, I also focus on the exact date, the day prior, and the day after a severe weather event and find that around one percent of stores sales are affected. Therefore, it is hard to find a correlation between severe weather and weekly store sales.

Second, precipitation is definitely related to sales, but adding precipitation to the models does not improve the predictive accuracy of sales. The incremental coefficient calculated by store level sales and precipitation is only 0.004. To avoid over-fitting issues, I apply a stepwise regression with precipitation, but precipitation is only included in a very limited number of cases. However, the correlation coefficient between the residuals derived from the model with peers and precipitations, is -0.18. It suggests that precipitation may have a value which peer stores cannot offer.

Weather variables have an explanatory value since sales patterns and weather patterns resemble each other. Nevertheless, since the firm used in this case study owns a number of stores, it is difficult to capture the influence of a weather event on the firm's sales.

CONCLUSION, LIMITATIONS, AND FUTURE RESEARCH

Retail sales are likely to be affected by weather conditions as such conditions affect customer behavior. For certain industries, weather variables are revenue-relevant information. By considering the relationship between sales and weather, the auditor can enhance analytical models to detect material misstatements. In addition, since weather information is timely and location-specific, it can provide sufficient audit evidence. Finally, since weather variables are external information and not affected by management, reliability is enhanced.

This research is an attempt to fill a gap in existing literature by suggesting ways to develop analytical procedures with external nonfinancial information. In particular, since the associations between weather variables and sales are nonlinear and seasonal, this work contributes to audit practice and academia by illustrating various approaches to link weather conditions to sales.

Results show that contemporaneous weather variables are correlated with revenue. Models with weather variables are more likely to reduce false positive rates but less likely to detect genuine errors than the models with the average sales of peer stores. When taking account of the contemporaneous sales of peer stores, weather variables are less likely to provide incremental information to forecasting revenue accounts. Nevertheless, in terms of false negatives which are connected to audit effectiveness the best model contains weather variables, peer sales, and lagged sales. Weather variables can enhance audit effectiveness even though they are less effective at forecasting revenue than peer sales. Even though contemporaneous peer sales play an important role in understanding sales accounts, peer store sales are less reliable than weather variables.

This study has some limitations. Because only a single firm is analyzed, generalizability is limited. Since the influence of weather on sales varies depending on industries, the careful utilization of weather variables is required. Accordingly, this study can be developed further through application to firms in various industries and locations. Also, this study uses a single account to develop a model, unlike other previous studies which generate models with other accounts. If other accounts such as cost of goods sold are added to the model, the incremental value of weather indicators might be limited further.

Chapter 5 Conclusions

DISCUSSION

The purpose of this dissertation is to investigate the significance of nontraditional audit evidence in improving audit quality. Specifically, this dissertation assess: 1) whether the tone of management qualitative disclosures influences audit fee decisions in initial engagements; 2) whether substantive analytical procedures developed with nontraditional audit evidence replaced by audit sampling improves audit quality for inspecting revenue accounts; and 3) whether weather variables improve substantive analytical procedures. By utilizing different forms of audit evidence, such as management qualitative disclosures and weather variables and by examining different stages of audit procedures, such as risk assessments in pre-engagement stage and substantive tests, the significance of nontraditional audit evidence is tested.

In the first essay, the importance of management qualitative disclosures in audit fee decisions in the pre-engagement stage is examined. Audit fee decisions are often understood as actions in response to a client's business risks. Numerical information regarding business risks are highly correlated with each other, but textual information might provide independent information that are not provided by numerical information (Li 2006). In particular, the tone of optimism in management qualitative disclosures, which might be related to client business and litigation risks, is examined and added to the traditional audit fee model. The results indicate that new variables play an important role in understanding the audit fee decision. In addition, this essay examines how the association between the tone of optimism in disclosures and audit fee decisions are modified by issuing a going-concern opinion, which is considered to be an indicator of an auditor's perceived business risks. The relationship between the tone of optimism in qualitative information in 10-K reports and audit fees becomes weaker in the presence of a going-concern opinion, indicating that auditors respond to the inconsistency between their perceived risks and management disclosures. On the other hand, the relationship between 8-K filings and audit fees is stronger if the client receives a going-concern opinion because auditors might put more weight on qualitative disclosures in situations with high uncertainty caused by significant client business risks. The results indicate that successor auditors are likely to use textual information for audit fee decisions in new engagements in order to examine client business risks and management fraudulent behavior.

This study contributes to the literature on the implications of qualitative information on a firm's risk evaluation. Even though auditors examine a client's business risks extensively in the pre-engagement stage, little is known about whether external auditors perceive these qualitative sources to be business and litigation risk indicators. In addition, this paper contributes to prior research with regard to audit fee decisions and risk evaluations. First, it provides additional variables related to developing a theory of audit fees. Hay et al. (2006) posit that prior audit pricing studies contain issues, such as inadequate control, variable proxies, and omitted variables. Adding variables from qualitative management disclosures creates new and independent components, which the existing numerical variables cannot contribute to the understanding of audit fee decisions. In addition, since considering qualitative management disclosures in audit fee decisions is based on audit practice (Louwers et al. 2013), this study attempts to develop realistic audit fee models.

The first essay demonstrates that auditors actually use nontraditional audit evidence for their audit fee decisions, and explores the audit procedures, other than risk assessment, in which nontraditional information could be effective and efficient audit evidence. The second essay addresses one of the current issues in audit practice, how substantive analytical procedures (SAPs) developed with nontraditional audit evidence can enhance audit quality. Based on a reading of the existing literature, this essay suggests that, in certain conditions, SAPs outperform audit sampling and *vice versa*. For instance, SAPs might be more effective than audit sampling when auditors examine large populations, when various sources of information should be considered, and/or when large errors are randomly distributed. On the other hand, in cases where client internal controls are not reliable and a number of small errors are evenly distributed, audit sampling might be preferred. Particularly, SAPs are beneficial for inspecting revenue accounts because often external factors such as industry competition are meaningful to examine revenue accounts and underlying transactions are large. Nevertheless, even if auditors develop expectations carefully and rigorously, it might be hard to get precise expectations for revenue accounts (Glover et al. 2015). Accordingly, in most cases, SAPs and audit sampling are complementary. Therefore, choosing the appropriate substantive test is a significant step in audit procedures and should be determines based on the audit environment and audit objectives.

This essay contributes to the audit literature, audit practice, and regulation by presenting the significance of SAPs. Prior studies have generally examined the effectiveness of audit sampling/SAPs in order to identify misstatements or to suggest how to improve each substantive test. This study identifies factors that might affect the effectiveness of substantive tests and suggests possible circumstances in which either substantive test might be more effective or the corroborations of both types of substantives test will outperform either test alone. Auditors might consider this when they choose appropriate substantive tests in certain circumstances and might be motivated to utilize more sophisticated SAPs. Finally, based on the arguments presented in this essay, the PCAOB might need to reconsider its opinion about the effectiveness of SAPs and find a way to resolve unintended effects on audit quality caused by their inspections.

The third essay provides specific examples of how to utilize nontraditional audit evidence. In particular, it examines weather variables, based on the existing literature on how weather variables can influence a retailer's sales. For revenue accounts in certain industries, weather variables can have the potential to become relevant audit evidence. In addition, weather indicators are easily accessible and available in a timely fashion, and are external data that cannot be affected by the client. Accordingly, in certain cases, weather variables can offer reliable and sufficient audit evidence. To test this hypothesis, this essay researches the extent that daily and location-specific weather information has explanatory value for the store-level sales account. Using daily store sales of a retailer operating about 2,000 stores throughout the U.S., the existing analytical models with and without weather variables are compared. As a control variable, the average sales of peer stores that share macroeconomic characteristics are added in the model. This study finds that the trends of weather and the movement of sales are similar, although when weather variables are added to the model as a control variable, the sales of peer stores may not provide sufficient incremental values to improve SAPs. In terms of false negatives which are related to audit effectiveness the best model contains weather variables, peer sales, and lagged sales. Accordingly, weather variables can improve audit effectiveness even though they are less effective at forecasting revenue than peer sales.

This study offers useful insights to both audit practice and to academia in terms of the significance of weather variables as audit evidence, and it suggests how to utilize these variables in audit procedures. This study also uses data comprising a large number of stores in the U.S., providing the necessary heterogeneity to improve expectations. Allen et al. (1999) do not find that store-level disaggregated data enhances analytical procedures and explain that this result is due to the homogeneous nature of the operation across the firm's thirty stores. They do not find substantial difference between the disaggregated and aggregated approaches. Nevertheless, even though individual stores may offer similar services and products, local customer characteristics and economic conditions for each store create a variance in sales that would be different from stores in other areas. This paper contributes to reevaluating the value of disaggregated data by location.

LIMITATIONS

As with any study, this dissertation contains some limitations that may limit the ability to generalize the findings. The potential limitations of the first essay are in rule-based textual analysis issues. Usually, the approaches to textual analysis are classified in one of two ways: rule-based approach and statistical approach. In the rule-based approach, a computer program reads the content and then classifies it into certain groups based on predefined rules, such as a dictionary. On the other hand, the statistical inference approach analyzes content that is pre-classified into certain groups (training) and then classifies the information based on the statistical inference. Although many studies employ rule-based approaches (e.g., Li 2006; Tetlock et al. 2008; Rogers et al. 2011), there are limitations to this approach. For example, the rule-based approach does not take into account the structure of sentences, so it could not examine the negation of words in the sentence, thereby providing biased results. In addition, the results could be different, depending on the dictionary used. For instance, Roger et al. (2011) use three different dictionaries to analyze the tone of optimism, but the results are somewhat different depending on the dictionary used.

This second study is a commentary and empirical test of the outcomes from prior studies. Such studies attempt to enhance the substantive analytical procedures that are examined in different data sets and models. Most of studies examine only a small number of firms except Hoitash et al. (2006). In addition, a meta-analysis was conducted in only seven of the previous studies. Accordingly, the outcomes of the previous studies might be difficult to generalize for other firms. Even though this essay provides a broad picture to determine appropriate substantive tests, the auditor might need to consider more factors. Therefore, it does not offer detailed information needed by auditors to support appropriate substantive tests.

The biggest limitation of the third essay is related to its research method, the case study. Since SAPs are often specifically developed for an individual client and it is difficult to examine a firm's detailed data, the previous literature commonly employs the case study methodology. In addition, a case study can provide the opportunity to analyze a firm's information. However, the results from this analysis are tentative because they are based on a single firm. Additionally, the influence of weather on sales could be different depending on the industry, and so these findings might not apply in all conditions.

FUTURE RESEARCH

The findings of this dissertation set the stage for additional research opportunities examining nontraditional audit evidence. One possible topic for additional study would be the examination of other form of audit evidence in different audit procedures. This study examines management qualitative disclosures and weather variables, but a number of relevant and reliable nontraditional audit evidence should be examined, such as: 1) How utilizing radio-frequency identification (RFID) improves the accuracy of examining inventory accounts; 2) How the tone of customers' products or service reviews on social media improve SAPs for revenue accounts.

In addition, the nontraditional audit evidence suggested in this dissertation could be analyzed in different ways and tested in different stages of audit. For instance, it might be an interesting extension to analyze textual information using both rule-based and statistical inference approach for the same research question and then compare the results. Moreover, it would be valuable to study the components in management qualitative disclosures as a variable in SAPs or the association between weather variables and risk assessments. Regarding the second essay, it would be interesting to test and compare SAP and audit sampling empirically. Finally, it would be interesting to see whether the association between weather and sales is modified by industry characteristics, as well as how to modify SAP to employ weather variables.

REFERENCES

———. 2002. Consideration of Fraud in Financial Statement Audit. In *AU Section 316*. New York: American Institute of Certified Public Accountants.

———. 2006a. Audit Evidence. In *AU Section 326*. New York: American Institute of Certified Public Accountants.

———. 2006b. Audit Risk and Materiality in Conducting an Audit. In *Statement on Auditing Standard No. 107*. New York, NY: American Institute of Certified Public Accountants.

———. 2006c. Understanding the Entity and Its Environment and Assessing the Risks of Material Misstatement. In *Statement on Auditing Standard No. 109*. New York, NY: American Institute of Certified Public Accountants.

———. 2012a. Analytical Procedure. In *AU-C Section 520*. New York: American Institute of Certified Public Accountants.

———. 2012b. Audit Sampling. In *AU-C Section 530*. New York: American Institute of Certified Public Accountants.

———. 2012c. Materiality in Planning and Performing an Audit. In *AU-C Section 320*. New York: American Institute of Certified Public Accountants.

———. 2012d. Performing Audit Procedures in Response to Assessed Risks and Evaluating the Audit Evidence Obtained. In *AU-C Section 330*. New York: American Institute of Certified Public Accountants.

Allen, R. D., M. S. Beasley, and B. C. Branson. 1999. Improving analytical procedures: A case of using disaggregate multilocation data. *Auditing: A Journal of Practice & Theory* 18 (2):128-142.

Bell, T. B., J. C. Bedard, K. M. Johnstone, and E. F. Smith. 2002. KRiskSM: A computerized decision aid for client acceptance and continuance risk assessments. *Auditing: A Journal of Practice & Theory* 21 (2):97-113.

Bell, T. B., R. Doogar, and I. Solomon. 2008. Audit labor usage and fees under business risk auditing. *Journal of Accounting Research* 46 (4):729-760.

Bell, T. B., F. O. Marrs, I. Solomon, H. Thomas, and K. P. M. LLP. 1997. *Auditing organizations through a strategic-systems lens: The KPMG business measurement process*.

Bell, T. B., M. E. Peecher, and I. Solomon. 2005. *The 21st Century Public Company Audit: Conceptuals Elements of KPMG's Global Audit Methodology*: KPMG International.

Beneish, M. D. 1997. Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance. *Journal of Accounting and Public Policy* 16 (3):271-309.

Bennett, G. B., and R. C. Hatfield. 2012. The effect of the social mismatch between staff auditors and client management on the collection of audit evidence. *The Accounting Review* 88 (1):31-50.

Bentley, K. A., T. C. Omer, and N. Y. Sharp. 2013. Business strategy, financial reporting irregularities, and audit effort. *Contemporary accounting research* 30 (2):780-817.

Biggs, S. F., C. C. G. A. R. Foundation, T. J. Mock, and P. R. Watkins. 1989. *Analytical review procedures and processes in auditing*: Canadian Certified General

Accountants' Research Foundation= Fondation de recherche de l'Association des comptables généraux licenciés du Canada.
Biggs, S. F., and J. J. Wild. 1985. An investigation of auditor judgment in analytical review. *Accounting Review*:607-633.
Blocher, E., and J. J. Willingham. 1985. *Analytical review: A guide to evaluating financial statements*: McGraw-Hill.
Bonini, S., F. Pavesi, and M. Scotti. 2011. Financial Analysts and Collective Reputation: Theory and Evidence. *Available at SSRN 1965451*.
Brazel, J. F., K. L. Jones, and D. F. Prawitt. 2014. Auditors' Reactions to Inconsistencies between Financial and Nonfinancial Measures. *Behavioral Research in Accounting*: Spring, Vol. 26, No. 1, pp. 131-156.
Brown, S. V., and J. W. Tucker. 2011. Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications. *Journal of Accounting Research* 49 (2):309-346.
Brumfield, C. A., R. K. Elliott, and P. D. Jacobson. 1983. Business risk and the audit process. *Journal of Accountancy* 155 (4):60-68.
Buhl, H. U., M. Röglinger, D.-K. F. Moser, and J. Heidemann. 2013. Big Data. *Wirtschaftsinformatik* 55 (2):63-68.
Carson, E., N. L. Fargher, M. A. Geiger, C. S. Lennox, K. Raghunandan, and M. Willekens. 2013. Audit Reporting for Going-Concern Uncertainty: A Research Synthesis. *Auditing: A Journal of Practice & Theory* 32 (Supplement 1):353-384.
Caster, P., D. W. Massey, and A. Wright. 2000. Research on the nature, characteristics, and causes of accounting errors: The need for a multi-method approach. *Journal of Accounting Literature* 19.
Cenker, W. J., and A. L. Nagy. 2008. Auditor resignations and auditor industry specialization. *Accounting Horizons* 22 (3):279-295.
Chen, L., G. Krishnan, and M. Pevzner. 2012. Pro forma disclosures, audit fees, and auditor resignations. *Journal of Accounting and Public Policy* 31 (3):237-257.
Chen, Y., and R. A. Leitch. 1998. The error detection of structural analytical procedures: A simulation study. *Auditing* 17 (2):36.
———. 1999. An analysis of the relative power characteristics of analytical procedures. *Auditing: A Journal of Practice & Theory* 18 (2):35-69.
Choi, J.-H., R. K. Doogar, and A. R. Ganguly. 2004. The Riskiness of Large Audit Firm Client Portfolios and Changes in Audit Liability Regimes: Evidence from the US Audit Market*. *Contemporary accounting research* 21 (4):747-785.
Choi, J.-H., C. Kim, J.-B. Kim, and Y. Zang. 2010. Audit Office Size, Audit Quality, and Audit Pricing. *Auditing: A Journal of Practice & Theory* 29 (1):73-97.
Christensen, B. E., R. J. Elder, and S. M. Glover. 2015. Behind the Numbers: Insights into Large Audit Firm Sampling Policies. *Accounting Horizons* 29 (1):61-81.
Cohen, J. R., G. Krishnamoorthy, and A. M. Wright. 2000. Evidence on the effect of financial and nonfinancial trends on analytical review. *Auditing: A Journal of Practice & Theory* 19 (1):27-48.
Colbert, J. L., M. S. Luehlfing, and C. W. Alderman. 1996. Auditing–Engagement Risk. *The CPA Journal*.
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*.

Davis, A. K., and I. Tama-Sweet. 2012. Managers' Use of Language Across Alternative Disclosure Outlets: Earnings Press Releases versus MD&A*. *Contemporary accounting research* 29 (3):804-837.

Dutton, J. A. 2002. Opportunities and priorities in a new era for weather and climate services. *Bulletin of the American Meteorological Society* 83 (9):1303-1311.

Dzeng, S. C. 1994. A comparison of analytical procedure expectation models using both aggregate and disaggregate data. *AUDITING-A JOURNAL OF PRACTICE & THEORY* 13 (2):1-24.

Eilifsen, A., W. R. Knechel, and P. Wallage. 2001. Application of the business risk audit model: A field study. *Accounting Horizons* 15 (3):193-207.

Eilifsen, A., and W. F. Messier. 2015. Materiality Guidance of the Major Public Accounting Firms. *Auditing: A Journal of Practice & Theory* 34 (2):3-26.

Elder, R. J., A. D. Akresh, S. M. Glover, J. L. Higgs, and J. Liljegren. 2013. Audit Sampling Research: A Synthesis and Implications for Future Research. *Auditing: A Journal of Practice & Theory* 32 (Supplement 1):99-129.

Engle, R. F., C. W. Granger, J. Rice, and A. Weiss. 1986. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association* 81 (394):310-320.

Ethridge, J. R., T. Marsh, and B. Revelt. 2011. Engagement risk: Perceptions and strategies from audit partners. *Journal of Business & Economics Research (JBER)* 5 (4).

Feinberg, E. A., and D. Genethliou. 2005. Load forecasting. In *Applied mathematics for restructured electric power systems*: Springer, 269-285.

FENG, M. and LI, C. 2014. Are Auditors Professionally Skeptical? Evidence from Auditors' Going-Concern Opinions and Management Earnings Forecasts. *Journal of Accounting Research* 52: 1061–1085. doi: 10.1111/1475-679X.12064

Francis, J. R., and D. Wang. 2005. Impact of the SEC's public fee disclosure requirement on subsequent period fees and implications for market efficiency. *Auditing: A Journal of Practice & Theory* 24 (s-1):145-160.

———. 2008. The joint effect of investor protection and big 4 audits on earnings quality around the world*. *Contemporary accounting research* 25 (1):157-191.

Ghosh, A., and R. Pawlewicz. 2009. The impact of regulation on auditor fees: Evidence from the Sarbanes-Oxley Act. *Auditing: A Journal of Practice & Theory* 28 (2):171-197.

Gibbins, M., and J. D. Newton. 1994. An empirical exploration of complex accountability in public accounting. *Journal of Accounting Research* 32 (2):165-186.

Glass, G. V. 1976. Primary, secondary, and meta-analysis of research. *Educational researcher* 5 (10):3-8.

Glover, S. M., D. F. Prawitt, and M. S. Drake. 2015. Between a Rock and a Hard Place: A Path Forward for Using Substantive Analytical Procedures in Auditing Large P&L Accounts: Commentary and Analysis. *Auditing: A Journal of Practice & Theory* 34 (3):161-179.

Griffith, E. E., J. S. Hammersley, and K. Kadous. 2015a. Audits of Complex Estimates as Verification of Management Numbers: How Institutional Pressures Shape Practice. *Contemporary accounting research* 32 (3):833-863.

Griffith, E. E., J. S. Hammersley, K. Kadous, and D. Young. 2015b. Auditor mindsets and audits of complex estimates. *Journal of Accounting Research* 53 (1):49-77.

Guy, D. M. 1981. *An introduction to statistical sampling in auditing*: John Wiley & Sons.

Hackenbrack, K. E., and C. E. Hogan. 2005. Client retention and engagement-level pricing. *Auditing: A Journal of Practice & Theory* 24 (1):7-20.

Hackenbrack, K. E., N. T. Jenkins, and M. Pevzner. 2014. Relevant but delayed information in negotiated audit fees. *Auditing: A Journal of Practice & Theory* 33 (4):95-117.

Hay, D. C., W. R. Knechel, and N. Wong. 2006. Audit Fees: A Meta-analysis of the Effect of Supply and Demand Attributes*. *Contemporary accounting research* 23 (1):141-191.

Henry, E. 2006. Market reaction to verbal components of earnings press releases: Event study using a predictive algorithm. *Journal of Emerging Technologies in Accounting* 3 (1):1-19.

Hirst, D., and L. Koonce. 1996. Audit Analytical Procedures: A Field Investigation*. *Contemporary accounting research* 13 (2):457-486.

Hoitash, R., A. Kogan, and M. A. Vasarhelyi. 2006. Peer-based approach for analytical procedures. *Auditing: A Journal of Practice & Theory* 25 (2):53-84.

Hoogduin, L. A., T. W. Hall, J. J. Tsay, and B. J. Pierce. 2015. Does Systematic Selection Lead to Unreliable Risk Assessments in Monetary-Unit Sampling Applications? *Auditing: A Journal of Practice & Theory* 34 (4):85-107.

Huang, H.-W., K. Raghunandan, and D. Rama. 2009. Audit fees for initial audit engagements before and after SOX. *Auditing: A Journal of Practice & Theory* 28 (1):171-190.

Ittner, C. D., and D. F. Larcker. 1998. Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research* 36:1-35.

Iyer, A. P. N., A. Chatterjee, and J. Kishnani. 2010. Compiler transformations to enable synchronous execution in an RIA runtime. *Internet Computing, IEEE* 14 (3):13-23.

Johnson, J. R., R. A. Leitch, and J. Neter. 1981. Characteristics of errors in accounts receivable and inventory audits. *Accounting Review*:270-293.

Johnstone, K. M., and J. C. Bedard. 2004. Audit firm portfolio management decisions. *Journal of Accounting Research* 42 (4):659-690.

Kachelmeier, S. J., and W. F. Messier Jr. 1990. An investigation of the influence of a nonstatistical decision aid on auditor sample size decisions. *Accounting Review*:209-226.

Kinney Jr, W. R. 1978. ARIMA and regression in analytical review: An empirical test. *Accounting Review*:48-60.

Kinney, W. R. 1979. Integrating audit tests: Regression analysis and partitioned dollar-unit sampling. *Journal of Accounting Research*:456-475.

———. 1987. Attention-Directing Analytical Review Using Ratio- A Case Study *AUDITING-A JOURNAL OF PRACTICE & THEORY* 6 (2):59-73.

Kinney, W. R., and G. L. Salamon. 1982. Regression analysis in auditing: A comparison of alternative investigation rules. *Journal of Accounting Research* 20 (2):350-366.

Knechel, W. R. 1988. The Effectiveness of Statistical Analytical Review as a Substantive Auditing Procedure - a Simulation Analysis. *Accounting Review* 63 (1):74-95.

Knechel, W. R. 2007. The business risk audit: Origins, obstacles and opportunities. *Accounting, Organizations and Society* 32 (4-5):383-408.

Kogan, A., M. G. Alles, M. A. Vasarhelyi, and J. Wu. 2014. Design and Evaluation of a Continuous Data Level Auditing System. *Auditing: A Journal of Practice & Theory* 33 (4):221-245.

Kopp, L. S., and E. O'Donnell. 2005. The influence of a business-process focus on category knowledge and internal control evaluation. *Accounting, Organizations and Society* 30 (5):423-434.

Koskivaara, E. 2004. Artificial neural networks in analytical review procedures. *Managerial Auditing Journal* 19 (2):191-223.

Kothari, S., X. Li, and J. E. Short. 2009. The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: a study using content analysis. *The Accounting Review* 84 (5):1639-1670.

Krishnan, G. V., M. Pevzner, and P. Sengupta. 2012. How do auditors view managers' voluntary disclosure strategy? The effect of earnings guidance on audit fees. *Journal of Accounting and Public Policy* 31 (5):492-515.

Krishnan, J., and J. Krishnan. 1997. Litigation risk and auditor resignations. *The Accounting Review*:539-560.

Krishnan, J., H. Sami, and Y. Zhang. 2005. Does the provision of nonaudit services affect investor perceptions of auditor independence? *Auditing: A Journal of Practice & Theory* 24 (2):111-135.

Larcker, D. F., and A. A. Zakolyukina. 2012. Detecting deceptive discussions in conference calls. *Journal of Accounting Research* 50 (2):495-540.

Larsen, P. H. 2006. An evaluation of the sensitivity of US economic sectors to weather. *Available at SSRN 900901*.

Lazo, J. K., M. Lawson, P. H. Larsen, and D. M. Waldman. 2011. U.S. Economic Sensitivity to Weather Variability. *Bulletin of the American Meteorological Society* 92 (6):709-720.

Lee, H. Y., V. Mande, and R. Ortman. 2004. The effect of audit committee and board of director independence on auditor resignation. *Auditing: A Journal of Practice & Theory* 23 (2):131-146.

Lehavy, R., F. Li, and K. Merkley. 2011. The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts. *The Accounting Review* 86 (3):1087-1115.

Leitch, R. A., and Y. Chen. 2003. The effectiveness of expectation models in recognizing error patterns and generating and eliminating hypotheses while conducting analytical procedures. *Auditing: A Journal of Practice & Theory* 22 (2):147-170.

Lev, B. 1980. On the use of index models in analytical reviews by auditors. *Journal of Accounting Research*:524-550.

Levi, S. 2008. Voluntary disclosure of accruals in earnings press releases and the pricing of accruals. *Review of Accounting Studies* 13 (1):1-21.

Levine, C. B., and M. J. Smith. 2011. Critical Accounting Policy Disclosures. *Journal of Accounting, Auditing & Finance* 26 (1):39-76.

Li, F. 2006. Do stock market investors understand the risk sentiment of corporate annual reports? *Working paper*.

———. 2011. Textual analysis of corporate disclosures: A survey of the literature.

Loebbecke, J. K., and P. J. Steinbart. 1987. An investigation of the use of preliminary analytical review to provide substantive audit evidence. *A Auditing: A Journal of Practice & Theory* 6 (2):74-89.

Lorek, K. S., B. C. Branson, and R. C. Icerman. 1992. On the use of time-series models as analytical procedures. *AUDITING-A JOURNAL OF PRACTICE & THEORY* 11 (2):66-87.

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

Louwers, T. J., R. J. Ramsay, D. H. Sinason, and J. R. Strawser. 2004. Auditing & Assurance Services in the 21 st Century: New York: McGraw-Hill Irwin.

Louwers, T. J., R. J. Ramsay, D. H. Sinason, J. R. Strawser, and J. C. Thibodeau. 2013. *Auditing and assurance services*: McGraw-Hill/Irwin New York, NY.

Louwers, T. J., R. J. Ramsey, D. H. Sinason, and J. R. Strawser. 2007. *Auditing & assurance services*: McGraw-Hill.

Luft, J. 2009. Nonfinancial Information and Accounting: A Reconsideration of Benefits and Challenges. *Accounting Horizons* 23 (3):307-325.

Lyon, J. D., and M. W. Maher. 2005. The importance of business risk in setting audit fees: evidence from cases of client misconduct. *Journal of Accounting Research* 43 (1):133-151.

Ma, P. 2012. Information or Spin? Evidence from Language Differences Between 8-Ks and Press Releases.

Merkley, K. J. 2013. Narrative disclosure and earnings performance: Evidence from R&D disclosures. *The Accounting Review* 89 (2):725-757.

Messier, W., S. Glover, and D. Prawitt. 2013a. *Auditing & assurance services: A systematic approach*: McGraw-Hill Higher Education.

Messier, W. F., C. A. Simon, and J. L. Smith. 2013b. Two Decades of Behavioral Research on Analytical Procedures: What Have We Learned? *Auditing: A Journal of Practice & Theory* 32 (1):139-181.

Murray, K. B., F. Di Muro, A. Finn, and P. P. Leszczyc. 2010. The effect of weather on consumer spending. *Journal of Retailing and Consumer Services* 17 (6):512-520.

Neter, J., and J. K. Loebbecke. 1975. *Behavior of major statistical estimators in sampling accounting populations: An empirical study*: American Institute of Certified Public Accountants.

O'Donnell, E., and J. J. Schultz Jr. 2005. The halo effect in business risk audits: Can strategic risk assessment bias auditor judgment about accounting details? *The Accounting Review* 80 (3):921-939.

O'Keefe, T. B., D. A. Simunic, and M. T. Stein. 1994. The production of audit services: Evidence from a major public accounting firm. *Journal of Accounting Research*:241-261.

Omura, T., and R. Willett. 2006. Using automated equilibrium correction modelling in analytic review. *Managerial Auditing Journal* 21 (2):207-223.

Palmiter, A. R. 2008. *Securities regulation: Examples and explanations*: Aspen Publishers Online.

Patelli, L., and M. Pedrini. 2013. Is Tone at the Top Associated with Financial Reporting Aggressiveness? *Journal of Business Ethics*:1-17.

PCAOB. 2004. PCAOB Standing Advisory Group Meeting: Financial fraud (September 8–9).

———. 2007. Auditing Standard No 5: Public Company Accounting Oversight Board

———. 2008. Report on the PCAOB'S 2004, 2005, 2006,and 2007 Inspections of Domestic Annual Inspected Firms Public Company Accounting Oversight Board.

———. 2010. Substantive Anlaytical Procedures: Public Company Accounting Oversight Board

———. 2011. PCAOB Release No. 104-2011-289 Public Company Accounting Oversight Board.

———. 2014. Staff Audit Practice Alert No 12: Public Company Accounting Oversight Board.

Peecher, M. E., R. Schwartz, and I. Solomon. 2007. It’s all about audit quality: Perspectives on strategic-systems auditing. *Accounting, Organizations and Society* 32 (4):463-485.

Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22 (1):435-480.

Pike, B. J., M. B. Curtis, and L. Chui. 2013. How Does an Initial Expectation Bias Influence Auditors' Application and Performance of Analytical Procedures? *The Accounting Review* 88 (4):1413-1431.

Pratt, J., and J. D. Stice. 1994. The effects of client characteristics on auditor litigation risk judgments, required audit evidence, and recommended audit fees. *Accounting Review*:639-656.

Puttick, G., S. Van Esch, S. D. Van Esch, and S. Kana. 2008. *The principles and practice of auditing*: Juta and Company Ltd.

Rittenberg, L. E., K. M. Johnstone, and A. A. Gramling. 2010. Auditing: A business risk approach.

Robson, K., C. Humphrey, R. Khalifa, and J. Jones. 2007. Transforming audit technologies: Business risk audit methodologies and the audit field. *Accounting, Organizations and Society* 32 (4-5):409-438.

Rogers, J. L., A. Van Buskirk, and S. L. Zechman. 2011a. Disclosure tone and shareholder litigation. *The Accounting Review* 86 (6):2155-2183.

Rothfusz, L. P., and N. S. R. Headquarters. 1990. The heat index equation (or, more than you ever wanted to know about heat index). *Fort Worth, Texas: National Oceanic and Atmospheric Administration, National Weather Service, Office of Meteorology*:90-23.

Schultz, J. J., J. L. Bierstaker, and E. O’Donnell. 2010. Integrating business risk into auditor judgment about the risk of material misstatement: The influence of a strategic-systems-audit approach. *Accounting, Organizations and Society* 35 (2):238-251.

Seetharaman, A., F. A. Gul, and S. G. Lynn. 2002. Litigation risk and audit fees: Evidence from UK firms cross-listed on US markets. *Journal of Accounting and Economics* 33 (1):91-115.

Simunic, D. A. 1980. The pricing of audit services: Theory and evidence. *Journal of Accounting Research* 18 (1):161-190.

Smieliauskas, W. 1986. Control of sampling risks in auditing*. *Contemporary accounting research* 3 (1):102-124.

Stanley, J. D. 2011. Is the Audit Fee Disclosure a Leading Indicator of Clients' Business Risk? *Auditing: A Journal of Practice & Theory* 30 (3):157-179.

Starr-McCluer, M. 2000. *The effects of weather on retail sales*: Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board.

Stice, J. D. 1991. Using financial and market information to identify pre-engagement factors associated with lawsuits against auditors. *The Accounting Review* 66 (3):516-533.

Stringer, K. W. 1975. A statistical technique for analytical review. *Journal of Accounting Research*:1-9.

Tabor, R. H., and J. T. Willis. 1985. EMPIRICAL-EVIDENCE ON THE CHANGING-ROLE OF ANALYTICAL REVIEW PROCEDURES: AMER ACCOUNTING ASSOC 5717 BESSIE DR, SARASOTA, FL 34233, 93-109.

Teitlebaum, A., and C. Robinson. 1975. The real risks in audit sampling. *Journal of Accounting Research*:70-91.

Tetlock, P. C., M. SAAR‐TSECHANSKY, and S. Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance* 63 (3):1437-1467.

Tol, R. S. J., C. Dorland, W. Lise, A. Olsthoorn, and F. Spaninks. 2000. *Weather impacts on natural, social and economic systems in The Netherlands*: Vrije Universiteit, Institute for Environmental Studies.

Trompeter, G., and A. Wright. 2010. The World Has Changed-Have Analytical Procedure Practices?*. *Contemporary accounting research* 27 (2):669-700.

Trotman, K. T., and W. F. Wright. 2012. Triangulation of audit evidence in fraud risk assessments. *Accounting, Organizations and Society* 37 (1):41-53.

Vadnais, K. 2012. " Does the Media Affect Mergers and Acquisitions?, New York University.

Valentine, J. C., T. D. Pigott, and H. R. Rothstein. 2010. How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics* 35 (2):215-247.

Vandervelde, S. D., Y. Chen, and R. A. Leitch. 2008. Auditors’ Cross-Sectional and Temporal Analysis of Account Relations in Identifying Financial Statement Misstatements. *Auditing: A Journal of Practice & Theory* 27 (2):79-107.

Venkataraman, R., J. P. Weber, and M. Willenborg. 2008. Litigation risk, audit quality, and audit fees: Evidence from initial public offerings. *The Accounting Review* 83 (5):1315-1345.

Wallace, W. A. 1982. Analytical review developments in practice: misconceptions, potential applications, and field experience.

Wheeler, S., and K. Pany. 1990. Assessing the performance of analytical procedures: A best case scenario. *Accounting Review*:557-577.

Whittington, R., and K. Pany. 2001. *Principles of auditing and other assurance services*: Citeseer.

Wilks, T. J. 2002. Predecisional distortion of evidence as a consequence of real-time audit review. *The Accounting Review* 77 (1):51-71.

Williamson, S. 2003. Report on wind chill temperature and extreme heat indices: evaluation and improvement projects. *Washington: OFCMS Supporting Research*.

Wright, A., and R. H. Ashton. 1989. Identifying audit adjustments with attention-directing procedures. *Accounting Review*:710-728.

Yoon, K., L. A. Hoogduin, and L. Zhang. 2015. Big Data as Complementary Audit Evidence. *Accounting Horizons*.

Zhan Shu, S. 2000. Auditor resignations: Clientele effects and legal liability. *Journal of Accounting and Economics* 29 (2):173-205.

APPENDIX A For Chapter 2

Table A.1 Variable Definitions

Variable	Description
LogAuditFee	natural log of audit fee (Audit Analytics)
Optimism10k	the tone of optimism calculated by the different between the number of positive words and the number of negative words divided by total number of positive and negative words in a firm's 10-K from the last fiscal year the replaced auditor was engaged
Optimism8k	the tone of optimism calculated by the different between the number of positive words and the number of negative words divided by total number of positive and negative words in 8-K coverage from the year prior to dismissed date to a day prior to the date of dismissal
ROAearnings	earnings, calculated as operating income after depreciation (OIADP) divided by total asset (AT)
Size	natural log of total assets (AT)
Invrec	inventory (INVT) plus accounts receivable (RECT) divided by total assets (AT)
NumSeg	the number of business segments
Foreign	1 if the firm has foreign operation, 0 otherwise
Merge	1 if the sum of special item responding to acquisition and merger (AQP), 0 otherwise;
Special	1 if the firm reported special items (SPI), 0 otherwise
Leverage	sum of long-term debt (DLTT) and debt in current liabilities (DLC) divided by total assets (AT)
Loss	indicator variable 1 if the firm's net income (NI) <0; 0 otherwise
BTM	book value of common equity (*CEQ*) divided by market value of common equity (PRCC_F × CSHO)
Growth	the percentage of change in sales (SALE) from period t-1 to period t, where period t is the last fiscal year the changed auditor was engaged
Big4	1 if a successor auditor is one of Big4, 0 otherwise
Resignation	1 if a predecessor auditor initiated auditor resignation, 0 otherwise
GC	1 if a successor auditor issues a going-concern opinion, 0 otherwise
IW	1 if a successor auditor indicates internal control weakness, 0 otherwise

APPENDIX B For Chapter 3

B.1 How to calculate weather indicators

1. Heat index [20]

Heat index is developed as follows:

$$\mathrm{HI} = -42.379 + 2.04901523 \times \mathrm{T} + 10.14333127 \times \mathrm{RH} - 0.22475541 \times \mathrm{T} \times \mathrm{RH} - 0.00683783 \times \mathrm{T} \times \mathrm{T} - 0.05481717 \times \mathrm{RH} \times \mathrm{RH} + 0.00122874 \times \mathrm{T} \times \mathrm{T} \times \mathrm{RH} + 0.00085282 \times \mathrm{T} \times \mathrm{RH} \times \mathrm{RH} - 0.00000199 \times \mathrm{T} \times \mathrm{T} \times \mathrm{RH} \times \mathrm{RH}$$

Where:

T = Air temperature, °F
RH = Relative humidity in percent

If the RH is less than 13% and the temperature is between 80 and 112 degrees F, then the heat index is calculated as follows:

$$\text{Adjusted HI} = \left(\frac{13 - \mathrm{RH}}{4}\right)\sqrt{\frac{17 - |T - 95|}{17}}$$

In case the RH is greater than 85% and the temperature is between 80 and 87 degrees F, then the heat index is adjusted as follows:

$$\text{Adjusted HI} = \left(\frac{\mathrm{RH} - 85}{10}\right) \times \frac{(87 - T)}{5}$$

If the temperature is below 90 degree F, then heat index is adjusged as follows:

$$\text{Adjusted HI} = 0.5 \times (\mathrm{T} + 61) + (\mathrm{T} - 68) \times 1.2 + (\mathrm{RH} \times 0.094)$$

2. Wind chill index

Wind Chill Index (WCI) is calculated as follows:

$$WCI = 35.74 + 0.6125 \times T - 35.75 \times V^{0.16} + 0.4275 \times V^{0.16}$$

Where:

V = Wind velocity, mph

T = Air temperature, °F

[20] http://www.wpc.ncep.noaa.gov/html/heatindex_equation.shtml

Figure B.1 Wind Chill Index

Temperature (ºF)

Wind (mph) \ Calm	40	35	30	25	20	15	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45
5	36	31	25	19	13	7	1	-5	-11	-16	-22	-28	-34	-40	-46	-52	-57	-63
10	34	27	21	15	9	3	-4	-10	-16	-22	-28	-35	-41	-47	-53	-59	-66	-72
15	32	25	19	13	6	0	-7	-13	-19	-26	-32	-39	-45	-51	-58	-64	-71	-77
20	30	24	17	11	4	-2	-9	-15	-22	-29	-35	-42	-48	-55	-61	-68	-74	-81
25	29	23	16	9	3	-4	-11	-17	-24	-31	-37	-44	-51	-58	-64	-71	-78	-84
30	28	22	15	8	1	-5	-12	-19	-26	-33	-39	-46	-53	-60	-67	-73	-80	-87
35	28	21	14	7	0	-7	-14	-21	-27	-34	-41	-48	-55	-62	-69	-76	-82	-89
40	27	20	13	6	-1	-8	-15	-22	-29	-36	-43	-50	-57	-64	-71	-78	-84	-91
45	26	19	12	5	-2	-9	-16	-23	-30	-37	-44	-51	-58	-65	-72	-79	-86	-93
50	26	19	12	4	-3	-10	-17	-24	-31	-38	-45	-52	-60	-67	-74	-81	-88	-95
55	25	18	11	4	-3	-11	-18	-25	-32	-39	-46	-54	-61	-68	-75	-82	-89	-97
60	25	17	10	3	-4	-11	-19	-26	-33	-40	-48	-55	-62	-69	-76	-84	-91	-98

Frostbite Times: 30 minutes; 10 minutes; 5 minutes [21]

Figure B.2 Heat Index

NWS Heat Index

Temperature (ºF)

Relative Humidity (%)	80	82	84	86	88	90	92	94	96	98	100	102	104	106	108	110
40	80	81	83	85	88	91	94	97	101	105	109	114	119	124	130	136
45	80	82	84	87	89	93	96	100	104	109	114	119	124	130	137	
50	81	83	85	88	91	95	99	103	108	113	118	124	131	137		
55	81	84	86	89	93	97	101	106	112	117	124	130	137			
60	82	84	88	91	95	100	105	110	116	123	129	137				
65	82	85	89	93	98	103	108	114	121	128	136					
70	83	86	90	95	100	105	112	119	126	134						
75	84	88	92	97	103	109	116	124	132							
80	84	89	94	100	106	113	121	129								
85	85	90	96	102	110	117	126	135								
90	86	91	98	105	113	122	131									
95	86	93	100	108	117	127										
100	87	95	103	112	121	132										

Likelihood of Heat Disorders with Prolonged Exposure or Strenuous Activity

Caution; Extreme Caution; Danger; Extreme Danger [22]

[21] Adapted from National Weather Service available at http://www.nws.noaa.gov/om/winter/windchill.shtml

[22] Adapted from National Weather Service available at http://www.nws.noaa.gov/om/heat/heat_index

B.2 How to select peer stores

Three factors are formulated by using selected macroeconomic indicators of Principle Component with Verimax rotation is conducted. The difference between each factor in a targeted store, compared to that of other stores, is calculated. The smaller difference means the stores are more similar. The ranks are assigned to the difference in ascending order. It is important to have a peer store sharing similar factors for Factor 1, Factor 2, and Factor 3. Since the Eigen value of Factor 1 is the highest among all three factors, Factor 1 is the most important, and Factor 2 and Factor 3 are then the next important. Therefore, I drop the stores not sharing a similar Factor 1 first, and then do the same thing for Factor 2 and Factor 3.

It is possible for a store having a close Factor 1, but not to have a close Factor 2 or Factor 3. Accordingly, depending on the decision of selected ranks, the number of peers would be different. For example, if lower ranks (i.e. rank $\leq$ 100) are selected, then a larger number of peers can be chosen. If t high ranks (i.e. rank $\leq$ 10) are selected, then some stores will not have any peer stores sharing close all three factors. However if low ranks are picked, then peers even not sharing macroeconomic characteristics can be selected, thereby contaminating the predictive values of peer stores

Let's assume that I am searching peers for the store A sharing similar Factor 1, Factor 2 and Factor 3 with the store A's factors among 1,000 candidates. If a store has Rank 1, that store has the closest factor of the store A. Similarly, if a store has Rank 1,000, then that store has the least closest factor of the store A among the all peer candidates. Suppose further that there are only three candidates: B store, C store, and D store, as shown as below table. If I decide the acceptable rank is three, then only D store is selected. In a case where the acceptable rank is 50, then B store can be selected as well. In a situation where the acceptable rank is only one, then there will be no peer store among the three.

Table B.1 Example of peer selection

	B store	C store	D store
Factor 1	4rd	100th	3st
Factor 2	50th	5th	2nd
Factor 3	10th	12th	1st

Table B.2 Relationshjp between the number of peers and stores losing

Rank Vs. missed stores		# of peers per store			
Rank	# of missed store	Mean	Std.	Min	Max
N/7	39	7.389	4.091	0	31
N/8	120	5.254	3.178	0	25
N/9	283	3.934	2.524	0	21
N/10	510	3.129	2.066	0	18

Table B.2 shows the relationship between the number of peer stores and stores losing by this peer selection procedure. N means the total number of operating units. As Hoitash et al.(2006) does, the rank is selected by the total number of units divided by certain integers. On the other hand, the number of peers per store is larger as well. As the table shows above, as selected rank is larger, the number stores not having peers missed decreased. Because loosing 39 stores is acceptable (2-3 percent of the total number of given stores), the rank is decided to N/7.

Table B.3 Results of predictive accuracy- by using daily state/store level data

Model #	Included variables	# of obs.	MAPE	Std.	Min APE	Max APE
Panel A: Stepwise Model– Daily State Level Data						
(1)	HDD & CDD	364	0.099	0.107	0.000	0.818
(2)	AT	364	0.099	0.107	0.000	0.812
(3)	SAT	364	0.099	0.107	0.000	0.811
(5)	Peer	364	0.039	0.043	0.000	0.262
(6)	Peer & HDD & CDD	364	0.041	0.043	0.000	0.274
(7)	Peer & AT	364	0.041	0.043	0.000	0.272
(8)	Peer & SAT	364	0.041	0.043	0.000	0.270
(10)	$Sales_{n-1}$	364	0.095	0.138	0.000	1.139
(11)	$Sales_{n-1}$ & HDD & CDD	364	0.095	0.138	0.000	1.107
(12)	$Sales_{n-1}$ & AT	364	0.095	0.138	0.000	1.106
(13)	$Sales_{n-1}$ & SAT	364	0.095	0.138	0.000	1.112
(15)	$Sales_{n-1}$ & Peer	364	0.073	0.112	0.000	0.938
(16)	$Sales_{n-1}$ & Peer & HDD & CDD	364	0.073	0.112	0.000	0.930
(17)	$Sales_{n-1}$ & Peer & AT	364	0.073	0.113	0.000	0.930
(18)	$Sales_{n-1}$ & Peer & SAT	364	0.073	0.113	0.000	0.931
Panel B: Stepwise Model – Daily Store Level Data						
(1)	HDD & CDD	364	0.101	0.108	0.000	0.808
(2)	AT	364	0.101	0.108	0.000	0.810
(3)	SAT	364	0.101	0.108	0.000	0.810
(5)	Peer	364	0.018	0.019	0.000	0.136
(6)	Peer & HDD & CDD	364	0.019	0.020	0.000	0.133
(7)	Peer & AT	364	0.019	0.020	0.000	0.133
(8)	Peer & SAT	364	0.019	0.020	0.000	0.134
(10)	$Sales_{n-1}$	364	0.089	0.105	0.000	0.875
(11)	$Sales_{n-1}$ & HDD & CDD	364	0.089	0.104	0.000	0.878
(12)	$Sales_{n-1}$ & AT	364	0.089	0.105	0.000	0.880
(13)	$Sales_{n-1}$ & SAT	364	0.089	0.105	0.000	0.880
(15)	$Sales_{n-1}$ & Peer	364	0.021	0.024	0.000	0.155
(16)	$Sales_{n-1}$ & Peer & HDD & CDD	364	0.022	0.024	0.000	0.153
(17)	$Sales_{n-1}$ & Peer & AT	364	0.022	0.024	0.000	0.155
(18)	$Sales_{n-1}$ & Peer & SAT	364	0.022	0.024	0.000	0.155

APPENDIX C For Chapter 4

Substantive Analytical Procedure and Audit Sampling

The standards specify not only required tasks related to each step but also the way to combine other audit procedures such as audit sampling. According to sampling standard these two types of independent substantive tests are considered as independent, detection risk equals the joint probability of analytical procedure risks (AP) and the test of detail risk (TD).Therefore,

$$AR = RMM \times AP \times TD$$

Where

AR = Audit Risk (AR)

RMM = Risk of Material Misstatement (RMM) which is Inherent Risk × Control Risk

AP = Analytical Procedure Risk which is the risk that the auditor's analytical procedures fail to detect a material misstatement

TD = Test of Detail risk which the risk that auditor's test of details procedures fail to detect the material misstatement.

In this line, AU350 contains the Table 1 as presented below:

Allowable Risk of Incorrect Acceptance (TD)
for Various Assessments of RMM and AP; for AR = .05

Auditor's subjective assessment of risk of material misstatement.	Auditor's subjective assessment of risk that substantive analytical procedures and other relevant substantive procedures might fail to detect aggregate misstatements equal to tolerable misstatement.			
RMM	*AP*			
	10%	*30%*	*50%*	*100%*
		TD		
10%	*	*	*	50%
30%	*	55%	33%	16%
50%	*	33%	20%	10%
100%	50%	16%	10%	5%

* The allowable level of AR of 5 percent exceeds the product of RMM and AP, and thus, the planned test of details may not be necessary unless specified by regulation or other Standards (e.g., confirmation or inventory observation procedures).

Note: The table entries for TD are computed from the illustrated model: TD equals AR/(RMM x AP). For example, for RMM = .50, AP = .30, TD = .05/(.50 x .30) or .33 (equals 33%).

This presents that after the auditor set Audit Risk (AR), evaluate Inherent Risk (IR) based on the vulnerability of account balance or class of transactions to misstatement, conduct test of controls to determine the level of Control Risk (CR), and evaluate Analytical Procedure (AP), Detection Risk (DR) is calculated.

TD is the a similar concept of the risk of incorrect acceptance (false negative) in terms of sampling risks associated with sampling in substantive tests of details (Louwers et al. 2004). Therefore, auditors establish the level of risk of incorrect acceptance after defining AR, IR, CR, and AP. Based on TD calculated by the statistical model provided by AICPA guidelines, and the number of the full population, I can determine the possible sample size auditors might need to collect in order to reach the auditor objective of audit risk.