ESSAYS TOWARD THE DEVELOPMENT, IMPLEMENTATION, TESTING, AND AUTOMATION OF RISK-BASED FULL POPULATION GENERAL LEDGER AUDITING SYSTEMS

By: Jamie William Freiman

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. Alexander Kogan and approved by

> Newark, NJ October, 2020

Copyright page: © 2020 Jamie W. Freiman ALL RIGHTS RESERVED

ABSTRACT OF THE DISSERTATION

Essays Toward The Development, Implementation, Testing, and Automation of Risk-Based Full Population General Ledger Auditing Systems

By: Jamie W. Freiman

Dissertation Director:

Professor Alexander Kogan

A business' general ledger (GL) contains a complete picture of all of its transactions and business dealings. In a modern era consisting of large multi-national corporations, this consists of millions, if not billions, of transactions within a day. In such an environment, the GL record system provides valuable insight into the day-to-day functionality of a business. In addition to legitimate transactions, the GL is likely to contain evidence of fraudulent or erroneous accounting practices should they exist. The importance of this is not lost on audit regulators. In 2002 the American Institute of Certified Public Accountants (AICPA) released Statement on Auditing Standard (SAS) no. 99, which requires auditors to "…design procedures to test the appropriateness of journal entries recorded in the general ledger and other adjustments." (AICPA 2002).

Traditionally, auditors have relied on sampling techniques to test large populations of data (Hall et al. 2000). However, this may not be the most effective approach to detecting low-frequency high-risk cases of fraud within the population (Neter & Loebbecke 1975). Academia has been quick to try and call for new analytic based methods to reconcile this and similar issues related to big data (Applebaum et al. 2017, Vasarhelyi et al. 2015). The following essays aim to close this research gap within the GL space by providing a risk evaluation framework and full population methodology for examining and raking individual GL updates based on their riskiness.

This collection of essays is designed to provide insight into a new approach for testing large populations of data and extend it into the GL space. Academics have chosen to move away from traditional sampling techniques in cases involving big data, instead advocating for selective suspicion scoring methodologies (No et al. 2018, Issa 2013, Kim 2011). These methodologies rank records based on a suspicion score derived from the application of analytic techniques to the total population. This approach enables an auditor to conduct a test of details examination on the same number of records as they would in a traditional sample. Unlike a traditional sample, however, these records will constitute the riskiest elements within the population. While such methodologies have been proven successful (Kim & Kogan 2012, No et al. 2018, Lee et al. 2019), such an approach has never been extended into the GL space. The GL is particularly important as it includes a comprehensive catalog of virtually all transaction events. To this end, these essays first establish an approach for determining the appropriate tests to apply to each individual GL dataset. This assessment breaks the GL down into risk categories with associated test recommendations. The remaining essays apply this methodology to a variety of audit environments. One application focuses on internal audit application to a large multi-national bank. The other essay focuses on working with external auditors to apply the approach to auditing a multi-national manufacturer.

The first original research essay provides a framework for evaluating the risks present in each individual GL called the General Ledger Adjustment Risk Evaluation

iii

(GLARE) framework. This risk assessment is used to determine the appropriate analytic techniques to use based on the risks being targeted by an auditor. This essay breaks GL risk down into seven key risk categories. Each category is then discussed in detail, and example tests and analytic procedures are discussed and suggested. The purpose of this essay is to provide practitioners and academics with insight and guidance into how to successfully evaluate and test different GL risks. This essay uses a combination of past literature and novel techniques to suggest potential solutions to mitigating different risk patterns present in a company or their GL records. It is crucial to the later implementation of a suspicion-scoring model that an auditor has a complete understanding of how to target specific risks. Without this understanding, it is likely that the methodology may be misapplied, in which case suspicious or risky records may not be detected. In addition to the framework itself, it is applied to a test dataset in order to generate ten potential risk problems, each of which is rooted in a GLARE risk category and audit assertions. Auditors provide feedback on their perceived importance of these risks to ensure that GLARE identifies risks that matter to auditors.

The second original research essay applies GLARE to aid in the construction and application of a full population filtering methodology in an internal audit environment. Manual entry GL data from a large multi-national bank is used to illustrate that this methodology is effective in detecting accounting irregularities and even an instance of fraud. In this case study, the company utilizes several internal ledger systems. This illustrates that these methodologies are robust to a variety of different datasets and structures when it comes to detecting issues. Additionally, each applied test was evaluated based on the results and insights that it provided.

iv

Once it is established that GLARE and full population filtering methodologies can successfully be applied to GL update datasets, the third original research essay applies a suspicion ranking methodology to examine the data of a large multi-national manufacturer from an external audit point of view. In this case, GLARE is utilized to build the filters. External audit partners were also consulted for the application of a formal methodology designed to rank suspicious adjustments to the GL for a final test of details sample. The resulting sample was designed to be comparable in size to a traditional audit but reflect the riskiest elements of the GL entries.

Additionally, a shorter forward-looking essay is included. This essay is designed to position this research in such a way that it is accessible to future academics wishing to apply it in a more automated continuous context. This essay discusses adaptations that must be made to make the utilization of the methodologies discussed elsewhere in this dissertation so that they fit appropriately within the continuous audit paradigm. In addition, an outline for a suspicion scoring dashboard to be used by auditors is developed. This is also illustrated in mockup renderings of how this would look from an auditor's perspective.

Together these essays are designed to move the audit of GL data in a more effective direction. In an era when traditional sampling techniques may not be as effective in populations that number billions of records per year, suspicion scoring may provide a fruitful alternative. The development of guidance, as well as proven success in the application of such approaches in a variety of different GL data sets, as seen in these essays, will hopefully aid in the generation of new standards and practices in the audit industry.

V

ACKNOWLEDGEMENTS

I want to take the time to acknowledge all the family, friends, and mentors who have taken the time to get me through this dissertation. My entire family, especially my mother, who from a young age taught me to believe in myself enough to reach such heights, my father for accidentally giving me a somewhat misguided motivation to aspire to such lofty academic heights, and my brothers who toughened me up for academic life. Thank you to Grandpa, and especially Grandma, the fear of going across your knee motivated me to finish. To my grandparents, Bobba and Zaida, providing me with a respite in South Africa to destress meant more than you will ever know. I also want to thank all of my aunts, uncles, and cousins who always provided me with a cheering section even when it meant being bored to death. To my chosen family of friends, you saw me through both my darkest and happiest days with steadfast resolve. I would not be here without you. I know many of you may have thought I was crazy, or that I couldn't do this but encouraged me anyway. Shout out to my first fire captain Brent Gregory for giving me the advice necessary to survive to see tomorrow and a nickname to motivate me to finish this journey. To my dissertation committee, I offer my most sincere thanks. It was not easy finishing this journey with me, and your help is greatly appreciated. I would like to mention Steve Schwartz for telling me about this profession and, more importantly, say thank you to Johnathan Ross for helping me get started on the path to Ph.D. and apply to Rutgers in the first place. Thank you to Miklos and Alex for mentoring me through my time at Rutgers and providing invaluable advice. A big thank you to Yongbum Kim for helping tremendously with this dissertation's studies and teaching me more than I thought I could

learn about some of these subjects. To Barbara Jensen, thank you for keeping the mood chipper whenever I was around the office and for your tireless efforts in organizing all of us at the CARLab. Thank you also, to the rest of the faculty and students I was around during my time at Rutgers.

I would like to give my greatest thanks to my wife, Dr. Christina Freiman. Without your love, support, and help, none of this would have been possible. Your offhanded suggestion of applying to Rutgers seems to have panned out nicely. It has been crazy completing our degrees together, but we did it! I would also like to thank Major Tom and Daphne, who both loyally sat at my side during every one of the numerous all-nighters it took to get to this point. Thank you also to Wyatt for keeping things interesting. Finally, I would like to dedicate this dissertation to the memory of those unable to see me reach this point but whose positive impacts on the world live on through their spirit that motivated me to complete this journey and kept me going at various points. Fiona Freiman, Daniel Weinberger, and Staci Jones. You may be gone, but you are far from forgotten.

TABLE OF CONTENTS

ABSTRACT OF THE DISSERTATION	ii
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	viii
LIST OF TABLES	X
LIST OF FIGURES	xi
Chapter 1: Introduction 1.1 Background and Motivation	1
Chapter 2: General Ledger Adjustment Risk Evaluation (GLARE) Framework
for Selecting Analytic Tests	
2.1 Introduction	9
2.2 Category Generation	
2.3 Value	
High/Low Values (Outlier Detection)	
Round/.99 Values (Digit Distribution Testing)	
Duplicated Values and Entries	
2.4 Frequency.	
2.5 1emporal	
2.0 HR Kelated	
2.7 Control Related	
Improperly applied extant controls	
Standard authorization-based controls	28
2.8 Accounting Estimation Based	30
2.9 Predictive	31
2.10 Example Application	
Evaluating the Dataset and Variables	
Generating and Evaluating Potential Risks	
2.11 Conclusion	
Chapter 3: Applicability of Full Population Examination of	of GL Update Datasets
3.1 Introduction	
3.2 Literature Review	
3.3 Model Development	
Test Development Framework	
3.3 Testing Procedures	
Duplicate Entry Testing	
Date Related Testing	
Reversal Entry Testing	
Basic Statistical Analysis	
Credit and Debit Testing	
Benford's Law and Digit Distributions	
Comment Keyword Mining	54

3.4 Description of Case Example	55
3.5 Results	56
Research Question 1 Results	57
Research Question 2 Results	60
3.7 Conclusion	61
Chapter 4: Application in an External Audit Environment	63
4.1 Introduction	63
4.2 Literature Motivation	66
Low Frequency Error Detection	66
Non-Statistical Method Biases	67
Modern Methodologies	
Utilization of General Ledger Datasets	68
4.3 Methodology	69
4.4 Overview of Instance Implementation	74
4.5 Data	75
The General Ledger	76
Alternative Data Sources	78
4.6 Filter Generation	79
4.7 Original Filters	83
4.8 Auditor Suggested Filters	
Employee Filters	85
Business Unit Filters	
Additional Filters	
4.9 Suspicion Scoring	
4.10 Prioritization and Sampling	
4.11 Evaluation of Results	
Results by Filter	
Results by Perceived Importance	
Results by Score and Materiality	
4.12 Impact	
4.13 Conclusion	100
Chapter 5: Future Directions: Continuous Auditing of General Ledge	r
Adjustments	
5.1 Motivation and Overview	
Necessity for Considering Continuous Auditing/Monitoring	
5.2 Potential Adaptation for a Continuous Audit Paradigm	
5.3 Auditor Dashboard	105
5.4 Concluding Thoughts	109
Chapter 6: Conclusion	
6.1 Summary	
6.2 Limitations, Considerations, and Future Research	114
6.3 Conclusion	115
REFRENCES	
APPENDIX	

LIST OF TABLES

Table 1: GLARE Categories	
Table 2: List of Variable in GLARE Application	
Table 3: GLARE Application Results	
Table 4: Internal Audit Test Categories	
Table 5: Reversal Adjustment Filtration Results	51
Table 6: Internal Audit Data Sources	
Table 7: Variable Descriptions	
Table 8: Omitted Variables	77
Table 9: Questionnaire Results	
Table 10: Tests and Target Objectives	
Table 11: Suspicion Score Results	
Table 12: Filter Results	
Table 13: Auditor Evaluation of Risks	
Table 14 Appendix: Audit Partner Suggested Risks	
Table 15 Appendix: Various MADS Sample Materiality Thresholds	

LIST OF FIGURES

Figure 1: Path of a Business Transaction	4
Figure 2: GLARE Framework	12
Figure 3: GL Testing Refinement Process	45
Figure 5: MADS Model Building Process	72
Figure 6: MADS Visual Example	74
Figure 7: Implemented Process Diagram	75
Figure 8: Filter Generation Process	79
Figure 9: Dashboard Outline	
Figure 10: Dashboard Mockup with Select Filters	108
Figure 11: Illustration of Summary Option for Each Test	
Figure 12: Weight Adjusting Feature	
Figure 13: Possible Charting Options	

Chapter 1: Introduction

This dissertation consists of three essays centered on providing a new approach to general ledger (GL) auditing. This approach involves applying risk targeting filters to each and every update made to the overall GL. The first chapter provides an introduction and describes the motivation and need for such an approach, along with a discussion of similar approaches applied in other areas of accounting. The second chapter examines the risks present in data updates to the GL. In doing so, it explores academic and professional approaches to targeting such risks in the full population of GL updates. The third chapter applies some of these methodologies utilizing the risk-based approach from chapter 2. This is applied to data from an internal audit perspective with the goal of answering: 1.) Can a risk targeting, full population filtering approach discover problematic or erroneous updates to a GL? And 2.) Is this approach adaptable to a variety of different systems? These questions were answered using internal data provided by a large multinational bank. The fourth chapter applies a comprehensive suspicion-scoring methodology that utilizes the full population application of risk-targeting filters to provide a sample for external auditors to examine further. The data for this study pertains to GL updates within a large multinational manufacturer. The fifth chapter provides a discussion on how these methodologies fit within the continuous audit paradigm of the future. In an attempt to add greater automation to the process, a dashboard system is outlined, and a mockup version presented. The sixth chapter provides a conclusion that discusses future research directions, summarizes the material covered in the earlier chapters, and discusses any limitations in the studies.

1.1 Background and Motivation

This dissertation offers relief on two critical tensions that must be resolved within the audit profession. The first of these pertains to how a company is audited. What may be some of the most effective data for auditors to examine within a company? It could be argued that the link between Journal Entries (JEs) and the GL may provide one of the most fruitful grounds for examination. However, when examined individually, the frequency and volume of such updates can generate populations of unwieldy size. The second key tension pertains to the issues surrounding the examination of large populations of data. This dissertation examines an alternative approach to traditional statistical and non-statistical sampling techniques to examine the full population of updates to the GL.

The journey of data within a company from its generation at the granular transactional level to the hyper-aggregated financial statement level provides internal and external auditors with a potpourri of data structures and formats to examine. On one end of the spectrum, there are the aggregated financial statements for which auditors are required to provide an attestation. While these financial statements can no doubt provide a wealth of information, at this level, data is aggregated, which presents some problems to auditors. This high-level data can be very useful for risk assessment or forming an initial or broad understanding of a company and potential issues but could cause problems if relied upon too heavily by auditors (Johnson et al. 1991). In such cases, auditors may miss both intentional and unintentional, material errors. On the other end of the spectrum is transactional level data. This data contains the most information about

business events; however, the volume of information and the number of data points can be problematic for time-crunched auditors.

To balance out these issues (aggregation vs. data volume), auditors have developed approaches that target the middle two processes in the spectrum. As a business event occurs, it is recorded at the transactional level. These transactions are represented as journal entries (JEs), which credit and debit each individual account pertaining to a given event. As JEs are made in the system, an update is generated in the general ledger (GL). The general ledger is a list of all accounts and their running balances. As a JE is produced, the changes to the accounts recorded in a JE are reflected in the updated balance to each relevant account within the GL. From the GL, the account balances are aggregated by financial statement preparers to generate income statements and balance sheets. A diagramed example of this information flow can be found below.



Figure 1: Path of a Business Transaction

As information flows through this process and is aggregated, a balance must be struck. On the one hand, aggregated data reduces information and therefore, "population" size of what can be examined. On the other hand, reducing the size of what is considered the testable data population reduces workload in terms of data points for examination. While some may argue that sampling may overcome this problem, how effective this may be is up for debate, as discussed in the following section. To regulate this issue, this dissertation focuses on examining a slice of business data that is little discussed in academic literature, the updates to the GL that occur with each JE.

Historically as JEs were completed and updates were made to the GL, there may have been no, or little, evidence of these changes. As the information age has dawned and companies are increasingly collecting data on every internal step, a dataset has emerged by which each individual change to the GL is available for auditors to examine. This practice is evidenced in the two case studies applied in this dissertation. On the one hand, not only is there evidence that companies are collecting this data internally (large multinational financial institution), but external auditors are also examining this data (final study in this dissertation). This is echoed in academic literature that calls for further inquest into such datasets (Gray and Debreceny 2014, Debreceny et al. 2005). Examining this link strikes the perfect balance in terms of data volume and aggregation. The data at this stage is fairly disaggregated. As a result, there is a variety of test procedures that can be applied based on risk profiling of the company and industry. This is explored in Chapter 2. At the same time, however, the information is not quite as granular as with JEs, which typically include descriptions involving products or individual customers pertaining to each individual transaction. This limits the volume of data that is being examined. While there is no argument that this data is not useful, it may be used later in an audit process during a test of details examination when a further investigation into exceptions is necessary.

Despite the selection of GL update data designed to limit population size, the volume of business transactions dictates that this will still be a large dataset. This means that some efficient methodology must be developed to examine this data, which is the second major focus of this dissertation. Due to constrained resources and time, external auditors typically use sampling methods to test large populations of data. When

sampling, auditors examine a small portion of data to draw conclusions on the entire population. Traditionally there has always been a debate between the merits of statistical and non-statistical sampling methodologies. Statistical sampling techniques utilize statistics and mathematics only to determine a sample population. Non-statistical sampling utilizes an element of auditor judgment when determining the sample from a population. Throughout much of the 1900s, statistical sampling methods reigned supreme. However, in the late '80s and throughout the '90s, there was a marked shift toward non-statistical sampling techniques (Hall et al. 2000). The increased reliance on auditor judgment involved in non-statistical sampling has been shown to result in lessthan-ideal sampling (Elder and Allen 2003, Hall et al. 2000).

Statistical sampling has one major shortcoming from an audit standpoint. Even in a comparatively small population (10,000 accounts), Neter and Loebbecke (1975) find that a variety of statistical techniques fail to detect high-risk, low-frequency events. As discussed in detail later on, these are often the issues of greatest concern to external auditors. While the logical solution may be to increase sample size, this has a limited impact (Hall et al. 2001) while resulting in a marked increase in auditor workload, making it an impractical solution at best.

Historically, the only logical alternatives to this methodology were non-statistical techniques that rely on auditor judgments. The goal with such methodologies is to utilize auditor expertise to draw samples designed to specifically target such high-risk issues. Such approaches are, however, not without limitations as well. The most widely examined issue with these methodologies revolves around auditor bias. The belief being that auditors utilizing such techniques may 1.) disregard potentially risky data because of

ignorance (unintentional or otherwise) or 2.) review the resulting sample through a lens bias toward examining the issues that they initially set out to seek, disregarding potentially serious issues that were not part of the original risk profile (Elder and Allen 2003, Hall et al. 2000, Elder and Allen 1998, Burgstahler and Jaimbalvo 1986, Blocher and Bylinski 1985).

To mitigate the disadvantages of these two types of sampling (statistical and nonstatistical), newer methodologies have been developed that seek to examine the full population of data points in a dataset to evaluate and further investigate high-risk exceptions. The most common modern solutions involve weighting and suspicion scoring (Issa 2013), and multi-tier filtering (Kim and Kogan 2014, Kim 2011). The applicability and feasibility of such approaches to GL update datasets are explored in chapter 3 of this dissertation. Upon determining that such methodologies show promise, chapter 4 applies one such methodology known as Multidimensional Audit Data Sampling (MADS) (No et al. 2018) to GL updates from an external audit perspective.

MADS as a methodology will be covered in more detail in later chapters. Broadly, however, it is a full population risk-based sampling methodology that is designed to produce a sample of the riskiest records for auditors to examine. It has been applied to a variety of different contexts within accounting but not yet in any GL related context (No & Huang 2019, Yoon et al. 2019, Lee et al. 2019). One key benefit of MADS is that it allows you to enumerate risks to develop filters that are applied to the full population. It, therefore, enables auditors to determine exactly what portion of the population is risk-free based on those filters that were applied. This is something that can only be extrapolated and estimated using statistical sampling techniques. Additionally, the application of a variety of filters for various risks enables some element of auditor judgment while limiting the impacts of auditor bias in the review of results.

The application of MADS in an external audit context is a culmination of the research developed in this dissertation. While the next chapter frames the potential risks associated with GL updates and suggests filters for such risks that are rooted in practice or academia, chapter 3 examines the applicability of modern full population risk-based testing to GL update datasets. Once this is established as a potentially viable avenue for exploration, chapter 4 represents the application of one such methodology (MADS) to a dataset as would occur in a live external audit. The final step in this process is to link this to the future of continuous audit.

Continuous auditing was first proposed by Vasarhelyi and Halper (1991). The concept of continuously monitoring audit data has been discussed in literature extensively and is the direction into which the future of audit seems to be headed (Applebaum et al. 2017, Kim 2011, King and Magnusson 2011, Curtis et al. 2009, Rezaee et al. 2002, Woodroof and Searcy 2001). To this end, the methodologies within this dissertation are placed within the forward-looking scope of application within a continuous monitoring environment. While the implementation-specific on-line version of the application is beyond the scope of this dissertation, there is both discussion and a mock-up of what a continuous monitoring solution or dashboard would look like within the context of utilizing full population risk-based filtering on GL updates.

Chapter 2: General Ledger Adjustment Risk Evaluation (GLARE) Framework for Selecting Analytic Tests

2.1 Introduction

Journal entries (JEs) and their impact on the General Ledger (GL) are an important and mandatory consideration for any external auditor. Statement of Auditing Standard (SAS) 99 requires that an auditor should "...design procedures to test the appropriateness of journal entries recorded in the general ledger and other adjustments." (AICPA 2002). These tests are conducted with risk assessment in mind as SAS no. 99 specifically requires the consideration of risks when conducting an audit. This is a crucial beginning step in the testing of updates to a GL. As a result, this essay is designed to present a formal framework for the systematic classification of General Ledger (GL) audit concerns based on risk with suggested procedures for targeting such risk areas.

While some academics have advocated different test procedures for GL and journal entries (Lanza & Gilbert 2007, Loraas & Searcy 2010, Fay & Negangard 2017), none have classified such tests based on risk. The AICPA Center for Audit Quality directly suggests that JE testing be conducted by first assessing risk areas (Practice Alert 2003-02), indicating the correct action for risk-based testing procedures. For this reason, it is believed that a risk classification approach to testing the updates to a GL is particularly necessary. Such an approach enables auditors to assess risk first and then maintain that link into selecting analytic procedures that are used in the following audit stages.

The GLARE system for GL audit issue classification is designed with two main goals in mind. Firstly, it provides a systematic overview of relevant risk areas. This outline will enable auditors to do an assessment of various aspects of risks associated with auditing updates to a GL without overlooking any major areas of issue. With this in mind, GLARE will serve as guidance to practitioners looking for greater formalization in the GL test selection process. Secondly, the world of auditing, as we know it, is moving toward a more autonomous environment. To facilitate this, the GLARE categorization of risks and associated tests can serve as a jumping-off point for those that wish to automate the GL audit. This may be of particular interest as automatable audit tasks often involve applying simple rule-based tests to a large volume of transactions such as those employed by generalized audit software (GAS). By providing this formalized classification framework, we hope to inspire insight into which of these GL audit processes fall into this category.

This framework was compiled based on experience and analysis of major historical audit failures. To further test the applicability of the designed framework, it is applied in the later sections of this dissertation successfully. Both companies to which this framework was successfully applied come from vastly differing industries (manufacturing and financial), providing insight into the diversity of application for the framework. Finally, we conducted discussions with audit practitioners to receive feedback and suggestions on the GLARE model. This is primarily reflected within feedback on the application of the framework for identifying to audit concerns and risks in the aforementioned companies. Their generally positive feedback illustrates that the GLARE framework should be adaptable to various types and sizes of companies. Not every test in this framework may be possible based on company-provided data. However, it is designed to be an extensive classification scheme that will provide auditors insight into potential risk areas.

The GLARE framework for GL update testing is divided into seven major classification areas: Value, Frequency, Timing, HR Related, Control Related, Accounting Estimation Based, and Predictive. Each of these areas is designed to, in totality, cover all the audit assertion objectives (completeness, cut-off, existence, rights and obligations, and valuation). Some of these categories incorporate traditional testing methodologies that may be familiar to even entry-level auditors. Others incorporate advanced analytic techniques in order to aid the auditor with error or fraud detection. Any violation within a classification branch may not necessarily indicate fraud or error. However, it would indicate anomalous behavior that should be further investigated by auditors.

The remainder of this essay is framed as follows. Figure 1 illustrates the GLARE framework visually. The next section discusses the development and selection of the 7 GLARE risk categories. The following seven sections each detail one aspect of this framework and issue classification system. Next, an example is generated by applying GLARE to data from a multi-national manufacturer. The results of this are evaluated by senior audit partners. Finally, a discussion and conclusion is provided, which includes any limitations.



2.2 Category Generation

To generate the variety of categories covered in the GLARE framework, two important factors were taken into account: data possibly available in GL updates and audit objective assertions. To develop these key areas, the process began by determining a list of possible risks that could be assessed within data pertaining to GL updates. Upon determining what risks may possibly be evaluated using such datasets, these risks were grouped, and each group grounded in at least one of the six key audit assertions (accuracy, completeness, cut-off, valuation, existence, rights and obligations). Conversely, it was ensured that in their totality, the broad risk categories that make up the GLARE framework, cover all six of these objectives.

To begin the process, the focus was on what type of data may be recorded in any companies updates to the GL. Keeping in mind that updating the GL consolidates some information found in the JE postings dropping some recorded information (event or detail description) and reflecting adjustments to individual accounts, resulted in the breakdown of data into five major data variables: adjustment value, account information, timing information, entrant data, and other. Each of these represents a category of nonoverlapping data that is typically recorded as JEs are reflected in the GL as an update to each individual account.

The first two pieces of data that are typically included are the most necessary to any GL update dataset. Namely, the account being adjusted and the value of the adjustment. As a result of these two categories being essential to all GL update datasets, a bulk of the risk assessment, and eventual GLARE categories, are centered around these types of data points. In addition to these two categories, there are typically two other variables that are collected. The first of these is some temporal data. This usually consists of a date or time as to when the adjustment to the GL was made. Depending on what type of temporal data is included in the dataset, some available tests may change. The second of these, entrant data, pertains to information or variables regarding who entered the adjustment to the GL as well as potentially who may have authorized this. This category of data can be utilized in a variety of different ways to examine numerous risks. Often times, other data such as a description, notes, or even business unit information may be recorded. For this reason, an "other" category of available data was included.

Once the five potential categories of data variables were enumerated, they were grouped into the seven GLARE categories in the framework. This was done by overlapping the available data categories and the six known audit objectives. Each of these audit objectives is designed to address a variety of risk areas. For example, valuation is designed to target the risk that a company has mis-valued a transaction. Each data variable category was linked to potential audit objectives and concerns. These links are then reflected in the seven GLARE risk categories. As an example, take adjustment value as a variable. This may address valuation or accuracy concerns upon examination. As a result, one GLARE category is designed to target this exact concern, value. However, this data category (adjustment value) may provide additional benefit by examining the same variable from a different perspective, namely how frequently the value appears. If there is a common frequency for example, this value occurs once a week, then a deviation from this pattern may present a concern. As a result, an additional audit objective could be examined; existence. Under this risk scenario, the value matters less than the frequency, hence the separate GLARE category of frequency. An outline of

the GLARE categories, the potentially useful data variables for each risk category, and the associated audit objectives are included in the table below. The following sections will break down each of these seven categories and discuss potential tests that may exist to target each of the risks and concerns for each category based on its associated data variables and audit objectives.

GLARE Category	Potential Data Variable(s)	Audit Objective(s)
Value	Adjustment Value	Valuation
	_	Accuracy
Frequency	Adjustment Value	Existence
	Entrant Data	Accuracy
	Timing Information	Valuation
	Account Information	
HR Related	Entrant Data	Existence
	Adjustment Value	Accuracy
	Account Information	
Temporal	Timing Information	Cut-off
		Completeness
Accounting Estimations	Account Information	Valuation
	Other	Rights and Obligations
		Accuracy
Controls	Adjustment Value	Completeness
	Entrant Data	Existence
	Timing Information	Accuracy
	Account Information	
	Other	
Predictive	Adjustment Value	Completeness
	Account Information	Existence
	Entrant Data	Accuracy
	Timing Information	Cut-off
	Other	

Table 1: GLARE Categories

2.3 Value

The value of transactions and the resulting update to the GL is a fundamental consideration in auditing. Auditors examine for outlying transactions and often use this as a jumping-off point for further investigation. This risk category primarily examines for correct valuation and accuracy within the dataset. Within this category five major

classifications of tests have been identified: unusually low value, unusually high value, round/.99 values, frequently duplicated values, values outside the employee's normal behavior pattern. It is important to note in this discussion that this is not to be confused with materiality setting. This would be covered separately in the audit procedures and may be applicable on an account-to-account basis.

High/Low Values (Outlier Detection)

Unusually high or low values are probably the most traditionally thoroughly practiced audit procedure. It is not uncommon for audit partners to scan sheets of entries looking for outlier values. Each of these two areas of testing is designed to detect such outliers. Hodge and Austin (2004) classify outlier detection techniques into three categories based on user knowledge and labeling. In type two and type three scenarios, normal values are labeled as such. The difference is that type two methodologies also require abnormal values to be labeled as well. In type one scenarios, none of the data is labeled, and the user has no prior knowledge. Since this is the case for external auditors, these are the techniques that are most applicable to GL auditing.

Some popular type one methods for performing outlier value detection include statistical modeling such as mean and standard deviation analysis on a univariate level (Rousseeuw & Hubert 2011, Rousseeuw & Leroy 1996, Solomon et al. 1994). An auditor may wish to determine standard deviations in value and examine all entries that fall outside say two standard deviations. Additionally, the use of more complex distancebased measurements and analytics rooted in machine learning algorithms such as clustering, linear regression or classification can also be used (Rousseeuw & Hubert 2011, Lu et al. 2003, Knorr & Ng 1997, Rousseeuw & Leroy 1996). In these cases, a machine-learning algorithm is applied to the data to generate some sort of class that can be described as being normal. From this, different measures can be used to determine what is abnormal or outlier. A comprehensive review of the merits of each of these methodologies can be found in extant literature dedicated to such a review (Nagi et al. 2011, Ben-Gal 2005, Hodge & Austin 2004). It is important to note that extremely high or low-value outlier entries may not necessarily be problematic. They may, however, alert an auditor to other issues that may occur. Several studies have employed such outlier detection techniques effectively.

Round/.99 Values (Digit Distribution Testing)

Round or .99 value items may also be indicative of a problem. It may be the case that accountants are fabricating numbers or do not know the actual value of an entry and are just entering a round estimate for something that should have a formal value. This issue is formally identified in earnings management by Carslaw (1988) and Thomas (1989). Additionally, employees may wish to avoid a threshold and enter a .99 value. While threshold avoidance is discussed later in this essay, here, the detection of this issue is addressed from a value standpoint.

The most commonly used technique used to test for a round number or .99 value issue outside of strictly searching and examining those specific records is based on Benford's Law. Benford's law was originally published in 1881 by Simon Newcomb and formalizes that in certain naturally occurring sets of numbers, the frequency by which the first digit occurring as any given number (1-9) follows the formula P(d) = Log10(1+1/d)where p is probability and d is the number (1-9). This contradicted prior belief that these random occurrences followed an even distribution pattern. Since its inception, this law has been extended to present formulae that cover numbers appearing in the 2nd place, 3rd place, 4th place, and 1st two digits combinatorial. By comparing the Benford's distribution of digits to the actual utilizing one of these formulae, auditors can identify potential issues and trends within a population where the distributions do not conform.

Accounting provides a fertile ground for Benford's Law analyses as research has indicated that as numbers are multiplied, divided, or raised to integer powers, processes that are typical within accounting, they are more likely to conform to Benford's Law (Boyle 1994). While Benford's Law has been shown to apply to many accounting scenarios (Nigrini 2012, Durtshi et al. 2004, Drake and Nigrini 2000, Nigrini and Mittermaier 1997, Nigrini 1996) it does not always apply to accounting scenarios that are as a result of intentional one-sided human thought (ex: ATM withdrawal) (Nigrini and Mittermaier 1997). However, when utilized under the correct conditions such as with revenue or accounts payable accounts, or on large accounting datasets such as the totality of updates to a GL over a year, this may provide valuable audit insights (Durtschi et al. 2004). The belief is that most malfeasant individuals are unaware of Benford's law or cannot keep track of their incorrect entries. By comparing the Benford distribution to the actual distribution of digits that should follow the distribution, auditors can see if there are any unexpected frequent occurrences of digits or digit combinations. Further investigation may then be conducted based on the nature of the account(s) being tested. For a more exhaustive analysis of Benford's law and its application to accounting, refer to Durtschi et al. (2004).

Duplicated Values and Entries

Frequently duplicated values or entries refer to identical entries or highly similar entries made within a short window. This may alert an auditor to several different issues. Firstly, an auditor may discover a batch of faulty or fraudulent entries. This is an important feature as duplicated entries have been directly linked to fraud detection (Ted et al. 1995). Additionally, it could alert an auditor to a failure of internal controls. Employees could generate several entries below a threshold to avoid a control process. Even if such entries are not fraudulent, it could be that in an automated system, an accountant accidentally hit the enter button one too many times, thinking the system was delayed. It could also be that these are all legitimate entries. In this case, such an examination would alert the auditor to a supplier that may provide a valuable confirmation letter explaining these transactions. Either way examining frequent entries over a short period of time could provide the auditor with leads or valuable audit evidence.

There are several ways of detecting duplicated entries or amounts. The AICPA recommends simply ordering entries chronologically and examining for these errors (AICPA 2008). Weiss and Naumann (2005), however, provide a more generalized framework that can be applied to detecting all kinds of duplicate entries. Their three-step process involves candidate description, duplicate definition, and finally, duplicate detection. The first step is selecting the field for duplicate consideration. Ideally, this would be a unique identifier. The second step is to define what is considered a duplicate. This may mean an exact match, or it may just be as similar records. The final step is to detect the entries. Several studies outline methodologies for evaluating what is and is not a match based on this three-step process. These methods are designed to eliminate some

of the auditor judgment. Some advocate for the use of probabilistic methods such as Fellegi and Sunter (1969). Others advocate for classification algorithms such as decision trees (Verykios et al. 2000), or clustering (Hassanzadeh et al. 2009). Each of these methodologies

This final area of examination classification with respect to value is the most modern, values inconsistent with employee behavior. The examination requires technological aid in order to deal with a potentially large number of accounting employees effectively. The goal in these tests is to either detect any employee malfeasance or to see if an employee's credentials have been stolen. By establishing a predictable range of employee entry values, an auditor should be able to detect entries made well outside that range. This type of testing can be duplicated and applied to various accounts as well.

2.4 Frequency

The frequency of journal entries is a factor that breaks down on several levels but may provide significant insight into potential issues. When examining the frequency of entries, it is important for auditors to use the correct techniques and examine within a specific scope. For example, they should examine the average frequency of entries within a specific account or account type. By establishing an expectation for the frequency of entries auditors would be better able to detect situations that may deviate from this practice, for example, channel stuffing. This is examined in the AICPA practice alert 2003-02 (AICPA 2003) which breaks down frequency issues into two categories. The first, standard entries, are frequent, recurring, and often automated. With respect to these entries, it is recommended that auditors establish the pattern with which they occur. Deviations from this pattern may indicate a potential problem. For example, if there is an entry made every two weeks that pertains to a particular client. If shortly before the end of the quarter, there are entries for several shipments within one week it may be an indicator of a problem.

The second category mentioned in the practice alert is those that are largely infrequent or those which should belong to the infrequent class but are occurring at frequent intervals. Such infrequent transactions are hard to detect and follow up on however detecting such cases can prove to be a windfall as with Gene Morse, the internal auditor who uncovered the WorldCom fraud by unraveling one freak transaction made in a typically inactive PP&E account (Lanza and Gilbert 2007). While less common it is also important for auditors to follow up on frequently occurring transactions where once such transactions were infrequent. The AICPA specifically outlines this as a characteristic of fraudulent entries or adjustments as per SAS 99 (AICPA 2003).

In order to examine these types of issues, it is crucial that auditors gain an understanding of computer systems. One such solution for examining such frequencies proposed by Lanza and Gilbert (2007) is the use of excel. Utilizing a system such as excel enables auditors to determine what the average frequency is for particular account activities. By examining the pattern with which certain adjustment amounts are made to an account, or how frequently adjustments are made to an account in the GL, auditors can discern a pattern of behavior. If that frequency or the value within that frequency changes week over week or month over month it would indicate to the auditor that maybe an investigation is necessary. A special note with respect to the impact of duplicate records on frequency analysis must be made. While this risk area does not specifically pertain to duplicate records, the discovery of duplicate records may be made. This illustrates the overlap of some of the risk categories in the GLARE framework. Such overlaps typically reflect areas that may pose a great impact on the financial statement. As a result, while duplicate entries are included within value risk on the GLARE framework they are not to be ignored while focusing on frequency testing. Alternatively put, by analyzing frequency certain duplicate records may be discovered which provides evidence as to the importance of frequency analysis.

2.5 Temporal

Temporal testing primarily revolves around the assertions of cut-off and completion. These are typically standard tests that are performed as part of any regular audit procedure and are addressed in the AICPA Practice Alert 2003-02 (AICPA 2003). Chief among these concerns are those adjustments made outside of business hours and those that are reversed after the end of a quarter. In addition, however, there is some element of testing for control effectiveness with respect to adjustments made outside of business hours.

The AICPA practice alert specifically mentions examining "entries made at unusual times of day, that is, outside regular business hours." (AICPA 2003). This belief is aligned with SAS No. 99 in its desire to enforce typical business control environments. As a result, it is clear that tests of controls relating to the time of activity are of special concern to auditors and audit regulators. In addition to timing with respect to standard controls, attention must be paid to cut-off and completion assertion testing. To this end, temporal risk behaviors on the GLARE framework also incorporate behaviors that include these assertions such as backdating or reversing adjustments across financial quarters. An example would be incurring a cost in Q1 and recording the cost in Q2 with an effective date from Q1. In this example, the expense account will be understated at the end of Q1 and only adjusted after the start of Q2, a violation of cutoff procedures. These behaviors are typically classified as earnings management practices that do not always conform to GAAP. Practice Alert 2003-02 specifically identifies these as problematic behaviors and notes that not only are these problematic on an individually occurring basis but upon aggregation can compound into a large systemic misstatement.

To test for timing risks there are several procedures that can be conducted with respect to the timing information variables that are contained in a large portion of GL update datasets. Typically, as seen in all seven datasets in this dissertation, some information is recorded with respect to the date an update was made, and the effective date on which the update should occur. Auditors would expect these two dates to match. Variations in this may produce interesting audit findings. To test for these factors Computer Assisted Audit Techniques (CAATs) could be used. While specific tests designed to target the individual cutoff concerns for a particular client based on the available data may vary, Yan (2015) specifically outlines the applicability for CAATs cutoff testing with respect to financial periods. Kuruppo (2012) utilizes CAATs in a classroom setting to prove that students can successfully apply them to detect cutoff issues, a fact that is reinforced in a later paper by Kuruppo and Oyelere (2017). In

addition, those individuals surveyed in the Kuruppo and Oyelere study believed that knowledge of such an application would make them more marketable as they were likely to employ such techniques on the job.

2.6 HR Related

HR-related testing involves looking at employee interactions. This can be broken down into two key types of interactions. The first type of interaction that must be considered is between employees and their role within the system. This approach primarily analyses whether employees are performing their duties correctly with respect to their assigned roles. The second type of interaction is with respect to other employees. This requires a more specialized set of examination skills. The goal is to identify various types of collusion that may occur between employees.

With respect to the first type of interaction, there are very simple tests that can be conducted by auditors. These predominantly focus on examining the role assigned to the entry preparer. The auditors can build a profile of expected entry behavior based on various organizational roles. These can be drawn from logical, corporate governance, and internal control procedures. For example, it may be considered unusual to see a factory employee entering marketing expenses. While there may be a logical explanation this would be considered a red flag. This may be more difficult for those "gatekeeper" roles such as upper management. In these cases, we advocate rule mining the data itself by examining the usual patterns of behavior by key employees. By examining the most common accounts that a manager interacts with the auditor can examine red flag instances that deviate from the management's normal pattern of behavior.
The second type of interaction involves more adaptive testing techniques. To examine employee interactions within general ledger data auditors can employ social network mining techniques. This can enable auditors to see how different employees between different departments interact on a regular basis. By examining how employees interact on a regular basis different patterns of behavior can be examined in GL data. One example of how this can detect a problem is if you find two managers from different departments are very close. It may be innocent. However, upon further examination, an auditor may discover that all the reimbursements for manager A are approved by manager B and vice versa. This may constitute a conflict of interest and be a method used to violate an internal control.

2.7 Control Related

SAS No. 99 specifically states that auditors must inquire about and subsequently test "Programs and controls the entity has established to mitigate specific fraud risks the entity has identified, or that otherwise help to prevent, deter, and detect fraud, and how management monitors those programs and controls." In addition, the PCAOB audit standards numbered 5 and 13 both specifically address a need for auditors to test internal controls further emphasizing it in their Staff Audit Practice Alert no. 3 (2008). While there is no doubt a lot of emphasis on testing of controls and analysis of controls-based risk, internal controls can vary wildly from audit to audit. To surmount this issue the GLARE framework focuses on three key areas of controls related risk that are either discussed specifically in standards and best practices or are standard controls expected to be seen in a vast majority of audit cases. These areas are: examination for human

intervention in updates to the GL (control override), improperly applied extant controls, and standard authorization-based controls (Lanza and Gilbert 2007).

Human intervention/control override

PCAOB standards require auditors to test for control effectiveness. In doing so they specifically suggest that auditors examine for management override with respect to JEs and their potential material impact on the GL and eventual financial statements (PCAOB 2008). While such fraud only accounts for 19% of reported fraud cases, the median loss was more than \$700,000 (AICPA 2016). In addition, such cases of fraud due to management override of controls took on average two times longer (24 months) to detect then general employee frauds (AICPA 2016).

While it is no doubt important to detect management override and human intervention with respect to reporting controls it may be challenging to do so. The primary focus for examining for such controls should pertain to data on the individual entering the adjustment to the GL. Suggested testing would be to examine if any of these updates were entered by senior management utilizing a simple search function. Of that population of updates some may be recurrent and explainable (closing out accounts at the end of a period) others however may require further investigation. Determining this requires an element of auditor judgment.

In order to further examine possible management override, the AICPA and PCAOB acknowledge that additional factors may need to be taken into account. For instance, the type of account that is being impacted by the manual override of a manager (AICPA 2016). It is therefore important to look beyond simply the individual entering the update to the GL and incorporate other factors and data that may be available. This for instance may include account information to detect if there is a little-used account that management is using to embezzle funds or commit some other type of fraud.

Additionally, it is suggested that auditors consider different factors such as business unit information that may be available in some GL update datasets. The concern here is over situations similar to how Enron hid financial losses in a variety of subsidiaries utilizing management override. Determining the frequency and best practices within a company can best prepare an auditor to detect such issues. Auditors may utilize month over month or year over year comparisons to determine normal patterns of behavior and weed out exceptions (Coderre 2000). More on this approach will be covered in predictive auditing.

Improperly applied extant controls

Auditors are not only required to make assertions over the control environment of a business but evaluate the effectiveness of extant controls. To this end this category of risks targets auditor concerns over extant controls. More specifically those controls that seem to be standard amongst companies across the board. Examples of such control rules would be: don't duplicate records, complete all the valid fields for a GL update (don't leave out a variable), and the maintenance of duty segregation.

While duplicate records are a topic covered earlier under value analysis it bears repeating in the control section of risk as it impacts both control assessment risk and risk of misstatement undervalue and existence assertions. As mentioned earlier simple CAATs such as excel search or match functions may be utilized to detect such issues by matching records or updates upon all available variables (Ciprian-Costal 2014, Coderre 2000). CAATs may also be utilized for detecting missing data within a GL update which may constitute a control failure (Coderre 2000). Detecting such issues as an auditor would be fairly simple and can be conducted by searching for missing fields or populating a data sheet with all records that are missing a particular required variable. As typically automated systems prevent such adjustments from being completed without all the required data, this may coincide with risks of management override which may be required to enter incomplete adjustments.

An additional category of risk involving internal controls would be segregation of duties (AICPA 2020, PCAOB 2007, AICPA 2006). In order to properly test for this, there are several approaches each of which depends on the data available. On the most fundamental level if there is approval needed for a transaction and GL adjustment a simple matching search can highlight cases in which the preparer and approver are the same individual (violation). In more complicated situations additional client firm data may have to be provided involving who has the authority to conduct what transactions. In these cases, a matrix of approved behaviors can be compared to the actual data using audit software (Lightle and Vallario 2003). Since modern models for segregation of duty controls require such data to be collected in ERP systems, it should not be problematic for auditors to obtain the necessary information (Kobelsky 2014).

Standard authorization-based controls

The final broad area of control-based risk in the GLARE framework is referred to as authorization-based controls. These risks traditionally fall into two main categories in cases of error or malfeasance. These are 1.) avoidance of control (transaction splitting to avoid necessary authorization, and 2.) subversion of the control.

The former of the two categories, avoidance of authorization controls, typically involves transaction splitting. This is a common practice by fraudsters who wish to avoid detection or more innocently by employees looking to shirk work (Kays 2018, ACFE 2012). The concept is that in order to avoid having additional authorization for a large transaction, an employee will split the one large transaction into two smaller ones. Since authorization limits are a commonplace control, this may be a typical concern for auditors. In order to detect such issues practices similar to digit distributions discussed in the valuation section earlier should be employed. By examining digit distributions of updates to the GL it may become glaringly clear that there is systemic avoidance by a population or an individual employee. For example, if an employee needs an authorization for an account update of more than \$50, you may expect to see a bunch of \$49 transactions in an attempt to avoid the threshold (Nigrini 2012).

The second of these two categories, subversion of authorization controls, is more difficult to detect. In these cases, flaws in the system may lead to one individual having an inconsistently applied limit, multiple limits (incorrect duplicates), or for these limits not to be enforced. Running CAATs on adjacent data sources to determine if individuals have duplicate limits or comparing authorization allowances to applied updates in a format similar to that discussed earlier with respect to segregation of duties, will highlight a majority of problematic cases. Dillaway (2010) proposes a comprehensive methodology by which auditors may utilize computer systems, build reliance on client data, and develop an algorithm to determine the appropriate application of authorization controls.

Under this framework, auditors can enter rules and conditions involving what requires authorization and who may give authorization. These rules can be mined from extant systems. The CAAT, armed with the rules and conditions, can then apply then to a total population, in this case, GL updates, and highlight problematic circumnavigations of the established authorization controls.

2.8 Accounting Estimation Based

Accounting estimations have been a hot topic for auditors to examine and have been extensively discussed (Smieliauskas 2012, PCAOB 2010, PCAOB 2010 (2) PCAOB 2008, PCAOB 2007, AICPA 2006, AICPA 2003). Impacts of accounting estimates can range from fraudulent material misstatements (Summers and Sweeny 1998) to earnings management (Albrecht et al. 2017). It is therefore of great importance that auditors give special attention to the valuation, rights and obligations, and accuracy assertions with respect to GL adjustments related to accounting estimates. To examine accounting estimates, auditors have to rely on account information, to determine what GL accounts may have accounting estimates.

Once accounting estimates and their related accounts have been determined it is up to the auditor to determine exactly how to evaluate the estimate. The result of this is typically a test of details. AU section 342 from the PCAOB provides guidance that auditors should evaluate the effectiveness of the procedures used to determine the estimate, and reperform or independently determine an estimate for comparison. While a great deal of time and energy has been put into deciphering the impacts of estimates, little attention outside of formal guidance has been given as to how to determine the validity of an estimate. It is largely understood that determining the appropriateness or fairness of an estimate is up to auditor judgment (PCAOB 2008). The guidance is clear that an independent determination of the estimate is necessary for comparison.

2.9 Predictive

While not strictly a risk area predictive auditing has been touted as the future of audit practice with many application areas ranging from Continuous Assurance (CA) or Monitoring (CM), to applications within the current static audit paradigm (Chan & Vasarhelyi 2018, Best et al. 2004, Koskivaara 2003, Koskivaara 2000). It is important to incorporate predictive audit testing within the GLARE framework because not only is it valuable to the examination of a variety of the aforementioned risk groups, but the evaluation of a client's predictive audit capabilities within their internal audit environment. Such information can be useful for determining a client's "tone at the top" or internal culture with respect to risk evaluation as per the standards.

Essentially predictive audit techniques can be divided into two key areas preventative and non-preventative (Kuenkaikaew & Vasarhelyi 2013). Preventative predictive audit techniques focus on detecting issues in an online manner as they occur. The key philosophy being "why allow a faulty transaction in the first place?". While there is no doubt a plethora of literature dedicated to painting the future of audit in continuous light (Chan & Vasarhelyi 2018, Malaescu & Sutton 2015, Searcy et al. 2003, Debreceny et al. 2003, Rezaee et al. 2002, Woodrif & Searcy 2001, Vasarhelyi & Halper 1991) the current audit paradigm is not quite there. As a result, this section will primarily focus on non-preventative predictive audit techniques which are those that are used to audit past data as with a current typical audit scenario. Kuenkaikaew (2013) outlines a framework for the application of predictive audit analytics to accounting data. This process begins with the evaluation of risks, a stage that should be facilitated by utilizing the GLARE framework. The second stage revolves around selecting data analytic techniques that would target said risks. This is also facilitated by utilizing the GLARE guidelines in this chapter. The final stages of utilizing predictive modeling and evaluating the prediction compared to reality will be discussed in this section in detail.

A variety of predictive modeling approaches exist. One standard approach would be time series modeling. This would involve building a model to compare data year over year, quarter over quarter, etc. Brühl et al. utilize Multiple Linear Regression modeling in this fashion to illustrate how this may be utilized to predict sales for German car manufacturers (Brühl et al. 2009). Another approach is probabilistic modeling which has successfully been utilized to predict transactional behavior (Cadez et al. 2001) and other audit-related data (Gaganis et al. 2007). Several studies have detailed how forecasting company behavior can be utilized in an audit review (Moon et al. 2003, Chiu 1994, Dugan et al. 1994). Dugan et al. (1994) test four traditional time series predictive models on income statement account balances: ARIMA, Census X-11, Holt-Winters Exponential Smoothing, and Random Walk. They find that while the ARIMA and Holt-Winters models preform best for predicting most accounts, a recommendation of Random Walk models is made as it performs well and is more cost-effective.

In a more modern context methodologies such as neural networks have been developed in an attempt to better predict audit behaviors. Koskivaara (2003) provides a comprehensive overview of the application of artificial neural networks (ANNs) to business and audit procedures with respect to forecasting. At the time of publication, this study identifies twenty-one academic articles detailing to the application of ANNs to auditing all of which fall under analytical review. After analyzing the methodologies, it is found that predicting year over year changes provide better results than any of the other time windows (Koskivaara 2003, Koskivaara 2000, Koskivaara et al. 1996, Wu 1994). It is for this reason that it is recommended that auditors utilize year over year comparisons for the purposes of predictive analytics with ANNs.

2.10 Example Application

To illustrate the GLARE framework in action an example case was utilized. In this example, the data is linked through the GLARE framework to construct possible risks for the company profile in question-based on audit objectives. These risks were used to determine a set of problems that auditors may wish to examine based on the dataset in question as would be expected to occur in a regular audit context. To validate the reasonableness of this application senior audit partners from 5 large audit firms were asked to review general information on the dataset and the list of potential issues that were generated in this example case. They were each asked to review their perceived importance of each risk concern on a scale of 0-3 with zero being irrelevant/no concern and 3 being very important. The averaged results of this review are reported in this section.

Evaluating the Dataset and Variables

The dataset used in this example consists of all updates to the GL of one large multi-national manufacturing firm. While the dataset used in this study only pertains to one operating segment of the firm, the operating segment operates as if independent and therefore illustrates the full gamut of possible GL update transactions that an independent company normally would.

The first step in the application of the GLARE framework is to examine the data and available variables. Based on this step an auditor can then determine what risks they are capable of targeting within the GL update population. With respect to the dataset in this example, each of the four main data classes (Account, Value, Timing, Entrant) and several other classes emerge. A breakdown of these variables and how they are classified with respect to the GLARE variable classes are included in the table below.

Variable	Description	Data Type	GLARE Class	
Account	The identifier of each account	Alphanumeric	Account	
Number	being credited or debited.	Alphandmene		
Amount (local currency)	The value of the credit or debit.	Numeric	Adjustment Value	
Description	A brief description of what the line item relates to. This usually includes a numeric code along with a few identifying words.	Alphanumeric	Other	
Effective Date	This is the date on which the journal entry became effective.	Date	Timing	
Entry Date	This is the date that the journal entry was entered into the system.	Date	Timing	
Preparer ID	This field identifies who prepared and entered an entry into the system.	Alphanumeric	Entrant Information	
Source	A two-letter code that classifies the type of entry being made.	Alphabetic	Other	
Document Number	The identifier of each journal entry. All line items from the same entry have the same document number.	Numeric	Other	
Account Type	A type of an account such as asset, liability, equity, revenue, and expense.	Alphabetic	Account	
Account class	Name of the account	Alphabetic	Account	

Table 2: List of Variable in GLARE Application

Generating and Evaluating Potential Risks

Based on the available data several questions were raised about potential risks in the dataset. Primarily they focused around the issues outlined in the GLARE framework and general understanding of how the manufacturer might operate. These concerns ranged from pedestrian and typical (duplicate updates) to high concern areas such as timing issues, to tests of controls. Based on the available data fields, consideration of audit standards, recommended best practices outlined in the previous sections, and the GLARE framework, judgment was made, and ten potential risk categories were developed. These concerns were outlined in the table below.

What Could Go Wrong	Data Variable Class(es)	GLARE Risk Category(s)	Assertion(s)	Avg. Score
Aggressive earnings management	Timing Account	Temporal	Timing Cutoff	2.8
Backdated postings	Timing	Temporal	Timing Cutoff	2.6
Search for duplicate updates	All	Value Controls	Occurrence	2.4
Incorrect Account Usage	Account Entrant Information	Controls	Accuracy Rights and Obligations	2.33
Systematic Department errors	Other Adjustment Value Entrant Information	Frequency	Accuracy Completeness	2.33
Credit debit balancing	Value	Value	Completeness	2
(Repeating payment) extreme values based on account	Account Adjustment Value	Value	Occurrence	2
Abnormal Patterns of Employee Behavior	Entrant Information	HR Based Predictive	Test of Controls	1.75

CLADED' 1

Entry outside of Business Hours	Timing	Temporal	Test of Controls	1.5
Description testing	Other	Predictive HR Based	Test of Controls	1.3

To determine the applicability and effectiveness of utilizing the GLARE framework to generate these concerns and to classify them, feedback was requested from senior audit partners. The audit partners each represented a different firm that makes up five of the largest audit firms. They were given a list of each of the potential risks as well as the assertion(s) that the risk was designed to target. Each of the audit partners was familiar with the dataset and case and asked to rate the relevance and importance of each risk from 0 (unnecessary) to 3 (most necessary). The results were then averaged and are reported in the table above.

The most noticeable feature of the results is that no risk received an average score lower than 1. This indicates that the GLARE framework was successful in being able to highlight issues that auditors in practice care about. The three issues that were ranked lower than average importance. Abnormal patterns of employee behavior received a score of 1.75 because while a majority of partners ranked it of average importance one rated it low. This illustrates that in general there was a consensus that such a test was of reasonable import. With respect to entries outside of business hours even though the senior audit partners did not unanimously regard this as important it is heavily covered in the standards and best practices. This more likely illustrates a misalignment between standards, suggested practices, and auditor perception of import. It is unlikely that even though the audit partners did not view this as a serious risk that they would violate suggested practices and standards and omit such a test. The final concern of description testing was a novel approach included in but not formally a portion of the GLARE framework. This illustrates that while there may be other tests available to auditors outside of GLARE, they are unable to properly value the importance as two out of the five respondents left this risk blank.

In addition to asking for a review on the ten established risks, the audit partners were given the opportunity to suggest potential tests or risks that they would be concerned about. Each of these recommendations was part of the extant GLARE framework. Suggestions included segregation of duties, incomplete records, and reversal adjustments. The difference in the opinioned necessity of these tests which were already part of the framework can be chalked up to auditor judgment. The survey instrument and additional results of the suggestions are included in the appendix.

2.11 Conclusion

This chapter is provided as an outline of the GLARE framework for auditors to apply when evaluating the risks of GL updates. This issue is little addressed in academia but is rooted in standards and best practice publications. The GLARE framework is developed for use in a three-stage process. Firstly, auditors must examine the available dataset. The types of variables that may be available are broken down into five categories: Account related, entrant related, value of adjustment, temporal, and other. Each of these categories allows for different types of risks to be evaluated and assessed. The GLARE framework has seven key areas of risk classification pertaining to GL updates. Each of these areas is linked to particular audit objectives. The bulk of this paper is dedicated to discussing the necessity for examining each of these seven risk areas along with potential tests that auditors may find useful in examining each particular type of risk. In addition to proposing the GLARE framework, an example is put forward of application to data from a large multi-national manufacturing firm. The variables that are made available are evaluated and linked to the various GLARE variable categories. These are in turn linked to audit objectives and risk groupings in the main GLARE framework. Based on these risk groupings and suggestions in this paper ten potential issues of importance are identified. To determine the validity of these issues five senior audit partners each from a different large audit firm is asked to give feedback. The average rated importance of each of the ten tests illustrates that the GLARE methodology identified risks that were either of average importance or more or are statutorily required for testing. The additional freeform feedback did not identify any risks or testing that fell outside of the GLARE framework. This illustrates that even given differences in auditor judgment the GLARE framework is comprehensive.

This study is not without limitations. The first of these is that the effectiveness of GLARE rests on an auditor's ability to effectively read and understand a businesses operating environment. While it removes some auditor judgment and provides guidance, the results of the example application illustrate it does not remove auditor judgment. Secondly, GLARE was only applied to the dataset of GL updates for one particular company in one industry. While there is no reason to believe that GLARE is not applicable to other firms in other industries it is up to future research to decide. Next, GLARE is based on the literature available at the time of writing. While this is extensive, there may be other techniques, approaches, or risks covered in literature not reviewed in this study that may generate additional GLARE categories unforeseen in this model.

Future researchers may wish to look at this or provide additional tests that are not covered in this essay.

Overall this study successfully produces a framework that was able to systematically identify risks that auditors can examine in detail. This framework serves to potentially eliminate auditor judgment. At the very least it serves as a unifying work that outlines the necessities behind targeting certain risks as well as tests that could be used to examine those risks. Practitioners may benefit from having access to a structured riskbased framework of test options. They may utilize this to learn new techniques or to justify their use of techniques. Academics can gain a general understanding of tests and risk and how they are linked in the GL adjustment environment. Additionally, there may be avenues for future publication with respect to tests that have not yet been applied to some of these risks in this environment.

Chapter 3: Applicability of Full Population Examination of GL Update Datasets

3.1 Introduction

There has been a call, in modern research, for academics to study the implementation of analytic techniques in populations of large data (Applebaum et al. 2017). Additionally, it has been documented that new full population testing techniques are being implemented by internal auditors (Freiman & Vasarhelyi 2019). Despite this, there has been limited work done to examine and implement such a system in accounting academia. This paper aims to fill this gap by providing a methodology and case study for the implementation of full population testing to internal General Ledger related data.

While not exogenous data, a population of millions of updates to General Ledger still qualifies as big data (Vasarhelyi et al. 2015). When examining this data internal auditors face a daunting task. This is especially difficult considering fraudulent or incorrect entries may only make up a small percentage of the overall population. In order to combat this problem, this paper proposes a system of full population testing.

First, a methodology for identifying audit goals and applicable tests is developed and rooted in a rule-based approach. This is primarily based on the GLARE framework. The approach for developing these full population tests is designed to produce smaller manageable results. To further combat the issue of excessive numbers of flagged items a suspicion scoring system is developed to identify the most problematic adjustments.

Next, the methodology is applied to one year's worth of manual entry data from a large multi-national financial institution. The data provided by this company came in the form of six separate systems organized into four separate and unique data sets. This

provided us with a novel opportunity to prove that these methodologies and the tests were robust enough to be applied to varying formats of manual General Ledger adjustments.

Once these methodologies and tests were applied to the data internal auditors from the firm were consulted to interpret the results. Most of the tests produced some sort of findings of irregularities that were deemed severe enough to be investigated by the internal audit teams. Some of these findings are reported in the results section of this paper. More often than not we were unable to find out what happened with these investigations. Despite this, we consider our approach a success since it provided enough evidence to launch a follow-up inquiry.

The rest of the paper is organized as followed. The next section discusses prior literature and motivations outlining the research questions the study aims to answer. Following that there is a section on the methodologies for model development. This covers the process for selecting and developing testing procedures as well as our suspicion scoring methodology. After that, we outline the testing procedures that were selected for implementation in our case study. The following section outlines the data that we received. The sixth section covers the results from implementing our tests on the four data sources. The final section concludes the paper with future research opportunities and limitations.

3.2 Literature Review

The issues pertaining to traditional statistical and non-statistical sampling are well documented (Deming 1954, Arkin 1957, Neter & Loebbecke 1975, Hall et al. 2001). Issues with traditional statistical sampling techniques primarily focus on two factors: sample size, and detection of low occurrence high-risk issues. The latter of these two issues is discussed in detail by Neter & Loebbecke (1975). Their study applies a host of sampling techniques to a variety of seeded accounting datasets. They find that none of the traditional techniques is particularly good at detecting low-frequency occurrence issues.

The former of the two issues raises some additional concerns. The logical solution to not finding low-frequency issues may be to expand the population. This approach however is disproven by Hall et al. (2001) in large datasets such as the ones examined in this study.

In order to detect low frequency, high-value anomalies auditors may wish to apply non-statistical sampling techniques to tailor their samples. Hall et al. (2000) find that this approach is applied in 85% of audit sampling tasks. These approaches however are not free of other concerns. By their very nature, non-statistical samples rely on some type of auditor judgment. These preconceived judgments and biases can impact the sample selection and the interpretation of the sample which may impact the methodology's ability to detect issues not foreseen by auditors. These phenomena are well documented in the academic literature (Elder and Allen 2003, Hall et al. 2000, Elder and Allen 1998, Burgstahler and Jaimbalvo 1986, Blocher and Bylinski 1985)

To resolve this tension, academics have turned to newer more novel approaches that bridge the gap between judgmental and statistical sampling. The result of this has been the birth of full population testing and suspicion scoring. These two approaches of full population filtering (Kim and Kogan 2014, Kim 2011), and suspicion scoring (Issa 2013) are combined in this study and applied to a large data source. Full population filtering involves the application of filters to a full population in order to detect problems within the population. Suspicion scoring is a risk-based approach to sampling whereby records are given scores based on their perceived riskiness. By combining filtering and risk scoring it is hoped that we can provide evidence that such modern approaches can successfully be applied to General Ledger data in an internal audit context.

These newer filter-based suspicion score models can be applied to a variety of different accounting paradigms. While these systems are typically developed or applied through the lens of the external audit there is an incentive for internal auditors to also adopt such practices. Several studies link increased internal audit performance to lower external audit fees (Mat Zain et al. 2015, Singh & Newby 2010, Goodwin-Stewart & Kent 2006). As such this study aims to apply a novel risk-filter based examination technique to internal audit data to determine whether or not it is capable of detecting issues in a population of data. One such paradigm that may be of great importance to auditors is the area of manual entries to general ledger systems. These systems serve as a link between journalized business events and general ledger compilations utilized to formulate financial statements and are therefore of interest to both internal and external auditors. From this we develop the first research question:

RQ1: Can a risk targeting, full population filtering approach, discover problematic or erroneous updates to a GL?

It is vital that as academics we are able to answer such a question in order to move the literature in a direction to incorporate these approaches to GL adjustment testing. By answering this question affirmatively academia can progress toward effectively testing such approaches to evaluate the efficiency and applicability in different environments. Ideally, such approaches would lend toward more automated and continuous monitoring paradigms. Should these approaches be unable to target problematic or erroneous updates to a GL then either new methods must be developed, or alternative approaches to selecting which parts of a company to target in an audit must be considered.

Additionally, businesses offer a variety of different challenges to data analysts. Within one large firm, you often have a plethora of systems with different data structures designed to fulfill the same function. Should the methodology be able to detect problematic issues, it should be adaptable to different environments. The methodology in this study must be versatile in order to cater to the wide variety of accounting datasets. As a result, we also postulate a further research question:

RQ2: *Is this approach adaptable to a variety of different systems?*

3.3 Model Development

Test Development Framework

Systematic auditing can be refined to a fundamental system of if-then rules. An auditor generates a binary condition and failure to meet that condition results in further investigation. Errors that amount to systematic or material levels without adequate explanation can amount to adverse audit opinions. It is this fundamental approach that we incorporated into this framework model.

Fundamentally this methodology and approach in this case study were to design a system of rules for which manual journal entries should follow. To mine and develop these rules we followed a cyclical five-stage process outlined in the diagram below. This cyclical process was repeated until we developed each test to a point at which it successfully recovered data associated with the auditor's rule. Because of the nature of this examination, the returned results were emblematic of a failure to meet adequate standards that the designed test was attempting to evaluate.



Figure 3: GL Testing Refinement Process

Step one in the process is to identify a rule that meets an audit objective. These rules were developed using a variety of approaches. Some rules were developed based on academic research. Another source for rule development was asking the auditors themselves what rules they used when examining manual journal entry data. Since internal control effectiveness is a material concern to auditors, we also used corporate policy to generate rules the entries should follow. The chief source for identifying these was the GLARE framework. After identifying the variables, and relevant risks filters were generated from ones associated with those risks in the previous chapter. The final

source we used for determining appropriate rules was data mining. By examining the data, we were able to find patterns of behavior that seemed interesting or abnormal. Based on these patterns of behavior we were able to discover simple rules that should apply based on specific parameters discovered using data mining.

Once we determined a rule was a valid concern for an auditor, we parameterized the rule in order to test it on a given population. These tests were designed to be adaptable in order to be tested on the variety of systems and related variables provided in these data sets. This step generates the machine test that will collect data on examples of failure to meet the rule generated in stage one.

After we generated parameters to test the rules, we ran these algorithms on the data. Once we received results, we worked with internal auditors to evaluate whether or not we were capturing data points that were truly emblematic of a failure to conform to the business rule we were hoping to test. Often times we found on this first attempt that we were receiving a large number of legitimate exceptions. In this case, we had to revise or add, parameters in order to account for such exceptions. Other times we would find that we would be missing instances that should have been caught by the system. In these cases, we sometimes had to delete business rules to widen the net. This process was repeated on a single data source until we developed an accurate test procedure. This was then applied to other data sources.

3.3 Testing Procedures

While we experimented with a variety of different testing procedures courtesy of the GLARE framework, we found that the following procedures answered a wide swath of questions and returned valuable results. The categories of tests that we used in application to all of our data sources were the following: Duplicate entry testing, daterelated testing, reversal entry testing, basic statistical analysis, credit and debit field tests, Benford's law and digit distribution testing, and comment keyword mining. These were generated utilizing the GLARE framework. Each of the various data sources included all five of the GLARE variable categories (Account, entrant, temporal, value, and other). Based on this and an evaluation of audit objectives the framework was consulted for potential tests. These come from each of the seven GLARE risk categories. Each of these test categories is linked to a particular audit objective or objectives as outlined in the table below. The details of these tests are outlined in the following subsections.

Test Category	Audit Objective	GLARE Risk Category
Duplicate Entry	Occurrence	Value/Controls
Date Related Testing	Accuracy	Temporal
Reversal Entries	Cutoff	Temporal/HR Related
Basic Statistical Testing	N/A	Value/Frequency/Estimation
Credit/Debit Testing	Accuracy	Value
Benford's Law	N/A	Value
Comment Keyword Mining	Test of Controls	Controls/Predictive

Table 4: Internal A	Audit Test	Categ	ories
---------------------	------------	-------	-------

Duplicate Entry Testing

This straightforward simple test was this logical start point. For the purpose of internal audit, businesses may wish to implement such a test to examine if employees or managers are trying to inflate numbers. Additionally, there may be some systematic error. In the case of manual entries, there may be some concern that employees accidentally enter the same information twice.

To implement this test, we looked for an identical match in all variable fields. The only field we did not include was a variable that was a chronological numbering of each entry. If this were included it would have been impossible to find any legitimate duplicate match. None of these data sources had a variable for time stamps. Therefore, we acknowledge that the application of this test to a population with such a variable may have to make allowances in that regard.

Date Related Testing

Previous studies have found that common financial statement frauds involve making journal entries that are temporally irregular (Wells 2001). As a result, this date related testing was initially designed to test for entries on holidays or weekends. We found out that this was almost impossible since the institution in question operated 24/7. Additionally, they operate in a wide enough variety of countries so as to prohibit the existence of uniform holidays. The discovery of this issue alone provided valuable insight.

We did however develop other testing procedures related to the dates the entries were made. All of these data sources had a variable that outlined the date of the credit and of the debit. As a result, we examined cases in which the same entry had debit and credit dates that were not the same. This presented an irregularity to auditors and at times a violation of the business's internal controls. In several instances, we discovered that these dates were not even in the same month.

Reversal Entry Testing

A subset of date-related testing we developed a host of procedures related to reversal entry testing. This concern related to reversal entries is that management or employees are manipulating accounts prior to quarter, month, or year-end dates and then reversing those accounts immediately afterward. There are still however legitimate reversal entries that occur. In this respect, the organization had time frames in which these entries should be reversed. We examined these "acceptable" cases with a test of that timeframe rule.

Primarily there was no variable or marker that specifically linked an original entry and its reversal entry in this data sets. As a result, we developed a procedure that identified reversal entries. We built into this system a methodology that highlighted specifically reversal entries that were suspicious. This was done to reduce this workload in identifying, with internal auditors, the cases that were unique cases of reversal entries.

To begin this process, we had to identify matches between reversal entries and original records. This can occur in one of four ways. The first is if one entry is reversed in one matching reversal entry (1:1). Next, one entry may be reversed in several corresponding reversal entries (1:N). Alternatively, several entries may be reversed in one common reversal entry (M:1). Or finally, several entries may be reversed in several non-matching entries (M:N). Because of the lack of data linking entries and their reversals it was almost impossible to examine cases other than 1:1 matches.

In order to determine cases in which this occurred this initial population prior to filtering, we matched records as follows. We selected records in which the credit half of the first entry matched the debited details from the second entry. Additionally, the debit details from the first entry had to match the credit details in the second. The variables that were used to do this matching included the account being credited/debited, and the value of the credit/debit. Additionally, we wanted to ensure a layer of continuity. As a result, we included the value for the enterprise segment in these matching criteria. If it is a true reversal entry this should be consistent in the originating entry and the reversal entry. We

acknowledge that this may have resulted in false matches, but this three-stage filtration process was able to produce true reversal entries that were also suspicious.

Once we had this initial population, we began this three-tier filtration process. The first step in this test was to highlight only cases in which the reversal and originating entry were in two separate months. We concluded that since quarters and fiscal years only change at the end of the month there was no real need to examine reversing entries within the same month.

The second tier of filtering involved four criteria. These were provided based on iterative work with this test along with information from the internal auditors. Firstly, we only examined cases in which the reversal entry occurred in the following month. Additionally, we stipulated that the reversal should be within ten days of the originating entry. These filters were designed specifically to target cases in which managers were attempting to inflate and reverse numbers in the following month. The belief was that they would want to reverse their false entries as soon as possible. Additionally, we removed any records that were reversed in a specific type of client account. We were told that there were frequent adjustments and reversals that occurred in these accounts that were part of normal practice. Finally, we eliminated any instances with a value of less than 1000. We worked with the auditors and after many iterations, we determined this was an acceptable level of cutoff for a large multi-national firm.

The third and final tier of filtering included three conditions designed to highlight the most suspicious cases of reversal entries. Firstly, this study made use of the employee identification variables. We only included cases where one employee was responsible for multiple reversal entries in this remaining population. This was done to identify employees that had made suspicious reversing entries multiple times throughout the year. Secondly, we only included cases in which a single business segment presented more than 10 of the remaining reversal entries. We concluded with auditors that after removing the aforementioned client account this would be considered an abnormally large number of reversal entries for one year. Finally, we highlighted from the remaining cases entries that belonged to a month in which we had originally identified 700 or more reversal entries reversing transactions in that month. We included this final filter because we wanted to examine cases in months during which many records ended up being reversed. We determined along with the auditors that the records in these months would seem suspicious if there is a high volume of reversals that emanate from these originating records.

Below is a table that outlines how this filtration process refined the number of 1:1 reversal matches to a small population of suspicious records for data source 1. We were able to use this methodology and discover an individual employee that was incorrectly or fraudulently using the reversal entry system. Upon following up with the company we found out that this particular employee had been fired for that exact reason. We present this as evidence that this filtration system worked to successfully identify a set of suspicious 1:1 reversal adjustments.

Initially Identified Population	30,177
Post-Tier 1 Filtration	7,955
Post-Tier 2 Filtration	246
Final Population After Tier-3 Filtration	136

Table 5: Reversal Adjustment Filtration Results

Basic Statistical Analysis

After examining the data and speaking with the auditors we decided there was some value to some basic statistical analysis of the data. We attempted a similar, if more simplistic, methodology to that employed by Arning et al. (1996). While we ran a host of basic tests to descriptively analyze the data, we did not find such interesting results. We still included this because of the impact that this statistical analysis may have on identifying audit outliers.

We used this analysis in the application of two types of tests. Firstly, we defined outliers as entries with values falling more than three standard deviations from the mean. These entries received an anomalous value weight in this final scoring. Additionally, we used this examination and discovered entries that had line items with a \$0 value. In these cases, we applied weights to the other records within that entry in order to draw attention to this abnormal behavior.

Credit and Debit Testing

A fundamental tenant of double-entry accounting is that credits and debits sum to the same value within an entry. We, therefore, included a test for this in this model. We did not find any failures. This test was still a valuable exercise to include because we were working with manual entries that may be entered mistakenly.

Additionally, we included testing on the accounts being credited and debited. When examining the data, we found cases in which the same account was credited and debited within one entry. Traditionally these should be separate accounts. As a result, we developed a test to examine how frequently this was occurring. These results were surprising. After discussing with the auditors, we found that there were occasions in which this was acceptable when two business units are transferring money within the same account. Despite this, however, they acknowledged the value of such a test. Noting that it should not be occurring with the frequency, which we discovered, was happening. As a result, we kept the test but agreed to only attribute a small weight to this in this final suspicion model.

Benford's Law and Digit Distributions

Benford's Law is a law of natural number occurrence. According to this law, any set of randomly naturally occurring numbers should follow a predictable pattern. This law has been applied to a host of accounting applications.

To our knowledge, the application of Benford's law to a large population of GL data has not been explored academically. As a result, we thought it would benefit academia and practice to examine whether GL data fit the predicted patterns. We believed that the holistic nature of GL data, along with the randomness of numeric values in a financial institution would mean that Benford's law should apply to this data. As a result, we planned on incorporating this into this suspicion scoring system.

Unfortunately, we found that this was not the case. We tested 1st, 1st two, and final digit distributions in all of the datasets and found no significant conformation to Benford distributions. Upon discovering that there was no confirmation with Benford distributions we decided to examine to see if there was any relevant pattern to this digit distribution that could be seen one year/month to the next. Again, we found no significant conforming pattern. We, therefore, conclude that perhaps Benford's Law does not apply to GL data. We leave the examination of this factor to future research.

The purpose of including Benford's law in this essay is twofold. Firstly, we wanted to demonstrate that not every test would work with every dataset. Additionally, we wanted to highlight the living nature of this methodology. We hope that by showing that some of these tests did not work out we will illustrate that this methodology is robust to the failure of one or more tests. This should not prevent auditors from developing new testing methodologies. Should they not be applicable to the dataset this methodology will still apply.

Comment Keyword Mining

Text mining is a technique that has become increasingly integrated into a variety of accounting applications. We decided to apply a basic text mining approach to examine whether or not comment data could be effectively used in examining manual entry data. This approach in this application was fairly unique but straightforward.

The first step in applying a text mining analysis is to select or develop a dictionary. In the case of this application, we formulated a dictionary of comment key words based on data given to us by the internal audit team. We developed a list of 24 important keywords. Consistent with text mining techniques we incorporated abbreviations, variations, or synonymous terms into one keyword category.

From this list of 24 keywords, a short-list of highly suspicious keywords for each individual ledger system was developed. Some of these keywords were consistent across all ledger systems such as "cancel". In this case, the internal auditors had told us that while it was acceptable to have canceling entries this preliminary results indicated this happened too often. As a result, they flagged all of these entries hoping to catch the truly problematic instances through this suspicion scoring algorithm. In other cases, certain keywords pertained to entries that may have been permitted in some systems but not in others. They were therefore only scored in the systems in which they should not have occurred.

3.4 Description of Case Example

For this case study, we examined a large multinational bank. This particular firm holds assets in excess of \$400 billion with annual revenue totaling over \$50 billion. As a large multinational it serves tens of millions of customers at branches in over 20 different countries. As a publicly-traded firm, it has a responsibility to ensure financial accuracy to its thousands of shareholders as well as the vast number of people it employs.

The bank provided manual entry data for the financial year 2016. As mentioned, decades of mergers and acquisitions have led to several legacy systems being in place at this firm. With respect to manual journal entries, there were 7 separate systems in which manual entries could be entered. Data on each of these systems was organized differently and not all of the systems contained the same data variables. This forced us to generate a methodology that was robust enough to adapt to different applications from company to company.

For the purposes of this study, we examined 4 different data sources, which covered 6 of the seven systems. Two pairs of systems were combined into a common data source (Systems 1 & 2, and systems 3 & 4). The data on the final system for manual entry was not provided, as it was part of a joint venture with an external firm. The description of the four data sources and their particular systems is outlined in a table below. Because the data was presented as four data sets, we ran four iterations of this methodology and compared the results.

Data Source	System	Description	Population	Number of Variables
	1	Manual entries can either be for specific (1) or generic		
1	2	(2) types of transactions. Entries must be made on an individual basis.	1,691,708	77
2	3	Entries can either be for specific (3) or generic (4)	2 722 472	77
2	4	entries may be entered in batches	3,722,472	11
3	5	This is the system used for entries pertaining to the wholesale department of the business. No batch entries are allowed.	33,686	34
4	6	This is the new system being implemented in the firm. No batch entries are allowed and any type of entry may be made into this one common system.	414,625	55
N/A	7	Entries for joint venture with non-participating third party	Not provided	N/A

Table 6:	Internal	Audit	Data	Sources	

3.5 Results

Overall, we found a host of issues that proved valuable insights for internal auditors of the firm. On a holistic level, this methodology was capable of discovering large scale system errors and violations of internal controls that may be considered material. While that may not differentiate this approach from traditional ones, these results carry to low-frequency high impact issues that may otherwise have not been discovered. In addition, there was one confirmed instance of an employee trying to publish deceptive financial data. While this employee had already been fired for his acts, the internal auditors acknowledged that our methodology might have detected this sooner than they were able to. Finally, these results were not limited to discoveries in any one data source. This proved that the outlined approach is adaptable to a variety of different datasets of various structures or sizes. Below are the results, broken down with respect to each of the research questions.

Research Question 1 Results

In this section, results will be reported, which paint a compelling picture that this methodology can in fact detect a variety of both systematic, and low-frequency high-risk problems in the datasets utilized in this study. Additionally, it should be noted that the applicability of all the tests to all the datasets varied. While this will be discussed further in the following section on RQ2, it is important to note that the absence of results from a test, in this case, is in fact a result itself. Because this methodology deploys full population testing, there is no need to extrapolate what portion of the population is "clean" with no issues. There is actual data that can be utilized by internal or external auditors to indicate which portion of the population has no issue. The materiality of this clean portion along with the problematic portion as well as the exact monetary impacts to the financial statements are known.

Systematic Issue Detection

With respect to systematic population issues, there are several results that indicate that this approach was capable of detecting the same issues that traditional or non-

statistical sampling approaches may have detected. Firstly, there was an issue discovered when comparing the account that was credited to the debited account within one entry. Typically, you would expect that these would be different accounts. This issue was predominant in data sources one (11% of entries) and two (7%). After consulting with the internal audit team from the firm we found that this may occur as a result of inter-departmental transfers but the rate at which it occurred may have indicated a system error that was previously unknown.

As well as the aforementioned issue, this methodology discovered a systematic lapse involving internal controls. The first of these pertains to simple date matching. Within one record the debit and credit dates should match. The filter for this was able to detect that within data source one this did not occur in an average of about 4,000 entries per month when applied to data source one. This is contrary to GAAP which requires that credits and debits occur simultaneously in the same entry. While taken independently each of these errors may be immaterial, together they reach a level that an external auditor may consider material.

Additionally, lapses in internal controls were also found utilizing text analysis on the data sources. Firstly, there were systematic issues in some of the data sources which had mandatory description fields left blank. More specifically, after filtering for keywords that were deemed to reflect important transactions by the internal auditors, several issues were found in data source one. One required control was an authorization approval for each entry. Utilizing a keyword filter, it was discovered that among others, no adjusting entries had any approval. Such records totaled over \$1 billion in value illustrating how material such issues can become when compounded.

Infrequent Issue Detection

In addition to systematic errors, this methodology was able to detect errors that were infrequent but potentially material that otherwise may have gone undetected utilizing traditional sampling methodologies. One example of such error that was discovered was with respect to duplicate entries. These were discovered primarily in data sources one, and three. In both of these cases, duplicate entries accounted for less than 0.25% of records (0.02% and .14% respectively) making it highly unlikely that the duplication would be detected utilizing traditional sampling methods. With respect to source one, the 365 duplicates were originally believed to be a system error. Upon looking into this issue further it was discovered that a vast majority of them took place during September and October with none occurring in Q1 or Q2. Even if this were to be a system error, internal auditors would consider this an interesting finding.

With respect to data source three, this trend raised some red flags. Out of the 23 cases of identical records, several were valued at tens of millions of dollars, the highest of which fell just shy of one hundred million. What was more concerning was the timing of the events. Almost all of the records pertaining to the duplications occurred on June 30th right at the end of Q2. This highly unusual behavior that may not have otherwise been discovered would be considered a valuable piece of information on an audit and certainly something an internal auditor would wish to discover.

Using the multi-tiered filtering and matching approach yielded additional results that may otherwise have gone undetected. The most notable of these that received substantive follow up review from the internal auditors were with respect to source two. In this case, the filtering and matching process for suspicious reversal entries returned a large number of cases (8,636). Almost 70% of these suspicious records were attributed to two employees. By the firm's admission, one of these employees had been fired for financial misdoings in the years since 2016. Had they employed this method in a continuous internal monitoring system it is likely that this individual would have been caught sooner.

<u>Research Question 2 Results</u>

Based on the findings outlined above, it is clear that this system of filtering and testing a data population is adaptable to a variety of different data structures and sizes. Not only did the methodology apply to all four sources, but the results varied with the real-world fluctuations that one might expect. This clearly indicates that with respect to RQ2, this approach to examining internal adjustment data is adaptable to different datasets.

Notable in this regard with respect to the other sources is data source four. Data source four had very few issues with respect to the filters that we applied. This is an important example that illustrates what applying this methodology to an ideal "well behaved" set of data may look like. One issue did potentially stand out. There were 12 cases in which entries were accepted by the system, but no data was included about a credit or debit account. While some of the values may be considered material (>100,000) without information linking the record to an account, it is unlikely they were reflected on a financial statement. As a result, this is indicative of full population filtering being applied to a largely unproblematic dataset successfully.
3.7 Conclusion

Previous studies have highlighted the failures of traditional statistical and nonstatistical population examination methodologies (Deming 1954, Arkin 1957, Neter & Loebbecke 1975, Hall et al. 2001, Elder and Allen 2003, Hall et al. 2000, Elder and Allen 1998, Burgstahler and Jaimbalvo 1986, Blocher and Bylinski 1985). Newer methodologies have begun to emerge that have slowly been applied to various populations of accounting data (Kim and Kogan 2014, Kim 2011, Issa 2013). This study aims to extend this literature into the domain of journal entry testing by answering the following research questions: 1.) Can a risk targeting, full population filtering approach, discover problematic or erroneous updates to a GL? 2.) Is this approach adaptable to a variety of different systems?

To answer these questions, seven different test conditions were examined and applied to four different systems within one large financial institution. The results indicated that a variety of issues could, in fact, be detected in this setting. These issues detected ranged from systematic failures of internal controls to low-frequency high-risk issues, including employee malfeasance. A variety of differing results were gleaned from four separate and unique data sources proving that the approach is, in fact, adaptable.

This study is not without its limitations. Firstly, the results are limited to adjustments to the general ledger. The adaptability of this approach to various datasets may not hold true in different accounting settings. Future research may wish to apply this methodology to different settings and find out. Secondly, the results pertain to a single financial institution. While each of these four data sources had different variables, keywords, and sizes, this does limit the assertion of adaptability. Future researchers may also wish to test the applicability of this methodology to datasets from other companies or industries.

While not exhaustive, this study provides a valuable stepping-stone in the realm of manageable full population examination of records. The successful implementation of the filtration and testing portion of the methodology outlined in this paper provides evidence that future researchers and practitioners can use to extend the accounting field's methods of practice. By providing tangible audit evidence in both the macro and micro scale, it is hoped that future research can apply this methodology in greater detail.

Chapter 4: Application in an External Audit Environment

4.1 Introduction

The General Ledger (GL) is a unique element, unlike anything else in accounting. This vital link serves as a connection of aggregated individual business events and the eventual formation of the financial statements that are released by publicly held companies. This unique situation poses both a significant threat and a great opportunity for auditors. Errors in the account balances represented in the GL may carry through to the final financial statement, possibly resulting in a material defect. Therefore, it is in an auditor's interest to provide assurance over these balances so as to thereby provide assurance over the final financial statements.

In order to provide a reasonable level of assurance over the entire accounting structure leading into financial statements, auditors examining the GL should not only examine account balances but rather all postings related to the changing of such balances. The vast volume of individual account changes reflected in these ledgers can make spotting these issues difficult for auditors. However, if auditors can accurately review changes to all accounts represented in the general ledger, they will be able to provide assurance with a higher level of confidence. Doing so would enable them to provide assurance and oversight over some of the most granular levels of business recording.

To tackle this monolithic problem, auditors have traditionally relied upon classic statistical and non-statistical sampling techniques. These techniques, however, are not without fault. Statistical techniques are poor at detecting low-frequency high-risk events, especially in large populations that exist today. Additionally, non-statistical techniques suffer from targeting biases as a result of the selective nature of the sampling methodology. Modern methodologies use techniques such as suspicion scoring and exception weighting to resolve such issues. One such methodology is the Multidimensional Audit Data Sampling (MADS) approach. This method has been employed in several different capacities; however, never to anything related directly to the unique instance of changes to account balances on the GL. MADS is a process of full population involving filtering, suspicion scoring, and then record priority.

This paper takes a design science approach to solving an accounting issue, as outlined by Kogan et al. (2019). In this case, the problem of ensuring the fair representation of a financial statement is refined through the lens of GL account updates. The artifact that is used to tackle this issue is the MADS methodology. This methodology was developed to surmount sampling issues in large populations (No et al. 2018) and has been successfully applied in other audit-related contexts in the past (Lee et al. 2019, No et al. 2019, Yoon et al. 2019). This application instance is unique as changes to accounts in the GL represent a link between business events and final consolidated financial statements. The effectiveness of this approach is evaluated through the importance of findings, and risk evaluations provided by senior audit partners. If the MADS methodology is capable of detecting issues or risks rated by senior partners as being of high concern or interest, this application may be considered successful.

The goal of this paper is twofold. Primarily the goal is to illustrate the effectiveness of MADS as an artifact in detecting issues within a GL related dataset. Secondarily, the paper contributes by illustrating how examining individual changes to accounts within the GL can lead to important audit-related discoveries. Throughout this paper, the methodology is applied to the GL postings of a large multi-national manufacturer. Several insights are gained concerning earnings management, low-frequency high-risk problems, and the effectiveness of internal controls.

To achieve this, the MADS methodology was applied to a portion of the postings related to changes of accounts within the GL of a large multi-national manufacturing company. This approach was ultimately designed to provide assurance over the final financial statements. The traditional audit link between financial statements and granular level data, such as GL postings, are the audit objectives. To this end, when implementing this approach, great care was taken to ensure that traditional audit objectives were being met over the course of application. To ensure the aggregated statements were correct, filters were designed to test and examine the GL at the posting level. Each of these filters targets specific audit objectives as rooted in the GLARE framework. Over the course of this study, completeness, accuracy, cutoff, classification, rights and obligations, occurrence, and valuation objectives are all tested, representing almost all of the traditional audit objectives. On top of this, while traditionally internal controls may not have been thought of as an audit objective or test of managerial assertions, they are also tested. This is because not only can GL internal control failures manifest as material misstatements on the aggregated financial statements, but auditors are required to perform such tests in a post-SOX era. By applying such an approach to the breadth of business data found within individual changes to accounts within the GL, insights can be gained over the accuracy or reliability of the combined financial statements that are generated from this data.

The remainder of the paper is organized as follows. The next section covers the literature motivation surrounding the impetus of this idea. The following section discusses the MADS methodology employed in this paper. The next section covers how that was

integrated with the example used in this study. The fourth section analyzes the results gathered from this study. The final section concludes.

4.2 Literature Motivation

This section will outline the major literature that motivates the necessity for this study. Firstly, the extant sampling methodology literature pertaining to statistical and non-statistical is examined. Predominantly, a discussion is provided on the issues that pertain to either of these sampling approaches rendering them potentially inadequate in a modern paradigm. These issues provide the stimulus for the development of the MADS methodology applied in this paper. Secondly, the modern literature surrounding the development of the applied MADS approach is reviewed. This establishes an avenue for the application of this alternative approach to examining a GL dataset as is done in this paper.

Low Frequency Error Detection

One key issue that exists with sampling techniques is their effectiveness (or lack thereof) in detecting low-frequency errors in a population. This problem has been well documented (Neter & Loebbecke 1975, Hall et al. 2001). Neter & Loebbecke's comprehensive analysis applied a variety of sampling techniques to four distinct accounting populations that range from accounts receivable to inventory. They find that none of the applied methodologies were particularly well suited for detecting lowfrequency errors in a population. This becomes more problematic when dealing with large populations that may have low-frequency high-risk areas of concern. The logical solution to this problem may be to increase the size of the sample. The sample size plays a vital role in an audit because errors can be detected only if they are included in the sample. Neter and Loebbecke tested sample sizes of both 100 and 200 and recommended in several cases that a larger sample size be used. It is important to note that the conclusion was made based on comparatively small population sizes (no population was larger than 10,000 observations). While audit partners interviewed for our study recommended a sample size of 200-300, our population was much larger (millions of records) in comparison. Combined with the information provided by Neter and Loebbecke, our study's large population indicates a need for either a larger sample size or a new approach not yet examined. Hall et al. (2001), however, found that the benefits of increasing sample size were limited, if statistically significant at all. This illustrates a further need for a new methodology in large population examination.

Non-Statistical Method Biases

Apprehensions about problems regarding non-statistical methods of sampling date back to the 1950's (Deming 1954, Arkin 1957). Despite concerns and the development of newer methodologies, a survey conducted by Hall et al. (2000) found that non-statistical sampling was used for 85% of audit sampling tasks. It is believed that this might be driven by an emphasis on inherent risk and auditor knowledge of a client (Elder et al. 2013). This theory acknowledges that such sampling methodologies are not free from auditor decision-making. Several studies have suggested that auditor expectations have potentially adverse effects on the way that these sampling methodologies are applied, impacting the resulting sample and interpretation of that sample (). Elder et al. (2013), and Akresh et al. (1988) provided a detailed examination of statistical and non-statistical sampling literature.

Modern Methodologies

Due to the aforementioned issues with traditional sampling techniques, several streams of research have been aimed at methodologies that are designed to develop samples and rank exceptions. These methodologies are aimed primarily at full population examination and exception prioritization for sampling. The common problem with the application of full population examination is the existence of a large number of exceptions. This has been discussed with the most common solutions being weighting and suspicion scoring (Issa 2013) and multi-tiered filtering (Kim and Kogan 2014, Kim 2011). MADS is a methodology that extends this literature and attempts to provide a new approach to sampling that resolves some of the aforementioned concerns (No et al. 2018).

Unfortunately, there is a lack of competing methodologies to deal with some of these concerns. As a result, this study applies the MADS methodology to a large GL related dataset to fill the gap in the literature and provide insight into resolving some of the issues with traditional sampling techniques.

Utilization of General Ledger Datasets

An additional contribution of this study is to examine the effectiveness of GL related datasets in drawing audit conclusions. This study takes into account not only the application of a novel GL related dataset but also those datasets that may provide additional information in a fashion similar to how auditors may conduct such a process in practice. Unfortunately, there is a dearth of extant literature on this subject. There is, however, a discussion on the necessity for such studies. Debrecheny et al. (2005) outline

how Embedded Audit Modules can be utilized with ERP systems to facilitate audits. They mention that metadata, similar to that used in this study, can be leveraged to provide better understanding and more insightful audit conclusions. They also detail the advantages of taking a holistic organizational approach to auditing data. This is echoed by Gray & Debreceny (2014), who assert that there is a need to audit the linkage between journal entries and the GL somehow. This study aims to overcome this by examining the incremental increases and decreases in account values that are represented in the GL.

This paper also utilizes additional datasets not directly related to the GL in a way consistent with how an auditor may approach for tests. The primary of these is the employee data. While this may not be strictly external to the audit client, they may be considered external to the GL portion of an audit. Several studies outline the benefits of using data in such a way. These arguments are best summarized by Earley (2015), who outlines that in order to leverage analytics properly, multiple data sources should be combined.

It is to these ends that this study will 1.) Utilize a novel dataset (data on the increase and decrease of accounts in the GL) that represents a link between journal entries and the GL postings 2.) Combine multiple largely unrelated data sources to enhance the understanding of the target data.

4.3 Methodology

In order for the financial statements to be considered a fair representation of a company's well-being, the information utilized in the compilation of these documents

must be fairly represented. Since a portion of this data is comprised of information reflected in the GL, this information must be verified first. To determine the accuracy of these balances, the changes to such balances must be examined. Due to the volume of changes that occur to all account balances within the GL, an additional sampling problem arises when examining such data. This paper will examine the effectiveness of the MADS methodology in solving this problem. The MADS methodology utilizes filters to highlight those updates that would be considered by senior audit partners to be risky. Based on the evaluated importance of these filters, and which filters a particular update failed, each is assigned a suspicion score which is used to rank account updates from most to least risky.

The first step in testing the MADS application to this environment is to obtain a dataset of all the changes to accounts within the GL. As individual business events occur, they are recorded as journal entries. These individual journal entries result in credits and debits that change the balances of various accounts. As this occurs, a posting is generated, which reflects the credit or debit and the impact on an individual account. Traditionally the final account balances are tabulated based on these postings and then represented in the GL. This process linking the events to financial statements is reflected in figure 1 from chapter one.

The methodology employed in this paper is designed to target the individual changes to each account as they occur. Such data is recorded using company-wide ERP systems. While not all transactional or business data is reflected in these individual changes to accounts, enough information is captured to examine several important audit objectives, as will be outlined later in this paper. Because such changes to accounts

happen in large volumes, a methodology must be applied to conduct testing over the population. It is imperative that such a methodology properly examines such changes to account balances.

The overarching methodology applied in this study was the MADS approach to sampling developed by No et al. (2018). The MADS methodology utilizes several stages aimed at filtering a total population of records down to a prioritized sample used for a test of details on an audit. This methodology has been replicated in other studies (No & Huang 2019, Yoon et al. 2019, Lee et al. 2019) successfully, and the original outlined framework is displayed in the figure below.



Figure 4: MADS Model Building Process

The MADS approach begins with the determination of potential problems that may occur in the existing dataset. Based on the determination of these problems and risks, filters are designed to highlight the existence of each problem. This relies predominantly on auditor judgment and is in line with established current standards as part of the audit planning process (AS 2101). In this application of MADS, we incorporate the GLARE framework in order to reduce and imitate more effectively some decisions that an auditor may make. It is designed to link available data to audit assertion backed risks. The GLARE framework incorporates recommended tests based on identified risks. These tests are rooted in academia, standards, and best practice guidelines.

As filters are each applied to the entire dataset, the results are evaluated, and the filters altered when needed. This iterative process is necessary to ensure that filters are only flagging risky postings and not normal business practices. This is pursuant to AS 14, which emphasizes the importance of collecting evidence related to "unusual or unexpected transactions, events, amounts, or relationships" (AS 14). In some cases, a secondary filter may be applied to a sub-population of items that have been flagged as a result of a primary filter. This secondary filtering is conducted to provide further insight into risky patterns within the data.

Once all the filters have been applied to the dataset, each filter and secondary filter is assigned a weight. This weighting is conducted in line with the guidance provided by Audit Standards 11, 12, and 14, which pertain to evaluating analytic results and material impacts of findings. Auditors assign these weights based on their judgments regarding the importance of the filter, the risks associated with that filter, and the results of the application of the filter. Each general ledger posting is then assigned a suspicion score. This score is based on the weights of each individual filter that a particular posting may have failed. These scores are used to rank the records for final prioritization. A visual representation of this process, including example filters, is displayed below.



Figure 5: MADS Visual Example

At any point during this process, additional analytic techniques may be applied to provide a further meta-analysis of the filter results. In the case of this study, an additional materiality analysis was conducted to ensure that the final prioritized sample contained postings of material nature. This will be discussed in detail in the following sections.

4.4 Overview of Instance Implementation

The overall implementation of the strategy employing MADS methodology toward examining changes to account balances is detailed in the figure below. MADS is designed as a flexible process and has never been applied to GL related data. As a result, it must be acknowledged that there are some differences between this application and those employed elsewhere in the literature. It must be noted that such differences are minor, necessary, and fit within both the spirit and overall guidance that the MADS framework provides, as well as the standards to which MADS was designed to conform.

The following few sections will outline how each stage was implemented within this

study.



Figure 6: Implemented Process Diagram

4.5 Data

The data used in this study pertains to one operating segment of a large multinational manufacturing firm. The external auditor of the firm provided this to us. As a result, there was very little ability to follow up with the firm itself if there were any issues, questions, or findings that occurred during this study. While the auditor was able to provide insight into some of these concerns, this feedback was limited. Despite this, there is sufficient evidence to suggest that this methodology did detect potentially material issues with account changes reflected within the final balances found in the GL. To reach these conclusions, three separate pieces of data were used: a log of account change postings affecting GL accounts, the chart of accounts, and a list of employees authorized to make said changes.

The General Ledger

The focus of this study is centered on the changes to accounts within the general ledger of one operating segment of a large multi-national manufacturer. This operating segment functions as an independent business, and the full gamut of business activities that would generally result in account balance changes are represented within this dataset. In total, there are 3,398,356 data points, which reflect some change to a GL account. These changes stem from 546,104 unique journal entries. Each of the changes to account balances was effective during the 2014 fiscal year. The average absolute value of an individual account change was \$2,322, and the total absolute value of all of the changes within the dataset is \$2,434,788,451.56. The table below outlines all 13 of the variables, including the eight present in the primary dataset.

Tab	le	7:	Var	iable	Descri	ptions
						_

Variable	Description	Data Type	GLARE Class
Account Number	The identifier of each account being credited or debited.	Alphanumeric	Account
Amount (local currency)	The value of the credit or debit.	Numeric	Adjustment Value
Description	A brief description of what the line item relates to. This usually includes a numeric code along with a few identifying words.	Alphanumeric	Other
Effective Date	This is the date on which the journal entry became effective.	Date	Timing

Entry Date	This is the date that the journal entry was entered into the system.	Date	Timing
Preparer ID	This field identifies who prepared and entered an entry into the system.	Alphanumeric	Entrant Information
Source	A two-letter code that classifies the type of entry being made.	Alphabetic	Other
Document Number	The identifier of each journal entry. All line items from the same entry have the same document number.	Numeric	Other
Account Type	A type of an account such as asset, liability, equity, revenue, and expense.	Alphabetic	Account/Other
Account class	Name of the account	Alphabetic	Account
Employee Status	Status of employment based on the employee data as of Mar 2019	Numeric	Entrant Information/Other

Five of the 13 variables were excluded from the examination in this study. This was done for one of the following four reasons: (1) all or most values associated with that variable were missing, (2) all values with respect to that variable are the same, (3) a variable is derived from the other in the same data set, or (4) a variable has 1-to-1 relationship with another variable in the same data set. Each of these omitted variables, along with a description of their omission is listed in the table below.

Variable	Description	Data Type	Reason For Omission
Amount (group currency)	The value of the line item in group currency.	Numeric	All records have zero values.
Business Unit	Denotes the business unit that the line item pertains to.	Numeric	We examined one business unit so that all records have the same value.

Table 8: Omitted Variables

Fiscal Period	Indicates the fiscal month that the event transpired in (effective date)	Numeric	This information can be derived from the other variable in the data.
Document	A brief description of the	Alphabetic	This was a 1:1 match
Description	the entry	Alphabetic	the data.
Preparer ID	A brief description of the	Alphanumaria	All records have missing
Description	preparer	Alphanumeric	values.

Alternative Data Sources

It is highly unlikely that in an actual audit environment, an auditor would only have access to a GL, and the account changes. Additionally, in order to run some filters on these individual changes, additional information beyond what is provided in the primary dataset may be necessary. In order to satisfy these two criteria, two additional sources of data were provided. The first of these was a chart of accounts (COA). This was used to decipher the encoded account numbers found in the primary dataset. Additionally, this was used to determine the account type and expected account behavior (contra vs. normal account) when analyzing account level aspects of various changes. This data source provided the account type and account class variables.

The second piece of additional information that was provided was a list of employees authorized to make account changes. The list included employee names and their corresponding ID number. This information was used to examine employee patterns of behavior. Additionally, it was used to generate the final variable, employee status. This binary variable denoted whether or not the employee was still authorized to make entries at the beginning of 2019.

Unfortunately, these two additional data sources do not match with the year of the primary dataset in this study. While the changes to accounts pertain to 2014, the COA

was from 2018, and the employee list was from early 2019. This had minimal impact on the integrity of the study. Few, if any, changes relate to accounts, not on the COA. Additionally, while there were discrepancies between those preparing postings in 2014 with those authorized to do so in 2019, an additional filter was designed to account for this.

4.6 Filter Generation

After reviewing the data, the first stage in the application process was the determination of potential risks and associated filters. The procedure implemented for this is illustrated in the five-stage diagram below. Based on a fundamental analysis of the GL data, along with a common understanding of both the industry and audit-based principals, a list of ten risks with proposed filters was generated. One key portion of the audit planning process is auditor judgment. This is represented in the final stage of the diagram.



Figure 7: Filter Generation Process

To begin the process and eliminate as much subjectivity and need for expertise as possible, the GLARE framework was utilized to satisfy the middle three steps of this filter generation process. The dataset in question contained variables that covered all four of the major necessary GLARE variable classes (adjustment amount, entrant information, account data, timing data) as well as providing several "other" category variables. This enabled the full breadth of the GLARE risk framework to be utilized. Based on this, ten initial filters were generated, which targeted six of the seven GLARE risk categories (the exception being accounting estimations).

In order to simulate the fifth subjective step and any necessary repetition of the process, a questionnaire was generated and distributed senior audit partners representing five of the largest audit firms. These senior audit partners were asked to rank the level of importance and concern associated with the risk as being high, medium, low, or irrelevant. These filters were generated to target specific audit objectives. Additionally, they were furnished with an opportunity to provide additional feedback in the form of unidentified risks or tests that they would suggest. The results of the questionnaire can be found in the table below. The additional feedback is included in the appendix.

What could go wrong	Potential Tests	Audit Objectives			Level of Importance: High (H), Medium (M),
		Occurrence	Completeness	Accuracy	Low (L)
Entries made multiple times	Search for duplicate entries			x	Firm 1: High Firm 2: Low Firm 3: High Firm 4: High Firm 5: Medium
Incorrect account credited/debit ed	Examine entries into an account that fall			x	Firm 1: High Firm 2: Medium Firm 3: Medium

Table 9: Questionnaire Results

	outside 1.5 times the Inner Quartile Range				Firm 4: *BLANK* Firm 5: *TBD*
Incorrectly entered amounts	Credit Debit Balancing			X	Firm 1: Medium Firm 2: Low Firm 3: Medium Firm 4: *BLANK* Firm 5: High
	T	Timing	Problems:	Ĩ	T
Entries outside business hours (potential fraud)	Date on weekend or holiday			x	Firm 1: Medium Firm 2: *Depends* Firm 3: Low Firm 4: Medium Firm 5: Low
Backdating entries	Effective date and entry date separated by greater than 35 days			х	Firm 1: High Firm 2: High Firm 3: Medium Firm 4: High Firm 5: Medium
Aggressive earnings management	Effective date or entry date within 7 days of quarter end/start and separated by 14 days or more.		X	X	Firm 1: High Firm 2: High Firm 3: High Firm 4: High Firm 5: Medium
	F	Fraud and Ir	ternal Controls		
Employee makes entries outside of their normal value range	Examine values outside 1.5 times the Inner Quartile Range for			x	Firm 1: Medium Firm 2: Medium Firm 3: Medium Firm 4: Low

	employee entries			Firm 5: *TBD*
Systematic error occurs within a department	Examine values outside 1.5 times the Inner Quartile Range based on the department or source	X	х	Firm 1: Medium Firm 2: Medium Firm 3: *BLANK* Firm 4: High Firm 5: *TBD*
Repeating payment error/fraud	Entries from one account with the exact same value made within the same day or the following day		х	Firm 1: Low Firm 2: Low Firm 3: High Firm 4: High Firm 5: Medium
Entries disguised as "normal" based on their descriptions	Examine entries based on description to determine entry type. Flag entries outside 1.5 times IQR		х	Firm 1: Low Firm 2: Low Firm 3:*BLANK* Firm 4: Medium Firm 5: *TBD*

Based on the feedback from the auditors, additional filters were necessary to target those risks they felt essential or ones they felt were missing. As a result, the GLARE framework was consulted, and additional filters (such as targeting for earnings management) were generated. The final list of filters based on this feedback, as well as additional data sources, can be found in the table below. In this final list of filters, each GLARE risk category is represented. These filters are designed to test a host of traditional audit objectives, and the link between the filters and their audit objectives, and

the designated GLARE risk category is outlined.

<u>Test</u> <u>Category</u>		<u>Test</u>	<u>Audit Objective</u>	GLARE Risk Category
	Dup	licate Entries	Occurrence	Control/Frequency
<u>itry</u>		Zero Sum	Completeness	Control/Value
ta En	Same accoun	at debited and credited and entry	Accuracy	Control/Predictive
Da	D	Description	Test of Controls	Control
	Reco	ords per entry	Occurrence	Frequency
	Backdated	Regular	Cutoff	Temporal
<u>ମ</u>	Records	Earnings Management Risk	Cutoff	Temporal
Timi	Advanced Entries	Regular	Cutoff	Temporal
		Earnings Management Risk	Cutoff	Temporal
	Wee	ekend Entries	Test of Controls	Temporal/Control
ted	Mu	ltiple Names	Test of Controls	Control
Prep8 Rela	Inacti	ve Employees	Test of Controls	HR Related/Control
	Intracom	pany transactions	Rights and Obligations	Control
her	Uni	usual Source	Classification	Predictive
õ	Cash or I	nventory Outflow	Rights and Obligations	Estimation/Value
	Unusu	ally high value	Valuation	Value

Table 10:	Tests and	Target O	bjectives
			-

4.7 Original Filters

Of the fourteen filters that were used, five came directly from the suggested list of filters that were proposed to the senior auditors (Duplicate records, Credit and Debit balancing, and three timing filters). These filters were straightforward and targeted issues that were identified as important risk areas by those senior audit partners. The timing filters consisted of one filter that examined for entries made on weekends, one filter designed to look for backdated alterations, and an additional filter that looked for advance

changes in balance. A majority of these filters were designed to make sure that adjustments to accounts are made in the appropriate time frame. These are designed to examine management assertions that records are altered in the correct periods and test the audit objectives involving cutoff. Should there be systematic issues involving backdated or advance changes, there may be a substantial impact on the consolidated financial statements. This may result in a material misstatement of under- or over-inflated revenues. Because of the importance and impact on earnings management, the latter two filters were supplemented with sub-filters.

A backdated change to an account is defined as a change made whose effective date occurs more than 28 days (one month) prior to its entry date. To distinguish instances impacting earnings management, a sub-filter was applied for additional insight. This population of backdated balance alterations was further examined and narrowed down by a filter that checked if this time gap occurred over a change of quarter. The 28day time window was maintained in order to avoid detecting standard practices at the end of the quarter that may occur immediately surrounding the quarter change. This same process was applied to adjustments made in advance. Advance changes are defined as having an entry date that proceeds an effective date by more than 28 days. By examining for systematic preferences to back date-specific types of account changes and enter others in advance, a pattern of earnings management may be detected. There is a high level of import-related to these filters. Systematic behavior to produce earnings management at the account level will have a direct impact on the aggregated financial statements. While some of these changes may be immaterial, the practice of conducting such an approach will result in a material skew in the fair representation of financial data at the end of each quarter.

4.8 Auditor Suggested Filters

In addition to these filters, five more filters were developed to target issues that were identified and proposed to the senior audit partners. These filters include the two employee filters (multiple preparers and inactive employees), two filters pertaining to business units (unusual source and intracompany transactions), and the filter designed to detect missing descriptions. While these filters do not match the survey's suggested tests directly, they are aimed at targeting those same risk areas.

Employee Filters

The employee-related filters utilized employee data to detect potential risks or irregularities. The first filter looking for multiple preparers is a test of internal control that requires that only one individual prepare an alteration to an account. A failure to meet this control may not directly indicate a problem, but it will indicate an irregular business event that may warrant further investigation. This would be a direct example of how this approach can not only be utilized to examine fairness and accuracy of financial representation, but also evaluate the effectiveness of internal control pursuant to the requirements of the Sarbanes-Oxley Act. While testing internal controls may not traditionally be considered an audit objective, in a post-SOX era during which management must assert the effectiveness of internal controls over financial statements, auditors are no doubt concerned with such functions. The filter for inactive employees was developed in conjunction with one of the senior audit partners surveyed in this study. This filter takes advantage of the fact that the GL data source predates the employee roster that was provided by five years. As a result, there are some preparers who made postings in the 2014 GL who are no longer on the 2019 employee roster. The belief is that while some of these employees left the firm to pursue other opportunities, some were fired due to substandard performance. As a result, the records prepared by these individuals should come under increased scrutiny.

Business Unit Filters

To detect abnormal entry practices within business units, two filters were developed. The first, unusual source, was used in conjunction with employee testing. This filter targets unusual behavior within a business source segment. Employees within departments typically utilize the same source as each department is assigned a specific source. It would, therefore, be unusual for an employee to utilize different sources from outside their normal operations. As a result, a filter was developed to search for instances where employees utilized a source that was abnormal based on their standard pattern of posting behavior. This filter is aimed at ensuring that these particular account changes are classified correctly and have not mistakenly been misclassified. The second of these filters was engineered to look for intracompany transactions between business units or segments. While this is a normal business practice to some degree, the presence of a large volume of these transactions may indicate an issue. The aim of this filter is to ensure that the rights and obligations of a particular asset or intracompany transfer are accounted for correctly. Such a filter is also aimed at ensuring a change has been entered into the correct account. Additionally, these types of transactions should be subject to further

scrutiny, as evidenced in the failure of auditors to find intracompany transference of debt in the case of Enron.

The final survey related filter examines the descriptions field. Initially, the aim was to analyze these descriptions for abnormal text patterns. This was included in the survey, and the feedback from the senior partners was very positive on this type of test. However, upon further examination of the posting descriptions, it was found that many changes lacked a description. The lack of data to establish a "normal" behavior pattern prevented this test from being conducted. However, since the lack of inclusion of a description represented an internal control violation, a filter was developed to examine for this type of error.

Additional Filters

The remaining four filters do not specifically relate to the survey responses; however, they are designed to tackle risks or issues similar to those represented on the survey. Two of these relate to data entry procedures and accuracy. This is in line with the risk outlined in the survey designed to make sure preparers are acting to make changes to accounts correctly. The zero-sum filter is designed to make sure that the preparer is correctly transferring all the details of a particular journal entry into account balance changes. This is designed to test for completeness and accuracy. In an ideal scenario, credits and debits to accounts as they pertain to an individual journal entry should balance. Failure in this regard would constitute a failure of internal controls and some level of misstatement. The next filter, records per entry, is designed to highlight groups of account alterations that represent an extensive journal entry. After relating account changes to their originating journal entries, there was a substantial number of entries with 100+ account alterations. The logic in this filter was that there might be an increased likelihood of user error when transferring such a volume of information. As a result, a filter was applied to examine these large batches of alterations. This filter targets management assertions of accuracy and occurrence, ensuring that no additional changes are being recorded that did not happen.

The final two filters are also designed to highlight high-risk areas. These were designed based on the feedback from the senior audit partners. The first of these was to examine those records that directly impact revenues and the income statement and the ownership of the rights and obligations that involve the related accounts, since no filters up until this point directly target this aspect of financial reporting. To this end, a filter was designed to highlight all changes in accounts related to a cash inflow or outflow. By adding emphasis to these changes, a direct line of audit data can be drawn to aspects of the income statement can be drawn. The final filter examines the account changes having the top 1% of value. While materiality is covered later in this methodology, this marginal materiality filter is applied in an attempt to make sure some of these high-value changes, which may have issues, make it into our final sample. This filter aims at ensuring that such high magnitude individual changes to account balances are accurate.

4.9 Suspicion Scoring

The development of a suspicion scoring function is pivotal to the success of the application of MADS to any domain. To this end, great care was taken to ensure that all relevant parameters were taken into account. The function that resulted was both simple and effective. It is denoted below in Equation 1. The equation sums the weights of each

individual filter that a particular change violates. The sum of these weights becomes the suspicion score for that change.



$$s_i = \sum_{j=1}^n x_{ij} w_j$$

- $x_{ij} = 1$ if the record i violates the filter j.
 - 0, otherwise.
- w_i = weight of the filter j.

The feedback from the senior audit partners was used to assign the various weights used in Equation 1. In addition to the primary weights, secondary weights were also applied. This occurred in cases where a filter may result in the flagging of one suspicious account balance change. However, the increased suspicion on that particular change to an account balance casts some element of suspicion on the changes made to other accounts, which occurred as a result of the same journal entry. In these cases, the secondary weights are designed to represent this small amount of increased risk present in problem-adjacent balance changes. The complete list of the primary and secondary weights for each filter can be found in the results section in Table 6.

4.10 Prioritization and Sampling

Once each alteration to balances was assigned a suspicion score, the population was ranked from highest to lowest score. In an absolute sense, these scores ranged from 13-0. The next stage would be to draw a sample by going down the list until a cutoff sample size is reached. After ranking all individual changes, two concerns remained. The first concern was the size of the sample. The same five senior audit partners who took part in the questionnaire unanimously recommended a sample size of 300 postings. The second concern was that the majority of this sample might be made up of immaterially low account alterations.

In order to address this second concern, a stage-two filter was applied to the final ranking of records pursuant to the original MADS framework. This filter removed changes to accounts with a value below a materiality threshold. A threshold of \$3,000 was selected. This was chosen after applying various other values of higher and lower amounts and analyzing the impact on the final list of suspicious changes. A threshold of \$3,000 was above the average adjustment value of \$2,322. Additionally, this threshold still provided a sample of changes that scored as part of the riskiest score range (\geq 5). The results of this prioritized sampling, both with and without the filter, are shown in the table below. Additional results using different thresholds can be found in the appendix.

	Total	records
Suspicion	No	
Score	amount	<u>></u> \$3,000
Score	filtering	
13	25	N/A
12	524	N/A
11	372	2
10	1,330	3
9	5,470	198
8	40,090	526
7	120,730	2,665
6	280,384	8,429
5	545,121	12,973
4	642,944	32,763
3	819,279	43,155
2	127,738	7,334
1	247,685	23,526
0	566,664	25,738

Га	<u>ble</u>	11:	Sus	<u>picion</u>	Score	Resu	lts

4.11 Evaluation of Results

Due to a lack of ability to perform a test of details to compare different techniques or evaluate the effectiveness of the ranked sample, the evaluation results in this application study were conducted in a variety of ways. Firstly, the overall results of the various filters are examined and discussed. Secondly, the results of particular filters are compared to the evaluated importance of corresponding risk factors as reviewed by senior audit partners in their questionnaire. Finally, the results of the final scored population are discussed with respect to materiality, account type, and risk.

Results by Filter

The results of each filter can be found below. There are several useful insights that can be gleaned from evaluating this information.

Test Category]	ſest	Weight	Number of Records	Portion of Population
Data Entry	Duplicate Entries		3 for duplicate records.	151,114	4.45%
			1 for records associated with a duplicate.	759,191	22.34%
	Zero Sum		2 for all records within the entries.	0	0.00%
	Same account debited and credited in one entry		3 for records that fail.	52,373	1.54%
			1 for records in the same entry.	204,674	6.02%
	Description		3 for records with missing description.	2,304,049	67.80%
			1 for all other records in the entry that contains two or more records with missing descriptions.	35,630	1.05%
	Records per entry		1 to all records in a journal entry that contains more than 100 records.	696,423	20.49%
Timing	Backdated Records	Regular	1 for all records of an entry backdated by more than 28 days.	2,408	0.07%
		Earnings Management Risk	2 for all records of an entry backdated across a quarter and related to revenue or expense accounts.	17,395	0.51%
	Advanced Entries	Regular	1 for all records entered 28 days or more before becoming effective.	5,273	0.16%
		Earnings Management Risk	3 for all records of an early entry made within the last week before	941	0.03%

Table 12: Filter Results

		the end of a quarter, with an effective date in the following quarter, and related to revenue or expense accounts.		
	Weekend Entries	1 for all records entered on a weekend.	149,881	4.41%
Preparer Related	Multiple Names	1 for all records entered by an Employee ID associated with multiple different employee names.	0	0.00%
	Inactive Employees	2 for all records entered by an employee that was terminated after the accounting period.	464,468	13.67%
Other	Intracompany transactions	1 for all intracompany transaction records.	128,223	3.77%
	Unusual Source	1 to all records of an entry if the preparer is using a source for the first time.	5,107	0.15%
	Cash or Inventory Outflow	1 for all records pertaining to cash or inventory outflows.	637,933	18.77%
	Unusually high value	1 for all records of journal entries in the top 1% of value.	65,848	1.94%

The first most important feature to note is the lack of results on two of the filters. Each of these two filters (multiple employees with the same ID and zero-sum balanced entry) are designed to check for basic control effectiveness. This example highlights that the lack of results is, in and of itself, a result. Auditors should not discount a zero result when applying the MADS methodology, especially to a dataset such as the GL. While the functionality and effectiveness of the filters should be checked, this good news illustrates that these controls are operating effectively and perfectly.

This is not, however, apparent with all the controls related filters. Two other filters were designed to detect the functionality of basic controls and failed to varying degrees. Firstly, a duplication filter was applied that matched instances where records of changes were identical across all fields. This occurred in over 4% of the population. While this may not seem like a material breach of a control, this was rarely a case of one accidental duplicate. There were many instances where there were 100 or more identical matches in the population for one individual account alteration. While this may be explainable as a systematic error, the potential exists for a more sinister explanation. Regardless, this should be of concern and warrant additional auditor follow-up.

The other control violation was of a more dominating and systematic nature. This was the filter designed to detect a missing description field. Initially, the auditor explained this should be a mandatory field as per the internal controls of the client firm. As a result, there was a variety of text mining-based filters that were to be applied. After running a preliminary analysis, it was found that more than two-thirds of the account changes were missing a description. This must surely constitute a material breach of an internal protocol. Several explanations may exist, which range from improper employee training to a system-based recording error. Regardless, it is imperative that this information is passed on to a client's management team, who may incorrectly believe that this control is operating effectively.

These results indicate that not only can the MADS methodology be applied to detect misstatements or suspicious records for further tests of details, but that it can be used to evaluate the effectiveness of internal controls. This should be of particular interest to auditors who can use this to zero in on the practical effectiveness of various controls and assess the impact that any shortcomings are having at a financial level.

Results by Perceived Importance

The second criterion that was used to evaluate the effectiveness of applying MADS in this context was the evaluation of the ability of filters to detect issues that auditors deemed most important. To do this, the results from the questionnaire were evaluated and combined into the Meta-analysis found in the table below. This analysis was conducted as follows. First, the scores of low, medium, and high were converted into a score of 1, 2, and 3, respectively. Then an average importance score was calculated for each potential problem presented to the senior auditors. In this way, the lowest possible concern was that of 1 and the highest that of 3, with the averages falling throughout that range. The results of this calculation can be found in the table below.

What Could Go Wrong	Avg. Score
Aggressive earnings management	2.8
Backdated postings	2.6
Search for duplicate postings	2.4
Incorrect Account Usage	2.33
Systematic Department errors	2.33
Credit debit balancing	2
(Repeating payment) extreme values based on account	2
Abnormal Patterns of Employee Behavior	1.75
Entry outside of Business Hours	1.5
Description testing	1.3

Table 13: Auditor Evaluation of Risks

Based on these results, auditors consider the two most significant risks to be backdated postings and aggressive earnings management. Both of these issues were detected using the MADS methodology. While neither occurred at a significantly high rate, there are additional factors that make this a worrisome concern.

Backdated account changes occur at a combined rate of just over half a percent. The average value for these changes is \$18,171.90, which is almost eight times the value of the population average. This introduces a concern of low-frequency high-value issues within the population. Such instances are difficult, if not impossible, to detect using traditional sampling methods, as mentioned earlier in this paper. An importance score of 2.6, however, indicates that this is still a significant concern for auditors who may not have to detect such issues using traditional sampling.

More alarming than the backdated alterations to accounts was evidence that highlighted a potential systematic approach to earnings management. In order to test for auditors' number one concern, earnings management, a set of secondary filters was designed. These filters were applied to backdated and advanced account changes with a lag of more than one week on either side of a change of quarter. Backdated earnings management risk entries occurred within 0.51% of the population, highlighting the infrequency of occurrence. These were, however, high magnitude changes with 37 backdated account changes being valued within the largest 1% of adjustments.

When analyzing the advanced changes, which occurred at a low rate of 0.03%, further examination was conducted. Additional analysis was done in conjunction with the COA data to determine what accounts were impacted and when. Upon conducting this
additional analysis, it was found that advanced changes were predominantly made to expense and liability accounts. Additionally, while this occurred in every quarter, it was most prevalent in Q4 and Q2. This, combined with the information that changes to revenue accounts were much more commonly backdated than any other account type, paints a picture of downward earnings management.

These factors support the use of MADS over other traditional methods. In this case, the factor with which auditors were most concerned, earnings management, was likely to be occurring. This concern is compounded given the value of the adjustments. Additionally, it seems unlikely that traditional sampling alone would have discovered these problems, all of which were occurring at very low rates.

<u>Results by Score and Materiality</u>

The final evaluation metrics that were used examine the overall scores of the various account changes and evaluate the materiality of each of these results. First, the materiality of various filters will be examined. Next, the scores will be evaluated. Finally, a combination of materiality, scores, and account types will be examined.

Materiality

With respect to the overall value of the changes that failed filters, three filters stood out as exceeding even a 10% materiality threshold. While most of these filters do not directly indicate a problem, they suggest suspicious behavior that may warrant further investigation. The first of these pertains to adjustments that were made by employees that were no longer active in 2019. This was deemed risky because while some employees may have left the company voluntarily, others may have been fired. In this case, the average value of an account change made by an inactive employee was \$21,096.23, over nine times the population average. The total value of these changes amounted to over \$1 bill. This may be explainable employee turnover; however, an auditor may wish to investigate such an irregularity. Especially concerning is that almost half of the value of all account changes in the population occurs within 13% of the overall population, all by individuals who were no longer with the company five years later.

The other filter to exceed a 10% materiality threshold was designed to target intracompany transfers. These occur when resources are moved around the company and are a legitimate part of doing business. Due to the high average value per event of over \$10,000, a small weight was applied to these records to move them up the priority list in case additional filters were violated. The final filter to exceed a 10% materiality threshold was the missing description filter. This fact lends credence to the claim that this was a material failure of an internal control.

Score

The most notable item that is gleaned when examining the scores is the 566,664 account adjustments that did not receive any score. This indicates that with respect to all of the filters and risks that were of concern, this portion of the population had no issues. This is a definitive claim that cannot be made using any traditional sampling techniques and should be considered a significant benefit of this approach.

Additional analysis was conducted to determine what should be considered a "risky" score. To do this, several one-dimensional K-means clustering algorithms were run to cluster the scores of each change reflected in the dataset. Despite running the algorithms with different numbers of clusters, a reasonably consistent trend emerged

whereby the rightmost (riskiest) cluster began at a score of 5. A total of 24,796 out of 157,312 alterations valued at \$3,000 or more fell into this category.

To further understand how these high score changes are impacting the financial statement, they were broken down into account type. Both liability and asset accounts consistently produce material values for the sum of the account changes that fall into the highest risk category. The value of asset account alterations that belong to the highest risk score group falls just shy of \$800 mill.

4.12 Impact

Taken individually, these findings may seem less consequential. It is worth bearing in mind, however, the impact that these individual account adjustments have on the financial statements. Low-frequency high-value changes are of the greatest concern to auditors because they may pose the most significant impact to the financial statements, yet they may be the hardest to detect. This is evident in the likely earnings management behavior that was discovered during the course of this study. Managing earnings downward before a quarter change means that released financial statements will be an inaccurate representation of the firm. In addition to earnings, management concerns the frequency of problems occurring within both asset and liability accounts will have direct impacts on the balance sheet. This illustrates clearly that examining account changes can provide a fruitful avenue for audit discovery and the analysis of a variety of audit objectives. It is essential to keep in mind when auditing changes to the GL the reason why accountants should bother. Changes to the accounts in the GL represent a record of all accounting events within a company. These alterations provide a direct impact to the numbers on financial statements. The fact that the MADS methodology detects problems successfully within this population should not be overlooked when evaluating the importance and impact of MADS on financial statements. This methodology was employed on a GL related dataset to detect real and impactful issues successfully. While these may be explainable, it would provide an avenue for auditors to further assert their effort in the determination of an audit opinion.

4.13 Conclusion

The list of adjustments to accounts within a GL provides a unique problem and set of conditions unlike anything else in the accounting world. On the one hand, it is the most crucial link in existence between the journalizing of business events and prepared financial statements. On the other hand, such datasets have only grown in size and data volume. This conundrum makes auditing this vital link a veritable uphill battle. Usually, auditors employ traditional sampling techniques to surmount this challenge; however, these come with their own sets of issues. This paper utilizes a new sampling technique known as MADS. It illustrates its effectiveness at detecting high-risk low-frequency events, material failures in internal controls, systematic issues such as earnings management, and portions of the population that are seemingly error free. The implementation of this methodology further proves the necessity of examining such data by auditors.

While no direct comparison was made to other approaches, the application of MADS, a methodology designed to solve the problems associated with traditional sampling techniques, was able to detect problems and make assertions that would likely go amiss had a traditional technique been used.

This paper is not without its limitations. Firstly, this approach was applied to one operating segment of a large manufacturing firm. As no two datasets are the same, it is a possibility that the insights gleaned from this application are the exception and not the rule. As a result, future research should apply this methodology to other industries or datasets from other companies within the same industry. Secondly, while the project was conducted in conjunction with auditors to develop our filters, the study was subject to data constraints that would not exist on a live audit. As a result, the findings and list of filters may have varied had access been given to the appropriate and contemporary data.

Overall the results in this paper are nothing if not encouraging. This study was successfully able to tie together multiple sources of data through the application of the MADS methodology. In doing so, several insights were gained, all of which would have had some impact on the published financial statements.

Chapter 5: Future Directions: Continuous Auditing of General Ledger Adjustments

5.1 Motivation and Overview

In today's modern audit climate, one would be remiss if they developed or extended a methodology without giving substantial consideration to its adaptation toward continuous auditing or monitoring. This chapter is, therefore, designed to provide that insight. It serves as a jumping-off point for the extension of the rest of this dissertation into that realm. It provides a brief discussion on the necessity for full population testing to move into the era of full continuity. This chapter will also address any potential adjustments that may have to be made to extend the methodologies from this dissertation into this paradigm. Additionally, to provide a vision of what this system may look like, a mock dashboard is provided, which would enable the continuous, or near-continuous, monitoring of GL updates.

<u>Necessity for Considering Continuous Auditing/Monitoring</u>

Beginning in the early 1990s (Vasarhelyi & Halper 1991) there has been a movement toward the continuous auditing and monitoring of accounting systems (Kogan et al. 1999, Woodroof & Searcy 2001, Rezaee et al. 2001, Rezaee et al. 2002, Murthy & Groomer 2004, Flowerday & von Solms 2005, King & Magnusson 2011, Singh et al. 2014, Applebaum et al. 2017). This has incorporated accounting data ranging from debt covenants (Woodroof & Searcy 2001) to internal controls (King & Magnusson 2011) to general ERP systems (Singh et al. 2014). It is evident, given the enormous stream of literature dedicated to this movement, that this is the direction that academia is headed.

Beyond this practitioner, bodies and standard setters have acknowledged that continuous auditing is the future of the accounting audit paradigm. The AICPA released an entire book entitled "Audit Analytics and Continuous Audit Looking Toward the Future," as well as publishing and funding several articles and studies (Shilts 2017). This book is dedicated to exploring the use of continuous auditing and examines several application case studies (AICPA 2015). It does not end there. In a 2018 speech at the 43rd World Continuous Audit Reporting Symposium, then PCAOB board member Kathleen Hamm acknowledged the importance of continuous auditing as a possible future of auditing. She further acknowledges that future directions of research take this into account when developing and applying methodologies. (Hamm 2018) As a result of this push, it would behoove those who develop methodologies such as those included in this dissertation to incorporate some directionality or discussion involving the continuous application of their audit methodology, system, or technique. This chapter, therefore, addresses this concern in a theoretical and loosely modeled discussion designed to provide a springboard for future research to this end.

5.2 Potential Adaptation for a Continuous Audit Paradigm

To adapt the contents of this dissertation to a continuous audit or monitoring paradigm would not seem to pose much of a hurdle. The GLARE framework discussed in the early stages of the dissertation would not need many, if any, changes. It was designed with continuous application in mind and is fluid to those concerns. The findings in chapter three should not change in a continuous environment. The ability for GLARE to be applied for the detection of errors and problems should not change whether the data is entering the dataset periodically (annual audit) or continuously. The primary changes would be on how the full population examination and scoring of records would be applied, as in chapter four.

The main difference in the static paradigm utilized in chapter four and a dynamic more continuous one is the volume of information flow and a change in the demands of sampling. As an audit becomes more and more continuous, the window of information between analyses decreases. As a result, the snapshot in question becomes smaller and smaller as the audit population batch reduces from annual data to monthly data, to weekly data, all the way down to the continuous level. While the frequency of these audits increases as the batch decreases, the automation of these procedures should eliminate any undue excess work. As this data window closes the emphasis on sampling changes. No longer is it applicable for practitioners to rank records and select the top n most suspicious records for detail testing.

The alternative approach that would be more appropriate for a continuous paradigm would be to set a threshold. In this environment, auditors would set a risk tolerance threshold. As records enter the system, the filters derived from the GLARE framework would be applied and a suspicion score calculated. This is much like what was done in chapter four. The difference would be in what is done with these scores. Instead of ranking them for sampling, any update to the GL that violates a preset auditor threshold score, would be flagged for review. Ultimately, the auditors would not have to review and potentially reverse these GL updates. Ideally, in a fully matured system, such updates that violate the auditor threshold would be blocked from occurring until a review is completed as per Kuenkaikaew (2013). This may generate a new type of control for use by internal auditors. A model of how such a system would look to an auditor is developed in the following section.

5.3 Auditor Dashboard

This section outlines a dashboard that can be utilized to implement the methodologies discussed in this dissertation and adapted to facilitate a continuous monitoring approach. The dashboard is designed to be customizable by the auditors with default settings based on the implementation of methodologies in this dissertation. An outline of the framework for our dashboard can be seen in the figure below. It has been designed for use by either internal or external auditors. The dashboard system will connect to whatever database source the auditor would like and is designed to be flexible in that regard. Additionally, the system will require the auditor to select the tests that they wish to apply to the data and the weights that they want to assign to the failure of each of those tests as per the GLARE framework. The final input required is the number of resulting entries the auditor would like as an output given a static audit environment. In a continuous environment, this can be adjusted to be a threshold value designed to flag what auditors may consider highly problematic or suspicious.



Figure 8: Dashboard Outline

As mentioned, the outline of the dashboard system includes three major data inputs. The first is raw data on the adjustments. This constitutes the dataset auditor wishes to examine. The next is a database of all preprogrammed filters as per the GLARE framework. This should include all filters enumerated in the earlier studies as part of this dissertation and those in the GLARE outline but not utilized in the two subsequent studies. The final one is a set of preprogrammed suspicion weights associated with each of the individual preprogrammed filters. These should be based on the findings presented in the earlier studies and sound auditor/developer judgment.

As the raw data is fed into the system, it is analyzed with respect to the GLARE framework variable classification scheme. While the raw data in its totality is incorporated into the dashboard system, this classification information is fed into the GLARE framework. Based on the available data classes, some GLARE risk categories may not be available for testing. For instance, if the raw data contains no temporal variables, such testing may not be possible. This information is relayed and informs which of the programmed filters are feasible.

The feasible filters and preprogrammed suspicion weights can either be directly fed into the dashboard system, or the auditors can make judgement-based adjustments. Based on these final inputs, the dashboard system will provide an output. This output can be in the form of a sample, the size of which may be either a default setting or auditor judgment. Alternatively, the output could be based on threshold testing mentioned earlier in this chapter. In this case, a default threshold could be applied, or auditors could manually adjust the threshold based on their risk tolerance.

The first figure in the series below is a mockup of the proposed dashboard system. The primary output of the dashboard system is designed to be a suspicion ranked list of suspicious journal entries and their scores. This output will also detail what tests were failed in order for them to end up on this list. Additional optional functions will also provide exploratory outputs. For example, the summary tab in the following figure allows the user to select a test and retrieve summary statistics on the numbers of entries that failed each individual filter. Alternatively, you can select descriptive statistics, and a host of descriptive statistics on the manual entry dataset will be returned as output. Additional exploratory filters should also be available, as seen in the charts tab in the final figure. This mockup is geared toward more immediate application and, therefore, does not include threshold value settings, as discussed in the continuous audit section above. Future iterations and implementations of this dashboard should do so.



Figure 9: Dashboard Mockup with Select Filters



Figure 10: Illustration of Summary Option for Each Test

Weights Apply Default Weights Edit Weights

Figure 11: Weight Adjusting Feature



Figure 12: Possible Charting Options

5.4 Concluding Thoughts

This essay is designed to provide forward-looking insight into the application of the themes, methodologies, and frameworks within this dissertation. The future of audit seems assuredly rooted in some type of more continuous paradigm, as evidenced by the large body of literature and comments from lead regulatory and practitioner organizations (Hamm 2018, AICPA 2015). As such, it is prudent to position any new methodologies or approaches in such a way that it fits into this developing paradigm.

The bulk of adaptation for a continuous environment centers around two factors. First of these is the fact that the GLARE framework is well adapted to handle transition and use in a continuous setting. While it is possible that new risk categories may emerge in this setting, the extant framework will largely be untouched. The methodologies for full population filtering and subsequent sampling will likely have to be altered. The belief is that since sampling will fade away in a pure continuous environment, the models have to be adjusted. To this end, a solution of threshold setting is proposed. This solution provides that auditors set a suspicion threshold. As updates are made to the GL filters, suspicion scores are applied in the same fashion in which they are done so the aforementioned chapters. Once an update has had all the applicable filters applied, a total suspicion score is calculated. If this total score is above the threshold value, it is flagged for review. Eventually, such records may not even make it into the system. Those that do not break the threshold do not get flagged.

The second focus of this chapter is to develop a prototype framework for how this can be incorporated into an audit dashboard. Such a dashboard could be used in the extant context to generate samples, or in a future threshold-based continuous audit setting. It is designed to incorporate auditor judgment and the GLARE risk framework to produce actionable outputs.

This exercise is not without limitations. While it is a forward-looking piece, it is not completely matured. For this reason, it serves as an exploration into future avenues of research. For example, others may wish to further explore the applicability of various GLARE framework elements in a continuous environment. Additionally, they may wish to build an active version of the proposed dashboard and test it out.

Overall this chapter is designed to provide additional forward-looking insight. While it achieves this, it also provides plenty of avenues for research. After all, the future of audit is coming, and we would all be remiss not to address the adjustments that may have to be made to fit approaches developed around static auditing practices, into ta dynamic continuous ecosystem.

Chapter 6: Conclusion

6.1 Summary

This dissertation is designed to fill two main gaps in the literature. The first of these contributions is to analyze a little-studied dataset that is common to all companies. Examining updates to the GL has the potential to provide auditors the perfect balance of detail and reduced information clutter. The second major contribution involves the extension of sampling literature. In this regard, the application of full population filtering and suspicion score sampling is applied to GL updates for the first time.

After chapter one introduces the dissertation and provides insight into its motivating factors, chapter two develops the General Ledger Adjustment Risk Evaluation framework. This framework is a necessary step in the following chapters as it provides auditors with a jumping-off point for filter determination. The chapter starts where any auditor would, breaking down the categories of variables that may exist within a GL update dataset. Five categories are generated: account information, value information, entrant information, temporal information, and other. Those variable classes are linked with audit assertions to form seven key risk areas: Value, HR, Predictive, Frequency, Control, Timing, and Estimations. Each of these areas is broken down within chapter two, and the necessity for each, as well as potential problems and targeting tests or filters, are proposed. Many of these filters are rooted in extant literature. Chapter two concludes with an example application of GLARE to ensure that the risks and associated filters generated by GLARE are of genuine concern to current auditors. A survey of 5 senior audit partners confirms this.

After a framework for risk evaluation and filter, generation is established in chapter two; chapter three seeks to discover the applicability of full population filtering

on GL update datasets. Four separate data sources, each with a different data structure, are utilized to answer the following two research questions: 1.) Can a risk targeting, full population filtering approach, discover problematic or erroneous updates to a GL? 2.) Is this approach adaptable to a variety of different systems? To answer these questions, the GLARE framework is used to generate filters to be used on the four data sources. These filters are then applied to the data sources, and the results are evaluated. After discovering several problematic issues, and even one case of employee fraud, it is confirmed that full population filtering informed by the GLARE framework can, in fact, discover errors or issues within a dataset of GL updates. As this process was applied to four separate and uniquely structured data sources, and a variety of issues were found in all four cases, the answer to research question two is that these approaches are adaptable to different scenarios.

Chapter four builds on the research in the previous chapters by applying a specific suspicion scoring risk-based full population testing sample methodology known as Multidimensional Audit Data Selection (MADS) to a dataset of GL updates from a large multinational manufacturer. In order to more accurately simulate audit conditions, data sources outside of strictly GL updates are used. In addition to the GL updates, a list of active employees, as well as a chart of accounts, is provided. Once again, GLARE is utilized to build a foundation of filters rooted in audit assertions and the available data. In addition, feedback from senior audit partners is also used to both generate additional filters, to assign suspicion weights to each of the established filters, and to provide a level of insight into the results. The results illustrate just how effective this approach can be. The MADS methodology, under the direction of GLARE risk evaluation, was able to

detect a wide variety of issues. Low-frequency high impact problems that are unlikely to be detected utilizing traditional sampling methods were detected. In addition, systemic issues with control effectiveness were discovered along with what is likely to be earnings management. The later of these two discoveries was particularly encouraging given the importance the panel of audit partners placed on earnings management risk. The final valuable insight provided with this methodology is that unlike traditional sampling techniques, auditors are able to determine the exact portion of the population that is errorfree with respect to the filters and risks that were applied. This may provide them with stronger confidence when asserting to the fair representation of financial statements.

The fifth chapter is a forward-looking and theoretical discussion about how these methodologies fit into the future continuous audit paradigm, with respect to the GLARE framework, little if any will change. Perhaps a continuous environment may result in additional risk categories, but it is not readily apparent that any existing category will be dropped or significantly changed. Full population filtering and risk-based suspicion scoring methodologies will have to change with respect to sampling. It makes little sense in a continuous frame ecosystem to maintain some element of hard sampling. As a result, it is proposed that this swap to a threshold system. Under this approach, every GL update is tested by all the filters and given a suspicion score as normal. If the summed score for the update breaches the auditor threshold, it is flagged for review. Eventually, such records will be blocked from formally entering the system until reviewed. In an effort to move these approaches toward a continuous audit paradigm, a more automated dashboard system is proposed. This system utilized raw data inputs, GLARE based filters, suspicion score weights, and auditor insight to process and generate output samples or output based

on a threshold. A mockup is also included of what this system would visually look like for sampling by auditors today.

6.2 Limitations, Considerations, and Future Research

This dissertation is not without its limitations. Firstly, while the utmost care was given to ensuring its completeness, the GLARE framework may miss some risks that are unique to certain firms, industries, or were overlooked in the creation of the framework. While several validations such as auditor evaluation of an example, and application to five different data sets were conducted, this pales in comparison to the millions of possibilities that are out there. Academics may benefit from examining this question further and subsequently appending the extant GLARE framework.

The second concern is in a similar vein. This being that the applications conducted in chapters three and four are not representative of real audit conditions. The concern would be that the filtering and suspicion scoring models included in each of these chapters are somehow inapplicable to other data sets because of some unknown factor, or the fact that they utilize historical datasets. Future researchers could apply these methodologies to a greater variety of GL update data sets to determine the actuality of this problem.

In addition to the limitations discussed above, there is a concern that perhaps the industries selected, manufacturing and finance, are not fairly representative of all firms within all industries. Furthermore, perhaps the datasets used to represent either the financial or manufacturing industries are not fair representations of those industries, to begin with. A final extension of this line of thought would be that the data set years in question are not even representative of the firms they are supposed to represent. For this reason, future researchers are encouraged to apply these techniques to year over year data

sets, data sets from different firms within the manufacturing or financial industries, or to firms from other industries.

6.3 Conclusion

Overall this dissertation is designed to extend new sampling methodologies to new hereto unexplored datasets. Full population risk-based suspicion scoring sampling techniques require an element of risk analysis to determine filters. To eliminate inexperience, some element of judgment, and to help formalize this initial process, the GLARE framework is developed. It serves to provide a touchstone in an otherwise ambiguous portion of the audit process. The GLARE framework is then utilized to determine if full population risk-based filtering is possible on GL update datasets. This is something that has not yet been attempted in academia. By applying this to four different sized and variously structured datasets, it is proven that not only does this work, but it is robust to changing circumstances.

Once it is determined that full population risk-based filtering rooted in GLARE will work suspicion scoring is added to the mix along with the use of data sets external to strict GL updates in order to more accurately reflect an actual audit climate. The results indicate that not only was this approach successful in detecting issues, you would expect standard sampling techniques to detect, but also those that it may not, such as lowfrequency high impact risky updates. In addition, it is shown that this type of approach utilizing GLARE and MADS will successfully detect those issues that auditors are most concerned about, such as earnings management or control failures.

In order to extend this research into the future, this dissertation also provides insight and analysis into how this can be adapted to the continuous audit/monitoring

ecosystem. To further ease this transition, a more automated dashboard system is outlined and illustrated in a mockup. To this end, this dissertation provides not only clear contributions to the existing academic field but also substantial guidance into how this line of research can further be extended.

REFRENCES

- Akresh, A., Loebbecke, J., and Scott, W., (1988). Audit approaches and techniques. In Research Opportunities in Auditing: The Second Decade, edited by Abdel-Khalik, A.,and I. Solomon, 32–49. Sarasota, FL: American Accounting Association.
- Albrecht, A., Kim, K., & Lee, K. (2017). Changes in Accounting Estimates: Are the Current Disclosure Requirements Sufficient to Deter Managerial Opportunism? *KAIST College of Business Working Paper Series*, (2017-022).
- American Institute of Certified Public Accountants (2020) Segregation of Duties. <u>https://www.aicpa.org/interestareas/informationtechnology/resources/value-</u> <u>strategy-through-segregation-of-duties.html</u> web
- American Institute of Certified Public Accountants (2016) Management Override of Internal Control: The Achilles' Heel of Fraud Prevention. *AICPA Antifraud Programs and Controls Task Force*, New York NY.
- American Institute of Certified Public Accountants (2015) Audit Analytics and Continuous Audit Looking Toward the Future. *American Institute of Certified Public Accountants Inc. New York, NY.*
- American Institute of Certified Public Accountants (2006) AU Section 314: Understanding the Entity and Its Environment and Assessing the risks of Material Misstatement.
- American Institute of Certified Public Accountants (2003) Members in Public Accounting Firms: Practice Alert 2003-02 *AICPA*
- American Institute of Certified Public Accountants (2002) Statement on Auditing Standards no. 99
- Appelbaum, D., A. Kogan, and M. A. Vasarhelyi. (2017). Big Data and analytics in the modern audit engagement: Research needs. *Auditing: A Journal of Practice & Theory* 36 (4):1-27.
- Arning, A., Agrawal, R., and Raghavan, P. (1996). A linear method for deviation detection in large databases. In *Proceedings of the 2nd International Conference* of Knowledge Discovery and Data Mining. 164–169.
- Arkin, H. (1957). Statistical sampling in auditing. *The New York Certified Public Accountant* (July): 454-469.
- Association of Certified Fraud Examiners (2012) Contract and Procurement Fraud. Association of Certified Fraud Examiners Inc.

- Ben-Gal, I. (2005). Outlier detection. In *Data mining and knowledge discovery handbook* (pp. 131-146). Springer, Boston, MA.
- Best, P. J., Mohay, G., & Anderson, A. (2004). Machine-independent audit trail analysis—a tool for continuous audit assurance. *Intelligent Systems in Accounting, Finance & Management: International Journal*, 12(2), 85-102.
- Blocher, E., and Bylinski, J., (1985). The influence of sample characteristics in sample evaluation. Auditing: *A Journal of Practice and Theory* 5 (1): 79–90.
- Brühl, B., Hülsmann, M., Borscheid, D., Friedrich, C. M., & Reith, D. (2009, July). A sales forecast model for the German automobile market based on time series analysis and data mining methods. In *Industrial Conference on Data Mining* (pp. 146-160). Springer, Berlin, Heidelberg.
- Boyle, J. 1994. An application of Fourier series to the most significant digit problem. *American Mathematical Monthly.* 101(9):879-886.
- Burgstahler, D., and Jiambalvo, J., (1986). Sample error characteristics and projection of error to audit populations. *The Accounting Review* 61 (2): 233–248.
- Cadez, I. V., P. Smyth, and H. Mannila. 2001. Probabilistic modeling of transaction data with applications to profiling, visualization, and prediction. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*
- Carslaw, C. A. P. N. 1988. Anomalies in income numbers: Evidence of goal oriented behavior. *The Accounting Review*. LXIII(2):321-327.
- Chandola, V., A. Banerjee, and V. Kumar. 2009. Anomaly Detection: A Survey. ACM Comput. Surv. 41 (3):1-58.
- Chiu, C. T. (1994, January). An intelligent forecasting support system in auditing: expert system and neural network approach. In 1994 Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences (Vol. 3, pp. 272-280). IEEE.
- Chiu, T., Vasarhelyi, M., Alrefai, A., and Yan, Z., (2018) Validating Process Mining: A Framework Integrating Auditor's Risk Assessment.
- Ciprian-Costel, M. (2014). Arguments on using computer-assisted audit techniques (Caat) and business intelligence to improve the work of the financial auditor. *Management Strategies Journal*, 26(4), 212-220.
- Chan, D. Y., & Vasarhelyi, M. A. (2018). Innovation and Practice of Continuous Auditing1. In *Continuous Auditing*. Emerald Publishing Limited.

Coderre, D. G. (2000). Computer assisted fraud detection. Internal Auditor, 57(4), 25-25.

- Cohen, Jeffrey, et al. "MAD skills: new analysis practices for big data." *Proceedings of the VLDB Endowment* 2.2 (2009): 1481-1492.
- Curtis, M. B., Jenkins, J. G., Bedard, J. C., & Deis, D. R. (2009). Auditors' training and proficiency in information systems: A research synthesis. Journal of Information Systems, 23(1), 79-96
- Debreceny RS, Gray GL. (2010) Data mining journal entries for fraud detection: An exploratory study. International Journal of Accounting Information Systems 2010;11:157–81.
- Debreceny, R. S., Gray, G. L., Ng, J. J. J., Lee, K. S. P., & Yau, W. F. (2005). Embedded audit modules in enterprise resource planning systems: Implementation and functionality. *Journal of Information Systems*, 19(2), 7-27.
- Debreceny, R., Gray, G. L., Tham, W. L., Goh, K. Y., & Tang, P. L. (2003). The development of embedded audit modules to support continuous monitoring in the electronic commerce environment. *International Journal of Auditing*, 7(2), 169-185.
- Deming, W. (1954). On the contributions of standards of sampling to legal evidence and accounting. *Current Business Studies* 19 (October): 14-32.
- Dillaway, B. B. (2010). U.S. Patent No. 7,814,534. Washington, DC: U.S. Patent and Trademark Office.
- Deloitte. 2012. Adding insight to audit -- Transforming Internal Audit through data analytics.
- Drake, P. D., & Nigrini, M. J. (2000). Computer assisted analytical procedures using Benford's Law. *Journal of Accounting Education*, 18(2), 127-146.
- Dugan, M. T., Shriver, K. A., & Silhan, P. A. (1994). How to forecast income statement items for auditing purposes. *The Journal of Business Forecasting*, 13(2), 22.
- Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of forensic accounting*, 5(1), 17-34.
- Earley, C. E. (2015). Data analytics in auditing: Opportunities and challenges. *Business Horizons*, *58*(5), 493-500.
- Elder, R., and Allen, R., (2003). A longitudinal field investigation of auditor risk assessments and sample size decisions. *The Accounting Review* 78 (4): 983–1002.

- Elder, R., & Allen, R., (1998). An empirical investigation of the auditor's decision to project errors. *Auditing: A Journal of Practice and Theory* 17 (2): 71–87.
- Fay, R., & Negangard, E. (2017). Manual journal entry testing: Data analytics and the risk of fraud. *Journal of Accounting Education*, *38*, 37–49.
- Fellegi, I. P., & Sunter, A. B. (1969). A theory for record linkage. *Journal of the American Statistical Association*, 64(328), 1183-1210.
- Flowerday, S., & von Solms, R. (2005). Continuous auditing: verifying information integrity and providing assurances for financial reports. *Computer Fraud & Security*, 2005(7), 12-16.
- Freiman, J.W., Vasarhelyi, M.A. (2019) The Current State of Continuous Auditing
- Gaganis, C., Pasiouras, F., & Doumpos, M. (2007). Probabilistic neural networks for the identification of qualified audit opinions. *Expert Systems with Applications*, 32(1), 114-124.
- Goodwin-Stewart, J., & Kent, P. (2006). Relation between external audit fees, audit committee characteristics and internal audit. *Accounting & Finance*, *46*(3), 387-404.
- Gray, G. L., & Debreceny, R. S. (2014). A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. *International Journal of Accounting Information Systems*, 15(4), 357-380.
- Hall, T., Hunton, J., and B. Pierce, B., (2000). The use of and selection biases associated with nonstatistical sampling in auditing. *Behavioral Research in Accounting* 16: 231–255.
- Hall, T., Herron, T., Pierce, B., and T. Witt, T., (2001). The effectiveness of increasing sample size to mitigate the influence of population characteristics in haphazard sampling. *Auditing: A Journal of Practice and Theory* 20 (1): 169–185.
- Hamm, K. M. (2018) Mexican Mangos, Diamonds, Cargo Shipping Containers, Oh My!
 What Auditors Need to Know about Blockchain and Other Emerging
 Technologies: A Regulators Perspective. 43rd World Continuous Auditing & Reporting Symposium, Newark, NJ
- Hassanzadeh, O., Chiang, F., Lee, H. C., & Miller, R. J. (2009). Framework for evaluating clustering algorithms in duplicate detection. *Proceedings of the VLDB Endowment*, 2(1), 1282-1293.
- Hodge, V.J. and Austin, J. (2004) A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22 (2). pp. 85-126.

- Issa, H. (2013). Exceptional exceptions *Diss. Rutgers University Graduate School at* Newark
- Johnson, P. E., Jamal, K., & Berryman, R. G. (1991). Effects of framing on auditor decisions. Organizational Behavior and Human Decision Processes, 50(1), 75-105.
- Kays, B. (2018) Purchase Order Fraud: How to Identify Red Flags in Procure-to-Pay. *Oversight.com*
- Kim, Y., (2011) Continuous Monitoring. *Diss. Rutgers University Graduate School* at Newark
- Kim, Y., and Freiman, J.W., (2019) An ontology and Framework for General Ledger Analytic Applications. *Working Essay*
- Kim, Y. and Kogan, A. (2012). Development of an Anomaly Detection Model for a Bank's Transitory Account System. *Journal of Information Systems*.
- King, N., & Magnusson, A. (2011). Continuous Audit Process Control Objectives. U.S. *Patent No. 8,005,709.* Washington, DC: U.S. Patent and Trademark Office.
- Knorr, E. M., & Ng, R. T. (1997). A Unified Approach for Mining Outliers. In *in Proc. KDD*.
- Kobelsky, K. W. (2014). A conceptual model for segregation of duties: Integrating theory and practice for manual and IT-supported processes. *International Journal of Accounting Information Systems*, 15(4), 304-322.
- Kogan, A., Sudit, E. F., & Vasarhelyi, M. A. (1999). Continuous online auditing: A program of research. *Journal of Information Systems*, 13(2), 87-103.
- Kogan, A., Mayhew, B.W., Vasarhelyi, M.A., (2019) Audit Data Analytics Research An Application of Design Science Methodology. *Accounting Horizons*, 33 (3): 69 - 73
- Koskivaara, E. (2000). Artificial neural network models for predicting patterns in auditing monthly balances. *Journal of the operational research society*, 51(9), 1060-1069.
- Koskivaara, E. (2003). Artificial neural networks in auditing: state of the art. *Turku Centre for Computer Science*.
- Koskivaara, E., B. Back, et al. (1996). Modelling Intelligent Information Systems for Auditing. *Intelligent Systems in Accounting and Finance, Huelva, Spain*, Papel Copy S.L. Plaza de la Merced. Huelva.

- Kuenkaikaew, S. (2013). *Predictive Audit Analytics: Evolving to a new era* (Doctoral dissertation, Rutgers University-Graduate School-Newark).
- Kuenkaikaew, S., & Vasarhelyi, M. A. (2013). The predictive audit framework. *The International Journal of Digital Accounting Research*, 13(19), 37-71.
- Kuruppu, N. (2012). A structured pedagogy for integrating generalized audit software into the auditing curriculum. *Business Education and Accreditation*, 4(1), 113-121.
- Kuruppu, N., & Oyelere, P. (2017). An Examination of Students' Attitudes and Perceptions Towards Incorporating Computer Assisted Audit Techniques in an Undergraduate Auditing Course. *Global Journal of Business Research*, 11(3), 55-71.
- Lanza, R., & Gilbert, S. (2007). A Risk-Based Approach to Journal Entry Testing. *Journal of Accountancy*, 204(1), 32–35.
- Lee, K., Li, Q., and Tojiboyev, N., (2019) Application of Multidimensional Audit Data Selection Framework: Payroll Process. *Working Essay*
- Lightle, S. S., & Vallario, C. W. (2003). Segregation of duties in ERP: an automated assessment tool enables internal auditors at MeadWestvaco to enhance their SOD control reviews throughout the enterprise. *Internal auditor*, *60*(5), 27-30.
- Loraas, T., & Searcy, D. (2010). Using Queries to Automate Journal Entry Tests: Agile Machinery Group, Inc. *Issues in Accounting Education*, 25(1), 155–174.
- Lu, C. T., Chen, D., & Kou, Y. (2003, November). Algorithms for spatial outlier detection. In *Third IEEE International Conference on Data Mining* (pp. 597-600). IEEE.
- Malaescu, I., & Sutton, S. G. (2015). The reliance of external auditors on internal audit's use of continuous audit. *Journal of Information Systems*, 29(1), 95-114.
- Mat Zain, M., Zaman, M., & Mohamed, Z. (2015). The effect of internal audit function quality and internal audit contribution to external audit on audit fees. *International Journal of Auditing*, *19*(3), 134-147.
- Montgomery, D. D., Beasley, M. S., Menelaides, S. L., Palmrose, Z. V. (2002). Auditors' new procedures for detecting fraud
- Moon, M. A., Mentzer, J. T., & Smith, C. D. (2003). Conducting a sales forecasting audit. *International Journal of Forecasting*, 19(1), 5-25.

- Murthy, U. S., & Groomer, S. M. (2004). A continuous auditing web services model for XML-based accounting systems. *International Journal of Accounting Information Systems*, 5(2), 139-163.
- Neter, J., & James, K., (1975) Behavior of major statistical estimators in sampling accounting populations: an empirical study; Auditing research monograph, 2, *AICPA Guides, Handbooks and Manuals*. 9.
- Neter, J., & Loebbecke, J., (1975) Behavior of major statistical estimators in sampling accounting populations: an empirical study; Auditing research monograph, 2, *AICPA Guides, Handbooks and Manuals*. 9.
- Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, *50*(3), 559-569.
- Nigrini, M. J. 1996. Taxpayer compliance application of Benford's law. *Journal of the American Taxation Association*. 18(1):72-92
- Nigrini, M. J. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud detection* (Vol. 586). John Wiley & Sons.
- Nigrini, M. J., & Miller, S. J. (2009). Data diagnostics using Second-Order tests of benford's law. AUDITING: A Journal of Practice & Theory, 28(2), 305-324.
- Nigrini, M. J., & Mittermaier, L. J. (1997). The use of Benford's law as an aid in analytical procedures. *Auditing*, 16(2), 52.
- No, W., and Huang, F., (2019) An Application of the Multidimensional Audit Data Selection (MADS): A Case Study of an Expenditure Cycle. *Working Essay*
- No, W., Li, K., Huang, F., and Li, Q., (2018) Multidimensional Audit Data Selection (MADS): A Framework for Using Data Analytics in the Audit Data Selection Process.
- Public Company Accounting Oversight Board (2008) Staff Audit Practice Alert No. 3. Audit Considerations in the Current Economic Environment
- Public Company Accounting Oversight Board (2007) Auditing Standard No. 5: An Audit of Internal Control Over Financial Reporting That Is Integrated with An Audit of Financial Statements.
- Public Company Accounting Oversight Board (2010) Auditing Standard No. 11: Consideration of Materiality Planning when Performing and Audit

- Public Company Accounting Oversight Board (2010) Auditing Standard No. 12: Identifying and Assessing Risks of Material Misstatement.
- Public Company Accounting Oversight Board (2010) Auditing Standard No. 13: Auditor's Responses to the Risks of Material Misstatement.
- Public Company Accounting Oversight Board (2010) Auditing Standard No. 14: Evaluating Audit Results
- Public Company Accounting Oversight Board (2010) Auditing Standard No. 2101: Audit Planning
- Public Company Accounting Oversight Board (2010) AU Section 342: Auditing Accounting Estimates.
- Rezaee, Z., Elam, R., & Sharbatoghlie, A. (2001). Continuous auditing: the audit of the future. *Managerial Auditing Journal*.
- Rezaee, Z., Sharbatoghlie, A., Elam, R., & McMickle, P. L. (2002). Continuous auditing: Building automated auditing capability. *Auditing: A Journal of Practice & Theory*, 21(1), 147-163.
- Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 73-79.
- Rousseeuw, P. J., & Leroy, A. M. (1996). Robust regression and outlier detection (Vol. 589). *John Wiley & sons*.
- Searcy, D., Woodroof, J., & Behn, B. (2003, January). Continuous audit: the motivations, benefits, problems, and challenges identified by partners of a Big 4 accounting firm. In 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the (pp. 10-pp). IEEE.
- Shilts, J. (2017) A framework for continuous auditing: Why companies don't need to spend big money. *Journal of Accountancy*
- Singh, K., Best, P. J., Bojilov, M., & Blunt, C. (2014). Continuous auditing and continuous monitoring in ERP environments: Case studies of application implementations. *Journal of Information Systems*, 28(1), 287-310.
- Singh, H., & Newby, R. (2010). Internal audit and audit fees: further evidence. *Managerial Auditing Journal*.
- Smieliauskas, W. (2012). Principles-Based Reasoning about Accounting Estimates. *Accounting Perspectives*, 11(4), 259-296.

- Summers, S. L., & Sweeney, J. T. (1998). Fraudulently misstated financial statements and insider trading: An empirical analysis. *Accounting Review*, 131-146.
- Solomon, C., Barnett, R., Packer, S. G., & Lewis-Barned, N. J. (1994). Amputations in the surgical budget. *The New Zealand medical journal*, 107(973), 78-80.
- Ted, E., Goldberg, H., Wooton, J., Cottini, M., & Khan, A. (1995). Financial Crimes Enforcement Network AI System (FAIS) Identifying Potential Money Laundering from Reports of Large Cash Transactions. *AI Magazine*, 16(4), 21–39.
- Thomas, J. K. 1989. Unusual patterns in reported earnings. *The Accounting Review*. LXIV(4):773-787.
- Vasarhelyi, M. A., & Halper, F. B. (1991). The continuous audit of online systems. In *Auditing: A Journal of Practice and Theory*.
- Vasarhelyi, M. A., Kogan, A., & Tuttle, B. M. (2015) Big data in accounting: An overview. *Accounting Horizons* 29.2: 381-396
- Verykios, V., Elfeky, M., & Elmagarmid, A. (2000). On the accuracy and completeness of the record matching process. *Proceedings of the 2000 Conference on Information Accuracy*, 54-69.
- Weis, M., & Naumann, F. (2005). DogmatiX tracks down duplicates in XML. Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, 442.
- Wells, J. T. (2001, Feb). Why employees commit fraud. Journal of Accountancy, 191, 89.
- Woodroof, J., & Searcy, D. (2001). Continuous audit: Model development and implementation within a debt covenant compliance domain. *International Journal* of Accounting Information Systems, 2(3), 169-191.
- Wu, R. C.-F. (1994). Integrating Neurocomputing and Auditing Expertise. Managerial Auditing Journal 9(3): 20-26.
- Yan, L. (2015). Computer Assisted Auditing Techniques for Financial Audit Effectiveness and Efficiency. *Journal of Next-generation Convergence Information Services Technology Vol*, 4(2), 95-104.
- Yoon, Y., Vasarhelyi, M., and Cho, S., (2019) Data Analytics in Auditing: Implementing the MADS framework in the Revenue Cycle. *Working Essay*

APPENDIX

Table 14 Appendix: Audit Partner Suggested Risks

Firm 1	Firm 2	Firm 3	Firm 4
Segregation	It is difficult to assign ratings to	For accounts that are	Rounded values
of	these without having the	expected to have	or consistent
incompatible	understanding of the business	regular entries	ending entries
duties -	process behind the data set.	within a given range,	
entries	We recommend for purposes of	it would be	Entries by User
posted to	this proof of concept, the	interesting to	ID
incompatible	"ratings" are "hypothetical",	statistically analyze	
accounts by	based on some assumptions	(based on risk	Entries on
the same	made being the sample data set	profile) and try to	unusual
preparer	used in the analysis, and used	identify outliers that	business hours
	to prove out the methodology	may indicate	
	behind the MADs research.	error/fraud. The	
		auditor would need	
	In a live project engagement,	to risk-profile	
	the auditor will gain a sufficient	accounts where this	
	understanding of the business	may be likely and	
	process (including journal entry	determine if that	
	process, flow of transactions	account has a certain	
	through the system, etc.) and	acceptable range of	
	the "ratings" would be assessed	size of entry, etc. but	
	based on that the understanding	this could be helpful	
	and the identified/assessed	in risk assessment	
	risks. The ratings for each	and in designing	
	"filter" will vary as such project	further substantive	
	to project.	procedures.	
	SUCCESTED	What about looking	Entries with
	SUGGESTED	at frequency of entries by certain	missing
	ADDITIONAL TESTS:		description
	Unusual combinations based on	personnel? For	Entries with
	understanding of typical journal	example, the CFO is	missing User ID
	entries for each business	not like Firm 1 to be	_
	process	making entries to the	
	Journal entry recorded in one	GL design test to	Manual entries
	period and reversed in the next	analyze entries made	vs Automated
	period (a non-recurring entry).	by unexpected	entries
	Journal entries created and	personnel.	1word phrase
	approved by same user, or		search
	where the user who created the		
	entry is blank, or approver is		
	blank.		

	Total records			
Suspicion Score	No amount filtering	≥\$1,000	≥ \$2,000	≥ \$3,000
13	25			
12	524			
11	372	2	2	2
10	1,330	19	5	3
9	5,470	1,356	370	198
8	40,090	2,361	855	526
7	120,730	7,171	3,734	2,665
6	280,384	22,717	12,141	8,429
5	545,121	34,903	18,434	12,973
4	642,944	104,135	53,369	32,763
3	819,279	153,488	73,156	43,155
2	127,738	14,729	9,197	7,334
1	247,685	48,682	32,455	23,526
0	566,664	97,868	42,706	25,738

Table 15 Appendix: Various MADS Sample Materiality Thresholds