EXCEPTIONAL EXCEPTIONS

By

HUSSEIN ISSA

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

Professor Alexander Kogan

and approved by

____

Dr. Alexander Kogan

____

Dr. Miklos A. Vasarhelyi

____

Dr. Helen Brown-Liburd

____

Dr. Richard Dull

Newark, New Jersey

October 2013

ABSTRACT OF THE DISSERTATION

Exceptional Exceptions

By HUSSEIN ISSA

Dissertation Director:

Dr. Alexander Kogan

The increasing utilization of computerized systems in businesses has led to the generation and storage of massive databases. In light of the availability of such big data, auditing is moving from the traditional sample-based approach to audit-by-exception. The literature is abundant with studies that propose various machine learning, statistical, and data mining techniques that have proved to be efficient in identifying exceptions. However, such techniques often inundate auditors and management with large numbers of exceptions. This dissertation, composed of three essays, attempts to help them overcome the human limitations of dealing with information overload by proposing methodologies to detect and subsequently prioritize such exceptions. These prioritization techniques can help auditors and management to direct their investigations towards the more suspicious cases, or exceptional exceptions.

The first essay evaluates the quality of auditors' judgment of business processes' risk levels using historic data procured from internal controls risk assessments of a multinational company. I identify the exceptions where auditor assessments deviate from the value predicted by an ordered logistic regression model. Subsequently, I propose two

metrics to prioritize these exceptions. The results indicate that the prioritization methodology proved effective in helping auditors focus their efforts on the more problematic audits.

In the second essay I propose a framework where I use a weighted rule-based expert system to identify exceptions that violate internal controls. These exceptions are then prioritized based on a suspicion score, defined as the sum of the risk weightings of all the internal controls that were violated by that specific record. Finally, the exceptions are ranked by decreasing order of suspicion score.

The third essay addresses the problem of data quality from a duplicate records perspective. I present the various techniques used to detect such duplicates, and focus on the issue of duplicate payments. I use two real business datasets as an illustration. Finally I propose a prioritization methodology where each duplicate candidate receives a cumulative score based on multiple criteria. The results show that my prioritization methodology can help the auditors to process duplicate candidates more effectively.

ACKNOWLEDGMENTS

I would like to gratefully acknowledge the contributions and guidance of the members of my dissertation committee, Alexander Kogan, Miklos Vasarhelyi, Helen Brown-Liburd, and Richard Dull. I would like to specially thank Dr. Kogan and Dr. Vasarhelyi for their mentorship during my four years at Rutgers University. I would also like to thank Dr. Brown-Liburd and Dr. Dull for their invaluable comments on my research.

I would like to specially thank my family for their love, support, and encouragement.

Table of Contents

LIST OF TABLES

List of Illustrations

CHAPTER 1: INTRODUCTION

With the current level of utilization of computerized systems in business around the globe, companies generate and store large amounts of data. The proliferation of Enterprise Resource Planning Systems (ERPs) and other management information systems have given companies access to real-time or close to real-time operational data. The existence of such disaggregate data, which is made readily available for companies to analyze and utilize, allows for and even necessitates a different type of assurance services conducted on a continuous basis. For instance, Section 409 of the Sarbanes-Oxley Act of 2002 requires the management to disclose any material changes in the financial condition of the company at the time of occurrence (U.S. House of Representatives, 2002). This change in the nature of expected services have progressively driven companies towards continuous assurance. The American Institute of Certified Public Accountants (AICPA) defines continuous auditing as: *"a methodology that enables independent auditors to provide written assurance on a subject matter, for which an entity's management is responsible, using a series of auditors' reports* issued virtually *simultaneously with, or a short period of time after, the occurrence of events underlying the subject matter"* (CICA/AICPA, 1999). Companies now have the option of auditing the complete population and identifying exceptions, rather than being forced to follow the traditional sample-based approach. Vasarhelyi and Halper introduced in 1991 the concept of audit by exception, where the complete population is audited to identify exceptions. Subsequently, the auditors and the company's management can focus their investigative efforts on these exceptions.

The literature is abundant with studies that apply information technology to real-time data in order to provide a certain level of continuous auditing (Dull, Tegarden, & Schleifer, 2006; Groomer & Murthy, 1989a; Kogan, Sudit, & Vasarhelyi, 1999; Vasarhelyi & Halper, 1991). Unfortunately, the objective of the majority of prior research in continuous auditing has been the detection of exceptions, which are instances that deviate from the expected, or normal, patterns in a dataset, but fail to address the subsequent phase of analyzing and processing these identified exceptions. While exception detection techniques proved efficient in capturing anomalous instances, they usually inundate the human user with an overload of exceptions to process (Alles, Brennan, Kogan, & Vasarhelyi, 2006; Alles, Kogan, & Vasarhelyi, 2008; Debreceny, Gray, Tham, Goh, & Tang, 2003), causing the overall continuous assurance efficiency to decrease due to the human limitation in performing complex aggregation and processing tasks, as prior research has shown (Iselin, 1988; Kleinmuntz, 1990).

This dissertation is motivated by the scarcity of studies that address the problem of processing large numbers of identified exceptions, thus filling a gap in the continuous auditing literature. It consists of three essays that provide illustrations on the problem of exceptional exceptions, by proposing various exception prioritization methodologies.

The first essay uses historic data consisting of control risk assessments procured from the internal audit department of a multinational consumer products company. It is used to infer an ordered logistic regression model in order to provide a quality review of internal auditors' and business owners' assessments of internal controls. The research questions this study attempts to answer are: 1) How can the quality of internal auditors' judgment in control risk assessments be reviewed and verified? 2) How can the quality of control risk self-assessments conducted by business owners be reviewed and verified? 3) And how can the exceptions that disagree from the norms be prioritized using a probabilistic model? This is accomplished by identifying outlying cases where the risk levels assigned by auditors and business owners differ from the expected values. However, the goal of this study is not only to identify these exceptions, but also to see how different auditors' assessments were from the predicted values of the ordered logistic regression model. Subsequently, this difference used to rank the detected exceptions using two measures of disagreement between the assigned and predicted levels.

In the second essay I develop a rule-based expert system based on business rules that test for internal control violations. An analytical model is proposed to determine the risk weighting of each rule based on auditors' pairwise comparison of different transactional records representing different control violations. Business rules do not have the same importance and violations of controls do not have the same risk. Consequently, it is important to assign the exceptions, i.e. transactions violating each rule, weights that correspond to this rule's importance and the degree to which its violation may increase control risk. An expert panel consisting of senior auditors is asked to participate in a survey and to examine a set of paired transactions such that each transaction violates one rule, and the two rules in each pair are distinct. The expert panel is asked to perform a pairwise comparison in order to provide a risk ranking within each pair, and subsequently justify their ranking by selecting the most appropriate reason for their assessment. The pairwise comparisons of all records are in turn used to infer a weighting system that is applied to the identified exceptions in order to prioritize them. Each exception, which is defined as a record violating one or more internal control, is assigned a ***suspicion score,*** defined as the sum of the risk weightings of all the internal controls that were violated by that specific record.

The third essay focuses on the degradation of data quality caused by the existence of duplicate records in databases, and applies different algorithms to illustrate the detection of duplicate payments. This essay is driven by the serious implications that low quality data can have on auditors' decision making and customer satisfaction, in addition to the considerable financial losses it can cause, which necessitate the continuous assurance of data quality. The research questions that this essay addresses are as follow: 1) How can duplicate records be identified using a rule-based system? 2) How can a methodology to rank the detected duplicates be devised in order to enable the human users to focus their attention on the more suspicious cases? This study uses two real business datasets from a telecommunication company that include payment transactions. First, the duplicate candidates are identified, using a three-way exact matching algorithm. Next a candidate prioritization methodology is proposed based on a cumulative score, which is calculated according to a set of criteria.

The remainder of the dissertation is organized as follows. Chapter 2 presents a review of the relevant literature. Chapter 3 presents the first essay, which reviews the quality of control risk assessments and proposes a prioritization methodology for the cases that do not conform to the predicted values. In Chapter 4 I develop a framework that identifies exceptions using a rule-based expert system, and propose an exception prioritization technique. Chapter 5 addresses the problem of duplicate records detection and proposes a ranking technique. Chapter 6 concludes the dissertation by listing the limitations of this dissertations and venues for future research.

CHAPTER 2: LITERATURE REVIEW

2.1. Continuous Auditing

Companies are increasingly relying on computerized systems, such as ERPs, to conduct their business processes. This electronization of business processes, coupled with today's real-time economy, encourages and even necessitates that companies generate timely data. In order to extract useful information and ultimately knowledge that can support decision making (Elliott & Kielich, 1985), it is crucial to ensure the quality of this data (Vasarhelyi, Chan, & Krahel, 2012). While the advancements in technology enabled real-time or close to real-time economy, assurance services evolved at a much slower pace. Most of these services are still conducted manually, an approach that is lengthy in time as well as expensive in cost. This problem is aggravated by the continuous generation of real-time information, compared to the periodical auditing of this captured data (Vasarhelyi, Alles, & Williams, 2010). This comes in contrast with the currently available technology, which can support a more continuous type of assurance. The concept of continuous auditing was first introduced by Groomer & Murthy (1989) and Vasarhelyi & Halper (1991). However, in order for this to succeed, an overall rethinking of the various aspects of auditing is required. For instance, while traditional audit takes a sample-based approach, mostly due to budget and time constraints, continuous auditing examines the whole population of records.

Businesses can benefit from increased automation and technology usage to improve audit efficiency and effectiveness through the implementation of continuous auditing systems. Elliot (1998) argues that companies can decrease the cost of labor associated with audits by taking advantage of technology and computerized systems (Elliott, 1998). Moreover, they can increase the efficiencies of their production (Menon & Williams, 2001).

Another aspect of auditing that needs to be reconsidered is data procurement. Data has to be made readily available to auditors at the time or close-to the time an event occurs. Data standards, such as XBRL and the Audit Data Standard that is newly proposed by the AICPA greatly facilitate fast, close to real-time, and even continuous procurement of data. Such data standards are not only enablers but even requirements for continuous auditing to succeed (Alles, Vasarhelyi, & Issa, 2012). The frequency and nature of tests conducted during an audit engagement must also be revisited to accommodate timelier audits. Examining the whole population in a timelier and more frequent manner increases the likelihood of detecting fraudulent or erroneous transactions as well as internal control violations. These successes will pave the way for traditional auditing to gradually evolve into a more timely form, the continuous audit (Vasarhelyi et al., 2010).

The adoption of continuous auditing has been spreading across various industries, albeit at a slower rate than developments in information technology. The Institute of Internal Auditors and ACL (software developer) conducted a joint survey and found that an increasing number of companies are gaining interest in continuous auditing. The results of the latter survey show that 36% of the companies who responded to the survey have already implemented continuous auditing, while 39% have planned to follow their leads in the near future (Alles et al., 2008). In fact, large companies such as AT&T, Siemens, Procter & Gamble, Itau-Unibanco, Metlife, to name a few, have all reported implementations of continuous auditing systems (Chan & Vasarhelyi, 2011).

PricewaterhouseCoopers conducted a survey in 2006 that concluded that approximately 50% of the companies in the United States have adopted some kind of continuous auditing systems, while 31 % of the remaining half reported planning to implement such systems. Unfortunately, the results of this study show that only 3% of the firms who have implemented or are in the process of implementing continuous auditing have fully automated continuous auditing systems. Instead, most of these firms conduct audits on more frequent basis, but not real-time continuous audits.

2.2. Exceptional Exceptions

Vasarhelyi and Halper (1991) implemented the first known continuous auditing system at Bell Laboratories. This implementation brought to light important issues, such as the quality of data, the optimal frequency of running tests, and the processing of the identified exceptions. Since this first successful implementation, numerous statistical and machine learning techniques and methodologies were proposed in the accounting literature, aiming to provide real-time or close to real time level of auditing (Dull et al., 2006; Kogan et al., 1999; Vasarhelyi & Halper, 1991). The majority of these methodologies use historic data at the transaction level to infer benchmarks (data modeling) against which new transactions are compared at a later stage (data analytics)[1] (Kogan, Vasarhelyi, & Wu, 2010). Alles et al. (2006) discussed the actual implementation of a continuous auditing system in a major multinational company following the continuous assurance architecture that was proposed by Vasarhelyi and

[1] These methodologies are based on the assumption that new data has the same patterns and behave similarly to the historic data used to create the benchmark.

Halper (1991). The main objective of the implementation was to identify exceptions, and the authors reported that the results yielded large numbers of exceptions. Alles et al (2006, 2008) and Debreceny et al. (2003) pointed out the problem of large numbers of identified exceptions associated with such continuous auditing systems. The alarms generated during the identification phase do not undergo any processing before they are sent to the auditors. Consequently, the overall efficiency and effectiveness of such continuous auditing systems is limited by the capabilities of the human users.

Continuous assurance services require performing complex tasks such as data aggregation and analysis. Unfortunately, as mentioned earlier, social sciences literature shows that humans do not perform well such complex tasks. They can be overwhelmed with large amounts of information, and have limited capabilities in collecting and processing information from multiple sources (Iselin, 1988; Kleinmuntz, 1990). As a result, it is crucial to provide a certain level of exceptions processing before presenting them to the human users if we want to take full advantage of continuous auditing systems. A system that can prioritize identified exceptions could greatly increase audit efficiency and effectiveness by drawing auditors' attention to the more suspicious exceptions first. This would allow for timelier reporting, and even addressing, of possible risks.

CHAPTER 3: A PREDICTIVE ORDERED LOGISTIC REGRESSION MODEL AS A TOOL FOR QUALITY REVIEW OF CONTROL RISK ASSESSMENTS

3.1. Introduction

Internal auditors play an important role in improving an organization's operations by providing assurance to 'management and the audit committee that risks to the organization are understood and managed appropriately (IIA, 2002). In fact, the main functions of the internal audit department were expanded after Sarbanes-Oxley Act of 2002 (SOX) to cover the areas of corporate governance, internal controls, and risk assessment (Sarens, 2009). The findings of risk assessment are usually used in the planning of annual audits on a macro level (schedule of annual audits) as well as on a micro level (degree of testing of internal controls) (Allegrini & D'Onza, 2003).

While internal controls have been used by the management to endure operational efficiency before the enactment of SOX in 2002, this act has imposed on companies the requirement of reporting on their internal control systems as well as the quality of the latter. In addition to that, Sox also mandated that external auditors must attest and report on the adequacy of the internal controls as well as the management's assessment itself. This requirement of evaluating internal control risks necessitated the development of new and more effective evaluation tools.

Control Risk Assessment is a widely used tool for the evaluation of business process control risk. It improves the efficiency of internal audit departments by permitting the auditors to develop a better understanding of business and consequently focus their attention on high risk processes (Allegrini & D'Onza, 2003; Wood, Brown, Howe, & Vallabhaneni, 1999). Internal controls-related PCAOB standards emphasize the importance of internal auditing by encouraging external auditors to take their work on internal controls into consideration (PCAOB, 2007). That being said, it is crucial to provide some assurance on the level of quality for internal audit work to be relied on by both management and external auditors.

Prior studies in the continuous auditing literature have shown how advancements in information technology can be utilized to transform assurance services into more timely and continuous services (Dull et al., 2006; Kogan et al., 1999). There is a plethora of information technology techniques in the literature that proved to be effective in detecting exceptions that do not conform to the general pattern in a certain dataset. However, the continuous auditing literature focuses on the detection phase, but fails to discuss the processing of these exceptions (Groomer & Murthy, 1989b; Vasarhelyi & Halper, 1991). This leaves the human users with the tasks of processing and analyzing the exceptions, tasks that social and behavioral research shows humans do not perform well (Iselin, 1988; Kleinmuntz, 1990). This is aggravated by the fact that the exceptions that generally result from the use of information technology techniques are too numerous to be investigated in totality by the users, who are then left with an overload of information beyond their capabilities to examine (Alles et al., 2006, 2008; Debreceny et al., 2003).

This chapter is based on a field study, and consists of two parts. In the first part the author uses a dataset of audit assessments, generated according to the following process. The internal auditors of a multinational corporation use control risk assessment surveys to identify and classify control issues related to business processes, to which they subsequently assign overall risk levels[2]. The second part of this study uses a dataset that is generated in a similar way, where business owners use control risk self-assessments (CRSAs) to evaluate their controls. After identifying weaknesses in their controls, the business owners classify the identified issues, and then classify the business processes associated with these identified issues according to their assessment of the overall risk level. The research questions that this study addresses are: 1) How can the quality of internal auditors' judgment in control risk assessments be reviewed and verified? 2) How can the quality of control risk self-assessments conducted by business owners be reviewed and verified? 3) And how can the exceptions that disagree from the norms be prioritized using a probabilistic model?

The objective of this study is two-fold. First, it uses historic control risk assessment data to evaluate and examine the quality of auditors' judgment in order to provide a tertiary quality assurance tool (after the management and the internal auditors). Netxt, an ordered logistic regression model is inferred from historic data of control issues that were identified and categorized by internal auditors and later on used to assign an overall business process risk level. This model is subsequently used to detect outlying instances where the risk level assigned by the auditor deviates from the norm or the expected value. The same methodology is followed using the CRSA dataset, where historic data are used to evaluate the judgment of the business owners with regards to the control risks. The

[2] The internal auditors evaluate business processes in different locations, and identify and classify the issues they detect, and assess the risk level by process and location.

historic data which consist of identified and classified control issues in addition to the overall risk level of the associated business processes is used to derive an ordered logistic regression model, which is employed in turn to identify all the cases where the business owners' assessment differs from the expected value predicted by the model. The second objective is to prioritize these outliers in order to help the auditors focus their efforts on the cases where the auditor's or business owner's judgments vary from the expected value significantly. Instead of simple binary classification where each case is considered as an exception or not, the proposed methodology examines how far off the auditor's (business owner's) judgment was from the expected value predicted by my model. The disagreement between the assigned and expected risk level is then measured using two metrics in order to rank these outliers based on their deviation from the expected value. The first metric is the ratio of the probability of the assigned risk level to that of the predicted risk level, as calculated by my model. The second disagreement measure is the difference between these two probabilities. The two metrics are cross-verified to ensure the consistency of the results. This method helps prioritize the outliers and focus investigative efforts on the most irregular instances first, advancing from mere detection of outliers towards prioritization of the results.

This study contributes to the literature by providing a methodology for the evaluation of control risk assessments conducted by internal auditors and business owners, which can be used to review the quality of their judgment in assessing controls risk levels. Additionally, it contributes to the continuous auditing literature by showing the added value of going beyond the exceptions' detection step, which has been the norm so far, and addressing the analysis of the detected exceptions by proposing two metrics that can be used to prioritize outlying instances. This will consequently impact overall audit efficiency by focusing auditors' efforts on the more irregular cases. The results indicate that the proposed models can in fact be used for quality review. For instance, the audit assessments model had an accuracy rate of 83% for the fitted model and 76.36% for the predictive model. This indicates that the auditors systematically assigned risk levels. The results of the control risk self-assessments, on the other hand, had the lower accuracy of 74.32% for the fitted model. This is to be expected as the business owners lack the experience of the internal auditors. In fact, the results show that business owners were systematically biased in favor of overestimating the level of risk. According to the feedback that the company provided, the proposed ranking metrics are shown to be effective in improving audit efficiency. A series of robustness tests are conducted to verify these findings.

The remainder of this chapter is organized as follows. Section 2 discusses the literature related to internal controls and quality of internal audits. Section 3 presents the auditors assessment business case utilized to test the proposed methodology. Section 4 describes the model using control risk self-assessments conducted by business owners. Section 5 concludes the chapter.

3.2. <u>Background</u>

3.2.1. Internal Controls

The enactment of Sarbanes-Oxley Act of 2002 has caused numerous changes to the auditing profession. One of the most significant implications of SOX, however, is mandating the reporting on the quality of internal controls. The Committee of Sponsoring Organizations of the Treadway Commission, or COSO, (Committee of Sponsoring Organizations of the Treadway Commission, 1992) defines an internal control as:

> *"A process effected by an entity's board of directors, management and other personnel, designed to provide reasonable assurance regarding the achievements of objectives in the following categories:*
> 1. *Effectiveness & efficiency of operations*
> 2. *Reliability of financial reporting*
> 3. *Compliance with applicable laws and regulations."*

Internal control systems are designed in a hierarchical manner. They consist of several internal control clusters that are associated with the various business cycles (Vasarhelyi, 1980) . These internal control clusters comprise of individual control measures known as internal control procedures (Cushing, 1974), which can be of a preventative, detective, or corrective nature. As the name suggests preventive controls aim to prevent the occurrence of errors and certain undesirable events. Detective controls, on the other hand, identify errors and exceptions after they have occurred. Lastly, corrective controls can be used to examine and correct the identified exceptions.

According to SOX requirements, the management and the external auditors must report on the adequacy of internal controls (Section 404, SOX) (U.S. House of Representatives, 2002). First, the management must report on the presence of an internal control system over financial reporting. Next, they are required to evaluate the system and identify any weaknesses associated with internal controls. Consequently, the management has to report whether their internal control system is adequate and complies with SOX requirements, in other words whether the controls deficiencies are material or not. Both the Chief Executive Officer and the Chief Financial Officer have to sign the report, as failure to do so will incur civil penalties (Section 302, SOX). SOX also requires external auditors to report not only on the existence and adequacy of the company's internal controls over financial reporting, but also on the management's assessment of this system. Subsequently, the external auditors would include their report on the quality of the internal control systems and the management's assessment in the yearly statements. This increased transparency gives external users access to information that used to be privileged to the management in the pre-SOX era.

The increased interest in internal controls in the post-SOX era is evident both in academia and in the auditing profession. The assessment of internal controls has a great impact on the level and scope of substantive testing. The higher the quality of the company's internal controls system, the less substantive testing is required. The problem that arises is the evaluation of the system's quality. The evaluation of an internal control system must take into consideration a myriad of qualitative information that is considered as part of the control environment, such as the company' overall attitude, philosophy and operating style, the functions of the board of directors and its committees, as well as many others (SAS No. 65, 1991). Such qualitative information, while may not directly affect the financial statements, can have a significant impact on the overall internal control system. The lack of adequate objective criteria that can be used to effectively evaluate the quality of internal control system presents a big challenge (Kinney Jr, 2000) that is aggravated by the absence of internal controls evaluation standards and norms. As a result, the most popular approach for internal control systems' evaluation is to resort to auditors' consensus on the quality of the systems (Srinidhi & Vasarhelyi, 1986; Srivastava, Dar, Jagadish, & Levy, 1996). Numerous studies in the literature examined the approach of auditors' agreement on the assessment of internal controls, and it was shown that auditors can achieve a high agreement level of over 60% (Ashton, 1974; Gaumnitz & Nunamaker, 1982; Srinidhi & Vasarhelyi, 1986). Due to the complexity of internal control systems, it becomes even harder to evaluate the overall system without examining the individual controls first (Wu & Hahn, 1989). In this chapter I propose a methodology that can provide an alternative to the auditors' consensus, as it uses the assessment of various components of the internal control system, then groups similar evaluation instances, and uses that as an assessment tool.

3.2.2. Quality of Internal Auditors Work

The functions of the internal audit department are best defined by the Institute of Internal Auditors (IIA) as: *"an independent, objective assurance and consulting activity designed to add value and improve an organization's operations. It helps an organization accomplish its objectives by bringing a systematic, disciplined approach to evaluate and improve the effectiveness of risk management, control, and governance processes"* (IIA, 2000). After the enactment of the Sarbanes-Oxley Act of 2002 following the accounting scandals such as Enron and WorldCom, the role played by internal auditors in organizations shifted focus from mere traditional assurance to consulting and value added services (Bou-Raad, 2000; Nagy & Cenker, 2002). In the immediate period after the implementation of SOX, the internal audit work became mainly oriented towards compliance with SOX and financial reporting (IIA, 2004) (Carcello & Hermanson, 2005). However this was a temporary shift, and there is evidence that the focus of internal auditing shifted again towards effectiveness of business operations and risk management (Protiviti, 2009).

Alegrini & D'Onza (2003) define risk assessment, the step necessary for an effective and efficient risk management, as a process where risk is identified and evaluated, and subsequently prioritized. They argue that risk assessment affects two stages or levels in the process of planning internal audits. Risk assessment is involved at a higher level during planning of the schedule of annual audits. It is also involved at a lower level during the planning of individual audit engagements, as suggested by the COSO model (Allegrini & D'Onza, 2003; Roth, 1995).

During the evaluation of the internal control systems' effectiveness, auditors gather and analyze large amounts of evidence and information. However, the complex relationships between internal controls variables make it difficult to decide on the amount of information needed for a proper assessment of internal control systems. To this complexity is added the lack of properly and clearly stated rules to help the auditors make that decision, where only general guidelines can be found (Davis & Massey, 1997). One of the widely used tools by many organizations is control risk assessments surveys (CRA), generally utilized as a way to determine and evaluate internal control risks. Wood et al. (1999) argue that control risk assessments enable auditors to broaden their coverage by focusing on riskier areas. The final decision on the amount of information necessary for the internal control systems assessment is in most cases based on the experience and judgment of the auditors themselves (Calderon & Cheh, 2002). Therefore, it is crucial to maintain a certain level of quality assurance of this judgment, in order for the work of internal auditors to be reliably taken into consideration by both external auditors and management. Despite the fact that the audit standards require external auditors to conduct the majority of the audit work, Auditing Standard No. 5 (and Auditing Standard No.2 before that) encourages the external auditors to use the work of internal auditors, conditional on its quality (PCAOB, 2004, 2007). The higher the quality of work conducted by the internal auditors, the more reliable and usable it is for the external auditors. Moreover, if the latter find the internal auditing function to be of a low quality, which in itself may constitute a material weakness in internal controls under Auditing Standard No. 5, the external auditors may be inclined to issue an adverse opinion on internal controls over financial reporting (PCAOB, 2004, 2007). The quality of internal audit work may affect external audits in three stages (Gramling, 2006). The first stage is risk assessment phase. The second stage affected by internal audit function's quality is when the external auditors understand, document, and test the internal controls system. Finally, the amount of substantive testing is also affected by the quality of internal audit work (Gramling, 2006).

The abundance of research that emphasizes the importance and effect of the quality of internal audit function necessitates that we understand the characteristics of a high-quality audit function. Both external as well as internal auditing standards provide some guidance in that regards. Statements on Auditing Standards 65 (SAS 65) lists competence (such as professional certificates, level of education, supervision of internal auditors' work), quality of work performance, and objectivity (policies regulating the relationship between auditors and auditees, reporting line) as the main characteristics of the quality of internal audit function (SAS No. 65, 1991). The characteristics of the internal audit quality as described by the IIA standards consist of independence, proficiency, due care, and objectivity. There is no doubt that a methodology that can verify the quality of internal audit work in the internal controls and evaluate the auditor's risk assessment would be very helpful. This study presents such a methodology.

3.3. Business Case 1: Audit Assessments (CRA by Auditors)

3.3.1. Data Generation Process

The business case in this section is based on data provided by the internal audit department of a multinational corporation. The collection of the data is conducted using control risk assessment surveys by location and for each business process. First the internal auditors examine a certain location, for example New York. Next they identify issues related to different business processes, and classify each issue as critical, major, or non-major based on the materiality thresholds set by the company. The auditors consider all the issues related to a business process, such as accounts payable, assess its overall risk, and finally classify the business process as low, medium, or high risk based on the identified issues and their severity.

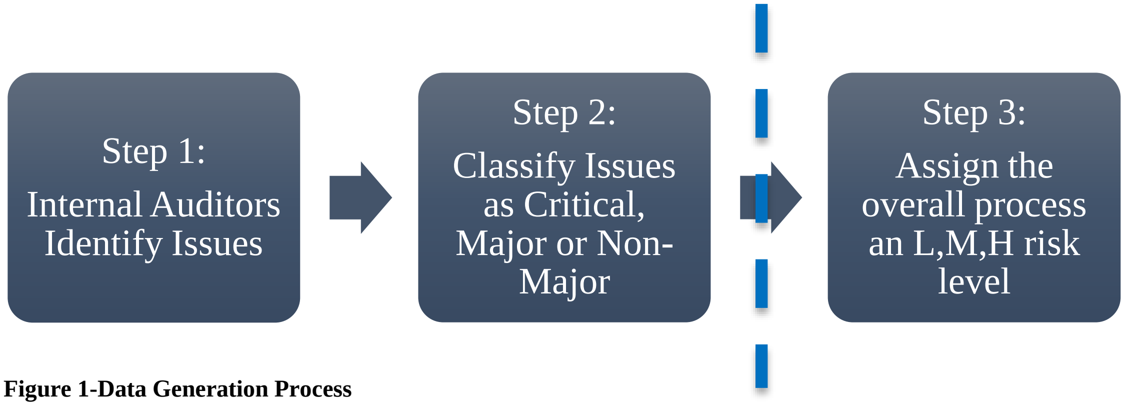

Figure 1-Data Generation Process

For example, the auditors visit the location of a chemical plant in New Jersey. They examine the controls related to various processes. As an example I will consider the Manufacturing process and the six sub-processes that it consists of, namely Cost Accounting, Planning, Plant Finance, Production/Warehouse, Storeroom, and Category Product Supply. In Step 1 the auditors will assess the controls of these sub-processes and identify any related weaknesses and issues. Next they categorize these issues in Step 2 as critical, major, or non-major issues according to their materiality and some other criteria. In Step 3 the auditors consider all the identified issues related to these six sub-processes and assign the Manufacturing business process an overall risk level for that chemical plant located in New Jersey.

3.3.2. Data Description

The model in this chapter emulates the auditors' judgment in Step 3 in Figure 1 and predicts the overall risk level of a certain business process based on the identified issues. I use a dataset that consists of the identification and categorization of audit issues for the sub-processes of each process as identified by the internal auditors. The dataset also includes the overall business process risk level assessments. The dataset is provided by the internal auditing department of a major multinational corporation, and covers the period extending from fiscal year 2008-2009 (FY08/09) to 2010-2011 (FY10/11). The locations cover several countries, and the business processes include (but are not limited to) revenue, accounts payable, distribution, computer installations, fixed assets, payroll, and purchasing.

During business processes audits, the internal auditors first identify audit issues and then classify them into three categories, namely non-major, major, and critical[3]. Subsequently, the auditors use the information they procure from the analysis of these issues to categorize the overall business process risk level as low, medium, or high. This procedure yields two separate files – one containing the issues identified during the audits while the other containing the overall business process risk assessment scores – which I combine in this study in order to obtain the final dataset that is analyzed and used to infer the ordered logistic regression model.

The dataset underwent several transformations. First, I aggregate the issues, originally recorded individually, to produce total counts per audit work-paper. Second, I transform issues scores and overall scores from text to ordinal numerical values, a format appropriate for the ordered logistic regression models used in this chapter and discussed in the following section. I then group the records by fiscal years, resulting in 344, 305,

[3] The three classes of identified issues are according to the materiality of estimated financial impact as specified by the company's policy. However other factors play a role in the classification process, which include the control objective, actions already taken to address the issue, repetition of the issue (i.e. issue identification in previous audits), as well as others.

and 275 records for the fiscal years 08/09, 09/10, and 10/11, respectively. The breakdown of the records in the dataset can be seen in Table 1.

Table 1-Breakdown of Audit work papers by fiscal year

	FY 08/09	FY 09/10	FY 10/11	Total
High Risk	31	16	17	**64**
Medium Risk	126	109	93	**328**
Low Risk	187	180	165	**532**
Total	**344**	**305**	**275**	**924**

The breakdown of the issues and their classification is shown in Table 2.

Table 2-Breakdown of Audit work papers Issues by fiscal year

	FY 08/09	FY 09/10	FY 10/11	Total
Critical Issues	13	1	2	**16**
Major Issues	294	216	219	**729**
Non-Major Issues	717	657	564	**1938**
Total	**1024**	**874**	**785**	**2683**

3.3.3. Methodology

3.3.3.1. Ordered Logistic Regression

Ordered logistic regression model is a non-linear model that can be applied in cases where the dependent variable has two or more nominal values that have a sequential order. The distribution of a Logit model is of a cumulative standard logistic type, where the events of the ordinal logistic regression model are cumulative rather than individual values. It is worth mentioning that an alternative to the ordered logit model is the ordered probit one, both yielding very similar results (Dijk & Pellenbarg, 2000).

One of the main objectives of this study is to prioritize detected exceptions, and the use of this probabilistic model enabled me not only to calculate the likelihood of each instance agreeing or disagreeing with the expected value, but to rank them depending on the level of disagreement. Moreover, the fact that my datasets contain the output values in the form of controls risk assessment scores allowed me to use supervised learning techniques. In addition to that, the values of the variables are categorized into three ordinal values. Ordered logistic regression fits both my purpose and my datasets perfectly. It is noteworthy to mention that although the values are ranked for the dependent and independent variables, the actual distance between these values cannot be measured. Going from Low risk to Medium risk does not necessarily involve the same change as moving from Medium to High.

The ordered logistic regression model is based on the following equation (Alpaydin, 2004):

Equation 1

$$\boldsymbol{y_i^* = logit = ln\left(\frac{prob(event)}{1 - prob(event)}\right) = \beta^T x_i + \varepsilon_i}$$

Where $\boldsymbol{y_i^*} = \textbf{logit} =$ log of the odds that a certain event takes place.

$\boldsymbol{\beta^T}$ is the vector of coefficients

$\boldsymbol{x_i}$ represents the vector of independent variables (with $\boldsymbol{x_0 = 1}$)

and $\boldsymbol{\varepsilon_i}$ is the disturbance term.

Normally the **logit** values, or $\boldsymbol{y_i^*}$, are not directly observable. Therefore we observe $\boldsymbol{y_i}$ such that

$$\boldsymbol{y_i = 0 \quad for \quad y_i^* \leq \mu_0}$$

$$\boldsymbol{y_i = 1 \quad for \quad \mu_0 < y_i^* \leq \mu_1}$$

$$\boldsymbol{y_i = 2 \quad for \quad \mu_1 < y_i^* \leq \mu_2}$$

$$\vdots$$

$$\boldsymbol{y_i = i \quad for \quad \mu_{i-1} < y_i^* \leq \mu_i}$$

where $\boldsymbol{\mu_i}$ are unknown parameters that are estimated using $\boldsymbol{\beta^T}$. While $\boldsymbol{y_i}$ shows us the log of the odds that a certain event takes place, the coefficients $\boldsymbol{\beta^T}$ inform us of the changes of the logit $\boldsymbol{y_i}$ that are caused by the independent variables (Kleinbaum & Klein, 2010; Norris et al., 2006; P. Warner, 2008).

3.3.3.2. Model

The model in this study is based on the ordered logistic regression discussed in the previous section. I hypothesize that the control risk assessment score is a function of the total counts of the three classes of audit issues. In other words, the dependent variable is the controls risk assessment score, while the independent variables are the aggregated counts of the three classes of issues, as shows in the following equation:

Equation 2

$$\boldsymbol{logit = ln\left(\frac{prob(event)}{1 - prob(event)}\right) = \beta^T x_i + \varepsilon_i = \beta_0 + \beta_1 CC + \beta_2 MC + \beta_3 NMC}$$

With **logit** = log of the odds that a certain event takes place.

$\boldsymbol{\beta_0}$ = Intercept

$\boldsymbol{\beta_i}$ = Coefficient

CC = Number of critical issues (identified by the auditor)

MC = Number of Major issues (identified by the auditor)

NMC = Number of Non-Major issues (identified by the auditor)

I estimate the values of $\boldsymbol{y_i}$ according to the following:

$$\boldsymbol{y_i = 0 \quad for \quad y_i^* \leq \mu_L \quad (Low\ risk)}$$

$$\boldsymbol{y_i = 1 \quad for \quad \mu_L < y_i^* \leq \mu_M \quad (Medium\ risk)}$$

$$\boldsymbol{y_i = 2 \quad for \quad \mu_M < y_i^* \leq \mu_H \quad (High\ risk)}$$

To test this hypothesis, this study used historic data from the previous two fiscal years to estimate the coefficients, which were later used to predict the scores for the observations from the subsequent year. Data from fiscal years 2008 through 2010 were used to calculate the coefficients, which were in turn helped calculate the probability for each observation from fiscal year 2010/2011 falling in each of the three classes of the dependent variable using the following formula (Alpaydin, 2004):

Equation 3

$$\boldsymbol{PredProb = \widehat{P}(C_i|x) = \frac{1}{1 + e^{-(\beta^T x_i + \varepsilon_i)}}}$$

where $\boldsymbol{\beta^T}$ is the vector of coefficients

$\boldsymbol{x_i}$ represents the vector of independent variables (with $\boldsymbol{x_0 = 1}$)

and $\boldsymbol{\varepsilon_i}$ is the disturbance term.

To illustrate how this formula is used, let us consider the following example. In order to calculate the predicted probability of record *123456* falling in the High Risk category, I use the intercept of the simple logistic model that separates High Risk Class from the other classes (Intercept_2 in the formula below) and the three coefficients (CC_Coeff, MC_Coeff, NMC_Coeff) which correspond to the three independent variables (CC, MC, NMC), respectively[4].

The formula used to calculate the predicted probability of record *123456* belonging to the High Risk Class (***Calc_H***) is:

$$\boldsymbol{Calc\_H} = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{-\{(\boldsymbol{Intercept\_2}+(\boldsymbol{CC\_Coeff}*\boldsymbol{CC})+(\boldsymbol{MC\_Coeff}*\boldsymbol{MC})+(\boldsymbol{NMC\_Coeff}*\boldsymbol{NMC})\}}}$$

The same procedure is followed to calculate the predicted probability of that record falling into the Medium risk class, and the final formula is:

$$\boldsymbol{Calc\_M} = \left(\frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{-\{(\boldsymbol{Intercept\_1}+(\boldsymbol{CC\_Coeff}*\boldsymbol{CC})+(\boldsymbol{MC\_Coeff}*\boldsymbol{MC})+(\boldsymbol{NMC\_Coeff}*\boldsymbol{NMC})\}}}\right) - \boldsymbol{Calc\_H}$$

[4] Critical issues count, Major issues count, and Non- Major issues count are represented in the formula as CC, MC, and NMC respectively. They represent the total count of critical (Major, and Non-Major) issues identified for each record.

The coefficients of the variables used are identical for both classes, with only the intercepts changing in the formulas. Due to the cumulative nature of the logit model, the part of the formula that is between brackets calculates the cumulative probability of the High and Medium risk classes, where the intercept of the simple logistic model that separates High risk and Medium risk classes from the Low Risk Class is used (Intercept_1 in the formula above). Therefore, one has to subtract the probability of High Risk class from the part in brackets to calculate the probability of the Medium risk class (***Calc_M***).

To calculate the predicted probability of the last class (Low risk) I exploit the probabilistic nature of this model, and use the following:

$$\boldsymbol{Calc\_L = 1 - Calc\_H - Clac\_M}$$

The class that has the highest calculated probability is then considered the ***predicted class***. A numerical example is provided in the following section.

3.3.4. Outliers' Ranking and Prioritization

The predicted values obtained in the previous step are compared to the assigned scores in order to identify outliers. Any record that is scored differently from the predicted score is classified as outlier. In other words, a record is an outlier if the predicted class differs from the true or assigned class.

In order to improve the efficiency of the model and focus my efforts on the most suspicious records, these outliers are ranked according to two criteria: ratio and difference.

The ratio of the calculated probability of the predicted class to that of the assigned class is then calculated and sorted in ascending order, based on the following equation:

Equation 4

$$\boldsymbol{Ratio = \frac{Calc.Class_{Assign}}{Calc.Class_{Pred}}}$$

where $\boldsymbol{Calc.Class_{Pred}}$ and $\boldsymbol{Calc.Class_{Assign}}$ represent the calculated probabilities of predicted class and assigned class, respectively.

A ratio of one indicates the predicted and assigned classes are the same. The lower the ratio, the wider the gap between the predicted and assigned scores is.

The other metric that I use for ranking is the difference between the probabilities of the predicted and assigned classes. The equation used to calculate this metric is:

Equation 5

$$\boldsymbol{Difference = Calc.Class_{Pred} - Calc.Class_{Assign}}$$

In this case, the bigger the difference, the less the predicted and assigned classes agree.

The final step in my ranking procedure involves cross-checking the two rankings of the outliers as a means of prioritization. This allows for the identification of the extreme or most suspicious outliers in the dataset. Consider the two sample records in Table 3.

Record *123456* has zero critical issues (CC), two major issues (MC), and three non-major issues (NMC). I used the model and the coefficients estimated in the previous step to calculate the probability of this record falling in the three classes (Calc_H for High risk, Calc_M for Medium risk, and Calc_L for Low risk), and I find it to be 0.60719, 0.39195, and 0.00086 for classes H, M, and L, respectively. As the highest probability is that of class H, the record's predicted class is H. However, the auditors had assigned record *123456* a medium risk level, i.e., categorized it as class M (Assign. Class). Therefore we need to measure the disagreement level between the model and the auditor's judgment.

Table 3-Numerical Example of Calculating the Predicted Probabilities and Ranking of records

Record	**CC**	**MC**	**NMC**	**Calc_H**	**Calc_M**	**Calc_L**	**Assign. Class**	**Pred. Class**	**Ratio**	**Diff.**
123456	0	2	3	0.60719	0.39195	0.00086	M	H	0.64551	0.21524
123457	0	1	1	0.001508	0.52778	0.47071	L	M	0.89186	0.05708

Where **CC =** Critical Issues Count

MC = Major Issues Count

NMC = Non-Major Issues Count

Calc_H = the calculated expected probability for the record to be in the High Risk Class

Calc_M = the calculated expected probability for the record to be in the Medium Risk Class

Calc_L = the calculated expected probability for the record to be in the Low Risk Class

Assign. Class = the audit score (risk level) that was assigned by the auditor for that record

Pred. Class = the audit score (risk level) that was predicted by my model for that record

Ratio = the ratio disagreement metric

Difference = the difference disagreement metric

First the ratio and difference are calculated based on the equations above:

$$Ratio = \frac{0.39195}{0.60719} = 0.64551$$

and

$$\text{Difference} = 0.60719 - 0.39195 = 0.21524$$

The same procedure is applied to record *123457*, and the ratio (difference) is found to be 0.89186 (0.05708), which is lower (higher) than that of record *123456*, indicating that the latter is more suspicious.

This procedure is repeated for all the outliers, who are ranked in increasing order when the ratio metric is used (the lower the ratio, the more anomalous that record is), and in decreasing order when the difference metric is utilized (the larger the difference, the more anomalous the record is). Finally the ranked outliers from both metrics were cross-checked to determine the most suspicious records.

3.3.5. Findings

I first tested the goodness of fit of the overall model. The p-values of less than 0.0001 in Table 4 suggest that the null hypothesis that the presence of the predictive variables in the model does not make any change can be safely reject. In other words, the model with the independent variables explains more than a simple intercept-only model.

Table 4-Test for the Null Hypothesis

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	618.6570	3	<.0001
Score	396.8161	3	<.0001
Wald	216.4014	3	<.0001

The results are then analyzed for parallelism, or the possibility of the coefficients for all the independent variables being the same. Results confirm the appropriateness of the methodological choice to use ordinal logistic regression for the purpose of this study (Norris et al., 2006). Table 5 presents the results of the maximum likelihood estimates analysis.

Table 5-Analysis of Maximum Likelihood Estimates

Analysis of Maximum Likelihood Estimates				
Parameter	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept 2	-9.5204	0.6339	225.5851	<.0001
Intercept 1	-2.9076	0.2255	166.2869	<.0001
CC	8.8490	1.4352	38.0138	<.0001
MC	2.5843	0.1842	196.8556	<.0001
NMC	0.4406	0.0620	50.5446	<.0001

As expected, the variable Critical issues (CC) has a high Standard Error compared to the other two independent variables. This is likely due to the fact that it occurred less frequently with only 16 occurrences, compared to major issues (MC) and non-major issues (NMC) with 729 and 1938 occurrences, respectively.

The results from the AS model indicated that internal auditors assigned the scores systematically, as the accuracy of the model is 83% on average. In other words, only 17% of the instances deviated from the scores predicted by the model. However, those 17% of the instances were in certain cases on the borderline of being outliers. In such cases, the likelihood of such a case being of low risk or medium risk was within 5%, indicating that the case could in fact be either one. This high accuracy is to be expected due to the expertise of the auditors. That said, the model can be used as a quality review tool, where any deviation from its predictions may invoke further investigation.

As mentioned before, the outliers were ranked based on the ratio and difference criteria, and the results show that the top 20 outliers were always the same, regardless of the ranking metric that was used. The top 20 suspicious observations were sent to the internal audit department of the multinational corporation for further investigation.
Notwithstanding the legitimacy of many of these outliers, the internal auditors found the model to be a useful tool to focus their effort and increase audit efficiency, as it provided them with a means to identify extreme outliers. In fact, it was sufficient in many cases simply to ask the score approver to explain the nonconformity with the model.

In order to check the consistency and robustness of the model, the sliding window technique was followed. Data from fiscal year 08/09 was first used to estimate the coefficients, which were used to predict the scores of the observations in fiscal years 09/10 and then 10/11. The second model used the records in 09/10 to calculate the coefficients and predict the scores of the instances in 10/11. The results were very close to the original model in this study. Although the coefficients were slightly different, the top 20 outliers were the same.
The predictive power of the overall model, however, decreased slightly with an accuracy of 76.36%. Confusion matrices for both fitted and predictive models are presented in Tables 6 and 7. Other stages of the sliding window technique yielded similar results, emphasizing the consistency and robustness of my model.

Table 6 presents the confusion matrix for the fitted model. This matrix is for the model that uses the data from fiscal years 2008/2009 and 2009/2010 to determine the coefficients, and then uses the same two year data to test the model. The diagonal cells present the true positives, i.e. the cases where the auditors' judgment conformed to my model's expectation. The other cases present the outlying cases where the auditors' assessment disagreed with the predicted value. There are two levels of outliers in this study. One-level outliers are the instances where the predicted and assigned values differ by one level, such as Low-Medium, Medium-High, or vice versa. The second level of outliers is the extreme outliers where the auditors' assessment differed from the model's expected value by two levels, in other words Low-High and High-Low.

The diagonals show that 83% of the instances were predicted and assigned to be of the same level, as shown in Cells L-L, M-M, and H-H. This nomenclature follows the pattern P-A, which stands for Predicted-Assigned. The first cell, designated L-L, shows that 88.38% of the instances that were predicted to be of Low Risk were classified by the auditors to be in the same risk class. The cell M-M shows that the auditors agreed on classifying 75.72% of the cases that were predicted by the model to be Medium risk. The last cell (H-H) shows 77.78% agreement between the model and the auditors' classifications of High risk cases.

The remaining cells correspond to the outliers, where auditors' judgment deviated from the expected value. We can see from Table 6 that there were no extreme outliers in the dataset, as shown in cells H-L and L-H. Cell (L-M) shows that 11.62% of the cases that were predicted to be Low Risk were assessed as Medium Risk by the auditors. On the other hand, 16.46% of the predicted Medium Risk cases were classified as Low risk by the auditors (cell M-L). Cell M-H shows 7.82% disagreement corresponding the instances predicted to be of Medium Risk but assigned a High Risk level by the auditors.

Finally, 22.22% of the cases that were expected to be of High Risk were in fact classified by the auditors as medium risk, as shown in cell H-M.

Table 6-Confusion Matrix (M0810-D0810)-Fitted Model

Confusion Matrix-Fitted Model				
Predicted Level	**Assigned Level**			
	L	**M**	**H**	**Total**
L	327 88.38%	43 11.62%	0 0.00%	**370**
M	40 16.46%	184 75.72%	19 7.82%	**243**
H	0 0.00%	8 22.22%	28 77.78%	**36**
Total	**367**	**235**	**47**	**649**

The confusion matrix of the predictive model is presented in Table 7. This model uses the data from fiscal years 2008/2009 and 2009/2010 to infer the coefficients then predict the probabilities for the data from fiscal year 2010/2011.

The diagonal cells that represent the true positives show an overall accuracy level of 76.36%. According to cell L-L, 85.53% of the time the auditors' judgment of Low Risk cases conformed to the model's expectation. Cell M-M, on the other hand, shows a lower percentage, as the auditors assessments agreed with only 68.24% of the cases predicted to be of Medium Risk by the model. This agreement becomes even lower with high risk cases, where the agreement rate drops to 50% (cell H-H).

As for the outliers, the results show the existence of three extreme outliers where the model predicted these instances to fall in the Low Risk class but the auditors assigned them a High Risk score. These records were especially interesting as they were classified as high risk by the internal auditors, but low risk by the predictive model. After further investigations, the company's internal auditors explained that due to fraud-related privacy concerns, the issues associated with these records were not documented. This is an example of the usefulness of the methodology proposed in this study, which was able to detect such instances that would have likely passed undetected. It is also clear that the low-medium and medium-low outliers occur with approximately the same frequency.

Of the cases that the model predicted to be Low Risk, the auditors assigned 14.71% to be of Medium Risk and 1.76% to be of High Risk (cells L-M and L-H, respectively). As for the instances with expected value of Medium Risk, the auditors classified 27.05% of them as Low risk and 4.71% as High Risk. The last row presents the cases where the model predicted the risk level to be high. While the auditors did not classify any of these cases as Low Risk, they assigned 50% of these cases a Medium Risk level.

Table 7-Confusion Matrix (M0810-D1011)-Predictive Model

Confusion Matrix-Predictive Model

Predicted Level	Assigned Level			
	L	**M**	**H**	**Total**
L	142 83.53%	25 14.71%	3 1.76%	**170**
M	23 27.05%	58 68.24%	4 4.71%	**85**
H	0 0.00%	10 50%	10 50%	**20**
Total	**165**	**93**	**17**	**275**

3.4. Business Case 2: Control Risk Self-Assessment (CRSA)

3.4.1. Data Generation Process

The second part of this chapter follows the same procedure as the audit assessment section but differs in the used data. Instead of internal auditors, this sections uses data that were collected by business owners using control risk self-assessments. Business owners at each location evaluate the controls of their business processes using self-assessment surveys. They start by identifying controls issues related to the sub-processes of each business process. Subsequently, they categorize these issues as critical, major, or non-major according to certain criteria, most notably materiality. Based on all the issues associated with the business process, the business owners classify the process as low, medium, or high risk. Another difference between the two parts is the frequency of such control risk assessments. Due to budget and time constraints, internal auditors can examine the controls in each location less frequently than the business owners, who tend to conduct such assessments at a more frequent basis. The data generation is illustrated in Figure 2.

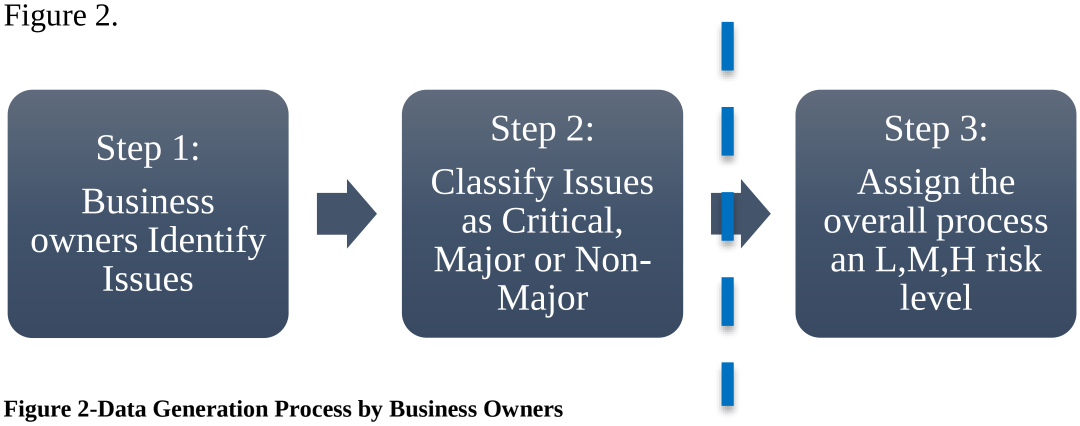

Figure 2-Data Generation Process by Business Owners

3.4.2. Data Description

The dataset used in this section is similar in structure to the one used in the auditors' assessment part. It includes issues identified and categorized by business owners, in addition to the location, business units, processes and sub-processes, as well as the overall risk level. This part of the study uses a dataset that extends over three fiscal years. Because of the nature of the data generation process, the dataset comes from two separate sources. Business owners document the issues that they find with various business processes in a file, and then use that information to assign the overall risk scores, which are stored in a different file. The logit model is inferred from a joint file that joins the information from these two data sources. This dataset comprised of a total of 9593 records. The difference in the number of observations in the two datasets is expected as controls risk assessments are conducted more frequently by business process owners than by internal auditors. The final breakdown of the risk levels of the business processes and the corresponding identified issues can be seen in Tables 8 and 9, respectively.

Table 8-Breakdown of Control Self Assessments Data by fiscal year

	FY 08/09	FY 09/10	FY 10/11	Total
High Risk	87	62	82	231
Medium Risk	1144	1137	1119	3400
Low Risk	2079	1939	1944	5962
Total	3310	3138	3145	9593

Table 9-Breakdown of the CSA Issues by fiscal year

	FY 08/09	FY 09/10	FY 10/11	Total
Critical Issues	11	15	7	33
Major Issues	821	787	1003	2611
Non-Major Issues	4478	4726	5520	14724
Total	5310	5528	6530	17368

3.4.3. Methodology

3.4.3.1. Model

This section follows the same methodology described in section 3.3.2 and infers an ordered logistic regression model from historic control risk self-assessment scores. Once again the hypothesis is that control risk self-assessment score is a function of the three classes of risk levels. Equation 6 presents the logit model used to determine the coefficients:

Equation 6

$$\boldsymbol{logit = ln\left(\frac{prob(event)}{1 - prob(event)}\right) = \beta^T x_i + \varepsilon_i = \beta_0 + \beta_1 CC + \beta_2 MC + \beta_3 NMC}$$

With **logit** **=** log of the odds that a certain event takes place.

$\boldsymbol{\beta_0}$ **=** Intercept

$\boldsymbol{\beta_i}$ = Coefficient

CC = Number of critical issues (identified by the auditor)

MC = Number of Major issues (identified by the auditor)

NMC = Number of Non-Major issues (identified by the auditor)

As the actual logit values cannot be observed directly, I estimate the values of $\boldsymbol{y_i}$ such that:

$$\boldsymbol{y_i = 0 \quad for \quad y_i^* \leq \mu_L \quad (Low\ risk)}$$

$$\boldsymbol{y_i = 1 \quad for \quad \mu_L < y_i^* \leq \mu_M \quad (Medium\ risk)}$$

$$\boldsymbol{y_i = 2 \quad for \quad \mu_M < y_i^* \leq \mu_H \quad (High\ risk)}$$

Control risk scores from fiscal years 2008/2009 and 2009/2010 are used to infer the model, and that information is utilized to predict the risk levels for the business processes from fiscal year 2010/2011 using the number of corresponding issues and their categories. The predicted probabilities were calculated according to the formula mentioned before:

$$\boldsymbol{PredProb} = \widehat{\boldsymbol{P}}(\boldsymbol{C_i}|\boldsymbol{x}) = \frac{\mathbf{1}}{\mathbf{1}+\boldsymbol{e}^{-(\beta^T x_i+\varepsilon_i)}} \qquad \textbf{(Y)}$$

where $\boldsymbol{\beta^T}$ is the vector of coefficients

$\boldsymbol{x_i}$ represents the vector of independent variables (with $\boldsymbol{x_0 = 1}$)

and $\boldsymbol{\varepsilon_i}$ is the disturbance term.

Once the probabilities of each record belonging to each of the three risk levels, that record is assigned to the predicted class whose probability is the highest for that record.

3.4.4. Outliers Ranking and Prioritizations

The two disagreement measures discussed earlier were also used to prioritize the outliers where the class assigned by the business owners deviated from the predicted class. The first metric is the ratio of the probability of the assigned class to that of the predicted class. When this metric increases, the two values come closer, and the disagreement decreases. On the other hand, when the ratio decreases, the business owner’s judgment grows further away from the model’s expected value. The second metric is the difference between the probabilities of the assigned and predicted classes. This metric acts in the opposite direction to the ratio, in other words the bigger the difference, the bigger the disagreement between business owners’ assessment and my model. These two disagreement measures are calculated for all the exceptions, and are subsequently used to sort the exceptions in decreasing order of disagreement. The sorted list of exceptions is then provided to the auditors for evaluation. I argue that this methodology increases audit efficiency by allowing the auditors to address the cases that are furthest from the expected value first.

3.4.5. Findings

Similar to the Auditors' Assessment part of the study, the null hypothesis (β=0) tests were conducted. Table 10 shows the results of testing the joint effect of the independent variables that I have included in the model, namely issues counts of the three categories. The small p-values support the rejection of the null hypothesis that all β are equal to zero.

Table 10-Test for the Null Hypothesis

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	2314.0419	3	<.0001
Score	1713.4213	3	<.0001
Wald	1296.3287	3	<.0001

The results of the analysis of maximum likelihood estimates (Table 11) indicates that all the independent variables are statistically significant, as the p-values for all variables are less than 0.0001. In this part of the study, as in the auditors' assessment part, the standard error for the critical issues counts is higher than the two other variables. Once again this is to be expected because the occurrence rate of these critical issues is much lower than those of the other variables. However, the higher coefficient of the critical issues count indicates a stronger effect.

Table 11-Analysis of Maximum Likelihood Estimates

Analysis of Maximum Likelihood Estimates

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 2	-6.8194	0.1737	1541.7570	<.0001
Intercept 1	-1.4294	0.0381	1405.3668	<.0001
CC	6.2924	1.3193	22.7495	<.0001
MC	1.3292	0.0564	554.6390	<.0001
NMC	0.4786	0.0171	784.8168	<.0001

While internal auditors assigned scores systematically with 83% accuracy, only 74.32% of the CRSA's scores were consistent with the fitted model. This can be seen from the diagonal cells of the confusion matrix of the fitted model as presented in Table 12. The business owners assigned a Low Risk score to 74.38% of the cases predicted to be of Low Risk as can be seen in Cell L-L. They also classified 76.42% of the Medium Risk instances in accordance to the fitted model (Cell M-M). The lowest conformity to the expected value was the High Risk cases, where only 45.45% of the instances with a High Risk predicted level were also classified as such by the business owners (Cell H-H).

Unlike the auditors' assessment model, 18 extreme outliers were identified by the CRSA model, in other words 0.279% of the total population. As previously mentioned, extreme outliers are the instances where the disagreement between the assigned and predicted risk levels is of two levels. Out of these 18 outliers, 17 were expected to be Low Risk but were assigned High Risk levels by the business owners, and that constituted 0.33% of the predicted Low Risk cases (Cell L-H). On the other hand, only one H-L extreme outlier was found (Cell H-L), which amounted to 1.01% of the cases predicted to fall in the High Risk class. The remaining cells in the confusion matrix describe the one-level outliers. Cell L-M indicates that 25.29% of the instances that were expected to be of Low Risk according to the model were in fact classified by the business owners as Medium Risk. On the other hand, this disagreement decreased to 16.43% and 7.15% for the cases expected by the model to be of Medium Risk but assigned by business owners a level of Low Risk and High Risk, respectively (Cells M-L and M-H, respectively). The highest rate of disagreement was 53.54%, and that was for the cases that were expected to be High Risk but were evaluated by the business owners as Medium Risk (Cell H-M).

Table 12-Confusion Matrix-Fitted Model

Confusion Matrix-Fitted Model				
Predicted Level	**Assigned Level**			
	L	**M**	**H**	**Total**
L	3817 74.38%	1298 25.29%	17 0.33%	**5132**
M	200 16.43%	930 76.42%	87 7.15%	**1217**
H	1 1.01%	53 53.54%	45 45.45%	**99**
Total	**4018**	**2281**	**149**	**6448**

Table 13 presents the confusion matrix for the control risk self-assessments predictive model. The data used for the estimation of the coefficients were from the fiscal years 2008/2009 and 2009/2010. The model was used to predict the risk level of the cases from the fiscal year 2010/2011.

Surprisingly, the overall accuracy of the predictive model was found to be slightly higher than the fitted model. This could be a sign of a learning process, especially that the fiscal year 2008/2009 was the first year the company utilized control risk self-assessments on a large scale. Therefore, a plausible explanation is that the business owners utilized control risk self-assessments more effectively as they gained experience using them over the years. As it can be seen from the diagonal cells (L-L, M-M, H-H) the accuracy is 76.5%. The business owners showed high agreement with the cases that were predicted to be Low Risk and Medium Risk, with conformity levels of 77.76% and 76.63%, respectively (Cells L-L and M-M, respectively). On the other hand, this agreement level dropped to 37.97% with regards to the High Risk level as depicted by Cell H-H.

The total number of extreme outliers from the predictive model are much lower than their counterparts from the fitted model. There are five extreme outliers in total, amounting to 0.159%. For instance, there are only two outliers (i.e. 0.09%) that were predicted to be of Low Risk but were assigned by the business owners a High Risk score (Cell L-H). On the other hand, I found three outliers (3.8%) that were predicted to be High Risk, yet they were classified by the business owners as Low Risk (Cell H-L).

As for the remaining outliers, the results were similar to the fitted model. The business owners classified 22.15% of the instances that were predicted to be of Low Risk level as Medium Risk (Cell L-M). On the other hand, the agreement level was 16.46% for the cases predicted to be Medium Risk but assigned a Low Risk level by the business owners and 6.92% for those classified as High Risk (Cells M-L and M-H, respectively). This agreement level went down to a low value of 58.23% for the High to Medium outliers (Cell H-M).

Table 13-Confusion Matrix-Predictive Model

Confusion Matrix-Predictive Model				
Predicted Level	**Assigned Level**			
	L	**M**	**H**	**Total**
L	1822 77.76%	519 22.15%	2 0.09%	**2343**
M	119 16.46%	554 76.63%	50 6.92%	**723**
H	3 3.80%	46 58.23%	30 37.97%	**79**
Total	**1944**	**1119**	**82**	**3145**

The results from both confusion matrices indicate that the lowest level of agreement between the business owners and the predictive model are the ones that are classified as

Medium Risk by the former and High Risk by the latter. As we recall from Tables 6 and 7, this was also the case in the first part of this chapter, i.e. the auditors' assessment. A plausible explanation is the reluctance of both auditors and business owners to assign high risk levels due to the possible ramifications of such classification. It is worth noting, however, that the incentives of the two parties are very different. While the auditors are responsible for reporting the risk levels and conducting substantive testing accordingly, their responsibility ends at that point. On the other hand, business owners have a lot more at stake. They are not only responsible for reporting these issues, but also held accountable for the actual control weaknesses. In fact, they will be required to correct these weaknesses with all the costs incurred by such correction.

The results also indicate a plausible systematic bias to overestimate the risk level by business processes owners who lack the auditing experience. Moreover, I believe they tend to overemphasize the risk associated with identified issues for fear of any bad surprises. One of the explanations could be that the business owners may overestimate risk levels out of conservatism, in the same way managers may resort to overbudgeting.

3.5. Conclusion

The work of internal auditors in the assessment of internal controls risk is gaining importance both to external auditors as well as management. Consequently, it is of great importance to ensure the quality of internal audits. This is especially true after the enactment of Sarbanes-Oxley in 2002, which emphasized the assessment of internal control risks by the management and external auditors. I propose a methodology to review the assessment of internal controls risk by internal auditors and business owners. I infer an ordered logistic regression model using historic data collected from internal controls risk assessments by internal auditors and business processes owners of a major multinational corporation. This model can be used for quality review of the auditors' and business owners' evaluation of controls risk, enabling the auditors to focus their efforts on the exceptions where the assigned risk level deviates from the norms. The predicted values of the proposed model are subsequently used to rank the detected exceptions using two measures of disagreement between the assigned and predicted levels in order to point the auditors in the direction of the more suspicious cases. This methodology improves audit efficiency by focusing on the concept of audit by exception, which was introduced in 1991 by Vasarhelyi and Halper. It can also be used as a teaching tool that allows non-expert users, for example business processes owners, to gain access to expert-like knowledge by using the data collected from experienced auditors to infer the model. Accordingly, I can present the business owners with a chance to explain cases with non-conforming risk levels. This would improve the preparer's ability to evaluate risk levels and enables the approver to verify the former's judgment in assigning those scores. In addition to that, the proposed model can be used as a consistency check and serve as a benchmark.

The results show that the null hypothesis (that the independent variables do not provide any significant explanation) can be safely rejected for both parts of the study, i.e. the auditors' assessments as well as the business owners' evaluations. For the control risk assessments conducted by internal auditors, the accuracy of the fitted model was found to be 83%, indicating a systematic assigning of scores. The remaining 17% which represented the exceptions were ranked using the two ranking metrics that I discussed earlier. The sorted list of the twenty most disagreeing exceptions were subsequently sent to the company's internal auditors for further investigation. The internal auditors confirmed the effectiveness of the methodology in detecting anomalous cases that required some explanation for the score disagreement.

This study has some limitations. First, the weights of the ordinal variables are unknown to us. When we move from critical to major to non-major, we do not know whether the distance from critical to major is the same as major to non-major. The company did not reveal the criteria of classifying the issues into these three categories. Access to this information would enable is integration in the prioritization of outliers. The same applies to the risk assessment scores. Another limitation is the unbalanced datasets, although this is to be expected in real life, as critical issues are less likely to occur than major and non-major ones. Future research can extend this study by addressing these limitations. Another possibility to improve this study is by developing a more sophisticated ranking technique and comparing its performance to the method presented in this study, based on a simulated dataset.

CHAPTER 4: IDENTIFYING AND PRIORITIZING IRREGULARITIES USING A RULE-BASED MODEL WITH A WEIGHTING SYSTEM DERIVED FROM EXPERTS' KNOWLEDGE

4.1. Introduction

The real-time economy and business electronization and globalization have caused a tremendous increase in the amount of data that is captured and stored. This change is also facilitated by the decreasing costs of storage. E-commerce and the increased use of information technologies in business, such as Enterprise Resource Planning Systems (ERPs), made huge amounts of disaggregate transactional data readily available for users to analyze and utilize. Companies recognize the importance of harnessing this captured data. A myriad of business intelligence systems has been developed to support decision-making, planning, and control, as well as monitoring organizational performance (Bernhard, 2012; Vijayan, 2012). However, this phenomenon of Big Data necessitates that we take a different approach to audit it. The nature of expected assurance services have progressively driven companies towards a continuous type of assurance. Traditional periodical audits and the use of small sample techniques are proving progressively less effective when Big Data is involved. This problem is expected to escalade as more companies *"wire themselves up and connect to their business partners, they make the entire economy more and more real-time, slowly but surely creating not so much a 'new' but a 'now' economy"*("The Real-Time Economy," 2002). In their Report to the Nations (2012), the Association of Certified Fraud Examiners found after conducting a global fraud study that fraud costs the typical organization 5% of its revenues on a yearly basis. They also found that the median time it took to detect the reported frauds was 18 months. The report also recommends against relying on traditional external audits as the primary fraud detection technique, as they only detected 3% of the reported frauds on average. On the other hand, implementing controls that aim to detect fraud was found to be effective in decreasing both the cost and extent of fraud scenarios (Ratley, 2012). For instance, SAP has recently unveiled a new fraud management and detection product which continuously monitors transactions, identifies exceptions, and alerts the corresponding (Fineberg, 2013).

As such, continuous auditing and monitoring can help improve the efficiency of internal audit work through automation and adoption of an audit-by-exception approach. In this approach the overall population is analyzed and only exceptions are investigated. This is a type of auditing that can be conducted much more frequently. A major German company in fact runs a continuous auditing system that they have implemented on a daily basis. Exceptions are identified, and alarms are sent to the concerned business owners in order to rectify these errors. If they fail to fix the errors in a timely manner, the internal audit department is notified to take action. The problem is that the number of exceptions identified by analytic procedures is so large that auditors feel overwhelmed and inundated with the captured exceptions. The poor performance of humans with regards to complex tasks has been well documented in the social sciences literature (Iselin, 1988; Kleinmuntz, 1990). As a result of such human limitations, the processing of large amounts of information can lead to decreased continuous audit efficiency. This is in contradiction to the main purpose of continuous auditing, which is to increase the efficiency and improve the quality of audits.

While the literature is rich with studies that propose statistical and machine learning techniques to identify exceptions (Dull et al., 2006; Groomer & Murthy, 1989a; Kogan et al., 1999; Vasarhelyi & Halper, 1991), they fail to address the issue of helping the auditors in the post-detection processing stage (Perols & Murthy, 2012). In other words, the proposed methodologies are efficient in helping the auditors identify anomalies and exceptions, but leave the analysis of these results completely to the auditors (Alles et al., 2006, 2008; Debreceny et al., 2003). There is a real need to provide the auditors with a more comprehensive model that first identifies the exceptions, and consequently prioritizes and ranks these exceptions in order of suspicion. The majority of expert system models in the literature follow a generic approach where they assign the same weight to all pieces of evidence that indicate rules violations. However, business rules, and accordingly their violations, do not have the same importance. Consequently, it is critical to assign each rule a weight that corresponds to its importance and the degree to which its violation may increase the control risk.

Motivated by this idea, this chapter proposes a framework where it integrates the judgment of the domain experts (in this case auditors) in a rule-based expert system. Consequently, each piece of evidence is treated according to its significance, which enables the development of a weighting system for the various rules in that expert system. Such a model would first identify exceptions, and then calculate their aggregate suspicion scores based on the weight of each rule they violate, which are in turn used to prioritize identified exceptions. The proposed framework can assist auditors with targeting the records with the highest suspicion scores. By focusing auditors' efforts on the more suspicious records, audit efficiency is expected to increase dramatically.

To develop the weighting system that would be integrated with my rule-based expert system, I conduct a behavioral experiment where an expert panel compares violated rules from a control risk perspective. The framework uses an expert panel that consists of a group of senior internal and external auditors with several years of experience, in particular in the area of control risk assessments. The objective here is to identify the business rules that are perceived to have the strongest effect on control risks, which in turn helps me to gain a better understanding of the importance auditors assign to each rule. In order to do that, the expert panel participants are asked to compare paired records in a specially designed set. Each pair consists of two records such that each record violates one rule, and the two rules are distinct. The participants are asked to perform this pairwise comparison in order to identify the record presenting the highest risk within each pair. Subsequently they are asked to justify their assessment by selecting the best appropriate reason.

The rest of this chapter is organized as follows. Section 2 presents a review or the literature relevant to this study. Section 3 discusses the methodology and describes the proposed framework as well as the data. Section 4 presents the main findings of the study. Section 5 concludes with a summary of the proposed framework and the results, in addition to the limitations and venues for future research.

4.2. Background

Expert systems are defined in the American Heritage Dictionary of the English Language (4th edition) as: *"A program that uses available information, heuristics, and inference to suggest solutions to problems in a particular discipline."* There exist some complex and sophisticated expert systems that can incorporate statistical models and neural networks (which attempt to mimic the way human brains learn) (Oz, 2006). However, the most common type of expert systems is Rule-based decision support systems, which normally consist of a set of rules in the form of IF-THEN statements (Martin & Eckerle, 1991). The latter are mainly developed to explore and identify problems with decision processes. Successful design and deployment of auditing expert systems not only increases auditor's efficiency, but also help them deal with large amounts of data in a more effective way, eventually leading to better informed decisions. Expert systems can emulate auditors' behavior and judgment, thus enabling non-experts to gain expert-like knowledge and expertise (Turban, 1990). Accounting firms have used expert systems in their audit engagements for a long time. In fact, even before the 1990s, the Big 6 firms had developed and were using over 30 different auditing expert systems (Brown, 1991). They have gained their popularity in the auditing profession because of their simplicity and malleability. It is easier to understand and interpret an IF-THEN statement than it is to analyze a statistical model. In addition to that, rule-based expert systems are flexible, as modifying rules is not a complicated procedure, which would make them adaptable to any future changes based on the needs of the organization (Hayes-Roth, 1985). The accuracy and efficiency of such expert systems can be improved using the feedback provided by the domain experts.

4.3. Methodology

4.3.1. Framework Description

As discussed previously, processing large number of exceptions can prove to be problematic. Therefore, it is of great importance to develop an expert system that is capable of not only identifying exceptions, but also prioritizing them. The framework proposed in this study as a solution to the problem of exceptions prioritization is an iterative process and consists of six steps. The first step is the development of a rule-based expert system. The rules of this system are based on business analytics generated jointly from a data archive (historic data) and the knowledge of domain experts (generally used in the auditing profession). Simultaneously, the data archive is used in Step 2 to infer weights for each rule in the expert system. In Step 3 the new data, such as transactions from the current fiscal period, are tested against the rule-based system in order to identify exceptions, which are defined as records that violate one or more rules. Based on the continuous auditing literature (Alles et al., 2008; Chan & Vasarhelyi, 2011) as well as anecdotal evidence, the amount of exceptions captured is expected to be so large that auditors will be overwhelmed with the amount of information requiring investigation. To improve audit efficiency and effectiveness, the weighting system previously inferred from the archive data is then applied in Step 4 to the identified exceptions in order to prioritize them, and consequently present them to the auditors in a prioritized manner. Transactions that violate a certain rule will obtain a score equal to the weight of that rule. The individual scores that transactions gain from violating different rules are aggregated into one ***suspicion score***, which is used in descending order to prioritize the exceptions presented to the auditors. This is a crucial phase of the proposed framework. Instead of following the current norms in the auditing profession, which are to take a sample of the identified exceptions, it helps the auditors process the identified exceptions in a more efficient and systematic way. Subsequently, the auditors investigate these prioritized exceptions in Step 5 in order to identify the true problems. The actual number of exceptions to investigate will largely depend on the time and budget constraints of the audit team. The final step in the proposed iterative process involves feeding back the findings of the auditors' investigation into the rule-based system as well as the weighting system, for refinement purposes, while feeding back transactional data into the data archive. The framework is illustrated in Figure 3.

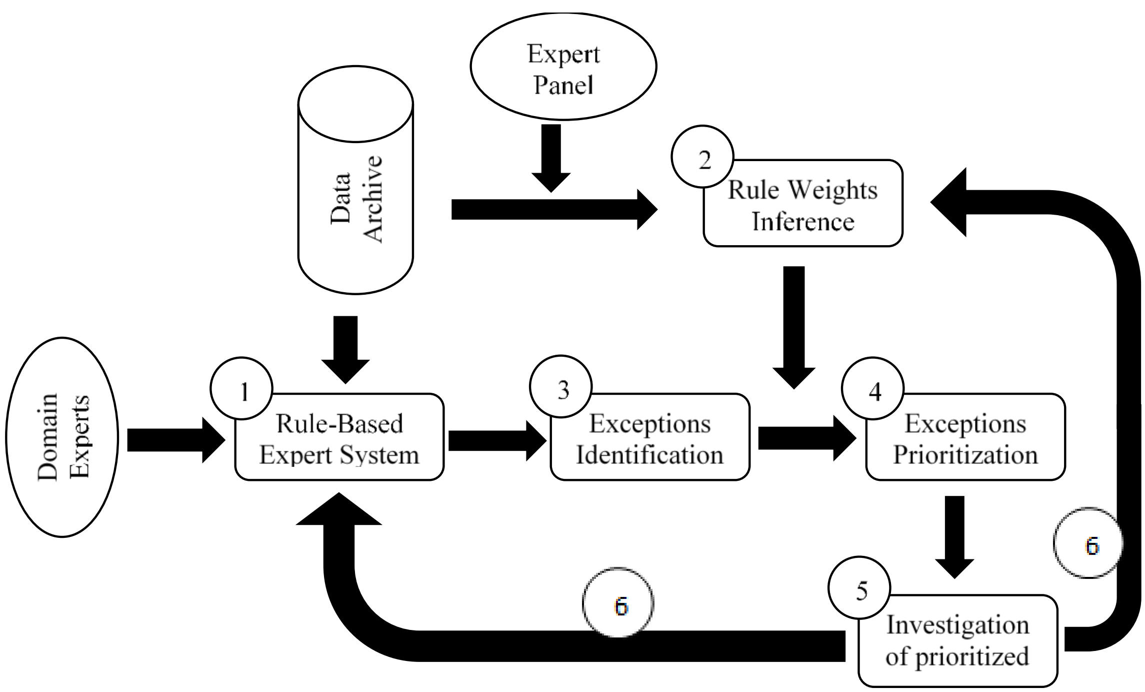

Figure 3-Proposed Framework

4.3.1.1. Rule-Based Systems

Rule-bases systems are popular machine learning techniques that owe their popularity to their simplicity and understandability by human users. They comprise of a set of IF-

THEN rules that classify records violating such rules as exceptions. These rules are easy to interpret by human users, facilitating any modification that the business needs may be required in the future. Their main objective is to train a model based on historic data (or archive data) against which future data is analyzed to identify any possible exceptions. They not only help in the automation of data retrieval and model integration, but also focus on aiding the users to make better informed decisions (Turban, 1990). Rule-based systems can be also viewed as expertise transference tools, as they emulate judgments of domain experts. In fact, part of the rules used in rule-based systems is usually inferred from the archive data and is supplemented by rules provided by domain experts, such as auditors, both external and internal. The analysis of new (or future) data in the context of continuous auditing and continuous control monitoring compares new records to internal control violation policies and/or benchmarks or patterns identified during the model testing phase in order to capture any exceptions (Chan & Vasarhelyi, 2011). Such exceptions could indicate errors that are either intentional (fraud) or unintentional (systematic).

In this chapter a historic Order-to-Cash (O2C) dataset is used both to infer certain business rules and to train the proposed model. In addition to that, the knowledge of domain experts (mainly external and internal auditors) is solicited in order to refine the set of rules. The expert system started initially with 33 analytics that were based on the common procedures used in auditing Order-to-Cash data in the audit profession. These analytics test for violations of controls in the six business processes involved in Order-to-Cash, namely Collections, Customers, Order Entry, Pricing, Receipts, and

Shipping/Billing. Four auditors experienced in auditing this type of dataset were asked to highlight 15 analytics that test for the highest risk areas. The reason behind this was to narrow down the analytics to make the experiment more manageable and to increase the response rate of the solicited participants. After receiving the feedback from those four experienced persons, 12 analytics were found to have 75% or more of the votes. These analytics were selected as they covered all the business processes except Pricing, which was selected by only one person. The explanation provided by the experts who did not select rules that cover Pricing was that they rarely test for Pricing, knowing that this is usually done automatically, unless the control system of a company is very weak. The 12 analytics can be categorized as tests of segregation of duties, unauthorized transactions, missing documents, and non-matching documents. A complete list of Analytics is presented in Appendix A.

4.3.1.2. Rules Weight Inference

Business rules, and accordingly their violations, do not have the same significance. Consequently, it is important to assign each rule a weight that corresponds to its significance and the degree to which its violation may indicate increased control risk. This is an integral part of my methodology as it helps to address a problem that has not been well studied in the continuous auditing literature, and that is exceptions processing. There is a plethora of studies in literature that propose various machine learning and data mining techniques that are highly efficient in identifying exceptions. There is a problem, however, that the number of these exceptions is so large that the auditors are overwhelmed with the amount of information they are required to process (Alles et al., 2008). Knowing that not all exceptions are equally suspicious, the weighting system can be employed to develop a ranking methodology that would allow the identified exceptions to be prioritized according to their level of suspicion.

This study uses an expert panel consisting of senior auditors with at least three years of experience to develop a weighting system for the significance of the rules in my expert system. Both internal auditors and external auditors participated via an experiment as part of my panel. The composition of this panel follows the Delphi technique guidelines, which state that the minimum size of an expert panel is seven. The typical panel size was found to fall between 15 and 40 participants, according to a study that surveyed published papers that followed the Delphi technique (Baldwin-Morgan, 1993; N Dalkey & Helmer, 1963). These techniques are especially beneficial in two situations: 1) cases where it is not feasible for the expert panel to meet and deliberate due to time or budget constraints and 2) when a subjective collective opinion is acceptable in the absence of more precise yet practical techniques. It has been shown in the literature that such a collective judgment yields more accurate results than random or individual judgments, within the concept of “two heads are better than one when exact knowledge is not available.” (NC Dalkey, Brown, & Cochran, 1969; Parente & Anderson, 1984).

The objective of this step is to identify the business rules that are perceived to have the strongest effect on control risks, which in turn helps to develop a better understanding of the importance auditors assign to each rule. In order to do that, the participants are asked to compare a set of paired records. Each record in a pair violates one rule that is different from the rule violated by the other record in that pair. After assessing each pair of records, the participants are required to select the record that indicates heightened control risk within each pair, and to specify the reason for their selection from a list of options. The participants are then asked to justify their judgment in order to verify whether they actually identified the violated rule, which in turn ensures that my use of various transactions correctly measure the risk associated with violating the corresponding business rule.

To better understand the procedure, consider the following example in Table 14. The auditors are asked to evaluate this pair of records and select the record that provides evidence of a higher control risk:

Table 14-Example of a Pair of Transactions

Customer Transaction ID	Customer ID	Invoice Number	Invoice Date	Created by	Inventory Item ID	Invoice Selling Price	Invoiced Amount	Shipment ID	Shipment document Created By	Shipment document Approved By
10034	1146	10001203	10/15/1997	1546	63	800	9600	10000877	1002	1546
4110	1000	10000262	3/1/1998	1117	628	600	1800			

In this example, the first transaction violates a segregation of duty rule where an individual can create/modify an invoice and approve the shipment. The second transaction, on the other hand, tests for orphaned invoices, in other words transactions in the invoice table that do not have matching shipping documents. The participants are asked to evaluate this pair and select the record that presents a higher control risk. If the participant chooses Record 2, the second rule is deemed to have a higher weight than the first one. The pairwise comparisons of all records result in a partial order of rules from the significance perspective. I use linear programming to solve this problem and infer the weight associated with each rule. This step is explained in more details in Section 4.3.3.2.

4.3.1.3. Exceptions Identification

The step that comes after the selection of the analytics that test for riskier rules violations and the development of the weighting system is the identification of the exceptions. The new data is applied to the rule based system to capture all the records that violate one or more rules. This step is run at the level of the total population, and not just a sample. Therefore, we just need to focus on this set of exceptions as they present a certain level of control risks. Assuming that the set of analytics utilized in my expert system is sufficiently complete, the remaining records present a negligible risk because they do not violate any of the used analytics.

4.3.1.4. Exceptions Prioritization

This step consists of the combination of the results from previous steps. At this stage, the expert system has identified the exceptions that violated one or more rules. The experiment was also used to infer the weight of each rule. Subsequently, the inferred rule weighting system is applied to the captured exceptions, each of which is assigned a ***suspicion score*** defined as the sum of the weights of all the rules it violates:

$$SS(X_i) = \sum W_{Rj} V_{Rj}$$

Where $SS(X_i)$ is the Suspicion Score of record X_i

W_{Rj} is the weight of rule R_j

V_{Rj} is the binary variable that equals one if record X_i violates rule R_j, and 0 otherwise

The higher the suspicion score, the stronger the recommendation to auditors to investigate that instance. In other words, the auditors are recommended to focus on the cases with high suspicion scores to improve audit efficiency. For instance, a score of zero means that this transaction did not violate any rules, and consequently is considered the least suspicious.

To better understand how exceptions prioritization process works, let us consider the following example shown in Table 15.

Table 15-Prioritization Process

Record	Weight	SOD Customers	Unauthorized	Unauthorized	SOD Credit	Match Shipping	Match Invoice to	Missing Sales	Unauthorized	SOD Invoices /	Orphaned	SOD Invoices /	Excessive Write	Suspicion Score
		1.00	2.58	1.23	1.23	1.91	2.36	2.81	2.36	1.91	2.58	1.00	3.04	24
1001		X				X								2.91
1002									X		X			4.94
1003								X				X		3.81
1004			X					X					X	8.43
1005														0
1006				X										1.23

According to this examples record 1005 has a suspicion score of zero, signifying that it does not violate any rule. On the other hand, record 1004 has the highest suspicion score amongst all the records. Since auditors are recommended to focus their efforts on the more suspicious records, they should prioritize the records in order of decreasing suspicion score, and consequently address them in the following order: 1004, 1002, 1003, 1001, 1006, and 1005.

4.3.1.5. Investigation of Prioritized Exceptions

Once the scores are assigned to each record and exceptions are prioritized, the result is a complete list of scored exceptions sorted by decreasing suspicion score. This list is provided to the auditors who conduct their investigation of a subset of the identified exceptions. The scope of this investigation will largely depend on the audit team's time and budget constraints, and the size of the subset of exceptions as well as the depth of investigation will change accordingly. The auditors will distinguish between true exceptions and false ones.

4.3.1.6. Feedback of Prioritized Exceptions

The results of the auditors' investigations are relevant to the fine tuning process. In fact, they can feed back into the rule-based system to refine it at two levels. First, this feedback can help adjust the rules that make up the expert system. Rules that are found to be misleading can be dropped from the expert system in the subsequent iteration. Second, the feedback from auditors' investigations enables us to modify the weights of the rules according to the audit teams' findings. This feedback will be treated as an additional prioritization of exceptions and will allow us to have a more general linear program with a difference between the weights from the experiment and the prioritization of the auditors. The feedback will be incorporated as a new set of constraints in the model, following the general case linear program (explained in more details in Section 4.3.3.2). In fact, when auditors provide a new feedback from their investigations, this study adapts the linear program by adding more terms to the objective function and more constraints. Moreover, the auditors' prioritization will be considered as vote with certainty one, unless they plainly express uncertainty in their judgment. This is different from the original experiment where the auditors may not reach a consensus, and therefore the certainty level can vary for each case. The effect of the original experiment on the weights determined by the linear program will decrease over time with more feedback from auditors incorporated in the linear programming model. The continuous linear program will include a progressively greater number of constraints, thus increasing its complexity. Finally, the data that is tested by the rule-based system will also become a direct input into the data archive for the succeeding periods, improving the overall accuracy of the expert system with each iteration.

4.3.2. Dataset Description

4.3.2.1. Order-to-Cash

While implementing continuous auditing is beneficial to companies, the benefit does not always outweigh the cost. This type of auditing can have a significant effect on the cost and effectiveness of business operations. As a matter of fact, continuous auditing is most beneficial when it targets high risk business processes such as Order-to-Cash (Chan &

Vasarhelyi, 2011). This study focuses on a high risk area, which is Order-to-Cash. Companies tend to pay special attention to this cycle as their objective is to generate profits. While payments can present an opportunity for fraudulent activities, collections can be worrisome to managements. Early collections enable companies to conduct their operations which are funded by the cash flows from such collections.

This chapter uses a simulated Order-to-Cash dataset based on the training dataset used by a major auditing software company. This type of data covers several business processes that involve receiving and fulfilling customer requests for products or services. Normally such data covers the *Customers, Order entry, Pricing, Ship-Bill, Receipts, and Collection* business processes. The Order-to-Cash cycle is comprised of several sub-processes that follow the subsequent steps:

1. A client order is received and documented
2. The order (service) is fulfilled (scheduled)
3. The product (service) is shipped (executed)
4. The invoice is created and sent to the client
5. The payment is received from the client
6. The payment is recorded in the general ledger

4.3.3. Data for Weighting System Derivation

4.3.3.1. Experiment and Data Collection:

The research data is collected using multi-item scales in a decision case setting. The survey is administered to senior auditors with experience in control risk assessment. They are presented with a sample of Order to Cash fictitious data (in the form of pairs of records) along with relevant facts. They are asked next to select the record that they believe to increase the control risk and subsequently to provide a rationale for their choice. Subjects are asked to respond to the questions as though they are in an actual audit engagement. The survey concludes by asking for demographic information including experience and education of the auditors.

The specific population of senior auditors was selected because (1) of the interest in auditors' judgments; (2) this population has the necessary expertise and experience in assessing control risks; (3) this population has sufficient knowledge about Order to Cash data auditing[5].

To conduct the experiment, the author solicited participation from senior auditors at accounting firms (for external auditors) and internal audit departments of several companies (for internal auditors).

As explained previously, the set of rules that are included in the rule-based system consisted of the twelve analytics that tested for the highest risk areas. Table 16-List of AnalyticsTable 16 presents these analytics and shows the business processes involved.

Table 16-List of Analytics

Analytic	**Business Process Involved**	**Rationale of the Analytic**	**Analytic Description**

[5] There is no need to worry about self-selection bias, as the expert panel used in this study comes from companies that deal with control risk assessments, whether for their internal use (internal auditors) or their clients (external auditors). Moreover, control risk assessment is basically the same across companies.

Analytic 1	CUSTOMERS	Segregation of Duties – Customers	Identify transactions where an individual created a Customer could also APPROVE the Sales Order.
Analytic 2	ORDER ENTRY	Unauthorized Transaction – Sales Order	Identify Single Sales Orders for Employees that are not authorized to create SOs in Authorization list.
Analytic 3	ORDER ENTRY	Unauthorized Transaction – Price	Identify transactions in the daily Sales Order table where the list price differs from the sales order price.
Analytic 4	ORDER ENTRY	Segregation of Duties – Credit Adjustment	Identify transactions where an individual CREATED/MODIFIED a Customer Credit Limit could also APPROVE the Sales Order.
Analytic 5	SHIP / BILL	Non-matching values – Shipping to Sales order	Identify transactions in the daily Shipping Documents table where the shipping price differs from the sales order price.
Analytic 6	SHIP / BILL	Non-matching values – Invoice to Shipping	Identify transactions in the Invoice table where the Invoice price differs from the Shipping Document price.
Analytic 7	SHIP / BILL	Missing values – Missing Sales Orders	Identify transactions in the Shipping Documents table where there are no matching Sales Orders.

Analytic 8	SHIP / BILL	Unauthorized Transaction – Shipments	Identify Single Shipping Documents for Employees that are not authorized to CREATE shipping lines from Authorization list.
Analytic 9	SHIP / BILL	Segregation of Duties – Shipment / Invoice	Identify transactions where an individual CREATED/MODIFIED an Invoice could also APPROVE records in the shipping document file.
Analytic 10	SHIP / BILL	Non-matching values – Orphaned Invoices	Identify transactions in the Invoice table where there are no matching Shipping Documents.
Analytic 11	RECEIPTS	Segregation of Duties – Invoice / Receipts	Identify transactions where an individual APPROVED an Invoice could also CREATE records in the Receiving file.
Analytic 12	COLLECTIONS	Unauthorized Transaction – Excessive Write Offs	Identify write off transactions where the write off amount exceeds a Percentage Threshold of the full Invoiced amount.

Next, pairs of analytical tests were matched for comparison purposes. Related as well as unrelated analytics were selected to be compared, in order to ensure that all analytics were eventually compared, either directly or indirectly. A comparison is considered to be direct when two analytics are compared against each other (represented by two transactions that violate them). Indirect comparison, on the other hand, indicates that two analytics are linked by one or more comparisons (e.g. analytic A is compared to B, and B is compared to C, therefore analytic A is indirectly compared to C). Figure 4 illustrates all the connections between various analytics.

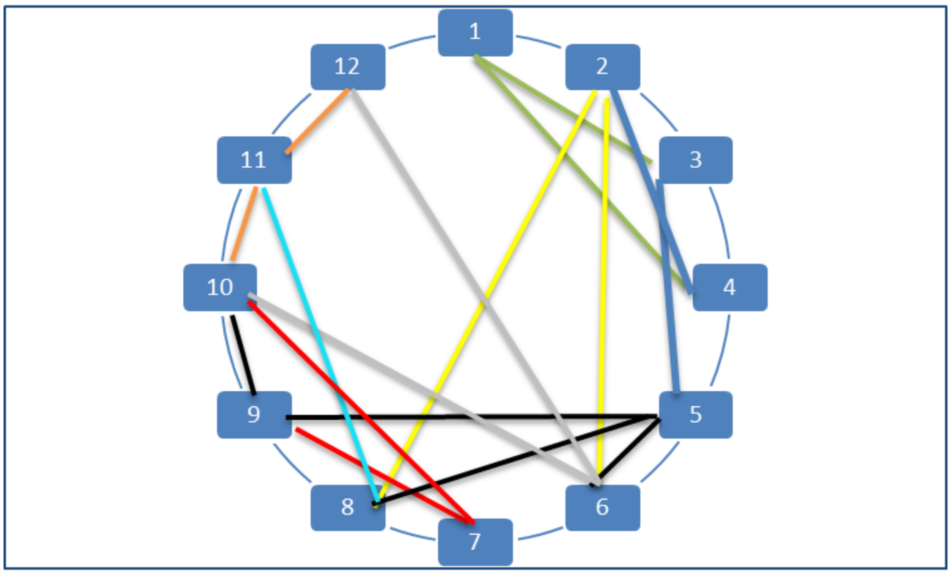

Figure 4-Connection - Analytic Comparisons

This figure depicts the connections of various analytics, showing that all the analytics were compared either directly or indirectly. For example, there is a direct line connecting Analytics 1 and 4 indicating a direct comparison, in which case a pair of transactions is created to compare the two analytics. On the other hand, Analytic 3 is indirectly compared to Analytic 10 as 3 is connected to 5, which is connected to 9. Analytic 9, in turn, is connected to analytic 10. Therefore, the comparability of all the analytics can be ensured.

After selecting the analytics and deciding on the pairs of analytics to compare, the variables that are required to conduct the two analytics were identified. These variables were combined in order to provide a uniform scenario where two transactions with the same set of variables are compared. Each transaction violated one of the business rules that were tested using that pair of analytics. The auditors were asked to select the transaction that presents the highest control risk, and the rationale behind their selection (see example above). There were 16 pairs in total, comparing 32 different scenarios. An extra pair of transaction was added where one of the transactions violated two rules to emulate real life cases. Transactional records may actually violate more than one rule in a real business environment and this technique can be utilized to test if that would present any difficulties. In the next section two linear programs are presented, a special case model and a general case model, which use the 16-pair and 17-pair scenarios, respectively. The sample was limited to 17 comparisons based on a pilot test that was conducted to determine the optimal number of pairwise comparisons that can be completed in an acceptable amount of time[6]. This was done in order to decrease the burden of experiment on the participants, and hence increase the response rate. For a complete list of analytics comparisons, please refer to Table 35 in Appendix A. A copy of the research instrument as it appears online can be found in Appendix B.

4.3.3.2. Derivation of Weights:

A. Special case model

[6] We conducted a pilot test with auditors, both internal and external, who are experienced in assessing internal control risks. Subsequently we made a few minor modifications to the survey based on the results of the pilot study.

The optimal weights of the different analytics are derived from the results of the experiment using linear programming. The objective function of the linear program is to maximize the differences in rules' weights (within each pair) weighted by the agreement (amongst the responses). First, I propose a special case model, where each transaction violates a single rule. This simplifies the calculation of the proportion of votes correctly identifying violated rules. As a result, each rule would have one weight W_{Ri} in each pair, amounting to two weights per pair.

The special case linear program is as follows:

$$Max \sum A_{ij}(W_{Ri} - W_{Rj}) + (M * S) \geq 0$$

Subject to $\left(W_{Ri} - W_{Rj}\right) \geq A_{ij} * S$

$$W_{Ri} \geq 1$$

$$W_{Rj} \geq 1$$

$$\sum W_{Ri} = 2 * N$$

$$S \geq 0$$

Where W_{Ri} and W_{Rj} are the weights of Rules R_i and R_j, respectively.

A_{ij} is the certainty about the ordering of the rules in pair P_{ij}, defined by the proportion of responses showing that transaction T_i presents a risk greater than or equal to that presented by transaction T_j

S is the scaling factor, a non-negative variable

M is a constant following the Big M method (or Big Component method)

N is the number of rules in the expert system.

The linear problem was solved using the simplex algorithm, where a feasible area in the shape of a polytope is defined by a set of linear inequalities. The simplex algorithm is an efficient algorithm that can guarantee reaching a global optimum, given that certain precautions are taken to avoid cycling. Consequently the algorithm starts at a starting vertex, moves along the polytope's edges until it reaches the optimum's vertex.

In order to ensure a feasible solution to the problem, the so-called Big M modification of a linear program is employed. In order to significantly maximize the difference between various weights, those weight differences are designed to be proportional to *S* and add the term M*S to the objective function. This linear program is then solved with progressively increasing values of M until a non-trivial solution is obtained. Without the introduction of M in the linear program, the result will always be the trivial optimal solution. In such scenario all the weights W_i would be equal to the lower bound of the weights, in other words one, except for the weight that has the greatest coefficient in the objective function, which would take the value of N+1. *S* in this case would always be zero. To avoid reaching this trivial solution, M is selected to be sufficiently big in order to force the problem to choose *S* as large as possible, which would make the pairs as separable as possible.

Control risk assessments can change from one auditor to the other. Consequently, it is important to introduce the certainty term A_{ij} to measure how certain I am that transaction T_j is riskier than transaction T_j. In this experimental setting, only the responses that correctly identified the rationale of the violated rules were in fact included in the analysis.

A_{ij} is expected to be equal to one if all the responses selecting T_i as the transaction presenting the higher risk also correctly identify the violated rule, unless the participants explicitly express a level of uncertainty. On the other hand, A_{ij} will equal 0 in case of a tie where the proportion of responses selecting T_i as the transaction presenting a heightened risk is equal to the proportion selecting T_j[7]. The certainty in this case that one transaction presents a higher risk is zero and the weights assigned to each rule (in that specific tie case) are the same. In such a case, an additional constraint needs to be added to turn the two inequalities involving T_i and T_j into an equality, and the result is the following two constraints to represent the tie:

$(W_{Ri} - W_{Rj}) \geq A_{ij} * S = 0$ (Constraint 1- Pij)

$(W_{Rj} - W_{Ri}) \geq A_{ij} * S = 0$ (Constraint 2-Pij)

In situations where mixed responses exist, A_{ij} is calculated such that:

$$A_{ij} = \frac{(\#\ correct\ T_i) - (\#\ correct\ T_j)}{(\#\ correct\ T_i) + (\#\ correct\ T_j)}$$

Where $(\#\ correct\ T_i)$ and $(\#\ correct\ T_j)$ are the numbers of responses selecting transactions T_i and T_j, respectively, as the transaction presenting the higher control risk, and at the same time correctly identifying the rule violated in each case.

The special case linear program consisted of 29 constraints in total. 12 constraints were included to ensure that each of the 12 rules has a minimum weight equal to one. In other

[7]We only consider the responses where the participants correctly identified the violated rule and ignore the remaining responses. The reason we do that is to ensure that the rankings within each pair actually reflect the rules that were violated.

words, the constraint that all the weights should be greater or equal to one, as the weights of the rules in the expert system should not equal zero is introduced. There were also 16 constraints representing the 16 compared pairs of transactions. There weren't any pairs with a tie between the two transactions, in other words where equal participants selected each transaction as the one presenting a higher risk. Consequently, the model did not include two inequalities to compensate for the tie in such cases, as explained previously. The last constraint ensures that the sum of the weights of all the rules has an upper bound of twice the number of rules (2*N).

B. General case model

While the assumption of single-rule violation in the special case model makes it easier to understand, this is not how transactions behave in real life. Because there exist transactions in real business data that violate multiple rules, the second model proposed in this study is more general and realistic, where a transaction is allowed to violate more than one rule. To examine this model, a new pair of transactions is added to the 16 pairs that were used in the special case model. This pair has one transaction that violates two rules that are different from the one violated by the second transaction in that pair. The general case linear program became the following:

$$Max \sum A_{ij}\left(\sum W_{Ri} - \sum W_{Rj}\right) + M * S \geq 0$$

Subject to $\left(\sum W_{Ri} - \sum W_{Rj}\right) \geq A_{ij} * S$

$$W_{Ri} \geq 1$$

$$W_{Rj} \geq 1$$

$$\sum W_{Ri} = 2 * N$$

$S \geq 0$

Where $W_{Ri}$and W_{Rj} are the weights of Rules R_i and R_j, respectively.

A_{ij} is the certainty about the ordering of the rules in pair P_{ij}, defined by the proportion of responses showing that transaction T_i presents a risk greater than or equal to that presented by transaction T_j

S is the scaling factor, a non-negative variable

M is a constant following the Big M method (or Big Component method)

N is the number of rules in the expert system.

This model is also solved using the simplex algorithm. The objective function here is to maximize the differences in weights of violated rules. It differs from the special case program's objective function in that it takes into accounts the weights of all the violated rules within a pair. This is the reason that this general model can be used for the succeeding iterations, as it is capable of incorporating auditors' prioritization of exceptions. With future feedback incorporated in the model in subsequent iterations, the auditors' prioritization is expected to be provided in the form of a ranking. In this case, A_{ij} would be one if T_i>T_j and 0 in case of a tie, unless otherwise stated by the auditors if they explicitly specify the level of certainty in their judgment. The study follows the same procedure of introducing a second constraint as in the special case model. Once again, the weights W_i and W_j would be equal in the case of a tie.

This linear program consisted of a total of 30 constraints, 12 of which are to set the lower bound of the weight of each rule to one. It included 17 constraints to represent the 17 pairwise comparisons, not requiring additional constraints to account for ties among the transactions within pairs. The remaining constraint was included to set the upper bound of the sum of all weights to *(2*N)*, *N* being the number of rules in the expert system.

4.4. <u>Findings</u>

4.4.1. Demographics

The results that are presented in this section are based on the responses of 17 participants. All of these participants completed the online survey, which was conducted using Qualtrics.

Among the 17 respondents, 11 were internal auditors from large multinational companies, while the remaining six were from public accounting firms. The internal iauditors had on average 2.3 years of experience in external auditing. Two of the participants reported that the highest degree they obtained was a Ph.D. Another two respondents mentioned they held an MBA degree. The number of participants who indicated they had a bachelor degree was eight, same as those who had a masters' degree. The most common professional designation was Certified Public Accountant (CPA) with 11 respondents, followed by four Certified Internal Auditors (CIA). There were two Certified Fraud Examiners (CFE), two Certified Information Systems Auditors (CISA), and one Certified Management Accountant (CMA).

The median years of professional experience of the participants was 12.5 years. However, the results showed a skewness equal to 0.98 and a standard deviation equal to 11.10. This was mostly the result of five participants in my panel who had over 25 years of experience. As for the experience in IT audit, the respondents had 3.18 years of experience on average. Once again, this result had a skewness value of 1.60. On the other hand, the average experience in auditing financial statements was 6.53 years. The results show that the participants had an average of 3.47 years in external auditing, compared to 5.71 years of internal auditing experience on average. This difference is mainly due to the fact that four out of the five participants who had over 25 years of experience are internal auditors. As this study is mostly related to control risk assessment, it is important to ensure that the panel met the experience requirements, and consequently checked for the participants' experience in conducting such assessments. The respondents had a median of eight years of experience in assessing control risks. This high value indicates that my panel in fact meets the requirements to be considered an expert panel, and consequently my participants are well qualified as experts in assessing control risks.

We asked the participants if they had ever worked on audit engagements in an online environment to examine the level of comfort in completing this experiment. The results show that the participants were split, where eight of them indicated that they had participated in online audit engagements, compared to nine who had never engaged in a similar activity before. It was interesting to find that out of the eight participants who had prior experience with online engagements, seven were internal auditors, and only one external auditor.

The median for the level of knowledge in audit analytics, a skill that is useful for the completion of my experiment, was five on a 7-point Likert scale. This value is high enough to ensure that the participants had in fact the expertise required for the completion of my experiment. As for the level of knowledge of continuous auditing and continuous monitoring, the median turned out to be 5 and 5.50 on a 7-point Likert scale, respectively. Table 17 presents the summary statistics of the demographic questions.

Table 17-Summary Statistics

Question	***Count***	***Sum***	***Median***	***Average***	***Variance***	***SD***	***Skewness***
Do you feel that you had enough data to perform the required task?	17	64	4	3.76	1.19	1.09	0.16
Did you find the task to be-Unmotivating-Challenging	17	71	4	4.18	1.40	1.19	-0.36
Did you find the task to be-Extremely easy-Extremely Difficult	17	68	4	4.00	0.88	0.94	-0.53
Please select degrees obtained-A.S./A.A.	17	1	0	0.06	0.06	0.24	4.24
Please select degrees obtained-B.S./B.A.	17	8	0.5	0.47	0.26	0.51	0.12
Please select degrees obtained-M.S./M.A.	17	8	0	0.47	0.26	0.51	0.24
Please select degrees obtained-MPA/MSA	17	0	0	0.00	0.00	0.00	N/A
Please select degrees obtained-MBA	17	2	0	0.12	0.11	0.33	2.71
Please select degrees obtained-Ph.D.	17	2	0	0.12	0.11	0.33	2.71
Please select the professional designation-CPA	17	11	1	0.65	0.24	0.49	-0.77
Please select the professional designation-CIA	17	4	0	0.24	0.19	0.44	1.46
Please select the professional designation-CMA	17	0	0	0.00	0.00	0.00	#N/A

Please select the professional designation-CFA	17	1	0	0.06	0.06	0.24	4.24
Please select the professional designation-CFE	17	2	0	0.12	0.11	0.33	2.71
Please select the professional designation-EA	17	0	0	0.00	0.00	0.00	N/A
Please select the professional designation-CISA	17	2	0	0.12	0.11	0.33	2.71
Years of professional working experience (in general)	17	275	12.5	16.18	123.15	11.10	0.98
Years of professional working experience in IT Audit	17	54	0	3.18	23.78	4.88	1.60
Years of professional working experience in auditing Financial Statements	17	111	4.5	6.53	65.01	8.06	2.54
Years of professional working experience in assessing controls risk	17	150	8	8.82	56.90	7.54	2.20
Years of professional working experience in external auditing	17	59	3	3.47	11.51	3.39	1.16
Years of professional working experience in internal auditing	17	97	3	5.71	64.35	8.02	2.06
How do you rate your knowledge of Audit Analytics?	17	85	5	5.00	1.38	1.17	-0.81
Continuous Auditing knowledge	17	85	5	5.00	1.13	1.06	-0.73
Continuous Control Monitoring Knowledge	17	86	5.5	5.06	1.18	1.09	-0.87

In order to check if the participants believed that they received enough information to complete the experiment, they were asked to rate the provided information. The results show that the respondents were neutral, with a median 4.0 on a 7-point Likert scale. The same thing could be said regarding the levels of difficulty as well as challenge presented by the experiment, as the median values for the answers of the participants were four on a 7-point Likert scale for both questions.

4.4.2. Rules' weights

As mentioned in a previous section, there was a total of 17 pairs, 16 of which were used for the special case model, while an additional pair was included in the general case model. The responses that misidentified the correct analytic that a transaction violated were excluded from the analysis, and only the cases where the participants selected the right rationales were taken into consideration. Overall, the responses of the expert panel show that 76.11% of the time on average the participants agreed on the transaction presenting a heightened control risk in each pair. This high agreement level indicates that the majority of the auditors evaluate control risk from the same perspective, yet the judgment involved allows for differences in assessments to exist.

In addition to that, the average correct identification of the intended rationale (i.e. the violated analytic) was a high 85% of all responses. After analyzing the responses that selected the transaction with the highest risk, the results indicate that 86% of them managed to identify the right analytic that was violated. This correctness level drops to 83% for the transaction that was least selected as the high risk one. Details about the agreements and the higher risk transactions, as well as correctly identified violations, can be found in Table 18.

Table 18-Agreements and Correctness

Pair #	Transaction 1	Transaction 2	T1 Correct	T2 Correct	Tmax	# correct Tmax	% Correct Tmax	# correct Tmin	% Correct Tmin	# correct General	% Correct General
P1	2	15	1	15	T2	15	100%	1	50%	16	94%
P2	5	12	3	9	T2	9	75%	3	60%	12	71%

P3	3	14	3	12	T2	12	86%	3	100%	15	88%
P4	10	7	8	6	T1	8	80%	6	86%	14	82%
P5	13	4	12	2	T1	12	92%	2	NA	14	82%
P6	6	11	6	9	T2	9	82%	6	100%	15	88%
P7	4	13	3	11	T2	11	85%	3	NA	14	82%
P8	3	14	3	12	T2	12	86%	3	100%	15	88%
P9	8	9	7	8	T2	8	89%	7	88%	15	88%
P10	5	12	3	10	T2	10	83%	3	60%	13	76%
P11	0	17	0	15	T2	15	88%	0	NA	15	88%
P12	6	11	6	10	T2	10	91%	6	100%	16	94%
P13	4	13	3	11	T2	11	85%	3	NA	14	82%
P14	7	10	7	8	T2	8	80%	7	100%	15	88%
P15	0	17	0	16	T2	16	94%	0	NA	16	94%
P16	3	14	3	14	T2	14	100%	3	100%	17	100%
P17	11	6	8	3	T1	8	73%	3	50%	11	65%
Average							86%		83%		85%

Where T1 and T2 represent the number of responses selecting Transaction 1 and Transaction 2 to present the heightened risk, respectively.

T1 (T2) Correct: is the number of responses selecting Transaction 1 (2) and correctly identifying the violated rule.

Tmax (Tmin) is the transaction that was chosen the most (least).

Correct Tmax (% Correct Tmax) is the number (percentage) of responses correctly identifying the violated rule for Tmax

Correct Tmin (% Correct Tmin) is the number (percentage) of responses correctly identifying the violated rule for Tmin

Correct General (% Correct General) is the number (percentage) of responses correctly identifying the violated rules for both transactions in a pair.

This study introduces two different linear programs to solve for the special case and general case models. These linear programs were solved using the responses obtained from the 17 auditors who participated in my experiment. The results show that while the majority of the rules kept the same ranking (order based on their weights), there were few changes in the order. The analytic that had the highest weight was Excessive Write-Offs for both the special and general case models. Moreover, the analytics that proved to have the lowest weights came in the same order, and showed a maximum 5% difference between the two models. Even for the cases where the order of analytics was different, the differences in the weights that were inferred from both models were less than 7%.

It was interesting to find that when all the responses were used, most of the analytics that test for violation of segregation of duties had the lowest weights amongst all the rules. In fact, the weights of three of these rules were absolutely at the bottom, while the fourth came seventh from a significance point of view.

A possible explanation of my results is that the auditors seem to weigh the rules whose violation leads to direct financial losses the highest. The results show that auditors placed higher weights on the controls that had direct impact on the company's financial statement compared to operational controls. For instance, excessive write-offs are a form of misappropriation of assets that entails direct loss of money, which cannot be recovered. Same thing applies to missing and unauthorized sales orders. Auditors keep assertions in their mind when they evaluate control risks. Missing shipping documents can indicate that the company is trying to boost their revenues. As for segregation of duties rules, auditors do not consider them equally important. Instead, they evaluate them according to their impact on the financial statements. Moreover, they take into consideration the presence of potential procedures that can mitigate the risk of violating these segregation of duties rules. For example, violating Analytic 1, which tests for the transactions where the sales order was approved by the same user who created the customer, can be mitigated by a second control such as credit check. Table 19 presents the weights of each rule inferred from both models.

Table 19-Wegihts Attributed to the Analytics (All respondents)

Analytic	**Rules Weights (Special case Model)**	**Rules Weights (General case model)**
Analytic_1_SOD_Customers	1.00	1.00
Analytic_2_Unauthorized_Sales_Order	2.67	2.60
Analytic_3_Unathorized_Price	1.53	1.51
Analytic_4_SOD_Credit_Adjustment	1.30	1.29
Analytic_5_Match_Shipping_to_SO	1.96	1.92
Analytic_6_Match_Invoice_to_Ship	2.30	2.25
Analytic_7_Missing_Sales_Orders	2.39	2.82
Analytic_8_Unauthorized_Shipments	2.32	2.27
Analytic_9_SOD_Ship_Invoice	2.00	1.96

Analytic_10_Orphaned_Invoices	2.63	2.56
Analytic_11_SOD_Invoice_Receipt	1.00	1.00
Analytic_12_Excessive_Write_Offs	2.91	2.83

A closer look at the results shows that internal auditors and external auditors do not always weigh the rules similarly. While excessive write-offs and missing sales orders were always significant, the analytics that test for segregation of duties violations were in general assigned higher weights by the external auditors than by internal auditors. A comparison of the ranks and weights of the analytics as assigned by the internal auditors, external auditors, and both groups can be seen in Table 20. These are under the general case model, which utilized 17 pairs.

Table 20-Internal vs. External Auditors

	Internal Auditors		**External Auditors**		**All Responses**	
Order	Analytic	Weight	Analytic	Weight	Analytic	Weight
1	Analytic 7 Missing Sales Orders	3.07	Analytic 12 Excessive Write Offs	2.71	Analytic 12 Excessive Write Offs	2.83
2	Analytic 12 Excessive Write Offs	2.93	Analytic 4 SOD Credit Adjustment	2.42	Analytic 7 Missing Sales Orders	2.82
3	Analytic 10 Orphaned Invoices	2.75	Analytic 7 Missing Sales Orders	2.36	Analytic 2 Unauthorized Sales Order	2.60

4	Analytic 2 Unauthorized Sales Order	2.65	Analytic 11 SOD Invoice Receipt	2.30	Analytic 10 Orphaned Invoices	2.56
5	Analytic 8 Unauthorized Shipments	2.36	Analytic 2 Unauthorized Sales Order	2.22	Analytic 8 Unauthorized Shipments	2.27
6	Analytic 6 Match Invoice to Ship	2.30	Analytic 10 Orphaned Invoices	2.22	Analytic 6 Match Invoice to Ship	2.25
7	Analytic 5 Match Shipping to SO	1.98	Analytic 6 Match Invoice to Ship	1.95	Analytic 9 SOD Ship Invoice	1.96
8	Analytic 3 Unauthorized Price	1.51	Analytic 8 Unauthorized Shipments	1.92	Analytic 5 Match Shipping to SO	1.92
9	Analytic 4 SOD Credit Adjustment	1.45	Analytic 9 SOD Ship Invoice	1.81	Analytic 3 Unauthorized Price	1.51
10	Analytic 1 SOD Customers	1.00	Analytic 5 Match Shipping to SO	1.68	Analytic 4 SOD Credit Adjustment	1.29
11	Analytic 9 SOD Ship Invoice	1.00	Analytic 3 Unauthorized Price	1.41	Analytic 1 SOD Customers	1.00
12	Analytic 11 SOD Invoice Receipt	1.00	Analytic 1 SOD Customers	1.00	Analytic 11 SOD Invoice Receipt	1.00

It is noteworthy to mention that the differences between the two could be driven by the small external auditors sample (6 auditors) compared to the sample of internal auditors (11 auditors).

4.5. <u>Conclusion</u>

With this shift in auditing towards an audit-by-exception approach, it is therefore crucial to develop systematic techniques that would allow the auditors to examine the entire population effectively and efficiently. There are plenty of methodologies in the literature that can efficiently identify exceptions. Rule based systems are popular tools that can accomplish that. They are easy to interpret, yet powerful enough to capture the transactions that violate rules and classify them as exceptions. However, the results of such systems are too numerous, giving rise to another kind of problem. Auditors have to process these exceptions, and due to their expected large numbers, might feel overwhelmed with this task. In fact, the human limitations with regards to processing complex and aggregate tasks have been well documents in the social sciences literature. It is therefore of great importance to provide the auditors with a methodology than can assist them in processing the identified exceptions.

I propose a framework that can address these issues by first identifying exceptions and then prioritizing them. I develop a rule-based expert system that consists of analytics that are commonly used by auditors. Furthermore, the expert system is refined by asking a panel of experts to select a subset of these analytics that test for high risk controls. The identified exceptions are then prioritized using a weighting system that inferred from an experiment involving senior auditors as participants. To develop this weighting system the auditors are asked to compare a set of paired records and then identify the record that presents a heightened control risk within each pair. Next, the participants are asked to justify their assessment in order to ensure that their judgment is indeed due to the violation of the rule I am testing for. The results of the panel's comparisons are then used to infer the weights of each rule by solving a linear program. The study incudes two linear programs. The first one is a special case model where each record within a pair can violate only a single rule. This restriction is relaxed in the general model, where there is a possibility of multiple rule violations by the same record. This general model has also the capability of incorporating future feedback from auditors' investigations of the identified and prioritized exceptions. This iterative process will lead eventually to a weaker effect of the original experiment, as the effect of progressively increasing feedback becomes stronger. Once the rules' weights are derived, they can be used to calculate an aggregate suspicion score for each transaction. This suspicion score can be used in turn to prioritize all the transactions in a dataset, and subsequently assist the auditors in dealing with the identified exceptions by pointing them towards the transactions with higher suspicion score.

The results obtained from 17 responses show a high agreement rate among the participants on the transactions presenting a heightened control risk. In addition to that, the participants were able to correctly identify the violated rule in each case with a high degree of correctness. The demographic questions I asked at the end of the experiment confirmed that the participants had the necessary experience to complete the task of comparisons, and consequently could indeed be considered as an expert panel. I was able to infer the weights of the analytics that were included in the proposed expert system. This weighting system makes it possible to rank and prioritize all the records. The results also show that the auditors assigned higher risks to the rules whose violations could impact the financial statements directly. Moreover, there are significant differences between the assessments of internal and external auditors.

This study has several limitations. For instance, the size of the expert panel is small, although within the range of the panel sizes as recommended by the Delphi technique. Future research can address this issue by soliciting additional participants. Another shortcoming is that the framework was not tested and applied to a real-business data, mainly because of the difficulty of obtaining such data. Moreover, the method followed to choose the analytics presents another limitation. Only a subset of analytics was used in the expert system, focusing on the analytics that covers the areas of highest control risk based on the recommendations of a small expert panel. Ideally we would want all the analytics to be included in the expert system to make it as comprehensive as possible. Future research can utilize a more comprehensive set of analytics that is not limited to areas of highest risk.

CHAPTER 5: DUPLICATE RECORDS DETECTION TECHNIQUES: A PRIORITIZATION APPROACH

5.1. Introduction

Companies generate and collect huge amounts of data every day. Information flows into the companies' management information systems, and in particular their accounting information systems, with every transaction that takes place. This phenomenon has increased exponentially with the wide-spread implementations of computerized systems in companies and organizations around the globe. The result is huge amounts of data collected, captured, and stored in companies' data warehouses. This Big Data is exploited for various purposes. For instance, operational databases store information generated by business transactions, which can be used to assess and improve the efficiency of the company's business operations. Moreover, the management often incorporates such data in the process of decision making related to business operations, hence viewing it as the cornerstone of these operations. On top of that, this data is used and audited by internal and external auditors to ensure the quality of the company's financial reports.

Given the importance and the prevalent usage of operational data, it is obvious that a certain level of quality has to be maintained. Inadequate data causes serious operational difficulties as well as direct financial losses. In addition to serious implications on decision making, the quality of the data may affect customer satisfaction, resulting in unnecessary and possibly high costs to repair damage caused by low-quality data (Redman, 1997; Wand & Wang, 1996). This issue is also aggravated by Sarbanes Oxley

Act of 2002 that requires companies to report on any material changes in their financial condition at or close to the time of occurrence of a certain event (Section 409, SOX). The real-time economy and the need for timelier reporting drive companies towards a more frequent and close to real-time auditing. In fact, there is an increasing trend to follow an audit-by-exception approach, which was proposed by Vasarhelyi and Halper in 1991. However, in order to rely on the results of such approach, it is crucial to maintain a high level of quality of data. The output can only be as good as the input data.

One of the issues that greatly diminish the quality of the data is the existence of duplicate records that represent the same real life objects. While the ideal situation is to have a global or unique identifier for every object or record in a database, which enables records to be identified, linked, and related across tables, this is not always the case in the complex databases in real-life situations. Many organizations have multiple data collection systems (e.g. SAP, Oracle, legacy systems) that may differ not only in assigning unique identifiers, but also in the format, structure, and schema of the underlying databases. As such, the quality of data will depend greatly on whether they are collected from single or multiple sources, as well as the compatibility of the latter. Additionally, data quality can be affected by human errors, including data entry errors and lack of constraints, for example allowing for incorrect entries like a person's age of 430 years (Chatterjee & Segev, 1991).

One of the problems that can result from such sub-optimal situations is the existence of duplicate records in the data. Duplication of records can occur when data are entered manually or gathered from multiple sources, whether different systems or simply different locations. Weis, Naumann, Jehle, & Lufter, (2008) describe duplicate records as *"all cases of multiple representations of same real-world objects, i.e., duplicates in a data source."*

Heterogeneous data often lacks a global identifier, or a primary key, which would uniquely identify real-world objects. This problem is not restricted to a certain line of business. Unfortunately, it can occur in census data, IRS tax information, accounts payable, medical records, and virtually any electronic database. In fact, even in every day operations we encounter duplicate records. For example, duplicate contacts and duplicate calendar events can occur when merging information from two different sources, such as a computer and a smartphone. This problem, in fact, is so prevalent that there exist companies whose sole business is developing solutions to fix duplicate records.

For the business world, an area of particular importance is the existence of duplicate payments. Duplicate payments can indicate various issues, from simple data entry mistakes, to intentionally fraudulent activities. No matter what the related intention or reason, duplicate payments can cause great losses to organizations. In 1998, for example, the Department of Health & Human Services estimated the duplicate payments made by Medicare to be $89 million (McMullan, 2001).

The computer science literature is abundant with papers that deal with the problem of duplicate records, mostly by proposing some domain-specific algorithms. However, the same could not be said about the accounting literature, where studies addressing this issue are scarce, despite the great interest that companies show in finding possible solutions to this problem. On the other hand, there exist companies and agencies who provide their own solutions to the problem of duplicate payments for a share of the payments collected from identified duplicates[8]. Unfortunately, these agencies use their proprietary detection algorithms and tend to refuse to divulge any information regarding their methodologies.

Motivated by the shortage of studies in the accounting literature that address the problem of duplicate records, this chapter attempts to fill in this gap by discussing various techniques employed in the detection of duplicate records. Next, this study illustrates the special case of duplicate payments, a problem of a particular importance to the business world. The proposed methodology identifies possible duplicates by looking at the level of similarity based on the combinations of some relevant variables. Two datasets, provided by the internal audit department of a telecommunications company, are utilized as an illustration of matching techniques used to capture duplicate payments.

The results confirm the existence of duplicate payments in the database of a multinational telecommunication company. Moreover, the duplicate detection techniques yield large numbers of duplicate candidates, which can be problematic. Limited by the auditors' time and budget constraints, the investigation of all the results is often prohibitively costly. Consequently, there is a great need to develop a methodology that can help the auditors in processing the numerous results. I propose a theoretical framework that can be used to prioritize the duplicate candidates based on multiple criteria, a step that can help solving

[8] A common commercial approach (recovery agencies) is a vendor firm offering data examination services at no cost, with compensation resting on savings from duplicate identification. Two main types of duplicates prevail: 1) unintentional errors that mainly imply a temporal cash flow drain and 2) collisional payments where an employee cooperates with the payee.

the problem of big numbers of candidates that are generally flagged by the duplicates detection algorithms.

The remainder of the chapter is organized as follows. Section 2 presents the literature related to the problem of duplicate records detection, and more specifically that of duplicate payments. Section 3 describes the methodology and the data used in this study to illustrate the problem of duplicate payments. Section 4 presents the findings. The proposed methodology for duplicate candidates' prioritization is described in Section 5. Lastly, Section 6 concludes the study and suggests future research.

5.2. Background

5.2.1. Duplicate Records Detection

5.2.1.1. Problem History

The issue of duplicate records has been discussed in academic and applied computer science literature for many years. While this problem is not unique to computer-based processes, it tends to be more pronounced in automated systems. Early papers refer to this matter as *record matching* or *records linkage* (Newcombe, 1988a; Tepping, 1968). Nowadays this problem is mostly referred to as *duplicate records detection*. The main objective of record matching is to detect multiple representations of the same real-world object in a database (Elmagarmid, Ipeirotis, & Verykios, 2007) (Chou, Du, & Lai, 2007). The first concerns raised about duplicate records regarded medical records used for epidemiological research (Newcombe, 1988b). Other sections of the society became concerned with the issue, such as tax agencies who wanted to gather information about tax payers with missing or incorrect social security numbers. Duplicate record detection was also found to be a very useful tool in the detection of fraud and money laundering (Ted, Goldberg, Wooton, Cottini, & Khan, 1995).

The existence of duplicate records is a very common problem and can be manifested in all aspects of our daily life. The scenarios that involve duplicates range from simple situations like duplicate contacts and calendars to the complex databases stored in companies' data warehouses. Table 21 is an example of a simple duplicate records scenario that can occur in a database due to the lack of a unified structure. All these representation refer to the same real world object, for instance the record of a customer. While a human user can immediately realize that they do in fact represent the same object, computerized systems will consider each one of them as a distinct object. As a result, the customer John B. Smith will have multiple records in the company's database.

Table 21-Example of Duplicate Records

Record	Name	Address	Age	Phone
1	John Smith	1 Washington Park	32 yrs	973-123-4567
2	J.B. Smith	1 Washington Park	32 years	1-973-123-4567
3	J. Smith	1 Washington Park	32 years	(973)1234567
4	John Smith	1 Washington Park Ave	32 years	+1-973-123-4567
5	John Smith	1 Washington Park Avenue	32 yrs	+19731234567

In real life, and especially in cases related to numeric objects such as dates and dollar amounts, identifying multiple representations of the same object becomes hard if not impossible, even for experienced human users. It would be very challenging to spot the similarities between two numeric records in more complex databases.

5.2.1.2. Duplicate Detection Process

The majority of the studies in the literature focused on domain-specific algorithms. These algorithms work best when they are applied to a particular area as they use production rules based on knowledge of that area. The main advantage of such algorithms is that the human expertise is incorporated in the algorithm itself during its design phase and as a result no subsequent human knowledge is required. It is this absence of a need for human expertise that enables such algorithms to be automated. On the other hand, the disadvantage is that this approach requires great efforts to keep the rules updated whenever non-conforming new data is introduced in the dataset. Few studies in the literature followed a more general approach, which allows algorithms developed based on this approach to be used across domains. They were predicated on the assumption that domain-specific knowledge will be provided by human experts in the phase that succeeds running the algorithm (Hernandez & Stolfo, 1995; Wang, Madnick, & Horton, 1989). This study follows the lead of domain-specific studies to address the issue of duplicate payments, as it is easier in this situation to incorporate experts' knowledge in the algorithm in advance rather than request such expertise at a later stage. Moreover, following this approach enables me to automate the process of identifying duplicate payments.

Weis and Naumann (2005) describe a generalized framework for duplicate records detection consisting of three phases:

- *Phase 1*- **Candidate Description**: the first step in detecting duplicates is to decide which objects to compare according to the relevance for the identification of the objects. The idea behind this step is that only elements that can represent the same real-world object should be compared, even if they are presented differently. Moreover, only the attributes that are relevant to the identification of an object need to be selected for comparison. As an example, let us consider the same example in Table 21. When comparing Records 1 and 2, there is no need to compare the ***Name*** from Record 1 to the ***Address*** of Record 2, but only to the Name value from Record 2. Moreover, at this stage we select the attributes that we deem relevant for my comparison, like the Name, Address, and Phone, but not Age.
- *Phase 2*- **Duplicate Definition**: In this step we decide on the criteria used to decide when two duplicate candidates are to be considered actual duplicates. The criteria depend on the description of the duplicate objects, in other words the collective relevant attributes that describe the objects. In addition to that, the criteria also include a similarity measure that would define how similar two candidates must be in order to be treated as duplicates. Back to my simplified example of customers' records, each record can be defined using the name, address, and phone number, and those attributes would constitute the description of the customer. As for the similarity measure, we may decide to consider as duplicates only the records that have identical values for any two out of the three attributes.

- *Phase 3*- **Duplicate Detection**: This phase is when the algorithm, which will detect duplicate candidates and subsequently identify real duplicates from a list of candidates, is selected. The first step is searching for two or more records that can be designated as candidates based on the criteria and definitions from the first two phases. Two main approaches are used in this searching step.
    - Blocking: where the entire dataset is divided into segments or *blocks* that contain records with the same values for a set of attributes.
    - Sorting: where the database is sorted, then candidates of duplicate records are sought in small segments of the database, usually comprising of neighboring records in the sorting order (Bitton, 1983). A variation of this approach uses a sliding window technique, where the number of compared records remains constant but the position of the window changes. There is, however, a tradeoff between the accuracy of the algorithm and the size of the window, which decides on the number of records to be compared. The larger the window size, the more accurate the detection process is, but at a greater computational cost. In fact, in order to capture *all* the duplicate records, all possible pairs of records must be compared, which would result in a costly combinatorial explosion (a quadratic number of comparisons).

The second step in Phase 3 is matching, where the possible duplicates that were identified during the searching step are compared, and then designated as matching or not matching. Either a probabilistic approach, which assumes the matching/non-matching patterns to be known in advance, or a machine learning/statistical model (for example based on decision trees) can be used in the matching step (Verykios, Elfeky, & Elmagarmid, 2000).

The first two phases (candidate description and duplicate definition) can be completed offline concurrently with system setup (Weis & Naumann, 2005). The company or the auditors can decide offline on the description and criteria to be used in the identification of duplicate candidates and the subsequent classification as matching or non-matching. The third step takes place when the algorithm is run and the actual detection is performed.

5.2.1.3. Duplicate Detection Methods

There are two types of duplicate records detection, based on how similar the records must be in order to be considered duplicates. According to the selected method, the matching algorithm would look for either exact duplicates where all the values for the relevant variables are identical, or fuzzy duplicates where the values can be similar rather than identical. One of the factors that play a major role in the selection of the method is the time and budget allocated for the task of investigating the resulting duplicate candidates. The selection of the desired method must be done before the detection process starts, as it has a direct effect on the results. More specifically, depending on which method the company decides to use, the number of false positives (i.e. the candidates that turn out to be non-duplicates) and the false negatives (i.e. the duplicates that were not classified as candidates) will vary considerably.

- *Exact matching*: This is the scenario where the set of candidates are exactly identical in the dataset, with respect to the examined variables. This standard procedure to identify exact duplicates begins with the standardization of the values by removing all spaces and changing all letters to the upper case. Subsequently, all the records in the table are sorted before the neighboring (or consecutive) records are compared. Records that have identical values for the relevant variables are classified as duplicate candidates. The sorting step is used as a preliminary clustering technique that groups possibly matching neighboring records prior to conducting a pairwise comparison (Fellegi, 1969; Newcombe, Kennedy, & Axford, 1959).

- *Fuzzy matching*: Also knows as *near-identical matching*, this technique aims at identifying sets of records that have "similar" values for the relevant as the duplicates candidates. Fuzzy matches may occur due to keypunch errors, different ways of entering values, or deliberate obfuscation. For example, fuzzy matches can be caused by using multiple address formats in the dataset. One system could record the address as *123 East Fourth Street*, while that address can be entered as *123 E. 4th St* in another system. The address format could be aggregate with the full address in one cell, or disaggregate and split into several cells describing the street address, city, state, zip code, etc. Unlike the exact matching method, where two records are considered duplicates only if they are identical, the fuzzy matching method classifies two records as duplicate candidates based on a certain threshold and some similarity criteria. Examples of similarity criteria include special characters, like hyphens and slashes, and leading and trailing character positions, like the location of a comma in numbers (e.g. 34,567 vs. 345,67), to list a few (Weis et al., 2008).

There are several similarity metrics used in the literature, such as character-based, token-based, and phonetic similarity metrics (Elmagarmid et al., 2007). This study focuses on character-based similarity metrics, which are designed for and work well with typographical errors. The concept of character-based similarity metrics relies on the distance between two values. There are several types of distances proposed in the literature of duplicates detection. The most common metric is the Edit distance, which measures the minimum number of edit operations (insertion, deletion, or replacement) needed to transform one record into the other. The Levenshtein distance is a special case of the edit distance, where each operation has a cost of one (Levenshtein, 1966). In other words, every modification, insertion, or omission of a character has a distance of 1, no matter where the position of this edit operation is. For example, the Levenshtein distance between "John Smith" and "J. Smith" is three, as we will need three operations to go from one to the other. Once again, all letters are capitalized and spaces and periods are removed before running the matching algorithm.

Another distance measure is the Affine Gap distance, which is similar to the Levenshtein distance except that it introduces two other edit operations, the open gap and the extended gap. This metric introduces a two-part penalty function that penalizes for the existence of a gap (length independent) and the extension of a gap (length dependent). Unlike the

Levenshtein distance, the Affine Gap distance works well when we have truncated records, such as “Bob E. Smith” and “Robert Edward Smith” (Waterman & Smith, 1976). A third type is the Smith-Waterman distance, which is an extension of the edit distance and the Affine Gap distance. It assigns higher costs to mismatches toward the center as compared to mismatches at the beginning and end of records. This works well with situations that involve prefixes and suffixes. For example, the penalties for mismatches in the prefixes and suffixes are lower than those in the middle when comparing the two records “Prof. Robert E. Smith” and “Prof. Robert E. Smith, Rutgers University” under the Smith-Waterman distance (Waterman & Smith, 1976).

The threshold and similarity metrics are situation-specific and consequently depend largely on the organization’s policy and needs. Some companies classify candidate duplicates based on pre-set profiles. An example of a company using this profiling technique is Schuffa Holding AG[9], which uses *k*-base classifiers for the matching procedure. Records are checked using multiple classifiers, and a point is added to the profile every time a classifier matches two candidates. The opposite is also true, where a point is subtracted from the profile when the record is classified as non-duplicate. At the end, all the points awarded to the profile are added. Subsequently the score is compared to a threshold, and a set of records are classified as duplicates if their score is higher than the threshold. On the other hand if they score below the threshold, they are classified as non-duplicates (Weis et al., 2008).

[9] Schuffa Holding AG is a credit information agency based in Germany, whose main line of business is to save and retrieve credit histories of more than 60 million persons.

In this study I adopt a similar technique where I classify the records as duplicates/non-duplicates according to a set of rules. I follow the exact matching approach because of the time and budget limitations of the company's internal audit team. The next section presents the issue of duplicate payments, and presents various methodologies proposed in the literature for the detection of such duplicates.

5.2.2. Duplicate Payments

The progressive evolution of information and telecommunication technologies led to the conversion of business processes from traditional paper-based into a digital form. This evolution encompassed accounting information systems which generated, computed, analyzed, and stored huge amounts of transactional data, all in digital format (Rezaee, Sharbatoghlie, Elam, & McMickle, 2002). Various types of systems were developed, and continue to be, in order to take advantage of these large databases. It became common for companies to implement systems ranging from simple automated accounting packages to complex Enterprise Resource Planning (ERP) systems, Decision Support Systems (DSS) and Knowledge-based Expert Systems (KES) (Chou et al., 2007). These systems assist auditors in numerous audit processes. In fact, one of the advantages of such systems is that they help to detect fraudulent activities.

However, the quality and efficiency of these systems depend greatly on the quality of the underlying data. Inconsistencies that occur due to the integration of different systems or due to human error may seriously affect data quality. An example of such errors is duplicate payments. These can be the result of simple human errors, (e.g. typing mistakes), object presentation (e.g. checks paid to *Rutgers* vs. *Rutgers University*), or more serious systematic errors like different structures from different sources (e.g. date format) (Chatterjee & Segev, 1991). Finally, they can be an indication of collusive fraud (Ngai, Hu, Wong, Chen, & Sun, 2010).

Duplicate payments are not rare or infrequent events. For instance, Medicaid identified more than $9.7 million in duplicate payments in a two-year audit period, and estimated the actual amount to be around $31.1 million (Novello, 2004). It is noteworthy that duplicate payments often go undetected. It is therefore important to implement techniques that would help in their detection[10].

Most studies in the academic literature and general practice follow the general framework described in the previous section, which uses three-way match on the *amount, date,* and *vendor*. One can think of these three as meta-variables, or combinations of several variables intended to form a unique identifier. For instance, the *vendor* can consist of one or more of the following: name, ID (or number), account, address, etc. The amount is generally the invoice amount; however, in some instances the paid amount is used instead, especially when the invoice is split into multiple payments. The date is included to differentiate between duplicate payments and recurring payments, such as monthly rent or services. Other studies include a fourth variable to refine the system depending on the

[10] The aforementioned recovery agencies charge from 15% to 50% of any recovered amounts (C. Warner, 2013)

situation. One study that dealt with duplicate Medicare payments used a 4-way match; however it added two more identifiers to describe the services (McMullan, 2001).

In practice, similar approaches are followed. The website of a recovery agency shows that they use the three-way match using four variables (*vendor number, invoice number, invoice date, and invoice amount*) as its duplicate detection criterion. The matching technique they use is a fuzzy matching. They identified the similarity threshold for the amounts of duplicate candidates within 3% of each other, half or doubles of each other. As for the invoice number, they state that if two candidates have the same first four digits of the invoice number or equivalent invoice numbers after removal of leading or trailing zeroes (e.g. invoices # 12300 and 00123RE), these candidates are considered duplicates. Recurring payments, such as monthly rent or installments, were excluded from the beginning, as they were considered legitimate and common (C. Warner, 2013).
The impact of duplicate payments on companies increases its importance to the business world, as illustrated by these studies and the numerous agencies that provide this kind of services.

5.3. Methodology

5.3.3. Data Description and Preparation

In order to test for the duplicate payments problem, this study analyzes two datasets that were provided by the internal audit department of a multinational telecommunications company. The two datasets were extracted from two non-overlapping data sources and covered payment transactions for the period extending from July 2008 to June 2010.

The datasets used in this study are real-life business data. Therefore, like real datasets, they require some preparation prior to analysis. Such datasets often contain missing values, erroneous formats, as well as other data issues. The data preparation method that is followed in this chapter involves the three following steps (Kimball, 2009):

- Parsing: identifying individual elements in the dataset
- Data transformation: making data conform to the data types of their corresponding domains. E.g. renaming a field, or converting a data element.
- Data standardization: standardizing the data into one format. For example, an address can be written as (*123 East Fourth Street*, or *123 E. 4th St.*) and these two may be seen by the system as two different addresses, increasing the number of false positives.

5.3.3.1. Dataset 1

The first dataset (henceforth Dataset 1) was procured from wire payments to other telecommunications carriers. It was fairly clean with no missing values. It consisted of 21,606 transactions, three of which had $0.00 for the Amount. Upon investigation, one of these three turned out to be a void payment, while the other two were a payment and its reversal. Consequently, these three payments were removed from the dataset. There were nine attributes in total, including the dates, the amount, payees' information, and information related to the payments as generated by the system. These variables were self-explanatory, with the exception of the two attributes that indicated the entered data and the effective date. After consulting the internal auditors of the telecommunication company, this confusion was resolved.

The records in Dataset 1 were divided as follows:

Table 22-Dataset 1 Description

	July 2008 /June 2009	July 2009 /June 2010	Total
Number of transactions	11611	9992	21603
Total Amount	$648,128,623	$586,682,198	$1,234,810,821
Average payment	$55,820.22	$58,715.19	$57,159.23
Negative payments (reversals)	$0	-$6,432,408	-$6,432,408
Highest payment	2858276.2	3000000	3000000

This table indicates that there were reversals for the amount of -$6,432,408 during the fiscal year 2009/2010, but none during the year 2008/2009. The average payment for the two fiscal years was similar, and same thing can be said about the number of transactions. The median payment amounts was found to be $10,815.1 for the total population. Other descriptive statistics of Dataset 1 can be seen in Table 23.

Table 23-Dataset 1-Descriptive Statistics

Basic Statistical Measures			
Location		**Variability**	
Mean	57159.2	Std Deviation	152170
Median	10816.9	Variance	23155600000

Mode	150000	Range	3800000
		Interquartile Range	39369

Due to the big standard deviation and range values, it was necessary to check for extreme outliers. The results confirmed the existence of some extreme observations. Table 24 shows the 5 highest values for the monetary amount as well as the 5 lowest values.

Table 24-Dataset 1-Extreme Observations

Extreme Observations	
5 Lowest Values (in USD)	**5 Highest Values (in USD)**
-800000	3000000
-789507	2858276
-500000	2786140
-500000	2574312
-500000	2570995

5.3.3.2. Dataset 2

The second dataset that was used in this study (henceforth Dataset 2) includes transactions made by checks. It consists of 47683 records, and involves 51 attributes that provide information on the dates, invoice, checks, and vendor, in addition to some information related to the source system. A complete list of variables can be found in Appendix C.

Contrary to Dataset 1, Dataset 2 required extensive cleaning. Several attributes had missing values to various extents. Some of these variables had to be excluded from the study either due to missing values or to irrelevance to the problem of duplicate payments detection.

Table 25-Dataset 2-Description

	July 2008 /June 2009	**July 2009 /June 2010**	**Total**
Number of transactions	9359	24710	34069
Total Amount	204159569.9	347095211.2	551254781.1
Average payment	21814.25	14046.75	16180.54
Negative payments (various)	-$424366.54	-242729.37	667095.91
Highest payment	8418242.66	14725000	14725000

Unlike Dataset 1, the average payment for the two fiscal years were not similar, with the fiscal year 2008/2009 averaging at $21, 814.25 as opposed to the $14,046.75 average from the fiscal year 2009/2010. However, the total number of observations from the second period was almost three times that of 2008/2009. Another difference between the two datasets is that there are negative payments during both fiscal periods in Dataset 2, unlike Dataset 1. While the negative payments indicated reversals in Dataset 1, the same could not be told about those found in Dataset 2, as they were the result of various activities, such as prepaid services.

As shown in the summary statistics table below, the median for the payments from Dataset 2 is $124.73. This is significantly lower than Dataset 1, which had a median value of $10,816.9. However, this significant difference is understandable as the payments in Dataset 1 are mostly to other telecommunication carriers, as opposed to Dataset 2 where the majority of the payments are for customers and smaller vendors.

Table 26-Dataset 2-Descriptive Statistics

Basic Statistical Measures			
Location		**Variability**	
Mean	16180.54	**Std Deviation**	192057
Median	124.73	**Variance**	36885800000
Mode	100.00	**Range**	14806390
		Interquartile Range	1357

Similar to Dataset 1, this dataset contains extreme outliers. Table 27 presents the five lowest and five highest extreme observations.

Table 27-Dataset 2-Extreme Observations

Extreme Observations	
5 Lowest Values	**5 Highest Values**

-81390.2	14725000
-51805	8915224
-51108	8418243
-50245	8344362
-50114	7761080

5.3.4. Detection Algorithms

The majority of the duplicate payments detection algorithms that are proposed in the industry follow the same logic, which conducts a three-way match. This logic attempts to identify a payment based on information related to the payee, the amount, and the date. Each one of these concepts could be described using one or more attributes. Table 28 presents a fuzzy matching algorithm that uses such an additional variable.

Table 28-Fuzzy Matching Algorithm

Vendor Number	Invoice Number	Invoice Date	Invoice Amount
Exact	Exact	Exact	Exact
Different	Exact	Exact	Exact
Exact	Similar	Exact	Exact
Exact	Exact	Similar	Exact
Exact	Exact	Exact	Similar
Exact	Similar	Exact	Similar
Exact	Similar	Similar	Exact

Exact	Exact	Similar	Similar
Different	Exact	Similar	Exact

This study uses several combinations of variables based on the three-way match logic described previously.

In order to uniquely identify each payment transaction in Dataset 1, three variables were included to identify the vendor, date of transaction, and the amount. The Carrier account number was used to identify the vendor. In order to check for consistency, the analysis was repeated with the Carrier account number substituted with the Carrier name. The results were consistent, and therefore the analysis continued using the Carrier account number. As for the date, the dataset contained two different variables. Both variables were included separately in the algorithm. The results showed significant differences when using these two date variables. This was expected as the two variables record different information, namely the date the transaction was entered in the financial system and the date it became effective. Consequently, the two combinations (each with one of the two date variables) were utilized and run them separately. In all the combinations, the amount of the transaction was included in the algorithm.

The same three-way matching technique was followed for the identification of possible duplicate payments in Dataset 2. The algorithm examined the vendor, date, and amount of the transactions. Because there were several variables that represented the same concepts, for instance the amount and the date, only the information that was obtained from invoices was included for consistency and relevance purposes. Therefore, algorithm consisted of the invoice date and invoice amount, in addition to the vendor's name. At a later stage the invoice number was introduced as a fourth variable in order for the detection algorithm to yield more "manageable" results, as the company auditors requested and expressed their wish to have fewer candidates.

The software used for the duplicates detection stage is ACL. This study followed the exact matching technique that was described previously, mostly because the number of false positive is expected to increase dramatically in the case of fuzzy matching. In fact, this increase could render the investigation of all the duplicate candidates prohibitively expensive.

Ideally, the effectiveness of a technique is generally measured using Recall (the correctly identified duplicates over all true duplicates) and Precision (correctly identified duplicates over all found duplicates), in addition to their harmonic mean, or the f-measure:

- Recall: $\frac{True\ Positives}{True\ Positives+False\ Negatives}$

- Precision: $\frac{True\ Positives}{True\ Positives+False\ Positives}$

- f-measure: $\frac{2.Precision.Recall}{Precision+Recall}$

The goal is to maximize the f-measure by maximizing precision and recall (Weis et al., 2008) .

Unfortunately, the datasets used in this study do not allow for the measurement of recall of the methods used, as it is not possible to identify the false negatives in real data. I had to rely on the investigation of the internal audit department of the telecommunication company that provided me with the data; however, due to time and budget constrains it was not possible to examine all the datasets. In order to do that, a labeled dataset is required, where the outcome of all transactions is known. In other words, where each transaction can be identified as unique or not. In fact, it was necessary to rely on the feedback of the company for the evaluation of my techniques.

5.4. Findings

5.4.1. Dataset 1 Findings

The duplicate detection algorithm that was applied to Dataset 1 was a three-way exact matching algorithm following the approach that was discussed in Section 3.3.2. Due to the existence of two date variables (entered and effective), the algorithm was ran twice, using one of the data variable each time[11].

The first set of variables consisted of the *Account ID, Effective Date,* and *Amount*. The duplicate detection process yielded 82 duplicate candidates, which were presented for the telecommunication company for further investigation.

[11] I ran the algorithm after replacing the carrier's Account ID with the Carrier's name. I got the same results, which indicates a consistency between Carrier Name and Account ID, eliminating the possibility of errors in matching carriers to account IDs.

For the second algorithm, the variable *Effective Date* was replaced with *Entered Date*, while the *Account ID* and the *Amount* variables remained unchanged. The use of this second set of variables returned 168 duplicate candidates. Similar to the results from the first set of variables, the results from the second set were presented to the internal audit department of the telecommunication company. While the majority of these duplicate candidates turned out to be false negatives, the company's internal auditors confirmed the existence of true positives, in other words real duplicate payments.

Set 2 identified 168 candidates and exhibited higher statistics in general. For instance, the median from Set 2 is \$63,000 as opposed to \$19,868.23 from Set 1. This is the result of the existence of some extreme payments in Set 2's results. The same thing applies to the range of the payments. In addition to that, the total amount for the candidates of Set 1 is \$4,506,354.14 for the 82 transactions, as opposed to \$25,141,759.1 for the 168 candidates from Set 2 (Table 29).

Table 29-Dataset 1-Descriptive Statistics for the Amount

Descriptive Statistics-Amount		
	Set 1	**Set 2**
N	82	168
Mean	54955.54	149653.3
Median	19868.23	63000
Mode	63000	63000

Range	811033	1500000
Std Deviation	125205.948	239051.306
Std Error Mean	13826.684	18443.2087
Variance	15676500000	57145500000
Skewness	5.15219017	1.87245077
Uncorrected SS	1517450000000	13305900000000
Corrected SS	1269800000000	9543300000000
Coefficient Variation	227.831357	159.736713
Sum Observations	4506354.14	25141759.1
Kurtosis	28.4931437	4.86962316

Table 30 presents the three highest and three lowest values of duplicate candidates in both sets of variables. Set 2 has three duplicate candidates (6 transactions) that amount for $1,000,000 each, driving the average candidate's amount to $149,653.3 for Set 2, compared to the average of $54,955.54 for Set 1's duplicate candidates. The results also indicate that Set 1 had one pair of candidates with negative payment, and another pair with a zero payment. These transactions turned out to be reversals. On the other hand, Set 2 included two negative pairs of candidates in addition to a zero payment pair of candidates. Once again, further investigation of these transactions showed that these were reversals as well.

Table 30-Dataset 1-Extreme Observations-Amount

Extreme Observations

Lowest		Highest	
Set 1	**Set 2**	**Set 1**	**Set 2**
-21526.46	-500000	789506.5	1000000
0	-21526.5	215590	1000000
16.65	0	172119	1000000

When the number of duplicate candidates from Set 1 was examined, the duplicate candidates corresponded to 23 vendors, the majority of which had only one duplicate candidate. However, one payee had 13 duplicate candidates (i.e. 26 transactions) that amounted for a total of $1,270,000. In fact, this vendor alone amounted for 31.7% of the number of duplicate candidates and 28.2% of the total amount from all the candidates identified by Set 1. However, that was not the highest total amount for a duplicate candidates, as there was one payee had two duplicate candidates that accounted for $1,579,013, which is twice the highest value for Set 1 in Table 30 ($789,506.5).

On the other hand, the vendors that exhibited duplicate candidates using variables Set 2 were 34 vendors. The results show that five vendors out the 34 vendors accounted for over 55% of the total number of duplicate candidates. The same five vendors accounted for approximated 74% of the total monetary amount for all the duplicate candidates from Set 2. The counts and amounts for all the vendors from Sets 1 and 2 can be found in Appendix D.

The variable *Invoice ID* was included in both cases to verify if this would have an effect on the results. The results of including this variable led to zero duplicate candidates, indicating that it was unique to each record. However, this did not prevent the existence of duplicate payments, as the company's internal auditors confirmed.

An interesting and even surprising finding was the presence of three classification errors, where the type of payment was categorized as *Commission payments*. This type of payment, in fact, is not allowed according to the company's policy, and consequently such payments should not exist. After further investigation by the company's internal auditors, they found that these transactions were data entry errors. *Commission payment* was underneath *Check payment* in the dropdown list under ***Transaction Type***. These were identified as not correctly reversed. The company was advised of the issue and took the necessary steps to rectify this control weakness by removing *Commission Payment* from the dropdown list and adjusting the reversals. This type of data problem identification, root cause identification, and feedback to management/system designers is common and important for corporate data quality.

5.4.2. Dataset 2 Findings

The second dataset was more challenging. The initial basic three-way match using *(Invoice Amount*, *Invoice Date*, and *Vendor Name)* resulted in 899 duplicate candidates. As the number of candidates was too high to be investigated in its entirety, the company's auditors requested more "manageable" results, in other words fewer results. In order to do that, *Invoice ID* was included as a fourth variable to identify each transaction. While an invoice number should act as a unique identifier or a transaction in a database and consequently should prevent the occurrence of any duplication, this was not the case with Dataset 2. In fact, the results still showed 33 duplicate candidates even after the introduction of the *Invoice ID* to the algorithm. After consulting the company's internal auditors about this matter, they explained that some invoices were split into multiple payments, and therefore may be related to multiple records in the database.

The median of the duplicate candidates identified by the second set of variables, i.e. the one that included the Invoice ID variable, was found to be much higher than that of the candidates from Set 1. On the other hand, the range from the two sets were very close. Table 31 presents a summary of the descriptive statistics for *Invoice Amount* from both sets.

Table 31-Dataset 2- Descriptive Statistics for Invoice Amount

Descriptive Statistics-Invoice Amount		
	Set 1	**Set 2**
N	899	33
Mean	3830.359	27747.17
Median	206.94	5724.5
Mode	250	25000
Range	136813	130483
Std Deviation	13206.6551	39410.6079
Std Error Mean	440.466608	6860.50625

Variance	174415739	1553196018
Skewness	5.831796	1.57646487
Uncorrected SS	169815000000	75109100000
Corrected SS	156625000000	49702300000
Coefficient Variation	344.788941	142.034706
Sum Observations	3443492.97	915656.53
Kurtosis	39.1054912	1.41501573

The duplicate candidates that resulted from using Set 1 contained negative payments, which were either adjustments or reversals. On the other hand, Set 2 did not yield any negative payments. The lowest value for Set 1's duplicate candidates was for a pair of transactions of the value of -$6250. The lowest value for a duplicate candidate identified by Set 2 was $80.35. On the other hand, the highest value from both sets was $130,563. The highest and lowest three duplicate candidates for both sets can be found in Table 32.

Table 32-Dataset 2-Extreme Observations-Amounts

Extreme Observations			
Lowest		**Highest**	
Set 1	**Set 2**	**Set 1**	**Set 2**
-6250	80.35	130563	130563
-500	100	101805	101805
-500	268.55	100000	75000

The duplicate candidates identified by Set 2 belonged to 13 vendors, each with one or two sets of candidates. However, the duplicate candidates that belong to two of these vendors accounted for over 50% of the total monetary amount from Set 2, which turned out to be $915,656.53. Set 2, on the other hand, resulted in candidates that belonged to 133 different vendors, with candidates ranging from one to fifty three duplicate candidates. One vendor alone accounted for approximately 25% of the total amount of all the duplicate candidates. The sum of the amounts for the top 5 vendors was over 50% of the total amount.

Surprisingly, approximately 13,000 records were identified in the dataset as refunds. This large number of refunds records could indicate systemic process problems or fraudulent activity. The majority of these refunds turned out to be for small customers with low monetary amounts, however there were few outliers that were paid to carriers. Some of these had high amounts, with one that was a $20,000 refund, in addition to four transactions that were in excess of $5000.

5.5. <u>Duplicate Candidates Prioritization</u>

One of the main issues in dealing with large datasets is the amount of information to process. As discussed in the previous chapters, the methodologies and techniques proposed in the continuous auditing literature to identify exceptions are numerous. Unfortunately, the same does not apply to the post-processing stage, where there is a shortage of studies that address that issue (Murthy et.al, 2012). This shortage, in fact, results in human users being inundated with exceptions, which decreases their efficiency tremendously due to their lack of processing and analyzing capabilities. This is especially true when we deal with such large amounts of information (Kleinmutz 1990, Iselin 1988).

As discussed in Sections 5.4.1 and 5.4.2, the company's auditors were overloaded with the duplicate candidates identified by the tests, and there was a clear need to prioritize these results. The method that is usually utilized in the profession is to include additional variables in the algorithm in order to render the results more manageable, in other words to yield less duplicate candidates. This approach is simple but not very efficient as it can lead to an increase in the number of false negatives, and consequently possible losses due to undetected duplicate records. Unfortunately, this is the method that is generally adopted by practitioners, and its popularity stems from its simplicity and ease of use. For instance, the internal auditors of the telecommunication company requested more practicable results, i.e. fewer results. In order to comply with their request, an additional variable was added to the original algorithm. This addition resulted in a dramatic decrease in the number of duplicate candidates, from 899 candidates to 33 candidates only. While this was not the ideal way of dealing with large numbers of exceptions, it was necessary in order to receive the company's feedback. The inefficiency of this approach made it clear that another methodology has to be devised to handle possible large numbers of duplicate candidates.

The prioritization approach proposed in this section is more complicated, yet more efficient and effective, as it addresses large numbers of exceptions by developing a composite score system. The latter is applied on top of the original detection algorithm.

After running the detection algorithm, I propose to use a scoring system in order to prioritize the identified duplicate candidates, rather than simply lowering their number. This composite score is a cumulative score based on certain criteria, some of which are from the original variables that are not usually included in the algorithm.

The first step in the prioritization process starts with the results of the three-way matching methodology generally used in duplicate detection algorithms. This technique includes three variables (or combinations of variables) aimed at identifying the vendor, the date, and the payment. The next step is to use certain criteria that would help to differentiate between possible duplicate candidates in order to rank them. Each candidate will be assigned a composite score defined as the sum of the weights of each criterion as they apply to each candidate. In other words, this composite score is a cumulative weight calculated from the weight of individual criterion.

$$CS_i = \sum W_{iCr_j}$$

Where CS_i is the Composite Score of the set of duplicate candidates *i*

W_{iCr_j} is the weight of criterion *j* when applied to the set of duplicate candidates *i*

Below are the proposed criteria for the prioritization technique:

5.5.1. Materiality

Materiality is simply the relative the monetary amount. This criterion is important in prioritizing the duplicate candidates as it bears a strong and immediate impact on the correctness of the financial numbers. Auditors are usually more interested in transactions with amounts that are close to the materiality threshold. The weight of this criterion is the relative amount, which is the ratio of the Invoice amount to the total amount of all duplicate candidates.

$$W_{i\_Materiality} = (Amt_i)/(\sum Amt_i)$$

where i is the set of duplicate candidates

$W_{i\_Materiality}$ is the weight assigned to the set of duplicate candidates *i* based on their materiality

Amt_i is the monetary amount of the set of supplicate candidates *i*

$\sum Amt_i$ is the total monetary amount for all duplicate candidates.

The ratio criterion is used instead of the absolute amount because relative numbers are better for comparison.

5.5.2. Missing Values

Duplicate candidates that result from missing values are less suspicious than the ones that have non-missing similar values. This is a binary variable that takes the value one if the set of duplicate candidates *i* does not have any missing values, and zero otherwise.

$$W_{i\_MissValue} = \begin{cases} 1/(\sum Count_i), & if \text{ the set of duplicate candidates } i \text{ does not have missing values} \\ 0, & Otherwise \end{cases}$$

Where $W_{i\_MissValue}$ is the weight assigned to the set of duplicate candidates *i* and is equal to 1/($\sum Count_i$) if the set of duplicate candidates *i* does not contain any missing values, and 0 otherwise.

$\sum Count_i$ is the total number of duplicate candidates in the dataset.

The rationale behind this criterion is that when two duplicate candidates are similar because of some missing values for certain variables, it is not certain that these values would have been the same if they were not missing. On the other hand, if the values are similar and not missing, it becomes definite that the candidacy of the two transactions is caused by an actual similarity in the values of those variables.

5.5.3. Count of Similar Candidates

This is the number of transaction that belong to the set of duplicate candidates. The rationale here is that when we have multiple transactions that are similar to each other, the higher the number of these transactions the higher the likelihood of actual duplication. Therefore, the following ratio is included as part of the prioritization methodology:

$$W_{i\_Count} = (Count_i)/(\sum Count_i)$$

Where W_{i_Count} is the weight assigned to the set of duplicate candidates *i* depending on the number of candidates in set *i*

$Count_i$ is the number of candidates that belong to the set of candidates *i*

$\sum Count_i$ is the total number of duplicate candidates in the dataset.

5.5.4. Frequency of the user

This is simply the number of candidates that were created by the same user. Its weight is equal to the ratio of the number of candidates created by that user to the total number of candidates in the dataset.

$W_{i\_FreqUser} = (Count_{U_j i})/(\sum Count_i)$

Where $W_{i\_FreqUser}$ is the weight assigned to the set of duplicate candidates *i* depending on the number of candidates that were created by the User U_j.

$Count_{U_j i}$ is the number of candidates that were created by User U_j, with duplicate candidate *i* being created by the same User U_j.

$\sum Count_i$ is the total number of candidates created by all users

5.5.5. Frequency of the Vendor

This criterion is similar to the previous one, where the weight of the frequency of the vendor is the ratio of the candidates that belong to Vendor V_j to the total number of candidates in the dataset.

$W_{i\_FreqVndr} = (Count_{V_j i})/(\sum Count_i)$

Where $W_{i\_FreqVndr}$ is the weight assigned to the set of duplicate candidates *i* depending on the number of candidates that were paid to Vendor V_j.

$Count_{V_j i}$ is the number of candidates that were paid to Vendor V_j, with the same Vender V_j being the payee of duplicate candidate *i*

$\sum Count_i$ is the total number of candidates created by all users

5.5.6. Duplicate Invoice Number

This criterion depends on whether the duplicate candidates remain as such after the invoice number is included, in other words if the candidates in the duplicate candidates set *i* show the same Invoice ID. It is a binary variable that equals 1/($\sum Count_i$) if the transactions are duplicate candidates with the Invoice ID included, and 0 otherwise.

$$W_{i\_InvID} = \begin{cases} 1/(\sum Count_i), & if\ the\ Invoice\ ID\ is\ the\ same\ for\ the\ candidates \\ 0, & Otherwise \end{cases}$$

Where W_{i_InvID} is the weight assigned to the set of duplicate candidates *i* and is equal to 1/($\sum Count_i$) if the set of duplicate candidates have the same *Invoice ID*, and 0 otherwise.

$\sum Count_i$ is the total number of candidates created by all users

To better understand how the composite score is calculated, let us consider the following example. Below is a set of duplicate candidates that emulates the results of the duplicates detection process.

Table 33-Sample Set of Duplicate Candidates

Record #	Vendor ID	Invoice #	Date	$ Amount	Created by
1001	619505	1241225	5/11/2009	268.55	JDoe
2034	619505	1241225	5/11/2009	268.55	JDoe

9418	619505	1241225	5/11/2009	268.55	JDoe
7430	203339		7/7/2009	4119.5	JSmith
6159	203339		7/7/2009	4119.5	JSmith
8332	552751	1325148	10/5/2009	80.35	JDoe
4723	552751	1279869	10/5/2009	80.35	JDoe

To calculate the Cumulative Score for each candidate, we must first calculate the weight of each criterion as illustrated earlier.

For Record 1001 I calculate the following weights:

- $W_{1001\_Materiality} = (Amt_{1001})/(\sum Amt_i)$ =268.55/ 9205.35 = 0.0292
- $W_{1001\_MissValue} = 1/\ (\sum Count_i) =\ 1/7 = 0.1429$ (as there are no missing values causing it to be a duplicate candidate)
- $W_{1001\_Count} = (Count_{1001})/(\sum Count_i) = 3/7 = 0.4286$
- $W_{1001\_FreqUser} = (Count_{U_j i})/(\sum Count_i) = 5/7 = 0.7143$
- $W_{1001\_FreqVndr} = (Count_{V_j i})/(\sum Count_i) = 3/7 = 0.4286$
- $W_{1001\_InvID} = 1/\ (\sum Count_i) = 1/7 = 0.1429$ (as the Invoice ID are the same for the candidates of set *i*)

As a result, the Composite Score of Record 1001 is the sum of all these weights: CS_{1001}=1.8863. This score will be the same to all the candidates in the set that includes Record 1001, i.e. Records 2034 and 9418

The Table below presents the weights and scores for all the records in the simplified example.

Table 34-Composite Score Calculation

Record #	Score - Materiality	Score -Missing Values	Score - Count	Score - Frequency by User	Score - Frequency by Vendor	Score - Invoice ID	Composite Score
1001	0.0292	0.1429	0.4286	0.7143	0.4286	0.1429	**1.8863**
2034	0.0292	0.1429	0.4286	0.7143	0.4286	0.1429	**1.8863**
9418	0.0292	0.1429	0.4286	0.7143	0.4286	0.1429	**1.8863**
7430	0.4475	0.0000	0.2857	0.2857	0.5714	0.0000	**1.5904**
6159	0.4475	0.0000	0.2857	0.2857	0.5714	0.0000	**1.5904**
8332	0.0087	0.1429	0.2857	0.7143	0.5714	0.0000	**1.7230**
4723	0.0087	0.1429	0.2857	0.7143	0.5714	0.0000	**1.7230**

According to this table, the set of candidates that consists of Records 1001, 2034, and 9418 has the highest Composite Score, and consequently the auditors are recommended to investigate it first. Next is the set that contains Records 8332 and 4723. The auditors would investigate Records 7430 and 6159 last as they had the lowest score.

This methodology would present the duplicate candidates to the auditors in a prioritized way.

5.6. <u>Conclusion</u>

The globalization and electronization of businesses have led to a greater reliance on databases. Numerous systems have been developed to handle and effectively use these databases. Moreover, the proliferation of automated audit and fraud detection systems has increased the popularity of such systems among companies. As a result, huge amounts of data are generated and captured in order to be used in decision support systems. However, the output quality of these systems remains largely dependent on the quality of the underlying data.

Notwithstanding the obvious advantages of these large databases and their importance, databases fed by multiple sources are not costless. Different database systems have different formats, structures, and identifiers. They may also change from a country or region to another. Consequently, there may exist multiple representations of the same real-world object in a company's database, a problem generally known as duplicate record. One type of duplicates records is of special interest to the business world in general, and the accounting world in specific, is the problem of duplicate payments. These can have a serious impact on the quality of audit and fraud detection systems. They can signify the presence of fraud, systematic errors arising from different database systems incompatibilities, or simply human error. There is a plethora of cases in the literature showing the data quality problems and financial losses due to duplicate payments. However, the literature that studies this problem follows an academic approach to the problem of duplicate payments is scarce, despite its significant impact on the business world.

This study described the general problem of duplicate records and presented various techniques used in their detection. Next, it elaborated on duplicate payments and used two datasets from a telecommunication company to illustrate how the duplicate payments detection works and the factors that play a role in this process. The results show that the techniques used for the detection of duplicate payments, while effective in identifying possible duplicates, often generate large numbers of duplicate candidates. Due to such large amounts of candidates and the limited budget and time that companies can allocate for their auditing, it is often impractical to investigate the candidates in their entirety. To solve this issue, a methodology is developed to prioritize the identified candidates based on multiple criteria, prior to presenting them to the auditors. Each candidate is assigned a composite score, which is calculated as the sum of all the scores that each candidate received based on the aforementioned criteria. Subsequently, the auditors can focus their efforts on the cases that have a higher composite score, in other words that are more suspicious.

This study has several limitations. First, it depends on the feedback from the company's internal audit department for the evaluation of the duplicate candidates. This sub-optimal approach, although greatly limited by the time and budget constraints, made it possible to utilize a real business dataset rather than a simulated one. Another related limitation is the fact that the datasets were not labeled, which would have helped to evaluate and compare the performance of the algorithms. Moreover, the lack of labeled data did not allow the validation of the performance of the prioritization methodology.

This study can be extended for future research in various ways. A more general and powerful fuzzy matching technique can be used instead of the simple exact matching used in this study. This would allow for the identification of a wider range of duplicate candidates, such as multiple-payment items, or items for which the date is slightly different. Another possibility for future research is to acquire labeled data in order to compare the performance of various algorithms and the prioritization methodology, even if such data consist of less records.

CHAPTER 6: CONCLUSION

Companies nowadays capture and store large amounts of data up to the most disaggregate transactional data. These readily available huge data warehouses are exploited for various purposes, ranging from support for decision making to exchanging knowledge. However, as a requirement for reliance on such data, it is necessary to provide more frequent and timelier auditing. This type of continuous auditing can ensure a high level of quality of data. There are numerous studies in the continuous auditing literature that propose statistical and machine learning techniques. However, these methodologies often yield large numbers of exceptions, thus inundating the human users with information.

This dissertation is an attempt to fill the gap in the continuous auditing literature with regards to providing some assistance to those users in handling excessive numbers of exceptions. It is comprised of three essays, each addressing a different problem.

The first essay of this dissertation evaluates control risk assessments conducted by internal auditors and business owners. First, it identifies the cases that deviate from the expected value of a logit model. Subsequently, a prioritization technique is proposed based on two disagreement measures. Results show that the null hypothesis, where the presence of our independent variable does not have any effect, can be safely rejected. In other words, the inclusion of the auditors and business owners' categorization of control issues has a strong impact on the overall control risk level. The results also indicate that the model was effective as a review tool, as it exposed all the cases that did not conform to the expected value. Moreover, it showed the level of disagreement between the auditors or business owners' assessment and that of our predictive model. The exception prioritization methodology proposed in this study can improve audit efficiency as it helps auditors and business owners to focus their investigations on the cases with higher disagreement levels. In addition to that, the model can be used as a learning technique. It can provide non-experts with the knowledge extracted from the knowledge of experts, as it can give the probability of the level of control risk as assigned by the experts. Lastly, the methodology proposed can be used as a consistency check by serving as a benchmark.

In the second essay I develop a framework that uses a rule-based expert system to identify Order-to-Cash records that violate certain analytics generally used by auditors. As business rules do not have the same importance, the suspicion a record earns from violating a certain rule should depend on its importance. Therefore, the framework describes a prioritization technique that proposes to rank exceptions according to a suspicion score, which is calculated from the weights of the rules that these exceptions have violated. This study uses an expert panel that consists of 17 auditors, both internal and external, with experience in control risk assessments. Using an online survey, the participants are asked to conduct pairwise comparisons of transactions that violate internal control rules and to provide a ranking within each pair. These rankings are used in turn to infer the weights of the rules that make up the expert system. Subsequently, the suspicion score of each exception is calculated by summing the weights of the analytics it violated. The exceptions are then prioritized and ranked in order of decreasing suspicion score prior to presenting them to auditors for further investigation. The last step in the framework is the feedback from the auditors' investigations into the expert system and the weighting system. The results indicate that the participants agreed most of the time (76.11%) on the riskier transaction within each pair. Moreover, the high level of correctly identifying the violated rule (over 85%) indicates that the panel can in fact be considered an expert panel. The rules that were considered the riskiest were the ones whose violations had a direct impact on the financial numbers of the company and entailed direct losses, as opposed to the violation of operational controls. The results also show differences in the way internal auditors and external auditors weighed the rules. However, due to the small size and unbalanced sample size (eleven internal auditor vs. six external auditors) may not be statistically significant. Still, it was interesting to find that internal auditors assigned the violations of segregation of duties the lowest weights.

The third essay of this dissertation discusses the problem of duplicate records and presents various techniques used to identify them. More specifically, it focuses on the problem of duplicate payments, using two real life datasets from a telecommunication company, covering two years of payment transactions. A three-way exact matching algorithm is used to identify payments that had the same date, amount, and vendor. The results indicate that while such algorithms are effective in identifying duplicate candidates, the latter are usually too numerous for auditors to investigate in their entirety. To solve this problem, this study proposes a methodology that prioritizes the identified candidates according to certain criteria, including materiality, duplication due to missing values, number of candidates in a set of candidates, frequency of the user who created these candidates, frequency of candidates for a certain vendor, and finally records that are duplicate candidates despite the presence of the invoice number. Based on these criteria, each candidate is awarded a composite score that can be used to provide the auditors with ranked candidates in order of decreasing suspicion, hence assisting them by directing them to the more problematic cases. The results also confirmed the existence of duplicate payments, which proves that the controls often implemented by companies to mitigate such violations can fail, especially when multiple systems are involved.

This dissertation has several limitations. First, Essay 1 shows unbalanced datasets where the number if critical issues is a lot less than major and non-major issues. Moreover, the criteria used by the company to categorize identified control issues is unknown. Finally, the weight of the ordinal variables, such as the difference between non-major and major issues, are unknown. Such information, if present can be included in the model to improve its accuracy.

Second, the second essay had a small expert panel system, which can affect the statistical significance of some of my results, although the 17 participant-panel falls within the acceptable range as recommended by the Delphi methods. In addition to that, the expert system consisted of only a subset of the analytics, which were considered as the most important by experienced auditors. Future research can address these issues by using a larger expert panel size and including the complete set of analytics in the expert panel. Another research venue is to test the framework using a real-business dataset to evaluate the overall performance of the expert system and the prioritization technique.

Finally, the results of Essay 3 depended largely on the feedback from the company's internal auditors, which proved to be lengthy and limited in scope due to time and budget constraints. Future research can address this issue through the use of a labeled data, although this solution may prove to be complicated. However, such solution can address another shortcoming of this essay, which is testing the effectiveness of the prioritization technique.

Despite these limitations, this dissertation fills a gap in the continuous auditing literature, and addresses a problem that has great implications on the auditing profession, especially with databases exponentially increasing to huge sizes and the growing need for timelier auditing of these data. This dissertation contributes to the literature by showing the necessity of prioritization techniques. Moreover, it proves that developing such methodologies can improve overall audit efficiency by pointing the auditors towards the more interesting exceptions, in other words, the Exceptional Exceptions.

REFERENCES

Allegrini, M., & D'Onza, G. (2003). Internal auditing and risk assessment in large Italian companies: an empirical survey. *International Journal of Auditing*, *208*(October 2002), 191–208. Retrieved from http://onlinelibrary.wiley.com/doi/10.1046/j.1099-1123.2003.00070.x/full

Alles, M., Brennan, G., Kogan, A., & Vasarhelyi, M. (2006). Continuous monitoring of business process controls: A pilot implementation of a continuous auditing system at Siemens. *International Journal of Accounting Information Systems*, *7*(2), 137–161.

Alles, M., Kogan, A., & Vasarhelyi, M. (2008). Putting continuous auditing theory into practice: Lessons from two pilot implementations. *Journal of Information*, *22*(2), 195–214.

Alles, M., Vasarhelyi, M., & Issa, H. (2012). Rethinking the Practice and Value Added of External Audits: The AICPA's Audit Data Standards (ADS) Initiative. Retrieved from http://systemsthinking.nl/ISAIS 2013/Alles - Issa - Vasarhelyi.pdf

Alpaydin, E. (2004). *Introduction to machine learning*. The MIT Press. Retrieved from http://books.google.com/books?hl=en&lr=&id=1k0_-WroiqEC&oi=fnd&pg=PR13&dq=Introduction+to+Machine+Learning&ots=p9-J_OgEuQ&sig=ZC_85OD6A4nclMONlxNmJ2IZSAk

Ashton, R. (1974). An experimental study of internal control judgements. *Journal of Accounting Research*. Retrieved from http://www.jstor.org/stable/10.2307/2490532

Baldwin-Morgan, A. (1993). The impact of expert system audit tools on auditing firms in the year 2001: a Delphi investigation. *Journal of Information Systems*. Retrieved from http://dialnet.unirioja.es/servlet/articulo?codigo=344163

Bernhard, A. (2012, September 20). How Big Data brings BI, predictive analytics together. Retrieved from http://www.cio.com/article/716726/How_Big_Data_Brings_BI_Predictive_Analytics_Together

Bitton, D. (1983). Duplicate record elimination in large data files. *ACM Transactions on database systems (*, *8*(2), 255–265. Retrieved from http://portal.acm.org/citation.cfm?id=319987&dl=

Bou-Raad, G. (2000). Internal auditors and a value-added approach: the new business regime. *Managerial Auditing Journal*. Retrieved from http://www.emeraldinsight.com/journals.htm?articleid=868461&show=abstract

Brown, C. (1991). Expert systems in public accounting: Current practice and future directions. *Expert Systems with Applications*.

Calderon, T. G., & Cheh, J. J. (2002). A roadmap for future neural networks research in auditing and risk assessment. *International Journal of Accounting Information Systems*, *3*(4), 203–236. doi:10.1016/S1467-0895(02)00068-4

Carcello, J., & Hermanson, D. (2005). Changes in internal auditing during the time of the major US accounting scandals. *Journal of Auditing*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1099-1123.2005.00273.x/abstract

Chan, D., & Vasarhelyi, M. (2011). Innovation and practice of continuous auditing. *International Journal of Accounting Information …*, *12*(2), 152–160.

Chatterjee, A., & Segev, A. (1991). Data manipulation in heterogeneous databases. *ACM SIGMOD Record*, *20*(4), 64–68. Retrieved from http://portal.acm.org/citation.cfm?id=141356.141385

Chou, C., Du, T., & Lai, V. (2007). Continuous auditing with a multi-agent system. *Decision Support Systems*, *42*(4), 2274–2292. doi:10.1016/j.dss.2006.08.002

CICA/AICPA. (1999). *Continuous Auditing Research Report*. Toronto, Canada: The Canadian Institute of Chartered Accountants.

Committee of Sponsoring Organizations of the Treadway Commission. (1992). Internal Control--integrated Framework: Evaluation tools. *Committee of Sponsoring Organizations of the Treadway Commission.*, *4*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Internal+Control--integrated+Framework:+Evaluation+tools#0

Cushing, B. (1974). A mathematical approach to the analysis and design of internal control systems. *The Accounting Review*. Retrieved from http://www.jstor.org/stable/10.2307/244795

Dalkey, N, & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management science*.

Dalkey, NC, Brown, B., & Cochran, S. (1969). *The Delphi method: An experimental study of group opinion*. Retrieved from http://www.rand.org/content/dam/rand/pubs/research_memoranda/RM5888/RM5888.pdf

Davis, J., & Massey, A. (1997). Supporting a complex audit judgment task: An expert network approach. *European journal of operational*, *7*(Ic). Retrieved from http://www.sciencedirect.com/science/article/pii/S0377221797001252

Debreceny, R., Gray, G. L., Tham, W., Goh, K., & Tang, P. (2003). The Development of Embedded Audit Modules to Support Continuous Monitoring in the Electronic Commerce Environment, *185*(February), 169–185.

Dijk, J. Van, & Pellenbarg, P. (2000). Firm relocation decisions in The Netherlands: An ordered logit approach. *Papers in Regional Science*. Retrieved from http://www.springerlink.com/index/PUYYEY4GW7YY6TFY.pdf

Dull, R., Tegarden, D., & Schleifer, L. (2006). ACTVE: a proposal for an automated continuous transaction verification environment. *Journal of Emerging Technologies in*, *3*(1), 81–96.

Elliott, R. (1998). Assurance services and the audit heritage. *CPA JOURNAL*, *68*(6), 40.

Elliott, R., & Kielich, J. (1985). Expert systems for accountants. *Journal of Accountancy*.

Elmagarmid, A., Ipeirotis, P., & Verykios, V. (2007). Duplicate Record Detection: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, *19*(1), 1–16. doi:10.1109/TKDE.2007.250581

Fellegi, I. (1969). A theory for record linkage. *Journal of the American Statistical Association*, *64*(328), 1183– 1210. Retrieved from http://www.jstor.org/stable/2286061

Fineberg, S. (2013, March 19). SAP Releases Fraud Management and Detection Engine. Retrieved from http://www.accountingtoday.com/accounting-technology/news/sap-unveils-fraud-detection-management-product66086-1.html?ET=webcpa:e6837:410064a:&st=email

Gaumnitz, B., & Nunamaker, T. (1982). Auditor consensus in internal control evaluation and audit program planning. *Journal of Accounting …*. Retrieved from http://www.jstor.org/stable/10.2307/2490898

Gramling, A. (2006). Assessing internal audit quality, (June). Retrieved from http://digitalcommons.kennesaw.edu/facpubs/1332/

Groomer, S., & Murthy, U. (1989a). Continuous auditing of database applications: An embedded audit module approach. *Journal of Information Systems*, *3*(2), 53–69.

Groomer, S., & Murthy, U. (1989b). Continuous auditing of database applications: An embedded audit module approach. *Journal of Information Systems*, *3*(2), 53–69. Retrieved from http://www.mendeley.com/research/continuous-auditing-database-applications-embedded-audit-module-approach/

Hayes-Roth, F. (1985). Rule-based systems. *Communications of the ACM*.

Hernandez, M., & Stolfo, S. (1995). The merge/purge problem for large databases. *Proceedings of the 1995 ACM SIGMOD*. Retrieved from http://portal.acm.org/citation.cfm?id=223807&dl=GUIDE,

IIA. (2000). Institute of Internal Auditors (IIA) (2000), Internal Auditing: Adding Value across the Board, Corporate Brochure, IIA.

IIA. (2002). The Institute of Internal Auditors, (2002), Professional Practice Framework (The Standards for the Professional Practice of Internal Auditing), The Institute of Internal Auditors.

IIA. (2004). Institute of Internal Auditors. 2004. Internal auditing's role in sections 302 and 404 of the U.S. Sarbanes Oxley act of 2002. Altamonte Springs: Institute of Internal Audit.

Iselin, E. (1988). The effects of information load and information diversity on decision quality in a structured decision task. *Accounting, Organizations and Society*, *13*(2), 147–164.

Kimball, R. (2009). The data warehouse toolkit. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:the+data+warehouse+etl+toolkit#0

Kinney Jr, W. (2000). Research opportunities in internal control quality and quality assurance. *Auditing: A Journal of Practice & Theory*. Retrieved from http://www.aaajournals.org/doi/abs/10.2308/aud.2000.19.s-1.83

Kleinbaum, D. G., & Klein, M. (2010). Logistic Regression. *Statistics*. doi:10.1007/978-1-4419-1742-3

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin; Psychological Bulletin*, *107*(3), 296–310.

Kogan, A., Sudit, E., & Vasarhelyi, M. (1999). Continuous online auditing: a program of research. *Journal of Information Systems*, *13*(2), 87–103.

Kogan, A., Vasarhelyi, M., & Wu, J. (2010). Continuous Data Level Auditing Using Continuity Equations. *Unpublished working paper, Rutgers …*.

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*. Retrieved from http://www.mendeley.com/research/binary-codes-capable-of-correcting-insertions-and-reversals/

Martin, J., & Eckerle, R. (1991). A knowledge-based system for auditing health insurance claims. *Interfaces*.

McMullan, M. (2001). Duplicate Medicare Payments by Individual Carriers. *Public Law*.

Menon, K., & Williams, D. (2001). Long-term trends in audit fees. *Auditing: A Journal of Practice & Theory*.

Nagy, A. L., & Cenker, W. J. (2002). An assessment of the newly defined internal audit function. *Managerial Auditing Journal*, *17*(3), 130–137. doi:10.1108/02686900210419912

Newcombe, H. (1988a). Handbook of record linkage: methods for health and statistical studies, administration, and business. Retrieved from http://portal.acm.org/citation.cfm?id=63465

Newcombe, H. (1988b). Handbook of record linkage: methods for health and statistical studies, administration, and business. Retrieved from http://dl.acm.org/citation.cfm?id=63465

Newcombe, H., Kennedy, J., & Axford, S. (1959). Automatic linkage of vital records. *Science*. Retrieved from http://www.festivalofdoubt.uq.edu.au/papers/Newcombe.pdf

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2010). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*. doi:10.1016/j.dss.2010.08.006

Norris, C. M., Ghali, W. A., Saunders, L. D., Brant, R., Galbraith, D., Faris, P., & Knudtson, M. L. (2006). Ordinal regression model and the linear regression model were superior to the logistic regression models. (Spss Inc, Ed.)*Journal of Clinical Epidemiology*, *59*(5), 448–456. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16632132

Novello, A. C. (2004). Duplicate Medicaid Transportation Payments, 1–4. Retrieved from http://www.osc.state.ny.us/audits/allaudits/093004/04f2.pdf

Oz, E. (2006). *Management Information Systems*. *multimedia systems* (5th Editio.). Boston, Massachussets: Thomson Course Technology.

Parente, F., & Anderson, J. (1984). An examination of factors contributing to Delphi accuracy. *Journal of …*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/for.3980030205/abstract

PCAOB. (2004). Public Company Accounting Oversight Board (PCAOB).2004. Auditing Standard No. 2. An Audit of Internal Control over Financial Reporting Performed in Conjunction with an Audit of Financial Statements.

PCAOB. (2007). Public Company Accounting Oversight Board (PCAOB). 2007. Auditing Standard No. 5 - an audit of internal control over financial reporting that is integrated with an audit of financial statements.

Perols, J. L., & Murthy, U. S. (2012). Information Fusion in Continuous Assurance. *Journal of Information Systems*, *26*(2), 35–52. doi:10.2308/isys-50216

Protiviti. (2009). Moving internal audit back into balance - a post SOX survey. Retrieved from http://www.protiviti.com/go/rebalancingsurvey#fourth.

Ratley, J. (2012). Report to the Nations: On occupational fraud and abuse. Retrieved from http://www.acfe.com/uploadedFiles/ACFE_Website/Content/rttn/2012-report-to-nations.pdf

Redman, T. C. (1997). *Data quality for the information age*. Norwood, MA,: Artech House, Inc. Norwood, MA, USA. Retrieved from http://portal.acm.org/citation.cfm?id=548570

Rezaee, Z., Sharbatoghlie, A., Elam, R., & McMickle, P. (2002). Continuous auditing: Building automated auditing capability. *Auditing*, *21*(1), 147–163. doi:10.2308/aud.2002.21.1.147

Roth, J. P. (1995). COSO Implementation Guide, Altamonte Springs, IIARF.

Sarens, G. (2009). Internal auditing research: Where are we going? Editorial. *International Journal of Auditing*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1099-1123.2008.00387.x/full

SAS No. 65. (1991). Ibid. and SAS No. 65 !AU Section 322), The Auditor's Consideration of the Internal Audit Function in an Audit of Financial Statements., (April).

Srinidhi, B., & Vasarhelyi, M. (1986). Auditor judgment concerning establishment of substantive tests based on internal control reliability. *Auditing: A Journal of Practice & …*. Retrieved from http://raw.rutgers.edu/MiklosVasarhelyi/Resume Articles/MAJOR REFEREED ARTICLES/M12. auditor judgment.pdf

Srivastava, D., Dar, S., Jagadish, H., & Levy, A. (1996). Answering queries with aggregation using views. *PROCEEDINGS OF THE*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.315&rep=rep1&type=pdf

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. Retrieved from http://www.aaai.org/ojs/index.php/aimagazine/article/viewArticle/1169

Tepping, B. (1968). A model for optimum linkage of records. *Journal of the American Statistical Association*, *63*(324), 1321– 1332. Retrieved from http://www.jstor.org/stable/2285887

The Real-Time Economy. (2002, January 31). *The Economist*.

Turban, E. (1990). *Decision support and expert systems: management support systems*. Retrieved from http://dl.acm.org/citation.cfm?id=542112

U.S. House of Representatives. (2002). The Sarbanes-Oxley Act of 2002. Public Law 107-204 [H. R. 3763]. Washington, D.C.: Government Printing Office.

Vasarhelyi, M. (1980). A Taxonomization of internal Controls and Hrrors for Audit Research. Retrieved from http://raw.rutgers.edu/MiklosVasarhelyi/Resume Articles/OTHER PUBLICATIONS/L13. a taxonomization of internal.pdf

Vasarhelyi, M., Alles, M., & Williams, K. (2010). *Continuous assurance for the now economy*. Sydney, Australia: Institute of Chartered Accountants in Australia.

Vasarhelyi, M., Chan, D., & Krahel, J. (2012). Consequences of XBRL standardization on financial statement data. *Journal of Information …*. Retrieved from http://aaajournals.org/doi/abs/10.2308/isys-10258

Vasarhelyi, M., & Halper, F. (1991). The continuous audit of online systems. *Auditing: A Journal of Practice & Theory*, *10*(1), 110–125.

Verykios, V., Elfeky, M., & Elmagarmid, A. (2000). On the accuracy and completeness of the record matching process. . *of the 2000 Conf. on*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.18.7629

Vijayan, J. (2012, September 12). Finding the business value in big data is a big problem. Retrieved from http://www.computerworld.com/s/article/9231224/Finding_the_business_value_in_big_data_is_a_big_problem

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 95. Retrieved from http://portal.acm.org/citation.cfm?id=240455.240479

Wang, Y. R., Madnick, S. E., & Horton, D. C. (1989). Inter-database instance identification in composite information systems. *Conference on System*, (June), 677–684. doi:10.1109/HICSS.1989.49184

Warner, C. (2013). Duplicate Payment Detection. Retrieved from www.autoaudit.com

Warner, P. (2008). Ordinal logistic regression. *The journal of family planning and reproductive health care / Faculty of Family Planning & Reproductive Health Care, Royal College of Obstetricians & Gynaecologists*, *34*(3), 169–70. doi:10.1783/147118908784734945

Waterman, M., & Smith, T. (1976). Some biological sequence metrics* 1. *Advances in Mathematics*. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/0001870876902024

Weis, M., & Naumann, F. (2005). DogmatiX tracks down duplicates in XML. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data* (p. 442). ACM. Retrieved from http://portal.acm.org/citation.cfm?id=1066207

Weis, M., Naumann, F., Jehle, U., & Lufter, J. (2008). Industry-scale duplicate detection. *Proceedings of the VLDB*, *1*(212), 1253–1264. Retrieved from http://portal.acm.org/citation.cfm?id=1454159.1454165

Wood, T., Brown, W., Howe, H., & Vallabhaneni, S. (1999). Control self assessment: For risk management and other practical applications. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Control+Self+Assessment:+for+Risk+Management+and+Other+Practical+Applications#0

Wu, F., & Hahn, R. (1989). A control-complexity and control-point orientation to the review of an entity's internal control structure in the computer environment. *Journal of Information Systems*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+control-complexity+and+control-point+orientation+to+the+review+of+an+entity's+internal+control+structure+in+the+computer+environment#0

APPENDICES

Appendix A

Pairwise Comparisons:

Following is a table showing all the pairwise comparisons as well as a graphical representation of the compared analytics:

Table 35- Comparisons of Analytics

Analytic Number	Comparison 1	Comparison 2	Comparison 3	Comparison 4
Analytic 1 SOD Customers	Analytic 3 Unauthorized Price	Analytic 4 SOD Credit Adjustment		
Analytic 2 Unauthorized Sales Order	Analytic 6 Match Invoice to Ship	Analytic 8 Unauthorized Shipments	Analytic 4 SOD Credit Adjustment	
Analytic 3 Unauthorized Price	Analytic 1 SOD Customers	Analytic 5 Match Shipping to SO		
Analytic 4 SOD Credit Adjustment	Analytic 1 SOD Customers	Analytic 2 Unauthorized Sales Order		
Analytic 5 Match Shipping to SO	Analytic 6 Match Invoice to Ship	Analytic 8 Unauthorized Shipments	Analytic 9 SOD Ship Invoice	Analytic 3 Unauthorized Price
Analytic 6 Match Invoice to Ship	Analytic 5 Match Shipping to SO	Analytic 10 Orphaned Invoices	Analytic 12 Excessive Write Offs	Analytic 2 Unauthorized Sales Order
Analytic 7 Missing Sales Orders	Analytic 9 SOD Ship Invoice	Analytic 10 Orphaned Invoices		
Analytic 8 Unauthorized Shipments	Analytic 2 Unauthorized Sales Order	Analytic 5 Match Shipping to SO	Analytic 11 SOD Invoice Receipt	
Analytic 9 SOD Ship Invoice	Analytic 5 Match Shipping to SO	Analytic 10 Orphaned Invoices		
Analytic 10 Orphaned Invoices	Analytic 6 Match Invoice to Ship	Analytic 9 SOD Ship Invoice	Analytic 11 SOD Invoice Receipt	Analytic 7 Missing Sales Orders
Analytic 11 SOD Invoice Receipt	Analytic 8 Unauthorized Shipments	Analytic 10 Orphaned Invoices	Analytic 12 Excessive Write Offs	
Analytic 12 Excessive Write Offs	Analytic 6 Match Invoice to Ship	Analytic 11 SOD Invoice Receipt		

Appendix B

Research Instrument:

`

Informed Consent

You are invited to participate in a research study examining control risk assessment. As you are aware, assessing control risk often requires substantial judgment. In the following section we ask you to analyse audit evidence to assess related control risk. The project is being conducted by Hussein Issa (Ph.D. candidate - Rutgers University) under the supervision of Helen L. Brown-Liburd (Assistant Professor - Rutgers University). You are asked to complete a realistic case that requires about 20-30 minutes of your time. Your participation is vital to the success of the study.

The case elicits your professional judgments from the perspective of an auditor; accordingly, please use the same high level of care you would exercise when actually resolving a financial reporting issue in practice. We are looking for YOUR INDIVIDUAL JUDGMENTS, so please DO NOT CONFER WITH COLLEAGUES when completing the case. Also it is necessary that you complete the case WITHIN ONE SITTING when you are able to do so uninterrupted.

In the following section, we provide you with multiple internal control scenarios designed to solicit your judgment regarding the level of control risk attributed to each scenario. We will provide you with important contextual information for each scenario and then ask you to make a risk assessment. Finally, we will ask some brief follow-up questions.

When reporting the results of the study your identity will remain anonymous; this means that we will not record any information about you that could identify you. Therefore, there will be no way to link your responses back to you.

Your participation in this study is voluntary, and you may choose to discontinue participation at any time. If you have any questions, please contact Hussein Issa at (732)600-5895 or via email at issah@rutgers.edu. If you have any questions regarding your rights as a research participant, please call the Rutgers University Office of Research & Sponsored Programs at 848-932-0150.

By taking this survey, you consent to being a participant in this research study.
Please complete the study within the next two weeks.
Thank you for your participation in this important research study.

Hussein Issa (Ph.D. candidate- Rutgers University)
Helen L. Brown-Liburd (Assistant Professor - Rutgers University)

I have read, understood, and printed a copy of, the above consent form and desire of my own free will to participate in this study.

○ **Yes**

In the following section, we provide you with multiple internal control scenarios designed to solicit your judgment regarding the level of control risk attributed to each scenario. The AICPA defines control risk as:

"Control Risk is the risk that a material misstatement will not be detected or prevented by the entity's internal control on a timely basis. The auditor must consider the risk of misstatement individually and in aggregate with other misstatements."

We will provide you with important contextual information for each scenario and then ask you to perform pairwise comparisons of a set of transactions in order to identify the transaction that represents the highest control risk within each pair. Subsequently, we ask you to justify your assessment.

Each scenario is to be evaluated INDEPENDENTLY from the other. In other words, previous scenarios should NOT be taken into consideration when assessing the control risk associated with the current transactions.

Thank you.

○ **I read and understood the task I am asked to perform**

Question 1

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Created by	Sales Order	Sales Order Date	Sales Order Approved by	Sales order Created by	Inventory Item ID	Quantity	Unit Selling Price	Amount
Transaction 1	CUID1146	1028	27624	6/6/1997	1028	1006	63	30	$800	$24,000
Transaction 2	CUID1006	1028	41766	8/4/1998	1547	1318	628	10	$2,327	$23,270

Background Information:

Price List

Inventory Item ID	List Price
63	$800.00
339	$26.00
490	$445.76
548	$1,393.00
628	$2,399.00
756	$82.19

○ Transaction 1

○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify)

Question 2:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Created by	Sales Order Number	Sales Order Date	Sales Order Created by	Sales Order Approved by	Ordered Amount	Customer Credit Limit	Credit Limit Created by	Credit Limit Updated by
Transaction 1	CUID1003	1547	27624	7/23/1998	1318	1547	$1,969,000	$10,000,000	1546	1070
Transaction 2	CUID1146	1028	41766	8/4/1998	1318	1068	$2,094,327	$2,500,000	1546	1068

Background Information:

Customer Credit Limit

Customer ID	Credit Limit	Created by	Updated By
CUID1000	$10,000,000	1028	1028
CUID1001	$2,500,000	1028	1028
CUID1002	$5,000,000	1546	1070
CUID1003	$10,000,000	1546	1070
CUID1004	$2,500,000	1068	1068
CUID1146	$2,500,000	1546	1068

○ Transaction 1

○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify) ____________

Question 3:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Sales Order Number	Sales Order Date	Sales Order Approved by	Sales Order Created by	Inventory Item ID	Quantity	Invoice Unit Price	Amount	Shipment ID	Shipping Document Unit Price
Transaction 1	CUID1146	10002391	7/5/1998	1068	1318	63	30	$800	$24,000	10000877	$860
Transaction 2	CUID1006	10007624	8/4/1998	1068	1547	628	10	$2,327	$23,270	10002391	$2,327

Background Information:

List of Users Authorized to Create Sales Orders

1068
1286
1318
1634
1952

○ Transaction 1

○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify)

Question 4:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Sales Order Number	Sales Order Date	Sales Order Created by	Sales Order Approved by	Inventory Item ID	Quantity	Unit Selling Price	Amount	Shipment ID	Shipment Document Created by
Transaction 1	CUID1146	10001203	6/6/1997	1547	1028	63	30	$800	$24,000	10000877	1286
Transaction 2	CUID1006	10000262	8/4/1998	1318	1634	628	10	$2,327	$23,270	10002391	1117

Background Information:

Users Authorized to Create Sales orders	*Users Authorized to Create Shipping Lines*
1068	1003
1286	1286
1318	1364
1634	1547
1952	1729

○ Transaction 1

○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify)

Question 5:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Created by	Sales Order Number	Sales Order Date	Sales Order Approved by	Sales Order Created by	Inventory Item ID	Quantity	Sales Order Unit Selling Price	Sales Order Amount	List Unit Price	Shipping Document Unit Price
Transaction 1	CUID1146	1028	10002391	6/6/1997	1326	1006	63	30	$800	$24,000	$800	$860
Transaction 2	CUID1006	1028	10007624	8/4/1998	1547	1318	628	10	$2,327	$23,270	$2,399	$2,399

○ Transaction 1

○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify)

Question 6:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Sales Order Number	Sales Order Date	Sales Order Approved by	Sales Order Created by	Ordered Amount	Customer Credit Limit	Credit Limit Created by	Credit Limit Updated by
Transaction 1	CUID1146	41766	7/23/1998	1634	1318	$1,969,000	$10,000,000	1952	1634
Transaction 2	CUID1006	27624	8/4/1998	1286	1547	$2,094,327	$2,500,000	1952	1068

Background Information:

List of Users Authorized to Create Sales Orders

1068
1286
1318
1634
1952

○ **Transaction 1**

○ **Transaction 2**

Please select from the list below the most appropriate reason for your assessment:

○ **Missing Values** ○ **Non-Matching values** ○ **Segregation of Duties** ○ **Unauthorized Transaction**

○ **Other (please specify)** ______

Question 7:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Transaction ID	Sales Order Number	Sales Order Unit Selling Price	Sales Order Date	Ordered Amount	Invoice Number	Invoice Unit Price	Shipment ID	Shipping Document Unit Price
Transaction 1	CUID1146	10034	10001203	$800	10/15/1997	$9,600	10001203	$860	10000877	$860
Transaction 2	CUID1000	14110	10000262	$600	3/1/1998	$1,800	10000262	$675	10002391	$600

○ Transaction 1

○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify)

Question 8:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Transaction ID	Sales Order Number	Sales Order Unit Selling Price	Sales Order Date	Ordered Amount	Shipment ID	Shipment Document Created by	Shipping document price
Transaction 1	CUID1146	10034	10001203	$800	10/15/1997	$9,600	10000877	1286	$860
Transaction 2	CUID1000	14110	10000262	$600	3/1/1998	$1,800	10002391	1117	$600

Background Information

List of Users Authorized to Create Shipping Lines

1068
1286
1318
1634
1952

- ○ Transaction 1
- ○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify)

Question 9:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Transaction ID	Sales Order Number	Sales Order Unit Price	Invoice Number	Invoice Date	Invoice Created by	Inventory Item ID	Invoiced Amount	Shipment Document Created by	Shipment Document Approved by	Shipping Document Unit Price
Transaction 1	CUID1146	10034	41766	$800	10001203	10/15/1997	1546	63	$9,600	1002	1117	$860
Transaction 2	CUID1000	14110	39269	$600	10000262	3/1/1998	1117	628	$1,800	1002	1117	$600

○ Transaction 1
○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction
○ Other (please specify)

Question 10:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Transaction ID	Invoice Number	Invoice Created by	Invoice Date	Inventory Item ID	Invoice Unit Price	Invoiced Amount	Shipment ID	Shipping Document Unit Price
Transaction 1	CUID1146	10034	10001203	1546	10/15/1997	63	$800	$9,600	10000877	$860
Transaction 2	CUID1000	4110	10000262	1117	3/1/1998	628	$600	$1,800		

- ◯ Transaction 1
- ◯ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

◯ Missing Values ◯ Non-Matching values ◯ Segregation of Duties ◯ Unauthorized Transaction

◯ Other (please specify)

Question 11:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Transaction ID	Invoice Number	Invoice Created by	Invoice Date	Inventory Item ID	Invoice Unit Price	Invoiced Amount	Shipping Document Unit Price	Entered in AR by	Adjusted in AR by	AR Adjustment Approved by	AR Adjustment type	AR Adjustment Amount
Transaction 1	CUID1146	10034	10001203	1546	10/15/1997	63	$800	$9,600	$860	1318	1008	1002	Write off	$10.05
Transaction 2	CUID1000	4110	10000262	1117	3/1/1998	628	$600	$1,800	$600	1068	1318	1405	Write off	$450.93

Background Information:
Write-offs greater than 15% must be approved by user **1002**.

○ Transaction 1

○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify) ______

Question 12:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Transaction ID	Invoice Number	Invoice Date	Invoiced Amount	Invoice Created By	Invoice Approved by	Receipt Number	Received Apply Date	Entered in AR by	Shipment ID	Shipment Document Created by
Transaction 1	CUID1002	10034	10001203	10/15/1997	$9,600	1572	1318	56897	2/10/1998	1318	10000877	1286
Transaction 2	CUID1146	14110	10000262	3/1/1996	$1,800	1894	1318	75963	4/2/1996	1068	10002391	1117

Background Information:

List of Authorized to Create Records in AR	*List of Authorized to Approve Invoices*	*List of Users Authorized to Create Shipping Lines*
1068	1152	1068
1286	1318	1286
1318	1431	1318
1634	1572	1634
1952	1894	1952

○ Transaction 1

○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify) ________

Question 13:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Transaction ID	Invoice Number	Invoice Date	Invoice Created by	Inventory Item ID	Invoice Unit Price	Invoiced Amount	Shipment Document Createdby	Shipment Document Approved by	Shipment ID
Transaction 1	CUID1146	10034	10001203	10/15/1997	1546	63	$800	$9,600	1002	1546	10000877
Transaction 2	CUID1000	4110	10000262	3/1/1998	1117	628	4600	$1,800			

○ Transaction 1

○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify)

Question 14:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Transaction ID	Invoice Number	Invoice Date	Invoiced Amount	Invoice Created By	Invoice Approved by	Receipt Number	Received Apply Date	Entered in AR by	Shipment_ID
Transaction 1	1146	10034	10001203	10/15/1997	$9,600	1546	1318	56897	12/10/1997	1318	10000877
Transaction 2	1000	4110	10000262	3/1/1996	$1,800	1117	1028	75963	4/2/1996	1068	

○ Transaction 1

○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify)

Question 15:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Transaction ID	Invoice Number	Invoice Date	Invoiced Amount	Invoice Created By	Invoice Approved By	Receipt Number	Received Apply Date	Entered in AR by	Adjusted in AR by	Adjustment Approved by	Adjustment type	Adjustment Amount
Transaction 1	CUID1146	10034	10001203	10/15/1997	$9,600	1546	1318	56897	2/10/1998	1318	1008	1002	Write off	$10.04
Transaction 2	CUID1000	14110	10000262	3/1/1996	$1,800	1117	1028	75963	4/2/1996	1068	1318	1045	Write off	$864.93

Background Information

Write-offs greater than 15% must be approved by user **1002.**

○ **Transaction 1**

○ **Transaction 2**

Please select from the list below the most appropriate reason for your assessment:

○ **Missing Values** ○ **Non-Matching values** ○ **Segregation of Duties** ○ **Unauthorized Transaction**

○ **Other (please specify)**

Question 16:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Transaction ID	Sales Order Number	Sales Order Date	Invoice Number	Invoice Date	Invoice Created by	Inventory Item ID	Invoice Unit Price	Invoiced Amount	Shipment ID	Shipment Document Created by	Shipment Document Approved by
Transaction 1	CUID1146	10034	41766	8/15/1997	10001203	10/15/1997	1546	63	$800	$9,600	10000877	1002	1546
Transaction 2	CUID1000	14110			10000262	3/1/1998	1117	628	$600	$1,800	10001324	1002	1318

○ Transaction 1

○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify)

Question 17:

Please select from the following two transactions the one that presents the highest level of control risk:

	Customer ID	Customer Transaction ID	Invoice Number	Invoice Date	Invoice Created by	Sales Order Number	Sales Order date	Invoiced Amount	Shipment ID	Shipment Document Created by	Shipment Document Approved by
Transaction 1	CUID1146	10034	10001203	10/15/1997	1546			$9,600	10000877	1002	1546
Transaction 2	CUID1000	14110	10000262	3/1/1998	1117	41767	8/16/1997	$1,800			

- ○ Transaction 1
- ○ Transaction 2

Please select from the list below the most appropriate reason for your assessment:

○ Missing Values ○ Non-Matching values ○ Segregation of Duties ○ Unauthorized Transaction

○ Other (please specify) ______

Post Experiment Questionnaire

The pairwise comparison of transactions section is completed. We will now ask you a few background questions.

○ **Continue to the background questions**

Do you feel that you had enough data to perform the required task?

Far too Little						Far too Much
○	○	○	○	○	○	○

Did you find the task to be:

Unmotivating						Challenging
○	○	○	○	○	○	○
Extremely easy						**Extremely Difficult**
○	○	○	○	○	○	○

Please select degrees obtained:

☐ A.S./A.A. ☐ B.S./B.A. ☐ M.S./M.A. ☐ MPA/MSA ☐ MBA ☐ Ph.D. ☐ Other ______

Please select the professional designation:

☐ CPA ☐ CIA ☐ CMA ☐ CFA ☐ CFE ☐ EA ☐ CISA ☐ Other ______

Professional working experience

Years of professional working experience (in general)	
Current position	
Years of professional working experience in IT Audit	
Years of professional working experience in auditing Financial Statements	
Years of professional working experience in assessing controls risk	
Years of professional working experience in external auditing	
Years of professional working experience in internal auditing	

Are you currently working as External or Internal Auditor?

- ◯ External Auditor
- ◯ Internal Auditor

Have you ever worked on audit engagements in an online environment?

- ◯ Yes
- ◯ No

How do you rate your knowledge of Audit Analytics?

Not at all Knowledgeable						Extremely Knowledgeable
◯	◯	◯	◯	◯	◯	◯

Continuous Auditing and Continuous Control Monitoring Knowledge:

	Not at all Knowledgeable						Extremely Knowledgeable
How do you rate your knowledge of continuous auditing?	◯	◯	◯	◯	◯	◯	◯
How do you rate your knowledge of continuous control monitoring?	◯	◯	◯	◯	◯	◯	◯

<u>Please express any other comments you wish here:</u>

Thank you for taking time out to participate in our survey. We truly value the information you have provided. Your responses are vital to our research.

Appendix C

List of variables for Datasets 1 and 2

List of variables in Dataset 1:

Account ID

Payment Id

Carrier Name

Transaction Type

Effective Date

Entered Date

Amount

Source Info

Payment Number

List of variables in Dataset 2:

BATCH_NAME

BATCH_DATE

ORG_NAME

ORG_ID

CHECK_NUMBER

CHECK_AMOUNT

CHECK_DATE

VENDOR_NAME

VENDOR_TYPE

VENDOR_PAY_GROUP

APPLIED_INVOICE_AMOUNT

INVOICE_ID

INVOICE_NUM

INVOICE_DESCRIPTION

INVOICE_DATE

AMOUNT_PAID

INVOICE_AMOUNT

PAYMENT_GL_DATE

PAYMENT_GL_PERIOD

ADDRESS_LINE1

ADDRESS_LINES_ALT

ADDRESS_LINE2

ADDRESS_LINE3

DISTRIBUTION_LINE_NUMBER

APPLIED_DIST_AMOUNT

DIST_DESCRIPTION

QUANTITY_INVOICED

DIST_GL_DATE

DIST_GL_PERIOD

DIST_CREATION_DATE

DIST_CREATED_BY

DIST_CREATED_BY_NAME

DIST_UPDATE_DATE

DIST_UPDATE_BY

DIST_UPDATE_BY_NAME

ASSETT_CATEGORY

SEGMENT1

SEGMENT2

SEGMENT3

SEGMENT4

SEGMENT5

SEGMENT6

SEGMENT7

SEGMENT8

SEGMENT9

DISTRIBUTION_ACCT_DESCRIPTION

BANK_ACCOUNT_NAME

BANK_ACCOUNT_NUM

TERM_NAME

CHECK_STATUS

CHECK_DESCRIPTION

Appendix D

Candidates Counts per Carrier - Dataset 1

Table 36-Candidates Counts-Set 1

Carrier	**Amount**	**Count**
1	1579013	2
2	1270000	26
3	431179.1	2
4	344238.5	2
5	196185.2	2
6	110000	2
7	100832.3	4
8	100000	2
9	91050	5
10	85860.26	2
11	58082.8	2
12	49539.84	6
13	27610.42	2
14	21526.46	5
15	12057.48	2
16	11053.82	2
17	7391.78	2
18	5707.5	2

19	3059.6	2
20	1217.56	2
21	715.2	2
22	33.3	2
23	0	2

Table 37-Candidates Counts-Set 2

Carrier	**Amount**	**Count**
1	7540000	29
2	5300000	16
3	3540000	14
4	2800000	7
5	1579013	2
6	1130000	24
7	1090000	10
8	520385.6	2
9	344238.5	2
10	267001.2	2
11	171157.3	2
12	114348.2	2
13	100000	2

14	98300.56	2
15	91050	5
16	85860.26	2
17	62336.18	2
18	58082.8	2
19	49685.62	2
20	49539.84	6
21	29711.68	2
22	27610.42	2
23	21756.22	2
24	21526.46	5
25	12057.48	2
26	11053.82	2
27	9801.98	4
28	5707.5	2
29	3059.6	2
30	2762.98	2
31	2762.98	2
32	1698.02	2
33	1217.56	2
34	33.3	2

Curriculum Vitae

Hussein Issa

1974	Born in Tyre, Lebanon
1987-1991	Cadmous High School, Baccalaureate
1992-2000	TAT-USA, Lebanon; Administrative Assistant
2000-2003	IUL University, Lebanon; Bachelor of Business Administration
2004-2006	Bargny Calcaire, Senegal; Production Assistant Manager
2006-2008	Emirates National School, UAE; Administrative Supervisor
2008-2009	United Software Consultants, USA; IT consultant
2009-2013	Rutgers University; Ph.D. in Accounting Information Systems