JOURNAL OF EMERGING TECHNOLOGIES IN ACCOUNTING
Vol. 9
2012
pp. 95–110

American Accounting Association
DOI: 10.2308/jeta-50411

A Model to Detect Potentially Fraudulent/ Abnormal Wires of an Insurance Company: An Unsupervised Rule-Based Approach

Yongbum Kim
Ramapo College of New Jersey

Miklos A. Vasarhelyi
Rutgers, The State University of New Jersey

ABSTRACT: Fraud prevention/detection is an important function of internal control. Prior literature focused mainly on fraud committed by external parties, such as customers. However, according to a survey by the Association of Certified Fraud Examiners (ACFE 2009), it was noted that employees posed the greatest fraud threat. This study proposes profiling fraud using an unsupervised learning method. The fraud detection model is based on potential fraud/anomaly indicators in the wire transfer payment process of a major insurance company in the United States. Each indicator is assigned an arbitrary score based on its severity. Once an aggregate score is calculated, those wire transfer payments whose total scores are above a threshold will be suggested for investigation. Our contribution is to report what we have learned and to document our findings using fraud/anomaly indicators to detect potential fraud and/or errors on real data from a major insurance company.

Keywords: continuous auditing; continuous monitoring; anomaly detection; fraud detection; unsupervised learning.

INTRODUCTION

The term fraud has been defined in many ways (American Institute of Certified Public Accountants [AICPA] 2002a, SAS 99; ACFE 2004; Winn 2004; Lectric Law Library). Despite slightly different viewpoints, one critical attribute found in common is that fraud is the intentional illegal theft from an organization for personal gain. Due to a series of financial scandals in the late 1990s and early 2000s, the AICPA issued SAS 99, which defined fraud and

We express our gratitude to anonymous reviewer(s), participants of the 18th Annual Research Workshop on Strategic and Emerging Technologies, the 19th World Continuous Auditing and Reporting Symposium, International Conference on Digital Forensics and Reporting Symposium in 2009, the 17th World Continuous Auditing and Reporting Symposium, the 20th World Continuous Auditing and Reporting Symposium, and the 18th Annual Research Workshop on Strategic and Emerging Technologies. We also thank the anonymous insurance company for providing the transactional data in this study.

Corresponding author: Miklos A. Vasarhelyi
Email: miklosv@rutgers.edu
Published Online: January 2013

categorized it into two types: fraudulent financial reporting and misappropriation of assets (AICPA 2002a). The former is misrepresentation of financial reports, such as earnings manipulation by falsifying accounting records and/or omitting transactions. In contrast, the latter occurs when assets are stolen or fraudulent expenditures are claimed. The ACFE's definition of fraud is broader than SAS 99. It includes bribery and corruption in addition to the two types defined by SAS 99 (ACFE 1996).

External fraud is committed by an external party, while internal fraud is by an employee. The framework of Jans et al. (2009) suggested three classifications for internal fraud: (1) statement or transaction fraud, (2) management or non-management fraud, and (3) fraud for or against a company. In this study, fraud or internal fraud will mean internal transaction fraud against a company committed by either management or non-management.

Fraud occurs only when fraudsters have incentives/pressures, opportunities to commit fraud, and rationalization to justify their behavior (SAS 99). Fraud-related activities are generally categorized into two groups: fraud prevention and detection. The former can be achieved by removing at least one of the conditions for fraud materialization. For example, fraud can be prevented by removing any incentives/pressures of a potential fraudster (also called fraud perpetrator). Also, if an enterprise's internal control system is sufficiently effective, it will be difficult for fraudsters to find an opportunity to commit fraud. Last, fraud can be mitigated by educating employees to have business ethics so that fraudsters cannot justify their actions easily. Most prevention methods are, however, difficult to implement and evaluate. For example, the incentives/pressures for fraud may be costly and difficult to control due to their qualitative characteristics. The effect of ethical education is difficult, if not impossible, to measure. As a result, a well-designed internal control system seems to be the only practical way to implement and evaluate fraud prevention and detection measures.

The Association of Certified Fraud Examiners (ACFE) estimated that the cost of occupational fraud and abuse (hereafter, "internal fraud") was approximately $994 billion in the U.S., which represents a loss in revenue of about 8 percent to businesses in 2007, and $660 billion (6 percent loss of revenue) in 2004. In 2009, the ACFE noted that intense financial pressures of the current economic crisis have caused an increase of fraud, and that employees posed the greatest fraud threat (48.3 percent increase in employee embezzlement from the previous year) (ACFE 2007). This increase in fraud may indicate ineffective internal controls and a lack of fraud detection/prevention systems. A company's internal control system (ICS) is a crucial factor for detecting and preventing fraud. A properly designed ICS facilitates reliable financial information by preventing, detecting, and correcting potentially material errors and irregularities on a timely basis. Fraud committed by employees has received little attention in the literature, while fraud by outsiders, such as customers, has been greatly researched. This might be partly due to lack of data and partly due to fear of losing competitive advantage (Bolton and Hand 2002; Phua et al. 2005). However, recent financial scandals have clearly showed that fraud by employees affects a company's revenue more adversely than that by outsiders does.

As fraud by employees (or internal fraud) becomes more emphasized, it is timely to examine how an enterprise can prevent and detect fraud. This study proposes and tests an unsupervised rule-based model that utilizes the transactional data of a major insurance company to check for fraud committed by employees.

The rest of the paper proceeds as follows. In the second section, we provide a literature review on fraud detection and prevention methods used in prior research. The third section, the methodology section, will discuss the data and the model used in this study, followed by the result and findings. And finally, the fourth section concludes by summarizing this study and discussing future research.

LITERATURE REVIEW

Fraud Detection and Prevention as a Way of Continuous Auditing

The Auditing Concepts Committee (American Accounting Association [AAA] 1972) defined auditing as "a systematic process of objectively obtaining and evaluating evidence regarding assertions about economic actions and events to ascertain the degree of correspondence between those assertions and established criteria and communicating the results to interested users." This definition clearly shows that the focus of auditing is verification of assertions made by managements, including the proposed financial reports (Alles et al. 2004). Continuous auditing (CA) broadens this concept by proposing timelier assurance, generally with less aggregated (or transactional) data than traditional auditing.

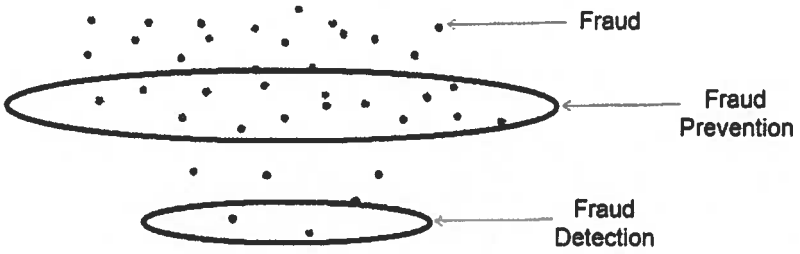
Vasarhelyi and Halper (1991) first introduced the concept of continuous auditing when they developed a monitoring tool in an online IT environment. Its rationale is to provide more timely assurance by continuously monitoring a company's entire transactional data. This suggestion did not draw much attention from either academia or practice for a decade due to skepticism about its feasibility and effectiveness. It is not a long time ago that continuous auditing became a hot issue in both academia and practice. After a series of recent financial scandals such as the WorldCom and Enron crises, and related audit failures such as Arthur Andersen, researchers, practitioners, and regulators have looked for possible solutions to prevent future financial disasters. Continuous auditing is believed to be the most promising means, so that recently, this area has drawn much attention from both researchers and practitioners.

Although much work has been published, the majority of papers on CA have focused on the technical perspectives of CA (Vasarhelyi and Halper 1991; Kogan et al. 1999; Woodroof and Searcy 2001; Rezaee et al. 2002; Murthy 2004; Murthy and Groomer 2004). A few papers discuss other aspects of CA, such as its concepts and research directions (Alles et al. 2002, 2004; Elliott 2002). Moreover, only a handful of papers (Alles et al. 2004, 2006) focused on empirical studies on CA due to a lack of available data. Traditional auditing studies can use publicly available aggregated data, while research on CA requires disaggregate/transactional data, which is typically kept private by companies in order to maintain a competitive position in a market. Hence, it is not surprising that companies are reluctant to provide their transactional data to researchers. However, empirical studies are an indispensable component of CA research in order to justify and verify its practical feasibility. The nature of fraud prevention/detection is similar to that of continuous auditing, whose aim is to detect and correct anomalies in a timely manner. In other words, fraud prevention/detection is a part of continuous auditing.

Fraud, undetected, can cause enormous losses to a corporation. Although finding fraud can be like looking for a needle in a haystack, companies have found that fraud control activities can more than repay the cost of initiating and running such a program (Major and Riedinger 2002). Fraud-related activities are generally categorized into two groups: fraud prevention and detection. A major challenge with fraud detection and prevention is that known positive cases are rarely documented (Major and Riedinger 2002). There are no public databases of known fraud available. This can be due to numerous factors, such as that companies who discover fraud do not want to disclose it to the public because such activity can cause reputational and/or financial damage. This may lead to loss of competitive advantage or a wrong conception of an easy target. Another problem with public disclosure is that if fraud perpetrators gain knowledge about what type of fraud is being monitored, they will proceed with alternate methods to elude detection. In fraud detection, it is desirable for fraudsters to make simple mistakes leading to their detection.

Fraud detection will be initiated to identify fraud when fraud prevention fails (Figure 1). Fraud is elusive, as a company can never be sure that no fraud exists within its business. Nevertheless, it is

FIGURE 1
Anti-Fraud Mechanism



prudent to reduce this risk by using fraud prevention and detection to actively monitor business processes. However, it is not cost effective to check each and every business transaction for fraud, given that companies have limited resources. As a result, a cost-efficient method such as mathematical algorithms applied to data might be an effective and efficient way to capture possible evidence of fraud (Phua et al. 2005). Data mining is a technique that is often discussed in research and used in practice to detect fraud by using mathematical algorithms. Data mining methods typically provide outliers/anomalies that can be investigated further by internal auditors. However, algorithms producing too many outliers/alarms can adversely affect their effectiveness because of too many false positives. On the other hand, a model producing too few alarms is not desirable because it may suffer from too many false negatives. In fraud detection, it is a well-known fact that a false negative error is usually more costly than a false positive error (Phua et al. 2005). Taken together, a fraud detection model should generate a reasonable number of exceptions for investigation by balancing its effectiveness and efficiency. In general, we are unaware that the fraud prevention control has failed. Consequently, fraud detection should be continuously applied regardless of existence of fraud prevention methods (Bolton and Hand 2001, 2002). Another notable attribute of fraud detection is that fraud detection methods must be updated and applied continuously. The relationship between fraud and fraud detection is like that between a computer virus and an antivirus program. While known computer viruses can be effectively detected and corrected by antivirus software, sooner or later, undetectable viruses will be introduced. Unless the antivirus software is updated to adapt the new computer viruses, its detection power will be in question. At the same time, the antivirus must keep the detection ability to capture the known viruses since a computer system can be attacked by the known viruses. In other words, a fraud detection method must be highly adaptive to detect the new types of fraud by existing fraudsters or newcomers while keeping the current detection ability to prevent the identifiable types (Bolton and Hand 2001; Winn 1996).

Supervised and Unsupervised Methods of Fraud Detection

Supervised Methods

Generally, there are two methods used in the literature to detect fraud: supervised and unsupervised methods. The most frequently used research methodology in academia is classification or supervised methods. The supervised methods utilize prior information (also called labeled information) that contains both legitimate and fraudulent transactions, while the unsupervised methods do not require any labeled data. Under the supervised methods, a database of known fraudulent or legitimate cases is used to construct models used to detect fraud (Bolton and

Hand 2002). The models are trained by prior labeled data, and then fraudulent and legitimate transactions are discriminated in accordance with those models. These methods assume that the pattern of fraud in the future will be the same as that in the past. In recent research, neural network models that use the supervised methods seem to be the most commonly used (Bolton and Hand 2002; Kou et al. 2004; Phua et al. 2005).

Although popular in research, supervised methods pose several limitations resulting from their heavy dependence on reliable prior knowledge about both fraudulent and legitimate transactions. This may be impractical, since prior information might be incorrect. This is mainly due to the fact that most companies do not have sufficient resources to examine every transaction to make sure they are all free from fraud. Consequently, some of those transactions labeled as legitimate are likely fraudulent, so that fraud detection models based on the information may be misleading (Bolton and Hand 2002). Another limitation of the supervised methods is that the results are often not easily understood. This may be a substantial obstacle to implement fraud detection models, since few enterprises can afford expertise for it (Sherman 2002). As a result, few enterprises would be interested in implementing the supervised methods in practice. This is similar to analytical review procedures used by auditors, where many sophisticated methods have been developed, but simple methods are dominantly used in practice. Another limitation is that a supervised model is not easily adjustable.

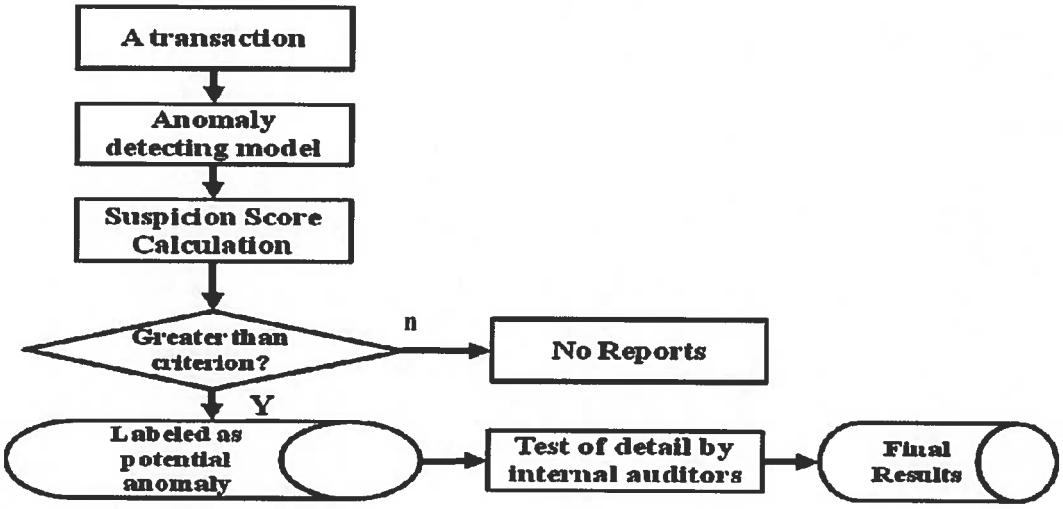
A major concern relative to fraud prevention/detection research is that models may work only for the data that are used in creating the models. The generalizability of fraud profiles is highly dependent on the context of the original model development and on the target environment. For example, if new data come into the dataset, those models may not work due to either over-fitting to the training dataset or unknown fraud types. In addition, the robusticity of models is of major concern in their extension, re-utilization, and adaptation. Considering that fraud perpetrators adapt to find loopholes of an enterprise's current fraud prevention/detection system, this can be a critical weakness. In order to adapt to unknown types of attacks, it is important that the systems should be extendable and adjustable dynamically. Last, but not least, the supervised methods suffer from uneven class sizes of legitimate and fraudulent observations. Generally, the number of fraudulent observations is greatly outnumbered by that of legitimate ones. About 0.08 percent of annual observations turned out to be fraudulent (Hassibi 2000). In other words, even if a model classifies all fraudulent transactions as legitimate regardless of their true identities, the error rate (= the number of correctly classified transactions/the total number of transactions) of the model is extremely small, which can be misleading.

Unsupervised Methods

Unsupervised methods have received less attention in literature than supervised methods. They focus on detection of changes in behaviors or unusual transactions (or outliers) by using data mining methods. Anomaly/outlier detection is the finding of patterns in data that do not conform to expected behavior (Chandola et al. 2009). The major advantage of unsupervised methods is that they do not require labeled information. Labeled information is generally unavailable due to censorship (Bolton and Hand 2002; Kou et al. 2004; Phua et al. 2005). The results are not disclosed in public, either to maintain an enterprise's competitive advantage or because of public benefits, such as the case of the U.S. crop insurance program (Little et al. 2002).

Unsupervised methods usually employ suspicion scoring systems that estimate the degree of departure from the norm. Suspicion scores may be calculated by utilizing if-then type of outlier rules. Rule-based systems are increasingly used to represent experiential knowledge. The criteria to be used to determine whether a transaction is an outlier may change for various reasons, such as cost and efficiency. Decision making by if-then rules is similar to a human's cognitive decision

FIGURE 2
Anomaly Detection Process



processes, which enables internal auditors to understand and adjust the models if necessary. However, this method has a major drawback because verification/evaluation of the newly devised models is often difficult, if not impossible, due to lack of testable data availability. To tackle this weakness, methods such as peer group analysis, where groups with similar profiles are compared, and break point analysis, where recent transactions are compared with past patterns, are used (Bolton and Hand 2001).

The results of unsupervised methods are not direct evidence that flagged transactions are fraudulent. Instead, the aim of unsupervised methods is to inform that flagged transactions have more propensities to anomalies that are either errors or fraud, based on the experience/analysis/preconceptions of the analysts. In other words, a flagged transaction can be legitimate, error, or fraudulent. This is clearly distinctive from that of supervised methods, whose outcome is either legitimate or fraudulent. As Jans et al. (2009) described, an outlier can occur by errors or mistakes. It can be said that unsupervised methods consider broader causes than the supervised ones do. Furthermore, a transaction will be worthy of further investigation if it is flagged by multiple criteria, since normal transactions are unlikely to be flagged by many indicators. Analogous to other rule-based systems, the actual examination of selected transactions allows for re-parameterization and improvement of the method. However, the verification of resulting flagged transactions requires internal auditors' direct examination (Figure 2).

Despite several drawbacks of unsupervised methods, they may be indispensable at the initial implementation stage, where prior labeled information is rarely available. In addition, considering that it is ultimately internal auditors that will use and maintain fraud prevention/detection models, and that only a few enterprises can afford the expertise necessary for them, a rule-based approach may be desirable for internal auditors (Sherman 2002). In this study, an unsupervised method is utilized due to the fact that the company does not have any prior records of fraud, its fraud detection system is at the initial stage, and the internal auditors do not have much knowledge in statistical analyses. The pros and cons of supervised and unsupervised methods are summarized in Table 1.

TABLE 1

Pros and Cons
Supervised versus Unsupervised methods

	Supervised Methods	Unsupervised Methods
Pros	1. Accurate for known fraud types 2. Results: fraudulent or not	1. Easy to apply and update 2. Unknown fraud can be found 3. Do not need both types of observations 4. Results: worthy of further investigation/ attention
Cons	1. Suffers from highly unbalanced class sizes (1 out of 1,200) 2. False negatives 3. May work only for the known fraud types 4. Highly dependent on historic data that may not be accurate. 5. Less understandability	1. Less accurate, but as accurate as complex methods in the long term 2. Needs verification processes by internal auditors

The next section discusses the implementation of these concepts in the context of the wire transfers of a major U.S. insurance company.

METHODOLOGY

Overview

This pilot study involves a major U.S. insurance company that is proceeding toward developing a continuous audit/fraud detection process. For this purpose, it was decided that a research team would cooperate with the internal audit organization to develop basic modeling and analysis methodologies in parallel with their internal audit process. The project plan entailed a set of progressive steps in the development of an automated discrepancy detection process. Once the process and models are developed, the data extraction process will be made frequently and systematically, progressing toward more frequent data screening to monitor potential fraud. The model proposed is similar to an external stand-alone system that is used to extract and analyze data for exceptions in continuous auditing (Vasarhelyi and Halper 1991; Searcy and Woodroof 2001; Rezaee et al. 2002; Murthy and Groomer 2004). The wire payment data are extracted from the production legacy information systems and analyzed externally. This is beneficial since running an automatic fraud detection system can be intensive on the production system, which might cause the system to operate suboptimally. In Pathak et al. (2005), they find auditing transactions in batches was more cost effective than initiating periodic audits after a certain period of time. Our model proposes that the fraud detection occurs in batches. Before an audit, the internal auditors can extract the desired data and run the fraud detection mechanism. Any resulting exceptions can be investigated during their regular audit.

The wire transfer process was chosen as a desirable first target due to: (1) data availability, (2) the frequency of the process and its importance as a major business activity, (3) the availability of knowledgeable and competent internal audit staff for knowledge engineering, and (4) the opportunity of the timing of the audit. According to the company's audit team, control over the wire transfer payment process was less systematically controlled as were the other processes in place, and they tried to find a more systematic and effective way to prevent and detect any fraudulent wire

TABLE 2

Potential Fraud Indicators
Trend Tests

Potential Fraud Indicators	Possible Screening Rules To Test
The payee transaction payment amount is out of the range of payment amounts.	Amount range for each payee (or all payees) & check outliers.
The payee transaction payment trend line over time has a positive slope.	Correlation between date (or sequence numbers) and payee amounts for each payee.
The payee is an outlier to payee baseline activity. (Send to a payee that normally does not send to.)	Payee frequency by each initiator and check the payees that have the least frequencies.
The initiator/approver transaction payment amount is out of the range of baseline payment amounts.	First, check the transaction amounts with their authorization amounts. Second, calculate 90, 95, or 99 PI. And then find the transactions that are beyond these bounds.
The transaction amount is out of range of normal activity from this bank account.	The 90, 95, and 99 PI amounts for each sending/ receiving bank account and check the exceptions.
The transaction initiator is not a normal sender from this bank account.	First, check the list of sender bank account, then create exception list of initiators by sending bank account.
The transaction payee is not a normal receiver from this bank account.	A list of payees by sending banks who have the least frequency.
Access to the bank account is commingled with many other types of transactions.	A list of bank accounts with wire types that have the least frequency.

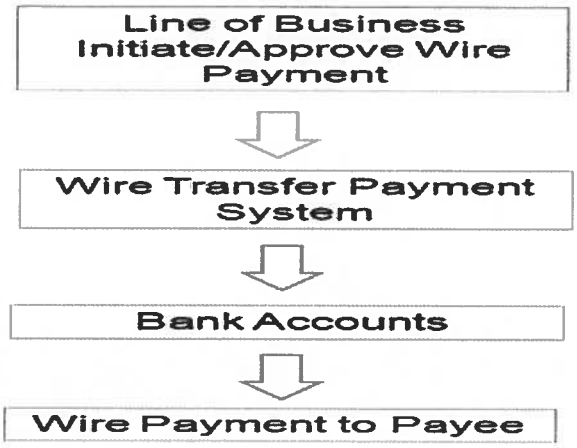
transfers. Furthermore, the company did not have documented historical information about past fraud occurrences. As a result, there is an assumption that there was no known fraud in the past. However, this lack of past experience about fraud existence does not necessarily mean that there was no fraud in the wire transfer process. Consequently, it was appropriate that the fraud detection and prevention model should be based on the unsupervised method. The objective was to create a statistical model to detect potential anomalies within wire payment transactions. Internal auditors would further investigate selected transactions for anomalies.

The indicators, which are individual fraud detection rules, were divided into target and trend indicators. The target indicators are pass/fail-type tests that do not require statistical calculations. For example, target tests examine whether the receiver is located in a country known as a financial safe harbor, and whether the transaction date by the initiator or approver is after the initiator's termination date or before the initiator's hire date. On the contrary, trend indicators are tests that are applied after statistical calculations. Our research objective was to focus solely on the trend type indicators by creating and running statistical algorithms that can be used to detect abnormal behavior or patterns in wire transfer payments (Table 2).

Wire Transfer Payment System

Generally, the wire transfer process in this company consists of three stages: initiation, approval, and settlement (or payment). Depending on the nature of the wires, certain types of wires

FIGURE 3
Wire Transfer Payment System



require more than one approval, while one approval is sufficient for others. Once the wire payment is approved, the wire payment is imported into the wire transfer payment system (Figure 3).

Wires processed within the wire transfer system can belong to one of four types: random, repetitive, concentration, and batch. These categories are based on the number of payments and on operational effectiveness and efficiency. The wire transfer system also has specific controls in place where users are assigned to specific bank account groups, and users assigned to specific transaction groups can only process certain transaction types.

Data Description

The dataset in this study consisted of wire transfer payments that were made by the company. The data spanned one year and consisted of over 225,000 wire payments paid to over 10,000 payees. Approximately 90 percent of the wire transfer payments belonged to approximately 10 percent of the payees. More interestingly, 62.82 percent of the payees were involved in only one transaction, while 93.84 percent of the payees were involved in less than 30 transactions. The insurance company provided the dataset in the form of seven tables (or files). The table primarily used in the study was the Wires table of 27 attributes whose records span from October 2007 to September 2008. The number of transactions becomes 229,531 after removing irrelevant records such as totals (12 of them). The other six tables are master files that are referenced by the Wires table. The master files kept employee information such as start date, status of employment, rank status, and authorization limits. These attributes are mainly used to check employees' authorization limits. The descriptive statistics for four numeric attributes—wire amounts, initialization limits, approval limits, and settlement limits—are shown in Table 3.

Model Development Process

The overall model development consists of five stages based on data mining methods. Once relevant information is collected, an initial brainstorming process will be performed by discussion with internal auditors. This process will determine the potentially relevant areas that need to generate possible indicators/rules. Once a model with a collection of rules is complete, transactional

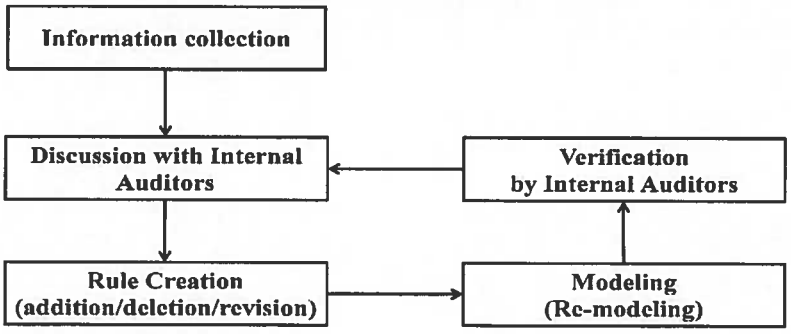
TABLE 3
Descriptive Statistics

	All_Wires	Transactions by Group		
	AMOUNT	*APPLIM*	*INITLIM*	*SETTLELIM*
n	229,531	8,239	8,239	8,239
Nmiss	0	0	0	0
Avg.	4,793,957	167,685,975	80,232,688	606,870
Median	70,242	10,000,000	0	0
Std.	79,213,746	452,077,500	325,524,629	24,628,745
Min.	0	0	0	0
Max.	13,260,787,693	9,814,999,869	7,806,759,586	1,000,000,000
Range	13,260,787,693	9,814,999,869	7,806,759,586	1,000,000,000

data will be tested with the model. Each indicator is assigned an arbitrary score based on the severity of the potential fraud indicated by the indicator. Once a particular wire transfer payment is washed through different indicators and scored, an aggregate total will be calculated and those wire transfer payments above a threshold will be suggested for investigation. The resultant observations will be verified by internal auditors. The verification results will be discussed and used to update the model. This process will be iterated until satisfactory results are found. The most important feature of the development process is that it is iterative and interactive. The overall process is shown in Figure 4.

The initial phase of the study encompassed obtaining a general understanding of the company's wire transfer payment system and the data. Understanding the system and its internal controls is important to facilitate the creation of indicators and algorithms to supplement and support the controls in place. To understand the data, data characteristics and layouts were obtained. In addition, descriptive statistics were calculated, including basic statistics such as mean, average, range (min/max), distributions, etc. This gave a general quantitative understanding of the data and the types of wire transfer payments being made. Next, the research team and the internal IT audit

FIGURE 4
Overall Model Development Process



( ): After the first round.

team brainstormed ideas for indicators that might potentially illuminate anomaly/outlier transactions. The indicators were based on the perceptions that the anomalies/outliers produced might be meaningful as fraud indicators. These indicators were transformed into statistical algorithms that utilized statistical data mining techniques. Generally, the indicators mainly consisted of three types of statistics: prediction, correlation, and frequency test. Using these types of statistics on the data allowed the determination of anomalies or patterns. Each indicator was assigned a score arbitrarily based on the severity of risk that the indicator might pose toward potential fraud: a score of one for low risk of fraud, three for moderate, and five for high risk of fraud. The assigned scores were based on the professional judgment of the internal audit department. After running the wire transfers through the different indicators, the suspicion scores were aggregated to determine what total score should be used as the cutoff/threshold for further investigation by the internal audit team. Upon completion of the investigation, the internal audit department will verify whether the flagged transactions were fraudulent. In addition, internal audit will suggest how to improve the model and the indicators. The model should constantly evolve to adapt to new findings from actual experience from application. Considering the persistent nature of fraud, the fraud detection/prevention process should be continuously run and updated. The target tests performed by the company are not discussed in this study in order to prevent harming the insurance company's fraud detection efforts.

Trend Indicators

The three statistical algorithms used are (1) prediction interval test, (2) correlation test, and (3) frequency test.

Prediction Interval Test

The prediction interval test involved stratifying payees into four categories: (1) payees with one wire payment, (2) payees with two wire payments, (3) payees with three to 29 wire payments, and (4) payees with 30 or more wire payments. Wires were stratified by the number of observations for statistical interpretations. The stratification was made for payees, initiators, and approvers. In addition, alternative alpha prediction intervals of 90 percent, 95 percent, and 99 percent were also considered. A higher alpha level will have fewer outliers and, *vice versa*, a lower alpha level will have more outliers. For payees with only one wire payment, a prediction interval was estimated by grouping the payee's wire payment together with other payees who have one wire payment in order to determine which payments were abnormal compared to the group as a whole. The prediction interval was applied to payees with 30 or more wires. For payees who have only two wire payments and for payees with three to 29 payments, the prediction interval test is not available for statistical validity. Future research may consider some type of statistical clustering method to detect outliers for these groups.

Correlation Test

The correlation test entailed the examination of how each payee's payment amounts increased or decreased in a logical manner that is not typical of other payee transaction history. The approach called activity monitoring (Fawcett and Provost 1999) is adopted for this type of test, which requires the maintenance of a usage profile for each payee or employee in order to determine any deviation in activity. In contrast to the prediction interval test, the minimum number of required observations to calculate correlation was three, so that stratification was made into two groups: more than or equal to three wires and less than three wires. The degree of overall increase of wire amounts is determined by the correlation value and its p-value for its statistical significance. In the

literature, various correlation values are suggested to decide whether observations are positively correlated. However, more than 0.3 is generally regarded as positively correlated. However, to be more conservative, this study uses the threshold 0.5 that is also commonly used.

Frequency Test

Depending on what are considered typical or normal activity patterns, we can determine which wire payments are anomalies or outliers. Use of the frequency test in statistics can be helpful in determining what typical or normal activity patterns are. Infrequent activities may indicate potential errors or fraud. The frequency tests indicator entailed examining each payee and the employee initiating wire payments to determine which pairs had unusual activity. For example, unusual activity can be a payee interacting with an uncommon employee or group of employees for the first time. Based on initial statistics during the initial phase of our study, payees typically encounter many different initiators and approvers in the company. It follows that encounters with only the same initiator or approver would not be considered normal.

Scoring System

The scoring of the indicators was developed with the assistance of the internal IT audit department. The knowledge engineering of experienced professionals (Vasarhelyi and Halper 1991) allows for the determination of types of indicators to be considered abnormal or potentially fraudulent in nature. An effective internal audit team may have the ability to identify indicators that suggest fraud (AICPA 2002b). Each indicator was assigned a score based on the perceived risk of the indicator. After each indicator is processed through the statistical algorithms, a total for each payee is tallied for indicators that were violated. Those payees' wire transfers that violated a certain total aggregate score were subject to investigation by the internal auditors. However, in practice, it is difficult for internal auditors to allocate a large amount of time to investigate exceptions. In running the initial algorithms, an enormous number of exceptions were found. Kogan et al. (1999) discuss the cognitive effect of information overload. An overload of alarms will have a negative effect on the internal audit department to adopt our fraud detection system. As a result, to make this pilot study feasible, it was necessary to increase the threshold score in order to investigate the transactions with the highest scores. The summary statistics of the aggregated scores are in Figure 5.

The number of exceptions was narrowed by applying thresholds 11 for the trend test score, 25 for the target test score, and 22 for the total score, and the resultant 47 flagged transactions are more manageable and feasible for internal auditors' investigation.

Results and Newly Emerging Issues

Results

The internal audit team investigated the 47 wire transfer payments during their normal audit. No evidence was found to support that the suggested wires were either fraudulent or erroneous. Although the fraud detection/prevention model did not find fraudulent wire payments, this does not mean that there were no anomalous payments. Instead, this may indicate that the model failed to detect an existing anomaly and, thus, calls for revision and fine-tuning of the model. The company intends to include the fraud detection process as a part of regular audit, so those indicators will be kept in place for future detection or preventive measures. In addition, the company is interested in refining the indicators and adding new ones to screen for anomalies. As a matter of fact, the company should consistently reevaluate and revise the model, considering the highly adaptive nature of fraud perpetuators.

FIGURE 5
Suspicion Scores and Their Thresholds for Further Investigation

score_trend	cnt_wires
0	183534
1	25933
2	3141
3	5179
4	1271
5	266
6	1005
7	8217
8	707
9	209
10	58
11	9
12	2

Score_target	cnt_wires
0	195948
5	18841
10	14401
15	334
25	7

Score_total	cnt_wires	Score_total	cnt_wires
0	163304	13	67
1	23204	14	18
2	2437	15	139
3	3562	16	187
4	964	17	7106
5	14962	18	538
6	2652	19	136
7	1490	20	41
8	1717	21	7
9	361	22	27
10	5406	23	2
11	897	24	1
12	299	25	6
		32	1

Internal Control Issues

During the study, the effectiveness of the company's internal control came into question. The major internal control issues that arose include: (1) certain controls meant to segregate the duties of employees were violated, (2) terminated employees were able to process payments, and (3) wire payment limits were circumvented because employees with $0 limits were able to process wire payments. These major internal control issues were brought to the attention of the internal audit department and were investigated. The internal auditors found that there were inconsistencies between the wire transfer payment process records and human resource records. The discrepancy is caused by the company keeping only the most recent information. For example, a terminated employee might have been an active employee when he initiated/approved a wire transfer. Although these internal control violations appeared to be clearly potential fraud indicators, investigation of their nature and frequency suggested that these violations were more likely due to systematic problems.

Scoring System Issues

During the investigation, the internal audit department noted that some of the 47 wires flagged were due to systematic reasons that were mainly attributed to the target tests. Each of the target tests' indicators was scored a risk level of high (the highest score) and, as a result, a flag by a few of these indicators will most likely hit the threshold amounts needed for investigation. This may suggest that an equally weighted scoring system may be more useful as a starting point than an unequally weighted one. However, it is also inherent that some indicators are clearly more important than others. As long as the indicators are subjectively assigned weights, this issue may come back again. Further deliberation will be necessary to find less subjective weighing methods for the indicators. The finding further illustrates that any fraud system must be further developed or updated as new flaws with the current system surface.

CONCLUSION

This paper provides a pilot study for fraud detection at a major U.S. insurance company. Although the literature has discussed numerous methods for fraud detection, there have been few that have implemented a pilot study on fraud detection by using real data from a company. Data mining was used as the approach to detect fraud. Statistical algorithms were created to detect abnormality or patterns in the data. Since much of the prior research uses complex methods such as neural networks and clustering to detect fraud, the use of simple statistics such as prediction interval, frequency test, and correlation test may seem trivial. However, it is not uncommon that simple methods are robust and powerful so that they beat the sophisticated methods. Therefore, simple statistics should not be overlooked. Also, they can provide a wealth of details and information. This study leaves out data analysis for payees where there are two wire payments and between 3 and 29 wire payments due to the lack of statistical significance. Future research can look into implementing other types of statistical methods such as clustering for detecting abnormal or patterned activity. The company plans to pursue data mining further as a continual effort to detect fraud. As the first step, our model will become part of their regular audit to examine whether their transactions are anomalous. Although the anomaly detection model in this study will need further fine-tuning in the future, its implementation clearly implies its usefulness in real practice.

This study provides a learning experience for academics by showing how a fraud detection and prevention model is implemented. In addition, this study shows that internal auditors can run fraud detection and prevention activities on a frequent basis instead of having a once-a-year audit. Several issues require further consideration: (1) the weighting of the indicators was problematic because most of the target indicators were scored as high risk and, thus, a violation of a couple of these indicators would trigger the threshold aggregate score for investigation, and (2) as the pilot study was progressing, some of the indicators needed to be adjusted. This was caused by the progressive increased understanding of the data and its characteristics.

REFERENCES

Alles, M. G., A. Kogan, M. A. Vasarhelyi, and J. Wu. 2004. *Continuity Equations: Business Process Based Audit Benchmarks in Continuous Auditing*. Proceedings of the American Accounting Association Annual Conference. Sarasota, FL: AAA.

Alles, M. G., A. Kogan, M. A. Vasarhelyi, and J. Wu. 2006. *Analytical Procedures in Continuous Auditing: Continuity Equations Models for Analytical Monitoring of Business Processes*. Proceedings of the American Accounting Association Annual Conference. Sarasota, FL: AAA.

Alles, M. G., A. Kogan, and M. A. Vasarhelyi. 2002. Feasibility and economics of continuous assurance. *Auditing: A Journal of Practice & Theory* 21 (1): 135–138.

Alles, M. G., A. Kogan, and M. A. Vasarhelyi. 2004. Restoring auditor credibility: Tertiary monitoring and logging of continuous assurance systems. *International Journal of Accounting Information Systems* 5 (2): 183–202.

American Accounting Association (AAA). 1972. Committee on Basic Auditing Concepts 1969–71. *The Accounting Review* 47 (4):14–74.

American Institute of Certified Public Accountants (AICPA). 2002a. *Consideration of Fraud in a Financial Statement Audit*. SAS 99. New York, NY: AICPA.

American Institute of Certified Public Accountants (AICPA). 2002b. *Management Antifraud Programs and Controls: Guidance to Help Prevent, Deter, and Detect Fraud*. New York, NY: AICPA.

Association of Certified Fraud Examiners (ACFE). 1996. *Report to the Nation on Occupational Fraud and Abuse*. Austin, TX: ACFE.

Association of Certified Fraud Examiners (ACFE). 2004. *Report to the Nation on Occupational Fraud and Abuse*. Austin, TX: ACFE.

Association of Certified Fraud Examiners (ACFE). 2007. *Report to the Nation on Occupational Fraud and Abuse*. Austin, TX: ACFE.

Association of Certified Fraud Examiners (ACFE). 2009. *Occupational Fraud: A Study of the Impact of an Economic Recession*. Austin, TX: ACFE.

Bolton, R. J., and D. J. Hand. 2001. Unsupervised profiling methods for fraud detection. In *Credit Scoring and Credit Control VII*. London, U.K.: Imperial College.

Bolton, R. J., and D. J. Hand. 2002. Statistical fraud detection: A review. *Statistical Science* 17 (3): 235–249.

Chandola, V., A. Banerjee, and V. Kumar. 2009. Anomaly detection: A survey. *ACM Computing Surveys* 41 (3): Article 15.

Elliott, R. K. 2002. Twenty-first century assurance. *Auditing: A Journal of Practice & Theory* 21 (1): 139–146.

Fawcett, T., and F. Provost. 1999. Activity monitoring: Noticing interesting changes in behavior. The Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego, California.

Hassibi, K. 2000. Detecting payment card fraud with neural networks. In *Business Applications of Neural Networks: The State-Of-The-Art of Real-World Applications*, edited by Lisboa, P. J., B. Edisbury, and A. Vellido. London, U.K.: World Scientific Publishing Co. Pte. Ltd.

Jans, M., N. Lybaert, and K. Vanhoof. 2009. A framework for internal fraud risk reduction at IT integrating business processes: The IFR framework. *The International Journal of Digital Accounting Research* 9: 1–29.

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

Kou, Y., C. Lu, and S. Sirwongwattana. 2004. Survey of fraud detection techniques. In *Networking, Sensing and Control*. 2004 IEEE International Conference on Knowledge Discovery and Data Mining. Falls Church, VA: Virginia Polytechnic Institute and State University.

Lectric Law Library. Fraud, to defraud. Available at: http://www.lectlaw.com/def/f079.htm

Little, B. B., Johnston, W. L., Jr., A. C. Lovell, R. M. Rejesus, and S. A. Steed. 2002. Collusion in the U.S. crop insurance program: Applied data mining. The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Major, J. A., and D. R. Riedinger. 2002. EFD: A hybrid knowledge/statistical-based system for the detection of fraud. *The Journal of Risk and Insurance* 69 (3): 309–324.

Murthy, U. S. 2004. An analysis of the effects of continuous monitoring controls on e-commerce system performance. *Journal of Information Systems* 18 (2): 29–47.

Murthy, U. S., and M. S. Groomer. 2004. A continuous auditing web services model for XML-based accounting systems. *International Journal of Accounting Information Systems* 5 (2): 139–163.

Pathak, J., B. Chaouch, and R. S. Sriram. 2005. Minimizing cost of continuous audit: Counting and time dependent strategies. *Journal of Accounting and Public Policy* 24 (1): 61–75.

Phua, C., V. Lee, K. Smith-Miles, and R. Gayler. 2005. *A Comprehensive Survey of Data Mining Based Fraud Detection Research*. Computing Research Repository, abs/1009.6119.

Rezaee, Z., A. Sharbatoghlie, R. Elam, and P. L. McMickle. 2002. Continuous auditing: Building automated auditing capability. *Auditing: A Journal of Practice & Theory*. 21 (1): 147–163.

Sherman, E. 2002. Fighting web fraud: Security: The Internet has made it easier for crooks to rip your company off. Here's how businesses can protect themselves and their customers. *Newsweek* (June 10). Available at: http://www.highbeam.com/doc/1G1-88125202.html

Winn, T. J., Jr. 1996. *Using Data Mining Techniques for Fraud Detection: A Best Practice Approach to Government Technology Solutions*. SAS Institute Whitepapers. Available at: http://www.ag.unr.edu/gf/dm/dmfraud.pdf

Winn, T. J., Jr. 2004. *Fraud Detection—A Primer for SAS Programmers*. SUGI 31 Proceedings. Available at: http://www2.sas.com/proceedings/sugi31/080-31.pdf

Woodroof, J., and D. Searcy. 2001. Continuous audit implications of Internet technology: Triggering agents over the web in the domain of debt covenant compliance. The 34th Hawaii International Conference on System Sciences.

Vasarhelyi, M. A., and F. B. Halper. 1991. The continuous audit of online systems. *Auditing: A Journal of Practice & Theory* 19 (1): 110–125.