Using “Data Mining” to Detect Frauds of Internal Audits

*Muhammad K. Ghanbari and **Mina Einakian

Fraud detection has become a critical component of financial audits. Iran as a developing country usually faces too many frauds in financial reports. The researches on important factories such as Iran Docharkh, Mazda Yadak and Bahman Group indicate that the production systems in Iran usually face problems which have kept the production system away from the right management techniques. Just in Time production (J.I.T) and Lean production. As a result accounting system will be far away from Lean Accounting and also from using the standard costs. On the other hand the most effective factor of finished value is inventory. The inventory contains raw materials inventory, work in process inventory (W.I.P) and finished products (F.P) inventory. These factors are the reasons which will increase or decrease the company profit and are able to indicate frauds in financial reports. Showing extra profit causes some problems for shareholders and lack of profit causes taxes problems.

This paper will present a new model (3TFP model) to help auditors for remote controlling of data and solve the problems mentioned. In this model first we will convert financial year to 12 months, after updating the bill of material we will physically count B.O.M. Work in Process and finished products in the end of each month and we will control them. Then the inventory of beginning and ending periods and consuming rate will be computed. After computing the factors mentioned, the cumulative consumption in each month for finished products, B.O.M and work in process products should be computed. Drawing diagrams for data gained and identifying critical regions are the next phases. Then we will use Fish-Bone method to analyze information. And finally we will use Pareto method to detect and clear the most suspicious factors causing frauds. It should be mentioned that implementing this model needs computer programming because of high volume of data.

I. Introduction

Fraud can be defined as a criminal activity, involving false representations to gain an unjust advantage [1]. Fraud is a million dollar business and it is increasing every year. "45% of companies worldwide have fallen victim to economic crime in 2004 and 2005. The average damage to the companies from tangible frauds (i.e. asset misappropriation, false pretences, and counterfeiting) was US$ 1.7 million". According to the Global economic crime survey 2005 of Price Waterhouse Coopers. Journal headlines and news topics indicate the same trend of increasing fraudulent behavior [4]. For example, the Association of Certified Fraud Examiners [15] estimates that total annual fraud losses in the United States are now US $994 billion, up from its fraud annual cost estimate of US $660 billion in 2006. KPMG (2006) highlights that fraud continues to be a major problem for organizations in Australia and New Zealand, with approximately half the organizations surveyed experiencing fraud annually and an average detection time of approximately one year [2].

*Dr. Muhammad K. Ghanbari, Computer Science Department, Alabama A&M University, Email: Muhammad.ghanbari@aamu.edu

**Mina Einakian, Computer Engineering Department, Urmia University, Email: Mina.einakian@gmail.com
Fraud occurs in a wide variety of forms and is ever changing as new technologies and new economic and social systems provide new opportunities for fraudulent activity. It is remarkable that 34% of these frauds are detected by chance. Fraud is detected in many ways, or at least one tries to detect it in many ways [4].

Traditionally, a company relies most on its internal control activities and the internal auditor to prevent and detect fraud. Data mining is different from the traditional data analysis techniques [4].

Internal fraud detection is concerned with determining fraudulent financial reporting by management: Hansen et al used a powerful generalized qualitative response model to predict management fraud based on a set of data developed by an international public accounting firm. Green and Choi presented a neural network fraud classification model employing endogenous financial data. A classification model created from the learned behavior pattern is then applied to a test sample. Fanning and Cogger also used an artificial neural network to predict management fraud. Beneish investigated the incentives and the penalties related to earnings overstatements primarily in firms that are subject to accounting enforcement actions by the Securities and Exchange Commission. Abbott et al. examined and measured the audit committee independence and activity in mitigating the likelihood of fraud [3].

If internal auditor isn’t sufficient, external audit (as far as this isn’t legally enforced yet), risk management systems, a whistle-blowing hotline, an investigations department, new technologies or other measures are installed, corresponding to the need the company experiences [4].

II. Concepts

What is Data Mining?

Data is a fundamental element in any organization's ability to manage its business; it is collected from a wide variety of sources, stored on many different systems, and is regularly used for marketing and sales activities. However, the use of this data in fraud detection is frequently overlooked.

The likelihood of identifying potentially fraudulent activity can be significantly enhanced through the regular application of data mining tools and techniques, although these are not foolproof and must be run in conjunction with other activities designed to reduce the threat of fraud.

Data mining is about finding insights which are statistically reliable, unknown previously, and actionable from data [10]. This data must be available, relevant, adequate, and clean. Also, the data mining problem must be well-defined, cannot be solved by query and reporting tools, and guided by a data mining process model [11] [5].

According to Witten and Frank (2000), data mining can be defined as:

“…The process of discovering patterns in data. The process must be automatic or (more usually) semiautomatic. The patterns discovered must be meaningful in that
they lead to some advantage, usually an economic advantage. The data is invariably present in substantial quantities.” [4]

Statistics and data mining methods have been applied successfully to detect activities such as money laundering, e-commerce credit card fraud, telecommunications fraud, insurance fraud, and computer intrusion etc.

What is Fraud?

There are many definitions for fraud, depending on the point of view considering. According to The American Heritage Dictionary, Second College Edition, fraud is defined as a deception deliberately practiced in order to secure unfair or unlawful gain’. Davia et al. [12] paraphrase this in a number of items that must be identified, when articulating a case of fraud:

• A victim

• Details of the deceptive act thought to be fraudulent

• The victim’s loss

• A perpetrator (i.e., a suspect)

• Evidence that the perpetrator acted with intent

• Evidence that the perpetrator profited by the act(s)

In a nutshell, ”fraud always involves one or more persons who, with intent, act secretly to deprive another of something of value, for their own enrichment ” [12]. Wells (2005) stresses deception as the linchpin to fraud. To exclude kinds of fraud we don’t wish to examine, the delineation of fraud to ‘occupational fraud and abuse’, as referred to by the Association of Certified Fraud Examiners, is of interest. Occupational fraud and abuse may be defined as:”The use of one’s occupation for personal enrichment through the deliberate misuse or misapplication of the employing organization’s resources or assets.” [13] This definition encompasses a wide variety of conduct by executives, employees, managers, and principals of organizations [4].

The term fraud here refers to the abuse of a profit organization’s system without necessarily leading to direct legal consequences. In a competitive environment, fraud can become a business critical problem if it is very prevalent and if the prevention procedures are not fail-safe. Fraud detection, being part of the overall fraud control, automates and helps reduce the manual parts of a screening/checking process [5].

Fraud Detection Requires both Internal and External Business Data

Most companies have some sort of internal data describing their business events (selling things or providing services). But the forms of data gathered for internal purposes most often are related to billing and account service purposes. Many potentially predictive variables are not gathered by internal systems (e.g., years in
business), but must be gathered from external sources. Information can be gathered from various data providers to enhance the corporate database, including

- Demographic data (available from Axciom, Experion, Equifax, Lexis-Nexis, etc.);
- Firmographic data (e.g., Dun & Bradstreet data and other business data sources);
- Psychographic data (inferences and classifications of people according to various measures of attitudinal and philosophical views) [1].

Internal control systems are not sufficient as a fraud detection mechanism. Since over one third of the fraud cases in the surveys are discovered by chance. Internal controls can be split into two groups: active and passive internal control systems. Active internal controls are signatures, passwords, segregation of duties etc. As Davia et al. (2000) put, these can be compared with fences. They may appear insurmountable at first sight, but like all fences, they have their weakness to be defeated by clever fraud perpetrators. And like a fence, once evaded, there is little or no continuing value in preventing or deterring fraud [12]. Passive internal controls operate at a different level. Instead of preventing fraud, like active controls attempt to, the emphasis here is on deterring. Passive internal control systems induce a state of mind in the would-be perpetrator that strongly motivates him “not to go there” [4].

Some common types of fraud include creating fictitious creditors, “ghosts” on the payroll, falsifying cash sales, undeclared stock, making unauthorized “write-offs”, and claiming excessive or never incurred expenses.

Internal Auditors

For a long time, there has been controversy over the role of the auditor with respect to the detection of fraud. It has been argued that an audit should be done by a competent, independent, individual and involves the collection and assessment of evidence about information to decide and report on the degree of correspondence between the information and certain established criteria [18][6].

The internal auditor’s role in fraud risk management continues to gain prominence in contemporary governance frameworks. Deloitte (2005) emphasized this role in the Sarbanes-Oxley era, noting that “given its prominence and potential magnitude, fraud needs to be on the radar screen of every internal audit function.” Similarly, The IIA, the American Institute of Certified Accountants (AICPA), and the ACFE recently highlighted that internal auditors “should exercise professional skepticism when reviewing activities and be on guard for the signs of fraud” [16, 17]. While internal audit standards prescribe professional responsibilities related to fraud, little is known about the extent that internal auditors perceive responsibility for fraud detection and the factors that affect their sense of responsibility [2].

The IIA’s Frequently Asked Questions repository states that “(internal audit) O plays
an important role in fraud detection” [17]. More specifically, The IIA’s *International Standards for the Professional*

*Practice of Internal Auditing (Standards)* provides specific authoritative guidance for internal auditors in the area of fraud. For example, Proficiency Standard 1210.A2 states:

> Internal auditors must have sufficient knowledge to evaluate the risk of fraud and the manner in which it is managed by the organization, but are not expected to have the expertise of a person whose primary responsibility is detecting and investigating fraud.

Proposed revisions to internal audit standards highlight a push to increase fraud standards for internal auditors. Furthermore, the new risk management standard (2120.A2) promotes that “the internal audit activity must evaluate the potential for the occurrence of fraud and how the organization manages fraud risk.” These standards clearly link the internal audit function to fraud risk management [2].

The triangle model of responsibility [19] provides an integrative framework for evaluating perceived responsibility and links among accountability, responsibility, and performance. In this context, perceived responsibility connects individuals to performance standards and to events covered by those standards. Schlenker (1997, 241) suggests that responsibility is “the psychological glue that connects an individual to a set of prescriptions for conduct and to an event that is governed by those prescriptions.”

Schlenker’s triangle model [15] identifies three excuses people use to avoid taking responsibility after failure: that one had no control in the situation, that the obligation was unclear, and that it was not really one’s obligation. Three retrospective studies tested the presumed negative association between excuse making and responsibility taking. The studies also examined the effects of self-determination theory’s concept of motivational internalization upon these variables. A complex but replicable pattern emerged, such that responsibility taking and motivational internalization correlated with adaptive outcomes such as future commitment and positive expectancy and excuse making did not. Of particular interest, perceiving that the person levying the obligation internalized motivation predicted responsibility taking, in all three studies [2], [7].

Schlenker and colleagues describe three common excuses that people use to pardon themselves when they have been associated with a failure of some kind. They define excuses as statements or attributions that allow one to minimize personal responsibility for events, both for oneself and with others. In this way, excuses both regulate emotions and serve as impression management.

The model describes three aspects of responsibility, for which excuses are made (Figure 1):

- **Prescription**: What is supposed to be done
- **Identity**: The sense of self
- **Situation/Event**: relevant to the prescription
The prescription-event (task clarity) link is considered strong to the extent that the prescriptions are specified in advance, pertinent to the situation, not subject to alternative interpretations, and not in conflict with other prescriptions that might be applied in the situation. Alternatively, this link is weaker when prescriptions are ambiguous, conflicting, difficult to prioritize, or questionable in terms of relevance to the event. The prescription-identity (professional obligation) link is strong to the extent that prescriptions apply unambiguously to the individual. This link is weaker when prescriptions are ambiguous, unclear, or conflicting. The identity-event (personal control) link is strong when an individual intends to produce specific consequences and had the ability and freedom to do so. The link is weaker when an individual’s will to act is diminished because action consequences are unforeseeable, accidental, or influenced by uncontrollable factors.

The triangle model of responsibility suggests that accountability and responsibility are related but distinct constructs. For example, Schlenker notes that “responsibility is not identical to accountability, rather, responsibility flows from accountability.” This suggests that accountability is a type of pressure put on internal auditors by other parties (i.e., senior management, the audit committee, internal audit standards, etc.), and perceived responsibility is the internal response of the internal auditor to these external pressures. To the extent that these external pressures highlight the importance of internal auditor responsibility for fraud detection, we expected accountability pressure to increase internal auditors’ perceived responsibility for fraud detection.

Internal audit standards do not prescribe different detection responsibilities for fraudulent financial reporting, misappropriation of assets, and corruption frauds that have a direct and material effect on the financial statements. However, the research literature provides evidence suggesting that professionals’ perceived responsibility for fraud detection differs across fraud type. DeZoort and Harrison (2008) found that external auditors perceive more responsibility for detecting fraudulent financial reporting than they do for misappropriation of assets and corruption. Furthermore, the ACFE [13] report highlights that fraudulent financial reporting schemes have a higher median per fraud cost (US $2 million) than misappropriation of assets (US $375k) and corruption (US $150k) schemes. However, this same report and others indicate that misappropriation of assets is much more prevalent and that internal auditors are much more likely to deal with this type of fraud in the workplace [13, 14] [2].
Appraising Internal Control

An entity’s internal control consists of five interrelated components:
- Control environment.
- Risk assessment.
- Control activities.
- Information and communication.
- Monitoring.

Technology as a Tool

People commit frauds, but as technology plays an increasingly important role in business life, the fraudster often leaves warning signals of his activity in an organization’s systems.

Each transaction will leave a trail. Increasingly, in order to enhance the way an organization does business, databases have been developed to store huge amounts of transactional and standing data from accounting, sales, purchasing, and payroll functions. This is used for marketing, forecasting, and reporting but rarely for detecting and predicting fraud. Also, this data can be a key factor in developing and implementing a fraud risk management strategy.

Use of Databases

The next stage in data mining is the use of databases to run complex queries. Microsoft Access is an extremely powerful tool which many fraud examiners will be able to use. More complex databases include ACL and WinIdea. These may require specialist knowledge. However, they can analyze large amounts of data and produce complex queries that can be automated. The following chart (Figure 2) illustrates that data mining has identified a series of transactions just above $50,000, which is the authorization limit for the company. Databases can also be used to identify suspicious transactions around points in time. [9]

![Figure 2: Series of Transaction Identified by Data Mining]
Tests of Controls

SAS 55 defines tests of controls as tests directed toward the design or operation of an internal control to assess its effectiveness in preventing or detecting material misstatements in financial statement assertion. Inquiry of company personnel, inspection of client documents and records, observation of client activities, and reperformance of controls represent some of the procedures used in performing tests of controls. In performing tests of controls, the auditor seeks answers to the following questions:

- Who performed the control?
- When was the control performed?
- How was the control performed?
- Was the control consistently applied?

Statistical Sampling

Statistical sampling helps the auditor to design an efficient sample, to measure the sufficiency of the evidence obtained, and to evaluate the sample results. The standard of field work requires auditors to obtain sufficient competent evidence. Sufficiency relates to the design and size of the sample. Statistical sampling permits the auditors to measure sampling risk and therefore to design more efficient samples, that is, samples of a size necessary to provide sufficient evidence.

The results of statistical (probability) sampling are objective and subject to the laws of probability. Hence, sampling risk can be quantified and controlled, and the degree of reliability desired (the confidence level) can be specified. Sampling risk is the risk that the sample selected does not represent the population.

There are a number of statistical (probability or random) sampling methods that will aid auditors in performing tests of controls. They include attribute sampling, variables sampling, discovery sampling, multistage sampling, and stratified sampling [9].

Sampling Risk

Sampling risk is the probability that a properly drawn sample may not be representative of the population; that is, the conclusions drawn from the sample may differ from those drawn if all the items in the population are examined. Sampling risk is to be measured and controlled. The auditor controls sampling risk by specifying the acceptable level when developing the sampling plan.

Analytical Techniques for Fraud Detection

Getting started requires an understanding of:

- The areas in which fraud can occur
- What fraudulent activity would look like in the data
- What data sources are required to test for indicators of fraud [8]
The following techniques are effective in detecting fraud. Auditors should ensure they use these, where appropriate. They include the following:

- **Calculation of statistical parameters** (e.g., averages, standard deviations, high/low values) – to identify outliers that could indicate fraud.
- **Classification** – to find patterns amongst data elements.
- **Stratification of numbers** – to identify unusual (i.e., excessively high or low) entries.
- **Digital analysis using Benford’s Law** – to identify unexpected occurrences of digits in naturally occurring data sets.
- **Joining different diverse sources** – to identify matching values (such as names, addresses, and account numbers) where they shouldn’t exist.
- **Duplicate testing** – to identify duplicate transactions such as payments, claims, or expense report items.
- **Gap testing** – to identify missing values in sequential data where there should be none.
- **Summing of numeric values** – to identify control totals that may have been falsified.
- **Validating entry dates** – to identify suspicious or inappropriate times for postings or data entry. [8]

**IT and Data Mining Used To Detect Corporate Fraud**

IT tools such as data mining can be effectively employed by fraud auditors. This involves the analysis of data stored in an information system (e.g., a database) to identify patterns that indicate unexplained or potentially questionable transactions. The advantage of such a computer analysis technique is that large numbers of transactions can be evaluated in a relatively short period of time. Further, multiple analyses of individual data elements can be performed to provide different evaluations of potential patterns or trends. Once a pattern is identified, the auditors must further investigate specific transactions to determine whether an improper transaction actually occurred. Data mining can also be very useful in other high-risk corporate processes. Two such areas include executive travel and contract and consulting services.

**III. Methodology**

Iran as a developing country usually faces too many frauds in financial reports. The producer systems in Iran usually face these problems:

- Lack of easy and inexpensive access to sources and raw materials.
- Economic boycotts.
- Exchange rate fluctuation.
- Purchasing high volume batches which aren’t economical.
- Managers fear to indicate real profit and costs because of quick changes of managers due to the wrong policy of companies and lack of time to implementing their policies.
These problems mentioned a bow have kept away the production system from the right management techniques, Just in Time production (J.I.T) and lean production. As a result accounting system will be far away from lean accounting and follow using the standard costs. The most effective factor of finished value is inventory. This inventory contains raw materials inventory, work in process inventory (W.I.P) and finished products (F.P) inventory.

These factors are the reasons which will increase or decrease the company profit and are able to indicate frauds in financial reports. Showing extra profit causes some problems for shareholders and lack of profit causes taxes problems. So we will present the 3TFP model in this paper to help auditors for remote controlling of data and solve the problems mentioned. It should be mentioned that implementing this method needs computer programming because of high volume of data.

In industrial countries production system is based on “high variety and low volume” but in the opposite way, in Iran, because of lack of technology, this system is based on single products or “low variety and high volume”. Thus the companies have to have a safety stock even more than standard.

Steps

i. Converting financial year to 12 month sequential information months.
ii. Updating Bill of Material (B.O.M).
iii. B.O.M physically counting for each product and also working process products and finished products in the end of month and controlling them.
iv. Computing the inventory of beginning period and ending period and also compute consuming rate and converting it to the local money (Dollars).
vi. Drawing diagrams for the trend of data gained from “v” (Figure 3).
vii. Identifying critical regions.
viii. Using fish bone chart to analyze data (Figure 4).
ix. Using “Parato” method to detect and clear the most doubts which cause frauds.
Figure 3: Diagrams of Trend of Data’s and Critical Regions
The internal auditor gains information and calculates the percentage of each probable factor based on “Parato” method (Figure 5).

<table>
<thead>
<tr>
<th>Definition</th>
<th>The percentage of identification in critical region</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>%60</td>
</tr>
<tr>
<td>B</td>
<td>%15</td>
</tr>
<tr>
<td>C</td>
<td>%10</td>
</tr>
<tr>
<td>D</td>
<td>%5</td>
</tr>
<tr>
<td>E</td>
<td>%5</td>
</tr>
<tr>
<td>F</td>
<td>%3</td>
</tr>
<tr>
<td>G</td>
<td>%2</td>
</tr>
<tr>
<td>Total</td>
<td>%100</td>
</tr>
</tbody>
</table>

Figure 5
The auditor will get the team focus on factor “A” with percentage of 60 of doubt. So that the team can detect the fraud if it is existed then he will move to the next factor and check them by hierarchy. If the critical regions were more than one, the auditor will compare all the critical regions by their mutual factor (Figure 7).
Deductive Fraud Detection

In contrast, a deductive approach attempts to determine what kinds of frauds can occur in a particular situation and then uses data mining and other IT methods to determine if those frauds exist. It follows a five-step process:

1. Understanding the business or operations to be studied.
2. Understanding what kinds of frauds could occur (fraud exposures) in the operation.
3. Determining the symptoms that the most likely frauds would generate.
4. Using databases and information systems to search for those symptoms.
5. Following up on symptoms to determine if actual fraud or other factors are causing them.

IV. Conclusion

Data analysis technology enables auditors and fraud examiners to analyze an organization's business data to gain insight into how well internal controls are operating and to identify transactions that indicate fraudulent activity or the heightened risk of fraud. Data analysis can be applied to just about anywhere in an organization where electronic transactions are recorded and stored. Data analysis also provides an effective way to be more proactive in the fight against fraud.

V. References

Handbook of Statistical Analysis and Data Mining Applications, Fraud Detection, (2009), Elsevier Inc.
MiekeJans, Nadine Lybaert, KoenVanschoof. Data Mining for Fraud Detection: Toward an Improvement on Internal Control Systems?
Clifton Phua, Vincentlee, Kate Smith & Ross Gayler, A Comprehensive Survey of Data Mining based Fraud Detection Research
Fraud Detection Using Data Analytics In the Banking Industry. ACL Transforming audit and risk.

Internal Control And Fraud Detection, DELTA PUBLISHING COMPANY


