Fraud Detection: A Primer for SAS® Programmers

Thomas J. Winn, Jr.
Texas State Auditor's Office, Austin, Texas

(Note: SAS is a registered trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.)

Abstract

Prevention of fraud and abuse has become a major concern of many organizations. The SAS System includes powerful tools for data management, data analysis, and reporting, which can be used in battling against fraud and abuse. This expository paper will describe some basic methods for detecting possibly fraudulent activities, using Base SAS software. Additional remarks will mention using other SAS software components for fraud detection.

Introduction

Recent highly publicized corporate scandals have resulted in strict new laws, policies, and standards that affect most companies, as well as many other organizations. The Sarbanes-Oxley Act of 2002 specifies new requirements regarding the financial management of publicly traded companies. It requires companies to implement antifraud programs and controls. Also in 2002, the American Institute of Certified Public Accountants (AICPA) issued Statement on Auditing Standards No. 99, which specifies requirements for auditors concerning fraud considerations in financial statement audits. The new rules affect more than just auditors, they also impact the work of employees in most operational components of many companies. Efforts aimed at preventing or detecting fraud and abuse are becoming widespread features of business entities around the world.

The Association of Certified Fraud Examiners estimated that, in 2004, occupational fraud and abuse cost the U.S. economy about $660 billion, which represents a loss of about 6 percent of revenues. Fraud is a significant and growing financial risk, which threatens the profitability and reputation of companies around the world. Fraud can occur anywhere, and it has become a major concern of many organizations.

Consideration of the possibility of fraud and abuse now is a part of *every* audit performed by the Texas State Auditor's Office (S.A.O.). Moreover, the S.A.O. is developing and implementing special tools and procedures for its own auditors to use in detecting fraud and abuse in the following high-risk areas: payroll, contracting and procurement, credit cards, revenues and receipts, assets management, and construction.

What is Fraud?

Fraud is an intentional act meant to induce another person to part with something of value, or to surrender a legal right. It is a deliberate misrepresentation or concealment of information in order to deceive or mislead. Fraud can range from minor employee theft and unproductive behavior to misappropriation of assets and fraudulent financial reporting. In different situational contexts, fraud can take somewhat different forms – for example, bribery, embezzlement, securities fraud, health care fraud, money-laundering scams, insurance fraud, software piracy, internet fraud, telemarketing fraud, mortgage foreclosure scams, and identity theft -- these all have their own special characteristics. There are at least as many types of fraud as there are types of people who commit it. But in each instance, fraud involves deception. Someone knowingly lies in order to obtain an unlawful benefit, or an unfair advantage.

Some examples of fraud include:

- any dishonest or fraudulent act;
- forgery or alteration of a check, bank draft, or financial document;
- misappropriation of assets;
- deliberate impropriety in the handling or reporting of money or financial transactions;
- wrongfully using influence in a business transaction to receive a benefit (such as bribery, kickbacks, and bid-rigging);
- profiteering as a result of insider information;
- disclosing insider information to another person in order for them to secure unlawful gain.

Abuse is behavior that is deficient or improper when compared with behavior that a prudent person would consider reasonable and necessary business practice given the facts and circumstances. Instances of abuse are not fraud or illegal acts, but they are harmful, and they need to be minimized.

Determining the difference between improper acts and fraud is a matter for decision by the judicial system. But the detection of acts which may turn out to have been fraudulent is something that all administrators, private and public, need to be involved in. Perhaps the best way to prevent fraud is for many people in every organization to become vigilant in detecting and reporting possibly fraudulent activities. Moreover, failure to report observed fraudulent activities may expose the observer to disciplinary action or to prosecution.

SAS Programmers Can Help

Modern technology has created new opportunities for communications and commerce but, at the same time, it also has created opportunities for fraud and abuse to occur. Organizations of all types, private and public, need to asses their vulnerability to fraud and abuse, and take appropriate action. An important step is to subject databases containing financial transactions to a systematic search for patterns which might be indicative of fraud. SAS programmers can help in this effort, because of their familiarity with enterprise data, and because of their access to powerful programming tools. While SAS programmers would not be expected to have the fraud expertise of a person whose primary responsibility is detecting and investigating fraud, they can assist investigators by gathering and analyzing information which may be pertinent. Their attentiveness to data can be a vital part of an organization's culture of ethical behavior. However, if fraud is suspected, care must be taken to secure the evidence, without informing someone that they are under suspicion.

Importance of Internal Controls

An internal control system is a framework of requirements and procedures which pertain to the recording, processing, and reporting of financial data, and the safeguarding of an organization's assets (cash, inventory, equipment, etc.). Internal controls ensure the reliability of the operational data, and provide assurance for attainment of the organization's goals and objectives.

Here are a few important principles concerning internal controls:

- Duties need to be distributed among employees so that no one person will have too much control in any financial area. No single employee should be able to handle any transaction from start to finish. Whenever several people handle different aspects concerning transactions, cross-checking of data occurs, mistakes and irregularities are reduced, and the chance that a fraud may go unnoticed decreases.
- Authorization procedures must be established and observed, to ensure the propriety and validity of certain activities and transactions. There should be written policies and procedures governing who may authorize transactions. Authorizations must be obtained *before*, and *not after* the fact.
- Documents must be properly designed and used. Retain all written records. Forms should be pre-numbered sequentially, and carefully accounted for. Spoiled documents should be voided, and retained.
- Reconciliation procedures should be followed, in which related sets of data are compared by a third party, in order to identify and examine any differences. There also should be regular inventories of physical assets.
- Internal controls also include provisions regarding physical security of assets and records. Access to records should be based upon need, and this must be reviewed regularly.

There need to be periodic reviews of the internal controls, to make sure that they are working. Fraud can occur whenever there are weaknesses in internal controls.

There usually are three conditions present when fraud occurs: incentive (pressure), opportunity, and rationalization (finding a justification for committing a fraudulent act). Be alert to the possibility that management or certain other key employees may have the ability to override at least some internal controls. Furthermore, they generally will take steps to conceal fraudulent acts, which will make it difficult to detect.

Professional Skepticism
The detection of possibly fraudulent activity requires that the analyst must possess an attitude that includes a questioning mind and a critical assessment of information. It is okay to assume good faith on the part of all parties to any situation or transaction, but be alert for evidence that might negate that assumption. Set aside any prior beliefs about the honesty and integrity of the individuals involved. If you don't expect to find any wrongdoing, then you probably won't. Remember, fraud can occur anywhere.

How You Might be Able to Detect Fraud – Some Generalities:
First of all, use your knowledge of the business to look for vulnerabilities which could provide an opportunity for someone to enrich themselves, if they thought that they might "get away with it."

Be on the lookout for unauthorized transactions, unexplained pricing exceptions, excessive payments to vendors, significant increases in expenditures, large transactions involving petty cash, and the inability to trace invoices and payments with available records. Are there any employees with both processing and approval authority for expenditures?

Fraud indicators include:

- Missing or altered documents to support transactions,
- Excessive voided documents or transactions without supervisory approval,
- Transactions with inappropriate authorizations,
- Excessive complaints from customers or other employees,
- Unusual billing addresses or arrangements,
- Payments based on photocopied invoices or fabricated invoices,
- Vendor payments sent to an employee's address,
- An employee who:
  - is living beyond his or her means,
  - can't manage money,
  - doesn't take vacations,
  - is dissatisfied with work,
  - is a take-charge person,
  - has expensive habits,
  - has close relationships with customers or vendors,
  - has, or is developing, outside business interests that are closely related to his or her employment.

The presence of fraud indicators does not mean that fraud is occurring, but fraud will not occur without at least some of them.

Follow up on all hints and rumors. Check the details of documents. Carefully examine unusual transactions. And assess the reasonableness of transactions. Look carefully for improper payments -- improper payments include inadvertent errors, such as duplicate payments and miscalculations; payments for unsupported or inadequately supported claims; payments for services not rendered; payments to ineligible beneficiaries; and payments resulting from outright fraud and abuse by clients and/or employees.

Data Mining and Model Building

Fraud detection is even more difficult than looking for that proverbial needle in a haystack! In the haystack problem, at least the needle and the hay don't look alike, and they don't change much over time!   Typically, we should expect that data pertaining to fraudulent acts would constitute a very small portion of very large data bases.   Because the class distribution is highly skewed, the application of data mining techniques for recognizing patterns indicative of fraud is particularly pertinent.

In general, data mining is the iterative and interactive process of finding new and potentially useful knowledge from large quantities of data.   My use of that term encompasses sophisticated statistical techniques, as well as many simpler data analysis methods. In data mining, techniques are used to uncover hidden patterns and subtle relationships among data in databases, and also to infer decision rules that may be useful in predicting future outcomes.   Some data miners have access to heavy equipment, while others have only "picks and shovels" for their use.

A model is a generalization which is created for the purpose of describing or predicting some particular phenomenon. Descriptive models are developed to help us to understand the phenomena; that is, they attempt to explain the underlying processes or behaviors that are pertinent to the phenomena under consideration. Predictive models strive to foretell the occurrence of some event -- they seek to declare in advance the attainment of certain data values, based upon the incidence of other data values. Typically, models are mathematical or statistical functions, or perhaps a set of logical rules, whose parameters are estimated based upon existing data.

Besides making a distinction between types of models according to their purposes, the literature also distinguishes between models according to the methods that were used for estimating their parameters.   In artificial intelligence, estimating model parameters is called *learning,* and learning methods are either *"supervised"* or *"unsupervised".* In supervised learning, the approach is to generate a function that maps inputs to specified outputs. Supervised data modeling techniques always involve the use of a well-defined dependent variable (the output / target).   Regression and classification methods are examples of supervised learning. Decision trees and artificial neural networks are examples of supervised learning; however, while neural networks often are more accurate for prediction than decision trees, they do not help us to understand the phenomena in the way that decision trees may. In unsupervised learning, there is no a priori output (no target), and the approach is to generate a model for a set of inputs. Unsupervised data modeling methods include various techniques for detecting clusters and recognizing patterns; including cluster analysis and self organizing maps. Most statistically-based data mining cannot be accomplished within the capabilities of Base SAS. You should learn how to make the best use of the tools that are available to you. Well, what <u>can</u> we do with Base SAS?

How You Might be Able to Detect Fraud – A Few Specific Suggestions:
The key to fraud detection is being able to recognize deviations from the normal pattern of activities. In examining data sets of transactions, use fundamental PROCedures (SORT, FREQ, MEANS, UNIVARIATE, TABULATE, SQL, PLOT, CHART) to determine customary values and to identify outliers. If fraud is suspected, determine the norms for an interval of dates that precedes the suspected fraud period. Then, examining data for the suspected fraud period, look for changes in the customary patterns. Prior to the suspected fraud period, during what days, dates, and times were the transactions most likely to occur? Is there evidence of activity outside of the customary pattern? Has the customary pattern changed?

Since deviations from the normal pattern may signal fraud, identify outliers in the data. Here is a simple procedure. Use PROC UNIVARIATE to generate descriptive statistics regarding data under consideration, and then use them in a DATA step to identify the observations that are outside 1.5 times the interquartile range (i.e., those that are beyond the "inner fence" values, one step outside, which are just beyond the "whiskers" in a box plot). Or, if further refinement is desired, use the outer fences -- 2 steps outside of the hinges -- instead of the one-step inner fences. These outside values are the "far out" data values in a box plot. (Note: 1 step equals 1.5 times the H-spread).

Using Benford’s Law, look for evidence of contrived numbers among data pertaining to claims for payment *[pertinent Base SAS code is included in my 2003 papers dealing with digital data analysis].*

Look for duplicate payments for the same goods or services. Look for vendor payments without a purchase order or invoice. Look for evidence of false refunds. Look for invoices from the same vendor with similar amounts, in round number amounts, dated the same or in close proximity, consecutively numbered or with nearly sequential invoice numbers, or with similarities in goods or services described – these may be evidence of split purchases, or of double billing. Look for evidence of cost mischarging.

Scrutinize billings, payroll, and expense reimbursements for improper payments made to employees. Look for payments that exceed authorized amounts. Look for multiple payroll payments covering the same payroll period. Compare addresses of vendor-payment payees with employees’ addresses to identify fictitious vendors, or possible conflict-of-interest situations *[Descriptions of some* *computer programs, which use Base SAS and "fuzzy logic" for matching records from two files on the basis of similarities in names and addresses, are provided in my papers about record matching.].* Compare reimbursements between employees performing similar work to find possibly excessive reimbursements. Compare reimbursements over time to establish norms, and to identify exceptions to the customary pattern.

Perform validity checks on Social Security Numbers: to determine if there are duplicate SSNs, if the numerical information meets authorized parameters for construction of SSNs, if the SSN was issued before the individual's birth date or from a location that would be inconsistent with the individual's personal history, or if the SSN was issued to someone who is now deceased. *[The author has written some (unpublished) Base SAS programs for social security number validation, which use Social Security Administration data. The code is straightforward, if one understands the principles behind the data.]*

Other Helpful Software Tools

There is no magical software package for fraud detection, certainly nothing that would continue to be useful into the future, as fraud schemes are dynamic and adaptive. There simply is no substitute for creative, and sometimes tedious, analytical and investigative effort. But enhanced tools do exist, which can be very helpful in the hands of individuals with analytical expertise and an investigative mentality.

The SAS System includes SAS/STAT, which is state-of-the-art software for performing statistical analysis. Among its many procedures, SAS/STAT includes procedures for least squares regression, discriminant analysis, logistic regression, and cluster analysis – which could be very useful in analytical explorations concerning fraud.

The SAS System also includes SAS/EnterpriseMiner, which offers a comprehensive set of data mining tools. Like all components of the SAS System, SAS/EnterpriseMiner builds on the foundation of Base SAS and it integrates fully with other SAS software products. SAS/EnterpriseMiner includes capabilities for modeling with clustering, neural networks, and decision trees, which are especially well-suited for use in fraud detection. However, SAS/EnterpriseMiner is very expensive software. I'm sure that one could make a very strong cost / benefit argument for licensing SAS/EnterpriseMiner; nevertheless, its price tag is beyond the discretionary software cost threshold of many information technology managers.

There is an abundance of non-SAS software which may be used to supplement SAS, including *Tiberius*, free neural network software available for download from http://www.philbrierley.com , and *AnswerTree*, a relatively inexpensive decision tree tool, that is available from http://www.spss.com/answertree. Other suggestions may be found on the "Fraud Detection Solutions" web page, at the KD Nuggets website, http://www.kdnuggets.com/solutions/fraud-detection.html

Conclusion

- Fraud is the intentional misrepresentation or concealment of information in order to deceive or mislead.
- Fraudulent activities include fraudulent financial reporting, misappropriation of assets, and corruption (corruption includes bribery and other illegal acts).
- The SAS System includes a comprehensive suite of tools which can be used in fighting fraud and abuse.
- In this presentation, we have described, or referred to, a few of the basic uses. And we also have listed several references for further study.

Suggestions for Further Reading

- AICPA Antifraud & Corporate Responsibility Resource Center, from the American Institute of Certified Public Accountants, http://www.aicpa.org/antifraud/
- 2002 Report to the Nation on Occupational Fraud and Abuse, from the Association of Certified Fraud Examiners, http://www.cfenet.com/pdfs/2002RttN.pdf
- 2004 Report to the Nation on Occupational Fraud and Abuse, from the Association of Certified Fraud Examiners, http://www.cfenet.com/pdfs/2004RttN.pdf
- Richard J. Bolton and David J. Hand, “Statistical Fraud Detection: A Review,” Statistical Science, Vol. 17, No. 3 (August 2002), pp. 235-255.
- Richard D. Cannon, “Fraud Detection Techniques Applied to the Analysis of Data Sets,” Proceedings of the Texas Conference for Government SAS Users, pp. 47-51 (sponsored by the Texas State Auditor’s Office and SAS Institute Inc., on July 29, 2003).
- Fraud Detection Solutions, from KD Nuggets, http://www.kdnuggets.com/solutions/fraud-detection.html
- Thomas J. Winn Jr., “Implementing Digital Data Analysis for Detecting Fraud Using SAS,”
    - Proceedings of the Conference for Southern SAS Users, pp. 428-431 (sponsored by South-Central SAS Users’ Group and South East SAS Users Group, on August 19-22, 2001), and
    - Proceedings of the Louisiana SAS Users Conference, pp. 208-219 (sponsored by South-Central SAS Users’ Group, on June 16, 2003), and
    - Proceedings of the Texas Conference for Government SAS Users, pp. 180-191 (sponsored by the Texas State Auditor’s Office and SAS Institute Inc., on July 29, 2003).
- Thomas J. Winn Jr., “Record Matching of Names and Addresses Using SAS,”
    - Proceedings of SCSUG 2002, pp. 379-384 (sponsored by South-Central SAS Users’ Group, on October 6-8, 2002), and

    - Proceedings of the Louisiana SAS Users Conference, pp. 224-229 (sponsored by South-Central SAS Users' Group, on June 16, 2003), and
    - Proceedings of the Texas Conference for Government SAS Users, pp. 208-213 (sponsored by the Texas State Auditor's Office and SAS Institute Inc., on July 29, 2003).
- Tom Winn, with Kirby Cossey, Ernest Cuellar, Olin Davis, Dorvin Handrick, and Joyce Inman, "Using the SAS System as an Audit Support Tool," Proceedings of SCSUG 2002, pp. 375-378 (sponsored by South-Central SAS Users' Group, on October 6-8, 2002).

Author Information

- Tom Winn, Ph.D.
  Senior Systems Analyst
  Texas State Auditor's Office
  P.O. Box 12067
  Austin, TX 78711-2067

- Telephone: 512 / 936-9735
- E-Mail: twinn@sao.state.tx.us