FRAUD MAGAZINE
A Publication of the Association of Certified Fraud Examiners
May/June 2005



# Detecting Fraud in the Data Using Automatic Intervention Detection

## Part One

By David H. Lindsay, Ph.D., CFE, CPA, CISA; Paul Sheldon Foote, Ph.D., Annhenrie Campbell, Ph.D., CPA, CMA, CGA; and David P. Reilly

Automatic intervention detection – an advanced computer-based tool – can be used to distinguish companies with fraudulent data from those with no indication of fraudulent reports.

Financial managers, auditors, and fraud examiners are eager to find an automated, timely approach to discovering which publicly traded companies may be harboring financial frauds. Investors, corporate treasurers, and portfolio managers all want a dependable, automatic process to help them sift through voluminous financial data to find the few firms among hundreds that give investors bad information. Sadly, recent history includes many seemingly attractive investment opportunities that became worthless once fraud was discovered.

Imagine, for example, the tasks of the investment operations manager of a university endowment who administers, say, $2 billion in investments. This individual is responsible for analyzing and understanding the performance of all the university's publicly traded, fixed income, and real estate investments. Of course, this individual would find audit reports useful, but they take a lot of time to review and are issued months after the company's books are closed. However, SEC filings and other financial information about these investments is available electronically and could be processed automatically long before the audit is completed. While an automatic process may not definitively identify which investments display fraudulent activity, the process could help the manager choose which investments should be studied more closely than the others. Among the almost endless variety of possible frauds that impact publicly held companies, many are perpetrated by altering the firm's financial records. These direct modifications are termed “interventions” in the terminology of the “Box-Jenkins” time series analysis. This article summarizes a study we conducted, which evaluated whether automatic intervention detection (AID) based on Box-Jenkins can be effectively used to distinguish companies with fraudulent reported data from those with no indication of fraudulent reports.

### Fraud Risk Assessment Tools

We know that fraud is costly. The ACFE's current estimate of the annual direct cost of fraud exceeds

$600 billion [Albrecht and Albrecht, 2002]. The cost in lost investor and customer confidence is incalculable. Auditors face direct costs from fraud whenever shareholders sue them upon discovery of ongoing fraud [Wells, 2002].

Current auditing standards require independent auditors to evaluate the risk that financial statements may have been fraudulently altered. Internal auditors and fraud examiners are expected to search for possible frauds. The best audit tool for fraud discovery has always been the experience, background, and expertise of the individuals on each engagement [Moyes and Hasan, 1996]. Fraud detection skills are in great demand and universities are adding forensic accounting and fraud auditing courses to the curriculum [Davia, 2000].

Decision aids have long included “red-flag questionnaires” which direct auditors' attention to specific fraud indicators [Pincus, 1989]. The red flags are incentives which might lead employees of an audit client to commit fraud and environmental opportunities that may allow fraud to occur [Apostolou et al, 2001]. Additional decision aids are computer based. Discovery sampling, long used by auditors, is used to test for the existence of errors and inaccuracies in randomly sampled sets of transactions. Signs of fraud deserving further investigation may be unearthed through this simple, direct process.

Another statistical approach compares the frequency of digits in reported data to the Benford Distribution, the expected frequency of the digits in numbers naturally issued from some underlying process such as financial record-keeping. Purposefully altered numbers rarely conform to the Benford statistic so a deviation may indicate the need for further examination [Busta and Wienberg, 1998].

Other digital analyses of reported amounts use computers to search through data files for specific items such as even-dollar quantities or amounts just under approved limits. A kind of ratio analysis is available which compares the relative incidence of specific amounts – for example, the ratio of the highest to the lowest value in the data [Coderre, 1999].

Advanced computer-based tools are being developed for reviewing very large quantities of data. Data-mining software can be used to sift through entire databases and sort the information along various parameters to locate anomalous patterns that may require further investigation. Database programming expertise is costly, but less expensive off-the-shelf data mining software can be used in the audits of smaller companies. In all environments, outcomes tend to be better when a fraud examiner is available to review the program output [Albrecht and Albrecht, 2002].

An audit variant of data mining is data extraction in which auditors use software to collect quantities of specific information for review. Extensive training is also needed for effective use of data extraction tools. Such tools are frequently applied when a fraud is already suspected rather than as a routine screening process [Glover, 2000].

As financial reporting has evolved from a periodic, manual activity, tools and technologies for fraud detection are also becoming ongoing, high-technology processes. Continuous auditing processes would support ongoing real-time financial reporting. Such auditing processes may use warehoused data generated by mirroring live data for audit purposes [Rezaee et al, 2002]. Automatic, real-time testing for fraud using neural network technology already is successfully identifying fraudulent credit card and cell phone transactions [ Harvey , 2002]. Computerized neural network simulations are applied to differentiate between patterns inherent in a set of “training” data and sets of test or live data. Neural networks can find discrepancies in data patterns in situations with high quantities of similar and repetitive transactions.

Fraud detection tools are no longer limited to aids for audit specialists to use to review dated information. Newer techniques are more likely to operate on information as it is being produced or released. Automatic, unsupervised detection methodologies will integrate the detection of fraud and the review of financial information to enhance its value and reliability to the user [Pathria, 1999].

**Intervention Detection through Box-Jenkins**

If someone wants to commit fraud by modifying the accounting information system, such modifications would be termed interventions in the terminology of Box-Jenkins time series analysis. The purpose of this study was to evaluate whether intervention detection could be used successfully to distinguish between real companies with reported fraudulent data and real companies with no reports of fraudulent data.

This study proceeds as follows:

1. For a hypothetical entity, show what a spreadsheet user would see in the way of charts and statistics for correct data and for data containing a single fraudulent intervention.
2. For the hypothetical entity, explain how to analyze the results from using a software system designed for interven- tion detection.
3. Describe an experiment of using the data from real firms. Some of the firms in the experiment had reported cases of fraud. Other firms have had no reported cases of fraud.
4. Conclude whether or not intervention detection can detect fraud in the more difficult cases of real firms with data con- taining one or more fraudulent interventions.

**Hypothetical Cyclical Company**

Assume that a hypothetical cyclical company should have reported the following results in millions of dollars for the last 10 periods: 1, 9, 1, 9, 1, 9, 1, 9, 1, 9.

Using Microsoft Excel and an 80-percent confidence level, the summary statistics for a correct report would have been those in Table 1.

| Correct Statistics | |
|---|---|
| Mean | 5 |
| Standard Error | 1.333333333 |
| Median | 5 |
| Mode | 1 |
| Standard Deviation | 4.216370214 |
| Sample Variance | 17.77777778 |
| Kurtosis | -2.571428571 |
| Skewness | 0 |
| Range | 8 |
| Minimum | 1 |
| Maximum | 9 |
| Sum | 50 |
| Count | 10 |
| Confidence Level (80.0 percent) | 1.844038403 |

**Table 1**

Confidence Level, for Microsoft Excel users, is based upon:

= CONFIDENCE (alpha, standard_dev, size)

Estimate the confidence interval for a population mean by a range on either side of a sample mean. Alpha is the confidence level used to compute the confidence level. For this study, we used an alpha of 0.20, indicating an 80 percent confidence level. Standard_dev is the population standard deviation for the data range (assumed to be known). Size is the sample size (an integer).

For the hypothetical cyclical company, we used:

Alpha = 0.20
Standard_dev = 4.216370214
Size = 10

The resulting confidence level of 1.844038403 indicates a range for the mean of:

5 + 1.844038403 = 6.844038403
5 - 1.844038403 = 3.155961597

**Hypothetical Cyclical Company with a Single Outlier**

Suppose, instead, the hypothetical cyclical company issued only a single, fraudulent report in Period 7 by reporting $9 million instead of $1 million.

Using Microsoft Excel and an 80 percent confidence level, the summary statistics for an outlier report would have been those in Table 2, below.

| **Outlier Statistics** | |
|---|---|
| Mean | 5.8 |
| Standard Error | 1.306394529 |
| Median | 9 |
| Mode | 9 |
| Standard Deviation | 4.131182236 |
| Sample Variance | 17.06666667 |
| Kurtosis | -2.276785714 |
| Skewness | -0.484122918 |
| Range | 8 |
| Minimum | 1 |
| Maximum | 9 |
| Sum | 58 |
| Count | 10 |
| Confidence Level (80.0 percent) | 1.806781262 |

**Table 2**

**Hypothetical Cyclical Company with a Single Inlier**

Suppose, instead, the hypothetical cyclical company issued only a single, fraudulent report in Period 7 by reporting $5 million instead of $1 million.

Using Microsoft Excel and an 80-percent confidence level, the summary statistics for an inlier report would have been those in Table 3, below.

| **Inlier Statistics** | |
|---|---|
| Mean | 5.4 |
| Standard Error | 1.257864151 |
| Median | 7 |
| Mode | 9 |
| Standard Deviation | 3.977715704 |
| Sample Variance | 15.82222222 |
| Kurtosis | -2.300034267 |

| | |
|---|---|
| Skewness | -0.237276944 |
| Range | 8 |
| Minimum | 1 |
| Maximum | 9 |
| Sum | 54 |
| Count | 10 |
| Confidence Level (80.0 percent) | 1.739662351 |

**Table 3**

From the two cases (outlier and inlier example), we can see that the confidence level has been substantially unchanged from the correct report along with the standard deviation. Thus standard procedures are useless in detecting fraud.

**Automatic Fraud Detection for the Hypothetical Cyclical Company**

If a company had strong cyclical patterns and a single, large fraudulent intervention, then it would be easy to spot the suspicious interventions simply by generating charts of the data series. However, for real firms, there can be multiple interventions over time and no strong cyclical or seasonal patterns in the data. Fraud examiners and fraud auditors need tools for identifying suspicious interventions in the accounting data of any company.

Automatic ForecastingSystems'Autobox™ (www.autobox.com) and FreeFore™ (www.autobox.xom/freef.exe) are able to detect many types of interventions, including: (1) outliers (2) inliers (3) steps (4) level changes (5) local time trends. In this study, we tested the ability of AID to distinguish companies known to have had fraud cases from those companies with no published indications of fraud.

**Automatic Forecasts for the Correct Series**

Note that Autobox™ identified the strong cyclical pattern of the data and continued that pattern in forecasts for future periods. So, all future forecasts are either 1 or 9. [For those of you who are curious, the model is: Y(t)=10-Y(t-1) +a(t) .]

This is also called “lumpy demand” and is quite prevalent in logistical studies.

**Automatic Forecasts for the Outlier Series**

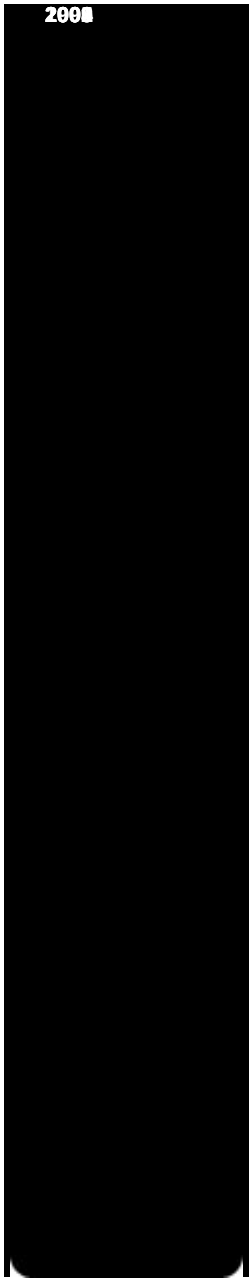
The insertion of one fraudulent report in a series of 10 years of reports has no major impact upon the automatic forecasts. The unusual value (“9”) at time period 8 (1998) is identified and replaced with the normative value of “1.” Forecasts are unaffected by this anomaly and are thus robust. (See Graph 5.)

**Automatic Forecasts for the Inlier Series**

The fraudulent reporting of one inlier value over a 10-year historical period at time period 7 (1997) of the value “5” creates an interesting problem. If the overall mean is approximately “5,” how can the value of “5” be identified as unusual? The answer is simple as the Expected Value isn't always equal to the mean and is based upon patterns in the series. If no pattern exists then the Expected Value is equal to the mean but not otherwise. Autobox™ identifies the anomaly, replaces it with the Expected Value and proceeds.

In the March/April issue: the methodology and applications of AID in the research study

***David H. Lindsay, Ph.D., CFE, CPA, CISA,*** *Professor and Chair of the Department of Accounting*



*and Finance at California State University , Stanislaus. His email address is: DLindsay@csustan.edu.* ***Paul Sheldon Foote, Ph.D,*** *is professor of accounting at California State University , Fullerton . His email address is: pfoote@fullerton.edu.* ***Annhenrie Campbell, Ph.D., CPA, CMA, CGFM,*** *is a professor of accounting at California State University , Stanislaus. Her email address is: acampbel@athena.csustan.edu.* ***David P. Reilly*** *is senior vice president at Automatic Forecasting Systems. His email address is: dave@autobox.com.*

## References

Albrecht, W.S. and C.C. Albrecht (2002) “Root Out Financial Deception,” Journal of Accountancy, (193)4, pp. 30-34.
Apostolou, B.A., J.M. Hassell, S.A. Webber and G.E. Sumners. 2001 “The relative importance of management fraud risk factors,” Behavioral Research in Accounting (13) pp. 1-24.
Box, G.E.P., and Tiao, G. (1975). “Intervention Analysis with Applications to Economic and Environmental Problems,” Journal of the American Statistical Association, Vol 70, 70-79
Busta, B., and R. Weinberg, (1998) “Using Benford's Law and neural networks as a review procedure,” Managerial Auditing Journal (13)6, pp. 356-366.
Coderre, D. (1999) “Computer assisted techniques for fraud detection,” The CPA Journal (69)8, pp. 57-59.
Davia, H.R. (2001) “Fraud 101: Techniques and Strategies for Detection” New York . John Wiley & Sons, Inc.
Downing, D.J., and McLaughlin, S.B. (1986). “Detecting Shifts in Radical Growth Rates by the Use of Intervention Detection,” Engineering Physics and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge.
Glover, S.M., D. Prawitt, M.B. Romney (2000) “The software scene,” Internal Auditor (57)4 pp. 49-57.
Harvey , F. January 12, (2002) “A key role in detecting fraud patterns: neural networks,” Financial Times. London . p. 3.
Moyes, G.D. and I. Hasan. (1996) “An empirical analysis of fraud detection likelihood,” Managerial Auditing Journal (11)3, pp. 41-46.
Pincus, K. (1989) “The efficacy of a red flags questionnaire for assessing the possibility of fraud,” Accounting, Organizations and Society, pp. 153-63.
Reilly, D.P. (1980) “Experiences with an Automatic Box-Jenkins Modeling Algorithm,” Time Series Analysis, ed. O.D. Anderson. ( Amsterdam : North-Holland), pp. 493-508.
Reilly, D.P. (1987). “Experiences with an Automatic Transfer Function Algorithm,” Computer Science and Statistics Proceedings of the 19th Symposium on the Interface, ed. R.M. Heiberger, ( Alexandria , VI: American Statistical Association), pp. 128-135.
Rezaee, Z., A. Sharbatoghlie, R. Elam, P.L. McMickle. 2002. “Continuous auditing: Building automated auditing capability,” Auditing (21)1, pp. 147-163.
Wells, J.T. 2001. “. . . And nothing but the truth: uncovering fraudulent disclosures,” Journal of Accountancy (192)7, pp. 47-52.