International Journal of Accounting Information Systems 11 (2010) 157–181

Contents lists available at ScienceDirect

International Journal of Accounting Information Systems

Data mining journal entries for fraud detection: An exploratory study

Roger S. Debreceny [a,*], Glen L. Gray [b]

[a] Shidler College of Business, University of Hawai'i at Mānoa, United States
[b] College of Business and Economics, California State University at Northridge, United States

ARTICLE INFO

Keywords:
Fraud
Journal entries
Data mining
Auditing
Accounting information systems
XBRL GL

ABSTRACT

Fraud detection has become a critical component of financial audits and audit standards have heightened emphasis on journal entries as part of fraud detection. This paper canvasses perspectives on applying data mining techniques to journal entries. In the past, the impediment to researching journal entry data mining is getting access to journal entry data sets, which may explain why the published research in this area is a null set. For this project, we had access to journal entry data sets for 29 different organizations. Our initial exploratory test of the data sets had interesting preliminary findings. (1) For all 29 entities, the distribution of first digits of journal dollar amounts differed from that expected by Benford's Law. (2) Regarding last digits, unlike first digits, which are expected to have a logarithmic distribution, the last digits would be expected to have a uniform distribution. Our test found that the distribution was not uniform for many of the entities. In fact, eight entities had one number whose frequency was three times more than expected. (3) We compared the number of accounts related to the top five most frequently occurring three last digit combinations. Four entities had a very high occurrences of the most frequent three digit combinations that involved only a small set of accounts, one entity had a low occurrences of the most frequent three digit combination that involved a large set of accounts and 24 had a low occurrences of the most frequent three digit combinations that involved a small set of accounts. In general, the first four entities would probably pose the highest risk of fraud because it could indicate that the fraudster is covering up or falsifying a particular class of transactions. In the future, we will apply more data mining techniques to discover other patterns and relationships in the data sets. We also want to seed the dataset with fraud indicators (e.g., pairs of accounts that would not be expected in a journal entry) and compare the

* Corresponding author.
E-mail addresses: roger@debreceny.com (R.S. Debreceny), glen.gray@csun.edu (G.L. Gray).

doi:10.1016/j.accinf.2010.08.001

sensitivity of the different data mining techniques to find these seeded indicators.

1. Introduction

This paper explores emerging research issues related to the application of statistical data mining technology to fraud detection in journal entries. The detection of fraud and particularly of financial statement fraud[1] has become an increasingly important component of the financial statement audit over the last decade. A number of important financial statement frauds have involved fraudulent journal entries or managerial override of controls that have utilized journal entries within computerized accounting information systems. These journal entries have often involved well-known examples of financial statement fraud including inappropriate revenue recognition, inappropriate capitalization of expenses and a wide variety of inappropriate accruals. Given likely fraudster response to known patterns of fraudulent journal entries such as *non-standard* journal entries[2] and the enormous volume of journal entries in typical computerized accounting information systems, it is questionable that direct auditor assessment of small samples of journal entries will effectively and efficiently detect likely patterns of fraudulent activity. Automated auditor analysis of journal entries has been increasingly mandated by auditing standards in the U.S. and internationally. Some degree of direct computerized analysis of journal entries is now part of the toolkit of audit teams on major audit engagements. There is, however, very little knowledge of the efficacy of this important class of audit procedures.

Although there are large bodies of literature regarding data mining in other domains, a broad search of audit literature did not locate any research literature on the data mining of journal entries.[3] Yet, auditing standards require that auditors consider fraud in their financial audits and those standards specifically require that auditors examine journal entries. Based on the successful applications of data mining to other domains, it would appear that data mining holds the potential to improve both the effectiveness and efficiency of the auditors in their analysis of journal entries and fraud detection. This is in line with recent calls for research on the role of journal entries in the audit process (Curtis et al. 2009).

In this paper, we set out the underlying issues that will guide effective and efficient data mining of journal entries. We review the standards from auditing regulators and guidance from the professional audit community and explore the potential for statistical data mining of large sets of journal entries. We then test the statistical properties of journal entries, in an exploratory study. We make first steps to data mining of such journal entries. These first steps are tested with a set of journal entries for 29 entities. We consider the essential elements of the journal entries. We explore their statistical properties, concentrating on their dispersion from known distributions. We identify some preliminary patterns within the journal entries. The paper makes an important contribution to the literature on data analysis, data mining and fraud detection within journal entries.

The remainder of this paper proceeds as follows: the next section provides general background material and then specifically addresses the role of journal entries in committing fraud and draws lessons from recent frauds that used journal entries. The section also summarizes the responses of standard setters to the heightened fraud risk environment since the late 1990s. In the third section, we explore the issues

[1] The Auditing Standards Board of the AICPA defined financial statement fraud in SAS 99 as: “Misstatements arising from fraudulent financial reporting are intentional misstatements or omissions of amounts or disclosures in financial statements designed to deceive financial statement users where the effect causes the financial statements not to be presented, in all material respects, in conformity with generally accepted accounting principles (GAAP)” (PCAOB 2002). There is a fine line between earnings management and financial statement fraud, but a line that is beyond the scope of this paper. We confine the discussion to deliberate and intentional material misstatements, typically undertaken by one or more members of senior management.

[2] According to the report of the Panel on Audit Effectiveness, “Non-standard entries is a term that is not precisely defined, although it is in common use among accountants and auditors. Such entries sometimes are referred to as “top-side entries,” “post-closing entries,” “manual adjustments,” “management entries” or “unusual adjustments.” In general, they are financial statement changes or entries made in the books and records (including computer records) of an entity that usually are initiated by management-level personnel and are not routine or associated with the normal processing of transactions.” (POB 2000, 83).

[3] This lack of published literature does not mean that the audit firms are not doing any journal entry data mining. Quite the contrary, the firms are deploying data mining technology, but what they are doing is proprietary and, as such, rarely gets published for public consumption.

involved in data mining journal entries. We discuss both the technical and the statistical properties of journal entries and how data mining can leverage the economic relationships embedded in the account combinations represented in the journal entry. In the fourth section, we introduce our data set. We then discuss our initial exploration of the statistical properties of the journal entries in our data set in the next section. In the final section, we draw conclusions and point to a research agenda.

2. Background

Over the last several years, there has been an increased emphasis on the detection of fraud as a key element of the financial statement audit. In 2000, the AICPA's Public Oversight Board's Panel on Audit Effectiveness pointed to a variety of necessary reforms to ensure the long-term viability of the audit (POB 2000). The significant frauds that involved manipulation of financial statements and disclosures in the late 1990s and early part of this century gave added impetus to a fundamental shift in the conduct of audits. A number of fraud schemes involved non-standard journal entries that were designed to make relatively simple adjustments between classes of accounts such that the financial statement results would show an improved position at the margin.

As a response to significant financial statement frauds over the last decade or so, there have been a number of changes to auditing standards and the regulatory environment governing the profession of auditing. The promulgation by the Auditing Standards Board (ASB) of SAS 99 (Consideration of Fraud in a Financial Statement Audit) (ASB 2003) and the enactment of Sarbanes–Oxley Act (SOX) by the U.S. government were central events. SAS 99 significantly increased the responsibility of auditors to address potential fraud as an integral part of financial audits (ASB 2003; CAQ 2008). For example, SAS 99 requires the direct assessment of journal entries for fraud risk. The International Auditing and Assurance Standards Board (IAASB) followed SAS 99 with similar language in their IAS 240 (IAASB 2009a). While individual frauds have been substantial and the range of fraud techniques employed broad, the proportion of frauds within the broader population of audit clients is minuscule. In support of the preparation of financial statements and accompanying notes and other disclosures, audit clients employ sophisticated information systems that generate vast quantities of electronic evidence. Finding evidence of fraud detection within this information milieu is challenging. Employing data mining has the potential to improve the efficiency and effectiveness of audit teams in the conduct of fraud-related audit tasks.

Modern accounting information systems increasingly record transactions in the general ledger at the atomic level. It is common for entities to have several hundred thousand journal entries in a given accounting period. Managers intent on committing fraud may also choose to conceal fraudulent transactions within other transactions in “mega-combined” entries. These factors make data mining of journal entries to detect fraud a challenging exercise.

The remainder of this section takes a closer look at financial statement fraud involving journal entries and the response of standard setters to heightened risk from fraud.

2.1. Financial statement frauds involving journal entries

As introduced in the previous section, the focus of this paper is data mining of journal entries within computerized accounting information systems. This potential class of substantive tests is designed to support the auditors' assessment of material misstatements in the financial statements arising from fraud. While misappropriation of assets is important, detection of financial statement fraud is of greater concern to investors and other stakeholders. This latter type of fraud usually has greater probability of giving rise to a material misstatement and to be committed by upper management. The focus of this study is with the use of journal entries to facilitate fraud and techniques to discover potentially fraudulent entries.

Amongst the panoply of financial statement frauds, the fraud at WorldCom Inc. is perhaps the most egregious. WorldCom provides a useful model of how financial statement frauds typically involve many adjusting journal entries. The WorldCom fraud was relatively straightforward, primarily involving adjustments from expense accounts to capital expenditure accounts. As the special report to the Board of Directors on the fraud shows, there was no justification in accounting principles or practice for these material adjustments (Beresford et al. 2003). The amounts were large and well known within the corporation. Some of these adjustments were even the topic of conversation between accountants in international operations and local auditors (Woods 2002). The WorldCom journal entries provide a useful example of the way in which frauds involve multiple signals that must be identified in the aggregate.

The special report described seven important characteristics of the adjustments and of the use of journal entries at WorldCom. First, the fraud involved straightforward and inappropriate accounting reallocations. These included transfers from flows to stocks. For example, significant transfers were made from what was effectively a suspense expenditure account, "Prepaid Capacity Costs," to the "Construction in Progress" account, which was treated as capital expenditure (Beresford et al. 2003). Second, journal entries also involved accounting treatments designed to influence disclosure rather than recognition. For example, line costs were transferred to accounts that rolled up into "Selling, General and Administrative Expenses (SG&A)." These adjustments did not change the reported profits, but did change the allocation between gross and net profit disclosures (Beresford et al. 2003). This change in disclosure was to influence the conclusions of analysts on WorldCom financial performance. Third, many of the suspicious journal entries were ill concealed, with large adjustments in rounded amounts that would be obvious to the most casual of inspections (Beresford et al. 2003, 126). Fourth, there were a large number of inappropriate or at best questionable journal entries. The special report noted that "[w]e found *hundreds* of huge, round-dollar journal entries made by the staff of the General Accounting group without proper support…" (Beresford et al. 2003, 244. Emphasis added.), Fifth, inappropriate journal entries were often accompanied by inadequate or no documentation and which circumvented normal internal controls. Sixth, the adjustments were almost universally at the corporate level. In many cases, however, these non-standard adjustments made at the corporate level required adjustments at operating divisions and international operations. Seventh, many individuals and groups within the corporation quickly became aware — or should have been aware — of the implications of fraudulent entries passed at headquarters, not the least of which was as the result of sweeping up after the aforementioned non-standard adjustments (Beresford et al. 2003).

Perhaps the most interesting aspect of the WorldCom case from the perspective of this paper is the statement that "WorldCom personnel also repeatedly rejected Andersen' s requests for access to the computerized General Ledger through which Internal Audit and others discovered the capitalization of line costs" (Beresford et al. 2003). There might have been a very different outcome to recent US corporate history if Andersen had more vigorously pursued electronic access to the General Ledger.

The WorldCom case was particularly egregious and, as the special report to the Board of Directors clearly describes, the potential red flags for the auditors were many and varied. Nonetheless, many of the same red flags exist in other financial statement frauds. The Cendant Corporation fraud that pre-dated the WorldCom fraud was almost a word-for-word transcription. A 1998 report to the Audit Committee of the then Cendant Corporation noted that in what "shows to have been a carefully planned exercise," a large number of "unsupported journal entries to reduce reserves and increase income were made after year-end and backdated to prior months; merger reserves were transferred via inter-company accounts from corporate headquarters to various subsidiaries and then reversed into income; and reserves were transferred from one subsidiary to another before being taken into income" (Willkie et al., 1998).

Perhaps what distinguishes the Cendant case from WorldCom was the wide range of accounts and accounting treatments that were involved in the fraud at Cendant. Hundreds of journal entries were required to achieve the desired impact on net income. The fraudulent entries impacted revenue, cash, accounts receivable and deferred revenue. Wallace (2000) noted that control violations within Cendant were highly disaggregated. Auditors and others charged with discovering these violations may need to aggregate these disaggregated transactions in order to see the broader picture and be in a better position to identify the control violations. A similar picture of journal entries at the heart of financial statement frauds can be drawn in many other exemplars of the last decade, including Health South (Weld et al. 2004) and Xerox, Enron and Adelphia (BFA) (DeVries and Kiger 2004).

2.2. Response of standard setters to heightened risks from fraud

The heightened recognition of the importance of financial statement fraud in the 1990s leads to an increased emphasis on fraud amongst auditing standard setters. The then Public Oversight Board (POB) of the AICPA provided some of the most influential guidance on how auditing should respond to this heightened risk environment in the report of their "Panel on Audit Effectiveness" (POB 2000). The Panel conducted reviews of working papers, which the Panel termed "Quasi Peer Reviews," for a significant number of audits. In addition, they reviewed the SEC's Accounting and Auditing Enforcement Releases (AAERs) over the previous two-year period, which was a particularly active period in SEC enforcement. When reviewing the response of auditors to high levels of fraud risk, the Panel made particular note of the failure of audit teams to assess "non-standard" journal entries. In some 15% of cases, auditors did not have a sufficient understanding of the client systems for preparing such entries. In nearly one third of other cases, the audit teams did not undertake substantive tests of non-standard journal entries (POB 2000). The Panel recommended to the ASB "develop stronger and more definitive auditing standards to effect a substantial change in auditors' performance and thereby improve the likelihood that auditors will detect fraudulent financial reporting" (POB 2000). The Panel made a series of detailed and integrated proposals; the most important of which for the purposes of this paper was that the audit contains a "forensic-type fieldwork phase." This was not to turn the audit into a forensic investigation, which would dramatically change the character of the audit. Rather, the proposals were to bring selected forensic techniques to the financial statement audit. Unsurprisingly, given their findings, the Panel made specific recommendations on direct examination of "non standard" journal entries. The Panel noted that "[a]ll or virtually all entities record non-standard entries. These entries can provide an avenue for management to override controls that could lead to fraudulent financial reporting. Consequently, auditors need to design tests in the forensic-type phase to detect non-standard entries and examine their propriety. This aspect of the forensic-type phase affects not only the extent of testing, but also its timing, because such entries can be recorded at various times during the year" (POB 2000).

The response of the ASB was SAS 99, "Consideration of Fraud in a Financial Statement Audit" (ASB 2003).[4] This standard notes that a material misstatement in financial statements can arise from fraudulent financial reporting, defined as "intentional misstatements or omissions of amounts or disclosures in financial statements designed to deceive financial statement users" and from misappropriation of assets (ASB 2003). SAS 99 requires that the auditor undertake a variety of analytical and planning tasks and substantive audit procedures to support the detection of errors arising from fraudulent financial reporting. The standard makes particular note of the role of journal entries and other adjustments in the conduct of financial statement fraud. SAS 99 imposed a considerably enhanced set of requirements on the auditor. The standard required auditors to "design procedures to test the appropriateness of journal entries recorded in the general ledger and other adjustments (for example, entries posted directly to financial statement drafts) made in the preparation of the financial statements" (ASB 2003).

SAS 99 provides detailed guidance on selection of entries and adjustments, requiring the auditor to assess the risk of misstatement from fraud, effectiveness of controls over journal entries and the nature and complexity of entries and accounts. The standard identifies markers of fraudulent entries including: "entries (a) made to unrelated, unusual, or seldom-used accounts, (b) made by individuals who typically do not make journal entries, (c) recorded at the end of the period or as post-closing entries that have little or no explanation or description, (d) made either before or during the preparation of the financial statements that do not have account numbers, or (e) containing round numbers or a consistent ending number" (ASB 2003). Auditors are cautioned that they should pay particular attention to non-standard entries and to other adjustments such as consolidation entries.

Finally, SAS 99 makes a number of explicit requirements for the auditor to undertake substantive tests of the detail of controls and transactions. The standard notes that fraudulent journal entries are likely to occur around the closing process and that, consequently, testing should concentrate on entries posted in the period leading up to the fiscal year end or during the preparation of the financial statements. Indicative tests of the journal entries data set include:

- Non-standard journal entries
- Entries posted by unauthorized individuals or individuals who while authorized do not normally post journal entries
- Unusual account combinations
- Round number
- Entries posted after the period-end

[4] SAS 99 is now an interim audit standard of the Public Company Accounting Oversight Board.

• Differences from previous activity
• Random sampling of journal entries for further testing

The detailed requirements of SAS 99 were a considerable augmentation to those of its predecessor, SAS 82. Because of the detailed requirements of SAS 99, a major thrust of audit firms has been to develop technologies, policies and procedures designed to enable them to fulfill these requirements. More recently, the ASB has also addressed the question of journal entries in their so-called "risk standards." These include SAS 109 on understanding the entity and its environment (ASB 2006a) and SAS 110 on audit procedures (ASB 2006b). SAS 109 requires that the auditor assess the manner in which information is moved to the general ledger from other systems, how system and non-standard journal entries are created and controlled and the role of consolidation and close processes (ASB 2006a). These requirements are incremental to those in SAS 99, arguably ensuring that the auditor develops a sophisticated understanding of the close process and the roles played by general ledger journal entries. The Center for Audit Quality has also provided guidance on the processes involved in selecting, acquiring, testing and analyzing journal entries for fraud detection (CAQ 2008).

The International Auditing and Assurance Standards Board (IAASB) have taken a somewhat more nuanced approach to the audit of journal entries. In 2003, the IAASB revised their fraud standard, ISA 240, at least in part as a response to the increased requirements of SAS 99. After recent redrafting, ISA 240 requires that "irrespective of the auditor's assessment of the risks of management override of controls," the auditor should "test the appropriateness of journal entries recorded in the general ledger and other adjustments made in the preparation of the financial statements" (IAASB 2009a).[5] Given that fraudulent managerial actions often take place at the end of accounting period and in the closing process, the standard requires that the auditor "select journal entries and other adjustments made at the end of a reporting period." The standard also requires that the auditor "consider the need to test journal entries and other adjustments throughout the period" (IAASB 2009a, para 32). Factors to take into consideration in the selection process identified by the Board are somewhat similar to SAS 99, including assessment of risk of misstatement from fraud, controls over journal entries and adjustments and the nature and complexity of the evidence environment. The Board also provides a checklist of markers of potentially fraudulent journal entries, again, similar to SAS 99. Similarly, the audit evidence standard, ISA 330, notes that the auditor should examine "material journal entries and other adjustments made during the course of preparing the financial statements" (IAASB 2009b, para 20).

Taken together, SAS 99, the two new risk standards (SAS 109 and 110) and ISA 240 considerably increased the requirements on the auditor to assess the controls over journal entries and consider the fraud risk environment as it impacts the creation of different classes of journal entries. These standards radically changed conduct of substantive tests on journal entries. While prior to SAS 99 auditors might have inspected suspicious journal entries in exceptional circumstances, direct tests of journal entries and adjustments is now a standard element of the audit at least for larger and higher risk clients. The proportion of audits that are subject to full analysis of journal entries is significant and growing. In a recent survey of auditors at national, regional and local firms, Janvrin et al. (2009, 106) show that Computer Aided Audit Tools (CAATs) are now used in nearly half of audits to identify journal entries and other adjustments to be tested. This change in the nature and extent of substantive tests has come at the same time as considerable changes in the information technology environment in which journal entries are processed. The following section addresses this new technology environment.

3. Understanding the properties of journal entries

In this section, we address how to build a systematic understanding of a set of journal entries that may contain deliberate deception or other signals of financial statement fraud or misappropriation of assets. As we will discuss in more detail later in the section, there is sparse research on the interrogation or data mining of journal entries. Given the parlous state of research, it is perhaps necessary to turn to first

[5] At the time of writing, there has been recent rule making at the ASB and PCAOB. The ASB has an exposure draft that more closely aligns their standards with those of the IAASB, including the fraud standard (ASB 2009). The PCAOB is also in the process of promulgation of new standards on audit risk. The new standards that will arise from this process will result in consequential amendments to SAS 99 (PCAOB 2008). In each case, there are changes in the language but not in the substance of their treatment of journal entries in the audit process.

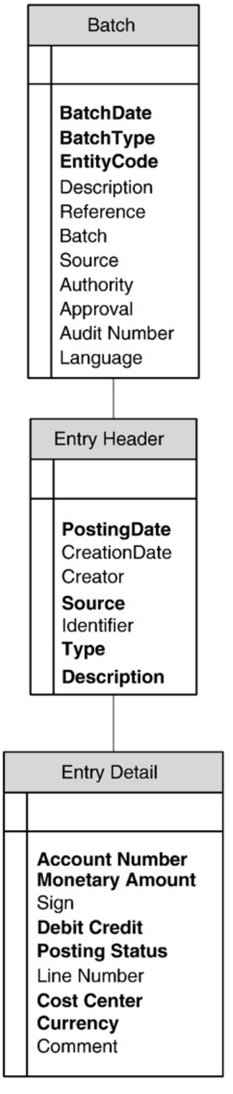

Fig. 1. Data structure of journal entry. (Based in part on the XBRL GL taxonomy and SAP R/3 general journal ledger entry).

principles and build a new research agenda and a new program for extracting knowledge from journal entries. An important element of any program of data mining is developing knowledge of the properties of the subject of the investigation.

As with many subjects of data mining, journal entries have a number of attributes that must be assessed both individually and taken together. Fig. 1 shows a simplified data model of a typical journal entry. There is journal entry header information that uniquely identifies the journal entry (JE). Some of that header information is entered by the user, but most of the information is automatically assigned by the software. The entry detail has one database record for each line in the body of the JE. In database terminology, there is a one-to-many relationship between the entry header information and the entry detail information. Since any JE, at minimum, would be expected to have at least one debit and one credit, we would expect at least two entry detail records for each entry record. In addition to the JE itself, sometimes a batch (or group) of

JEs are submitted and processes together. In that case there would be batch data elements similar to those listed at the top of Fig. 1.

3.1. The technology environment

The effective and efficient data mining of journal entries requires comprehensive understanding of likely markers (red flags) of fraudulent entries or adjustments, statistical properties of journal entries and the technological environment in which the client transacts the journal entries. In the following paragraphs, we address the last of these considerations. There are a number of important technological and policy issues to consider before deliberating on the most effective and efficient form of data mining of journal entries. First, to what extent are journal entries passed in subsidiary ledgers and only summarized in the General Ledger? Second, what automated controls exist over the passing of journal entries? Third, how does the client process accounting estimates, consolidations, adjustments and other fine-tuning into journal entries? Fourth, what are the implications of bespoke and third party analytical and consolidation software applications in generating "mega-combined" journal entries? We address each of these issues in turn in the following discussions.

3.2. Granularity in processing of journal entries

The General Ledger (GL) within the accounting information system is the final repository for the economic impact of all economic events that affect an organization and, by extension, the financial statements. The coupling of the General Ledger to the original processing of the economic event that triggers the event is an important factor in the design of data mining solutions. General Ledger systems typically process data that may have arisen from a variety of transaction processing systems, including: sales, purchases, logistics, maintenance and manufacturing sub-systems. The General Ledger could receive transactions from those systems at different levels of aggregation. In many older and legacy accounting information systems, data are coming in at a high level of aggregation. There may be a single journal entry each month that captures all the multitude of transactions that were processed by the particular sub-system. For example, sales systems will transmit a single entry each month that will have aggregated postings to revenue, cost of goods sold, accounts receivable, etc. Automated drill down from the General Ledger to the original transaction is usually difficult or infeasible in such systems. The mining of value-adding processes at the General Ledger level is then not necessarily particularly productive.

Conversely, non-standard journal entries, such as closing adjustments that may be markers of financial statement fraud, are likely to be relatively obvious to the auditor of these more traditional or legacy accounting information systems. In these systems, non-standard journal entries can be observed among a relatively small number of journal entries. They are likely to have clearly identified markers that can be observed by the data mining application.

On the other hand, in many client instantiations of ERP[6] systems, the General Ledger records the transactions at the atomic level. It is not uncommon for clients to have several hundred thousand journal entries. Certain transaction classes may be aggregated, but are exceptions to this rule. These include low transaction value retail sales and individual monthly billings for telecommunications clients. Nonetheless, the GL is typically a very rich information environment. Further, many clients maintain closely coupled transaction processing systems. A drill down is increasingly feasible from the GL to, for example, sales processing and customer relationship systems.

Whether data mining techniques can identify potentially fraudulent journal entries in a population of hundreds of thousands of journal entries is still an unanswered empirical question. Identifying non-standard journal entries within a large population of journal entries is also particularly challenging.

[6] Enterprise Resource Planning is a class of enterprise wide application software for tracking financial transactions, logistics, sales, purchases, inventory etc.

3.3. Adjusting journal entries

Adjusting journal entries for estimates to accounts (e.g., estimate for bad debt allowance) are a focal point for financial statement fraud and require a highly developed approach to data mining. There is an enormous variation in how clients calculate these many estimates that are used in financial reporting. Many adjusting journal entries will be as the result of spreadsheet analysis of accruals, reserves, impairments, etc. These spreadsheet analyses are an important source of control risk (Janvrin and Morrison 2000; Panko 2006; PricewaterhouseCoopers. 2004). In other cases, adjusting entries might result from stored procedures that are triggered in business intelligence data warehouses or from programs written in SAP's ABAP language. The logic of the journal entry will rest in the supporting application and require further inquiry or assessment. When assessing what may seen to be questionable journal entries the auditor must consider the risks associated with the particular method for generation of the journal entry.

3.4. Classes of journal entries and adjustments

There are several classes of General Ledger journal entries. The bulk of journal entries within General Ledger systems are so-called “system” entries. These entries will normally be posted as result of the conclusion of some phase in a business process such as the acquisition of inventory or at the delivery and billing of goods or services. Such journal entries are posted under the control of the application software. In most cases, these system entries result in an appropriate reflection of the nature of the business process in the accounting system and do not represent fraud. However, some of these entries may be because of systematic fraud. For example, in financial statement frauds that entail inappropriate revenue recognition may well involve system transactions that have been fraudulently entered into the accounting information system at the direction, intervention or insistence of some level of management within the enterprise.

A second class of journal entries are the “top-line” journal entries to the General Ledger (CAQ 2008, 12). As with system entries, the vast majority of these transactions are entirely appropriate. For example, a number of these entries results from analyses conducted by enterprise staff. Entries to adjust allowances for doubtful debts to reflect debtor payment histories at the end of the period; rollback of inventory to adjust standard costs to lower of cost or market and adjustments to account for impairment to the valuation of acquired goodwill are all examples of “top-line” journal entries that are normally valid and appropriate. Equally, any of these entries could be inappropriate and be evidence of fraud.

A third class of journal entries is comprised of “mega-combined” journal entries. These are entries are pushed to the General Ledger from analytical and consolidations systems. These systems include custom applications for managing entity-specific accounting accruals and estimates. Examples are applications for managing exposures, such as the valuation of financial instruments, warranty provisions and capital leases. Consolidation and rapid-close application, such as those from Hyperion,[7] are a second example of systems that give rise to “mega” journal entries. These mega entries will typically be few in number and impact upon many accounts — at worst, a single journal entry transferred from Hyperion might post to several hundred accounts in the General Ledger system.

Journal entries arising from these systems are likely to be high value from a fraud perspective given that they are by definition arising not from a particular internal or external atomic transaction, but from a review of an accrual or adjustment. Employing known markers of fraudulent entries may be highly debatable when the auditor views so many economic assumptions for accruals or adjustment through the lens of such a highly aggregated single journal entry.

A challenge also arises from the disparate locus of control between the General Ledger and the analytical and consolidations systems. These systems may or may not have their own systems of controls and transaction logs. The Oracle Hyperion Financial Management software package, for example, has sophisticated built-in controls and logs. The controls over custom, in-house developed systems are likely to vary widely. Nonetheless, controls over these accruals and adjustments are split between the general ledger and the analytical system, making data mining potentially difficult.

[7] Hyperion is now part of Oracle Inc.

3.5. Understanding the statistical properties of journal entries

At the most elementary level, we can see a set of journal entries as independent members of a population of events, corresponding to a known or expected distribution. Unfortunately, research into the statistical properties of journal entries appears to be a null set. An extensive literature review did not identify a single paper, other than the discussion of applying digital analysis (significant digit law) that we discuss in the next paragraphs. There is no literature that models the statistical properties of populations of journal entries. Nor is there a literature that takes exemplar databases of journal entries and tests the statistical properties of those databases. This is indeed surprising, given the public policy importance of financial statement fraud or the centrality of these assessments to the value adding characteristics of audit firms. The literature on audit sampling is of only limited value to the discussion of data mining journal entries. When assessing journal entries, the problem is not generating a representative sample since the auditor has the complete population available in electronic form. Rather, the task is to identify those journal entries that are anomalous and potentially indicative of fraud. What distribution will represent such a population?

Digital analysis is a generic term employed in the forensic accounting and auditing profession for investigations of leading digits within populations of interest. The Significant Digit Law,[8] which is at the heart of digital analysis, shows that leading digits in a variety of populations are not normally distributed (Hill 1995). To the contrary, they follow a logarithmic distribution. As Hill (1995) notes, empirical evidence for this law has been found in a wide range of natural, as distinct from artificially constructed, populations. An alternative approach using a Bayesian approach has been proposed by Ley (1996) and supported by evidence from extensive simulations by Geyer and Williamson (2004).

Digital analysis has been employed to determine fraudulent patterns of data in operations management (Hales et al. 2008), scientific publishing (Diekmann 2007) and earnings management (Guan et al. 2006; Skousen et al. 2004). Digital analysis has been strongly recommended by Nigrini and others as a vital, even essential, tool in fraud detection (Nigrini 2000; Nigrini and Mittermaier 1997). While there are clear challenges with the practical application of digital analysis (Cleary and Thibodeau 2005), it remains a very useful tool for detection of possible fraud within a large data set.

While the first digits of journal entries are of considerable interest in the detection of fraud, the final digits are also of considerable interest. Are there journal entries with significant numbers of zeroes or other indications of fraud, as suggested in SAS 99? We are particularly interested in the three to six digits from the right, as indicative of thousands to millions of dollars. After three digits from the left, we expect that digits will appear with equal probability. Detection of unusual patterns in the right-most digits can employ traditional parametric measures such as goodness of fit, skewness and kurtosis. Cheng and Hall (1998) note that these tests may influenced almost as much by the validity of the particular parametric model e.g. by the weight of the tails of the fitted distribution as by the hypothesis of homogeneity." An interesting solution to this problem is suggested by Hartigan and Hartigan (1985). Their "dip" statistic measures the maximum difference between the unimodal distribution function (worst case) and equal distribution, as the most extreme modality. The dip test allows us to see patterns in the right hand side digit that might indicate fraud.

3.6. Understanding the General Ledger structure

Each journal entry must be interpreted in light of the chart of accounts. The structure of the chart of accounts for the General Ledger is specific to the particular entity. The concept of journal entry with "unusual account combinations" requires matching the conceptual understanding of such combinations to the client's chart of accounts. Typically, the auditor develops an internal taxonomy to represent a generic chart of accounts. The generic chart of accounts allows analysis of the journal entries in terms of unusual patterns of activity. These patterns include abnormal volumes of transactions to particular classes of accounts; transactions to classes of accounts at atypical times in the closing cycle; and journal entries made to unusual combinations of accounts.

[8] This is commonly known as Benford's Law (Benford 1938) who independently determined relationships found decades previously by Newcomb.

The auditor and, particularly, those from audit firms that have centralized data collection and analytical functions must map these standard templates or taxonomies to the client's chart of accounts. There is clearly a significant time cost in matching the hierarchy of the client's chart of accounts to the standard taxonomy. Whenever the client modifies its chart of accounts, the data collection team must adjust the mapping of the client's modified chart of accounts to the audit firm's generic chart of accounts taxonomy. There is a potential role for XBRL GL in providing more sophisticated mapping of the client's general ledger to the generic taxonomy. The mapping may include not just the date, account and transaction amount but information on controls and data sources.

The significance of the double entry accounting system as the foundation for an *n*-dimensional matrix representation of the value adding activities of the companies is well-known (Ijiri 1975; Ijiri and Kelly 1980; Leech 1986; Mattessich 1964, 2003; Sampson and Olan 1992). At a minimum, each journal entry involves at least two accounts as well as a time dimension. In the case of systems journal entries, any one transaction may simultaneous post to several hundred accounts. The accounts involved within a journal entry form part of an accounting taxonomy. The journal entry line amounts must be interpreted in relation to this taxonomy as well as the amount, statistical properties and temporal characteristics of the transaction. In addition, each transaction is not an entity unto itself. Each transaction has to be interpreted in light of all the other transactions that impact upon an individual account or group of accounts. Taken in the totality, this set of attributes provides a very rich population to data mining software. Yet, how auditors can apply matrix techniques to the analysis of journal entries is still highly tentative and speculative (Arya et al. 2000; Sampson and Olan 1992). Much additional research is required to better explore how matrix representations could be better integrated into journal entry analysis.

3.7. Putting it all together

In summary, the questions that affect the application of data mining to journal entries in the audit are:

- What are the sources of the journal entries? How do those sources influence data mining for all enterprises? For the particular enterprise?
- Are there unusual patterns in the journal entries between classes of accounts?
- Are there potential overrides in controls over processing of journal entries?
- Does the class of journal entry influence the nature of the journal entry? For example, do adjusting journal entries carry a greater probability of fraud?
- Is there evidence of unusual patterns in the amount of the journal entries either from the left most digits (Benford's Law) or from the right most digits (Hartigan and Hartigan's dip test)?
- How can we triangulate and combine these various possible drivers of fraud in the journal entries to allow directed data mining?

4. Investigating populations of journal entries

We now move to the first stages of assessing the questions we set out in the previous section. One of the major difficulties for researchers in the financial fraud domain is obtaining access to real-world internal accounting data to test various hypotheses and models. For this paper, we were fortunate to have access to a large database that included data sets of journal entries for a wide variety of organizations. Specifically, an anonymous software vendor provided journal entries for 36 organizations. There are journal entries for two years for two organizations. The software vendor removed any identifying information from the files prior to their transfer. Unfortunately, there was no access to the opening trial balance for these organizations. Nor were there details on the source of the journal entries or the employees involved in passing the entries. These organizations were from both the for- and not-for-profit sectors. As evidenced from the style of the journal entries and charts of accounts, the underlying accounting information systems were all different. Eight of the data sets were for periods less than twelve months and were excluded from the analysis below. One organization's journal entry data set was incomplete and was also eliminated from the analysis.

4.1. Mapping individual charts of accounts

One of the challenges in analyzing this journal entry data was that each organization had its own chart of accounts. Using the disparate charts of accounts would have made cross-sectional analysis very difficult. As such, the first step in preparing this data for analysis was to create a comprehensive standard chart of accounts. Then each organization's chart of accounts was mapped to the standard chart of accounts. Table 1 shows the descriptive statistics of the 29 sets of charts of accounts of the organizations we studied. There are a relatively small number of accounts in use in most of the organizations, with a small number of organizations having complex charts of accounts.

We constructed a master chart of accounts with a "five–four" structure. The first five digits designate the primary account and the second four digits for the sub-account. In most cases, the four digits correspond to a particular sub-account for one of the entities in the sample. Conversely, the master accounts (five digits) are part of the logical structure of the master chart of accounts. There are 1672 accounts (five–four) in the master Chart of Accounts, with 343 primary (five digits) accounts. The resulting database that we used for our subsequent analysis had a total of 496,182 line items across the 29 organizations. There is considerable variation in posting to the various accounts. Table 2 shows the number of transactions per primary account:

There are ten accounts with more than 10,000 transactions each, including the usual suspects of Accounts Receivable (38,714 transactions), Accounts Payable (44,916 transactions) and Salaries and Wages (44,158 transactions).

4.2. Descriptive statistics

Table 3 shows some basic descriptive statistics for the 29 organizations. The first observation that arises from these descriptive statistics is the difference between the organizations. The first column lists the total number of journal entry line items within the fiscal year. The highest number was nearly 154 thousand and the lowest was less than one thousand. The dollar values of the journal entry lines also varied widely. The maximum journal entry line item for the *ChiEta* entity was $362,478,016. The smallest maximum was $34,929 for *Pi*.

The relatively larger maximum entries probably indicate that summary journal entries transferred information between accounting modules. For example, perhaps only one journal entry was used to transfer summary information once a month from the accounts payable module to the general ledger. Summary versus detailed journal entries will be part of the auditor's risk analysis and their subsequent develop of the audit program. For example, if only monthly summary journal entries are used, then the auditor is going to focus on the source modules (e.g., accounts payable). From a data mining perspective, it is probably better if details are transferred to the general ledger because the general ledger will then be essentially a large, comprehensive database. If, on the other hand, the details are kept in each module, then each module is its own isolated database. Being able to data mine across modules can be important in an audit. For example, a common search is to find any vendor addresses (stored in the vendor master file) that are the same as employee addresses (stored in the employee master file) (CAQ 2008). Matches could mean that an employee has set up a fake vendor that is subsequently receiving checks from the company.

The number of journal entries and line items per journal entry varies widely. Table 4 shows the number of individual line items that make up the various journal entries for each entity. The first column shows the number of distinct journal entries in the year (*N*). Then we show the mean, standard deviation and minimum and maximum number of line items per journal entry. There were organizations that have journal entries with very large numbers of line items (e.g. Beta, Chi and Zeta). These are examples of so-

Table 1
Active accounts in organizational chart of accounts.

Minimum	43
Maximum active accounts	1036
Median active accounts	107
Average active accounts	164

Table 2
Transactions per five digit accounts in master chart of accounts.

Minimum	1
Maximum	44,916
Median	86
Mean	1401
Standard deviation	4784

called "mega entries," where transactions are coming either from stored automatic journal entries that reverse prior-period adjusting journal entries or transfer data from subsidiary systems.

5. Statistical analysis of journal entries

In this section, we investigate the statistical analysis of our set of journal entries.

5.1. Digital law

Digital analysis (also first-digit law or Benford's Law) is a statistical technique regularly discussed in the professional guidance on fraud detection in general (Benford 1938; Nigrini 2000; Nigrini and Mittermaier 1997; Tackett 2007) and journal entries in particular (CAQ 2008). Digital analysis predicts that the first digit of a set of numbers will have the distribution shown in Table 5.

Table 3
Descriptive statistics for organizations.

(1) Entity	(2) Number line items	(3) Total line items $(000)	(4) Maximum line item $
Beta	40,617	$240,221	$2,927,854
Chi	153,800	$60,889,933	$250,650,816
ChiEta	18,572	$13,396,011	$362,478,016
ChiNu	2421	$27,374	$653,316
ChiPi	4871	$2712	$495,667
Delta	29,866	$78,716	$393,500
Eta	689	$215,237	$12,000,000
EtaNu	2318	$43,517	$489,822
EtaPi	1445	$23,463	$685,613
Gamma	4433	$38,823	$618,214
Kappa	7244	$97,421	$576,281
KappaXi	7210	$41,261	$464,904
MuXi	11,531	$1,730,387	$24,223,476
Nu	3182	$6533	$425,000
Omicron	5303	$195,516	$12,893,261
Phi	8410	$19,229,705	$549,332,992
PhiPsi	38,329	$41,455	$663,000
Pi	1426	$2140	$34,929
PiNu	6998	$13,940	$324,012
Psi	3258	$2569	$46,710
Rho	4579	$4529	$80,000
Sigma	1378	$840	$19,367
Tau	1377	$863	$15,353
Theta	1739	$1516	$41,785
Upsilon	4524	$9337	$129,566
Xi	30,174	$674,415	$9,016,084
XiNu	2781	$154,551	$1,637,364
XiRo	32,554	$5,381,983	$38,741,784
Zeta	62,638	$11,094,568	$38,232,288

Table 4
Line items per journal entry.

(1) Entity	(2) *N*	(3) Mean	(4) Std Dev	(5) Min	(6) Max
Beta	1097	37	48	2	642
Chi	5275	29	132	2	1375
ChiEta	2915	6	8	2	66
ChiNu	502	5	7	2	34
ChiPi	2422	2	1	2	32
Delta	3576	8	16	2	318
Eta	144	5	7	2	46
EtaNu	400	6	5	2	32
EtaPi	484	3	2	2	14
Gamma	541	8	13	2	118
Kappa	991	7	8	2	66
KappaXi	2325	3	4	2	45
MuXi	2790	4	4	2	36
Nu	536	6	10	2	65
Omicron	578	9	20	2	387
Phi	1433	6	6	2	35
PhiPsi	9215	4	3	2	85
Pi	360	4	4	2	44
PiNu	1086	6	9	2	67
Psi	859	4	6	2	76
Rho	826	6	8	2	58
Sigma	491	3	3	2	45
Tau	412	3	2	2	23
Theta	484	4	5	2	34
Upsilon	864	5	6	2	63
Xi	552	55	73	2	450
XiNu	986	3	3	2	21
XiRo	2637	12	36	2	525
Zeta	4682	13	42	2	736

Table 6 shows the number of journal line items for each organization with a particular first digit of the dollar amounts, the actual distribution (Act%) of those digits, and the variance from the expected distribution (Diff = Act% − Expected%). We show the Chi-square and *p*-value for each entity. For every one of the 29 entities in the study, the Chi-square distribution indicates that the observed pattern of first digits differs from that expected by Benford's Law. If we assume that Benford's Law should apply to a population of journal entries, then the variations in the table indicate many red flags that need further investigation by the auditor. For example, why is the number of 5s considerably greater than expected at Beta? Probably the interesting question becomes: how is the auditor going to investigate that question? Beta has 40,614 journal entry lines. Is the auditor going to pull every journal entry where there was a journal entry line where the dollar amount started with 5? That would be 5,902 journal entry lines. Instead, the auditor would want to determine ways to efficiently analyze patterns in those lines. For example, do 5s show up more frequently for specific account numbers, for specific combinations of account numbers, for specific employee ID's posting the journal entries, for specific time periods (e.g., near end of quarters, end of years, or just after the start of the next quarter or next year), or for other patterns that the data mining software discovers. Benford's Law builds on certain assumptions about underlying data, including that there are no

Table 5
Expected digit distribution under Benford's Law.

Digit	Probability	Digit	Probability
1	30.1%	6	6.7%
2	17.6%	7	5.8%
3	12.5%	8	5.1%
4	9.7%	9	4.6%
5	7.9%		

Table 6
Observed digit distributions in journal entry database.

Entity	Data	1	2	3	4	5	6	7	8	9
Beta	Count	10,366	6348	5182	4767	5902	2290	2128	1984	1650
$X^2=2911.7$	Act%	26%	16%	13%	12%	15%	6%	5%	5%	4%
$P=0.000$	Diff	−5%	−2%	0%	2%	7%	−1%	−1%	0%	−1%
Chi	Count	48,726	26,660	18,844	14,425	12,252	9656	8118	8008	7111
$X^2=272.0$	Act%	32%	17%	12%	9%	8%	6%	5%	5%	5%
$p=0.000$	Diff	2%	0%	0%	0%	0%	0%	−1%	0%	0%
ChiEta	Count	5752	3410	2208	1694	1453	1203	976	898	958
$X^2=49.9$	Act%	31%	18%	12%	9%	8%	6%	5%	5%	5%
$p=0.000$	Diff	1%	1%	−1%	−1%	0%	0%	−1%	0%	1%
ChiNu	Count	795	460	249	244	225	147	96	116	89
$X^2=44.5$	Act%	33%	19%	10%	10%	9%	6%	4%	5%	4%
$p=0.000$	Diff	3%	1%	−2%	0%	1%	−1%	−2%	0%	−1%
ChiPi	Count	810	1200	921	531	381	320	228	220	260
$X^2=618.3$	Act%	17%	25%	19%	11%	8%	7%	5%	5%	5%
$p=0.000$	Diff	−13%	7%	6%	1%	0%	0%	−1%	−1%	1%
Delta	Count	9410	4772	3592	2723	2755	1952	1559	1594	1509
$X^2=180.5$	Act%	32%	16%	12%	9%	9%	7%	5%	5%	5%
$p=0.000$	Diff	1%	−2%	0%	−1%	1%	0%	−1%	0%	0%
Eta	Count	219	103	119	46	50	46	18	55	32
$X^2=46.0$	Act%	32%	15%	17%	7%	7%	7%	3%	8%	5%
$p=0.000$	Diff	2%	−3%	5%	−3%	−1%	0%	−3%	3%	0%
EtaNu	Count	662	415	253	218	221	124	177	126	122
$X^2=37.0$	Act%	29%	18%	11%	9%	10%	5%	8%	5%	5%
$p=0.000$	Diff	−2%	0%	−2%	0%	2%	−1%	2%	0%	1%
EtaPi	Count	428	258	192	132	125	71	99	53	87
$X^2=24.4$	Act%	30%	18%	13%	9%	9%	5%	7%	4%	6%
$p=0.001$	Diff	0%	0%	1%	−1%	1%	−2%	1%	−1%	1%
Gamma	Count	1300	795	564	425	341	342	205	200	261
$X^2=38.9$	Act%	29%	18%	13%	10%	8%	8%	5%	5%	6%
$p=0.000$	Diff	−1%	0%	0%	0%	0%	1%	−1%	−1%	1%
Kappa	Count	2362	1188	1037	666	513	450	458	319	251
$X^2=81.2$	Act%	33%	16%	14%	9%	7%	6%	6%	4%	3%
$p=0.000$	Diff	3%	−1%	2%	0%	−1%	0%	1%	−1%	−1%
KappaXi	Count	2249	1298	813	627	655	489	359	398	322
$X^2=42.7$	Act%	31%	18%	11%	9%	9%	7%	5%	6%	4%
$p=0.000$	Diff	1%	0%	−1%	−1%	1%	0%	−1%	0%	0%
MuXi	Count	3297	1947	1311	1050	1122	969	624	615	573
$X^2=134.2$	Act%	29%	17%	11%	9%	10%	8%	5%	5%	5%
$p=0.000$	Diff	−1%	−1%	−1%	−1%	2%	2%	0%	0%	0%
Nu	Count	651	889	365	529	247	192	111	97	101
$X^2=523.3$	Act%	20%	28%	11%	17%	8%	6%	3%	3%	3%
$p=0.000$	Diff	−10%	10%	−1%	7%	0%	−1%	−2%	−2%	−1%
Omicron	Count	1602	925	699	446	418	341	298	301	273
$X^2=18.9$	Act%	30%	17%	13%	8%	8%	6%	6%	6%	5%
$p=0.015$	Diff	0%	0%	1%	−1%	0%	0%	0%	1%	1%
Phi	Count	2475	1512	1131	790	653	604	472	367	385
$X^2=21.8$	Act%	30%	18%	13%	9%	8%	7%	6%	4%	5%
$p=0.005$	Diff	−1%	0%	1%	0%	0%	1%	0%	−1%	0%
PhiPsi	Count	11,705	7057	4643	3937	3422	2064	2072	2049	1380
$X^2=275.6$	Act%	31%	18%	12%	10%	9%	5%	5%	5%	4%
$p=0.000$	Diff	0%	1%	0%	1%	1%	−1%	0%	0%	−1%
Pi	Count	450	251	167	116	129	122	65	71	55
$X^2=20.3$	Act%	32%	18%	12%	8%	9%	9%	5%	5%	4%
$p=0.009$	Diff	1%	0%	−1%	−2%	1%	2%	−1%	0%	−1%
PiNu	Count	2041	1322	1047	684	524	332	399	354	295
$X^2=86.2$	Act%	29%	19%	15%	10%	7%	5%	6%	5%	4%
$p=0.000$	Diff	−1%	1%	2%	0%	0%	−2%	0%	0%	0%
Psi	Count	1023	567	377	366	309	171	145	134	166
$X^2-50.9$	Act%	31%	17%	12%	11%	9%	5%	4%	4%	5%
$p-0.000$	Diff	1%	0%	−1%	2%	2%	−1%	−1%	−1%	1%

(continued on next page)

Table 6 (*continued*)

Entity	Data	1	2	3	4	5	6	7	8	9
Rho	Count	1286	856	628	464	390	254	250	244	207
$X^2=28.0$	Act%	28%	19%	14%	10%	9%	6%	5%	5%	5%
$p=0.000$	Diff	−2%	1%	1%	0%	1%	−1%	0%	0%	0%
Sigma	Count	390	174	185	151	127	160	64	55	72
$X^2=84.6$	Act%	28%	13%	13%	11%	9%	12%	5%	4%	5%
$p=0.000$	Diff	−2%	−5%	1%	1%	1%	5%	−1%	−1%	1%
Tau	Count	417	232	192	146	126	88	57	45	74
$X^2=24.4$	Act%	30%	17%	14%	11%	9%	6%	4%	3%	5%
$p=0.001$	Diff	0%	−1%	1%	1%	1%	0%	−2%	−2%	1%
Theta	Count	560	323	181	221	154	86	74	74	66
$X^2=47.7$	Act%	32%	19%	10%	13%	9%	5%	4%	4%	4%
$p=0.000$	Diff	2%	1%	−2%	3%	1%	−2%	−2%	−1%	−1%
Upsilon	Count	1241	915	491	432	344	285	304	232	280
$X^2=72.1$	Act%	27%	20%	11%	10%	8%	6%	7%	5%	6%
$p=0.000$	Diff	−3%	3%	−2%	0%	0%	0%	1%	0%	2%
Xi	Count	8864	5221	3102	3079	2611	1741	2196	1644	1716
$X^2=394.2$	Act%	29%	17%	10%	10%	9%	6%	7%	5%	6%
$p=0.000$	Diff	−1%	0%	−2%	1%	1%	−1%	1%	0%	1%
XiNu	Count	854	499	319	254	248	147	180	161	119
$X^2=20.6$	Act%	31%	18%	11%	9%	9%	5%	6%	6%	4%
$p=0.008$	Diff	1%	0%	−1%	−1%	1%	−1%	1%	1%	0%
XiRo	Count	9118	6254	4333	3199	3111	2130	1800	1427	1182
$X^2=326.0$	Act%	28%	19%	13%	10%	10%	7%	6%	4%	4%
$p=0.000$	Diff	−2%	2%	1%	0%	2%	0%	0%	−1%	−1%
Zeta	Count	17,809	10,841	8815	6193	5267	4233	3692	3071	2717
$X^2=222.5$	Act%	28%	17%	14%	10%	8%	7%	6%	5%	4%
$p=0.000$	Diff	−2%	0%	2%	0%	0%	0%	0%	0%	0%

systematic assignment of the numbers. So, as an alternative explanation, it may be that journal entries violate one or more of those assumptions. This critical question needs further research.

5.2. Last digits

Professional guidance also discusses journal entries that contain "round numbers or a consistent ending number" (CAQ 2008). Journal entries with these characteristics have abnormal distributions of *last* digits. Unlike the first digit, which is expected to have a logarithmic distribution, the last digits would be expected to have a uniform distribution. As a test of uniformity, Table 7 shows the distribution of the fourth digit for each organization for all dollar amounts greater than $999. By this position we expect a uniform distribution of the integers (the same number of 0s, 1s, etc.). Table 7 shows the distribution was definitely not uniform for many of the entities. For example, for *Beta*, some 19% of the fourth digit of journal entry line items ended in zero (10% was expected). Many of the entities had journal line items with fourth digit significantly greater than expected, ranging up to 58% for *XiNu*. Some eight of the 29 entities had one of the fourth digits being three times more than expected. However, there could be situations in organizations that make some numbers appear more often. For example, an appliance company might price plasma TVs at $1,598 etc. In that situation, 8s would be expected to appear more frequently in the population.

The Hartigan and Hartigan (1985) dip test of unimodality provides a test of the modal distribution against the base case of equal distribution. Table 8 shows the dip test for each of the organizations in the study.[9] The table shows the dip test value (3) and probability (4). Columns (5) and (6) show the low and high ends, respectively, of the modal interval for the best-fitting unimodal distribution. Some 21 of the organizations had statistically significant dip test values, where $p<0.01$. Given that these values range from zero to 999,[10] some of

[9] The reported test employs the revised algorithm of Cheng and Hall (1998). The reported statistic eliminates the 5398 line items of less than $1.

[10] Given that these are the last three digits, a value of $10, $100, $1000 and $10,000 would all have the same value of zero. The tests shown in this section were repeated with adjustments for line items greater than $1000. Essentially identical results were observed.

Table 7
Observed fourth digit distributions in journal entries database.

Entity	Digit	0	1	2	3	4	5	6	7	8	9
Beta	Count	1685	741	769	790	778	819	744	679	857	809
$X^2 = 881.8$	Act%	19%	9%	9%	9%	9%	9%	9%	8%	10%	9%
$p = 0.000$	Diff	9%	−1%	−1%	−1%	−1%	−1%	−1%	−2%	0%	−1%
Chi	Count	9923	7728	8234	7861	7539	8366	7700	7862	7327	7659
$X^2 = 606.8$	Act%	12%	10%	10%	10%	9%	10%	10%	10%	9%	10%
$p = 0.000$	Diff	2%	0%	0%	0%	−1%	0%	0%	0%	−1%	0%
ChiEta	Count	1875	1011	1058	944	1005	1110	1033	941	1016	1093
$X^2 = 613.3$	Act%	17%	9%	10%	9%	9%	10%	9%	8%	9%	10%
$p = 0.000$	Diff	7%	−1%	0%	−1%	−1%	0%	−1%	−2%	−1%	0%
ChiNu	Count	400	73	100	98	100	75	121	57	82	88
$X^2 = 756.7$	Act%	34%	6%	8%	8%	8%	6%	10%	5%	7%	7%
$p = 0.000$	Diff	24%	−4%	−2%	−2%	−2%	−4%	0%	−5%	−3%	−3%
ChiPi	Count	16	18	18	7	6	1	8	9	1	3
$X^2 = 44.65$	Act%	18%	21%	21%	8%	7%	1%	9%	10%	1%	3%
$p = 0.000$	Diff	8%	11%	11%	−2%	−3%	−9%	−1%	0%	−9%	−7%
Delta	Count	1500	815	1105	870	1031	863	1081	1005	1017	919
$X^2 = 334.1$	Act%	15%	8%	11%	9%	10%	8%	11%	10%	10%	9%
$p = 0.000$	Diff	5%	−2%	1%	−1%	0%	−2%	1%	0%	0%	−1%
Eta	Count	198	60	37	37	40	57	18	56	21	26
$X^2 = 449.4$	Act%	36%	11%	7%	7%	7%	10%	3%	10%	4%	5%
$p = 0.000$	Diff	26%	1%	−3%	−3%	−3%	0%	−7%	0%	−6%	−5%
EtaNu	Count	870	62	94	75	65	79	114	137	57	64
$X^2 = 3483..9$	Act%	54%	4%	6%	5%	4%	5%	7%	8%	4%	4%
$p = 0.000$	Diff	44%	−6%	−4%	−5%	−6%	−5%	−3%	−2%	−6%	−6%
EtaPi	Count	63	33	73	54	34	43	36	49	70	44
$X^2 = 39.8$	Act%	13%	7%	15%	11%	7%	9%	7%	10%	14%	9%
$p = 0.000$	Diff	3%	−3%	5%	1%	−3%	−1%	−3%	0%	4%	−1%
Gamma	Count	696	124	135	85	114	169	101	154	105	132
$X^2 = 1651.4$	Act%	38%	7%	7%	5%	6%	9%	6%	8%	6%	7%
$p = 0.000$	Diff	28%	−3%	−3%	−5%	−4%	−1%	−4%	−2%	−4%	−3%
Kappa	Count	823	191	295	282	214	398	329	272	259	192
$X^2 = 956.6$	Act%	25%	6%	9%	9%	7%	12%	10%	8%	8%	6%
$p = 0.000$	Diff	15%	−4%	−1%	−1%	−3%	2%	0%	−2%	−2%	−4%
KappaXi	Count	688	120	223	180	174	395	191	289	229	185
$X^2 = 926.7$	Act%	26%	4%	8%	7%	7%	15%	7%	11%	9%	7%
$p = 0.000$	Diff	16%	−6%	−2%	−3%	−3%	5%	−3%	1%	−1%	−3%
MuXi	Count	1566	316	351	352	303	451	326	329	324	288
$X^2 = 2986.2$	Act%	34%	7%	8%	8%	7%	10%	7%	7%	7%	6%
$p = 0.000$	Diff	24%	−3%	−2%	−2%	−3%	0%	−3%	−3%	−3%	−4%
Nu	Count	35	19	27	12	14	24	10	17	13	21
$X^2 = 28.3$	Act%	18%	10%	14%	6%	7%	13%	5%	9%	7%	11%
$p = 0.000$	Diff	8%	0%	4%	−4%	−3%	3%	−5%	−1%	−3%	1%
Omicron	Count	278	159	221	197	190	199	218	178	175	208
$X^2 = 49.3$	Act%	14%	8%	11%	10%	9%	10%	11%	9%	9%	10%
$p = 0.000$	Diff	4%	−2%	1%	0%	−1%	0%	1%	−1%	−1%	0%
Phi	Count	1179	610	583	607	588	696	579	640	552	591
$X^2 = 468.2$	Act%	18%	9%	9%	9%	9%	11%	9%	10%	8%	9%
$p = 0.000$	Diff	8%	−1%	−1%	−1%	−1%	1%	−1%	0%	−2%	−1%
PhiPsi	Count	1996	281	401	369	300	735	341	290	322	225
$X^2 = 4901.3$	Act%	38%	5%	8%	7%	6%	14%	6%	6%	6%	4%
$p = 0.000$	Diff	28%	−5%	−2%	−3%	−4%	4%	−4%	−4%	−4%	−6%
Pi	Count	43	31	31	65	29	28	40	44	27	42
$X^2 = 32.5$	Act%	11%	8%	8%	17%	8%	7%	11%	12%	7%	11%
$p = 0.000$	Diff	1%	−2%	−2%	7%	−2%	−3%	1%	2%	−3%	1%
PiNu	Count	206	110	94	141	130	119	101	110	116	129
$X^2 = 71.1$	Act%	16%	9%	7%	11%	10%	9%	8%	9%	9%	10%
$p = 0.000$	Diff	6%	−1%	−3%	1%	0%	−1%	−2%	−1%	−1%	0%
Psi	Count	195	43	39	36	54	30	39	37	73	38
$X^2 = 377.3$	Act%	33%	7%	7%	6%	9%	5%	7%	6%	13%	7%
$p = 0.000$	Diff	23%	−3%	−3%	−4%	−1%	−5%	−3%	−4%	3%	−3%

(continued on next page)

Table 7 (*continued*)

Entity	Digit	0	1	2	3	4	5	6	7	8	9
Rho	Count	179	63	91	89	57	49	72	65	45	77
$X^2=168.7$	Act%	23%	8%	12%	11%	7%	6%	9%	8%	6%	10%
$p=0.000$	Diff	13%	−2%	2%	1%	−3%	−4%	−1%	−2%	−4%	0%
Sigma	Count	13	16	6	22	11	14	40	15	10	29
$X^2=53.1$	Act%	7%	9%	3%	13%	6%	8%	23%	9%	6%	16%
$p=0.000$	Diff	−3%	−1%	−7%	3%	−4%	−2%	13%	−1%	−4%	6%
Tau	Count	30	8	9	63	26	18	33	17	13	21
$X^2=98.1$	Act%	13%	3%	4%	26%	11%	8%	14%	7%	5%	9%
$p=0.000$	Diff	3%	−7%	−6%	16%	1%	−2%	4%	−3%	−5%	−1%
Theta	Count	52	37	30	25	56	34	67	34	34	23
$X^2=47.3$	Act%	13%	9%	8%	6%	14%	9%	17%	9%	9%	6%
$p=0.000$	Diff	3%	−1%	−2%	−4%	4%	−1%	7%	−1%	−1%	−4%
Upsilon	Count	142	102	86	68	98	73	154	102	99	128
$X^2=68.1$	Act%	13%	10%	8%	6%	9%	7%	15%	10%	9%	12%
$p=0.000$	Diff	3%	0%	−2%	−4%	−1%	−3%	5%	0%	−1%	2%
Xi	Count	1809	1243	1393	1423	1223	1350	1561	1352	1166	1423
$X^2=221.4$	Act%	13%	9%	10%	10%	9%	10%	11%	10%	8%	10%
$p=0.000$	Diff	3%	−1%	0%	0%	−1%	0%	1%	0%	−2%	0%
XiNu	Count	1076	77	81	87	76	101	110	70	84	88
$X^2=4774.4$	Act%	58%	4%	4%	5%	4%	5%	6%	4%	5%	5%
$p=0.000$	Diff	48%	−6%	−6%	−5%	−6%	−5%	−4%	−6%	−5%	−5%
XiRo	Count	3461	2381	2212	2171	2246	2552	2212	1902	2179	2094
$X^2=705.6$	Act%	15%	10%	9%	9%	10%	11%	9%	8%	9%	9%
$p=0.000$	Diff	5%	0%	−1%	−1%	0%	1%	−1%	−2%	−1%	−1%
Zeta	Count	6717	3912	3942	3908	3852	4308	3996	3702	3859	3554
$X^2=1802.9$	Act%	16%	9%	9%	9%	9%	10%	10%	9%	9%	9%
$p=0.000$	Diff	6%	−1%	−1%	−1%	−1%	0%	0%	−1%	−1%	−1%

Table 8
Last three digits — dip test.

(1) Entity	(2) *N*	(3) Dip	(4) *P*	(5) Low	(6) High
Beta	40,193	0.017	0.000	5	5
Chi	152,654	0.003	0.000	15	15
ChiEta	18,501	0.008	0.000	0	0
ChiNu	2413	0.012	0.118	0	0
ChiPi	4717	0.069	0.000	2	2
Delta	29,400	0.006	0.002	1	12
Eta	686	0.032	0.004	0	0
EtaNu	2300	0.019	0.002	0	0
EtaPi	1437	0.024	0.002	0	82
Gamma	4412	0.012	0.010	0	0
Kappa	7231	0.007	0.101	0	0
KappaXi	7205	0.016	0.000	0	51
MuXi	11,497	0.009	0.000	0	0
Nu	3105	0.021	0.000	1	25
Omicron	5077	0.008	0.164	0	0
Phi	8345	0.011	0.001	0	0
PhiPsi	37,639	0.016	0.000	0	15
Pi	1413	0.018	0.041	1	31
PiNu	6923	0.019	0.000	0	25
Psi	3250	0.011	0.091	0	0
Rho	4558	0.007	0.328	8	12
Sigma	1372	0.021	0.011	1	69
Tau	1368	0.017	0.061	3	3
Theta	1701	0.017	0.026	1	2
Upsilon	4456	0.016	0.000	1	10
Xi	30,122	0.008	0.000	2	28
XiNu	2729	0.028	0.000	0	0
XiRo	32,063	0.006	0.000	0	0
Zeta	61,502	0.004	0.000	0	0

these distributions are particularly interesting. For example, Chi has an effectively unimodal distribution of the last three digits, with the value 15. There are, however, only 97 line items of greater than $1000 and with the last three digit value of 15. There are several similar transactions involving the "Severance, Bonuses & Fringe Benefits Account," which may warrant further investigation. Similar results are observed for other organizations of interest, including Gamma, MuXi and Theta.

Auditor investigation of journal entries with particularly high levels of rounded or other unusual patterns cannot rely, however, purely on those patterns as a screen as the number of entries may be too large to investigate. These patterns have to be considered in conjunction with the number of accounts involved. Fig. 2 illustrates these relationships. The extent of abnormal patterns, such as rounded journal entries, is the vertical axis. The number of different accounts involved in these unusual patterns is the horizontal axis. If an entity is in Quadrant A (high-small), there is a high proportion of unusual journal entries with a relatively small number of accounts involved in these journal entries. Quadrant B (high-large) also has a large number of journal entries with abnormal patterns but with a large number of accounts to which these entries post. Quadrant C (low-small) has both low abnormality and a small number of accounts. Quadrant D (low-large) has relatively low levels of abnormal journals but with a large number of accounts.

If an entity is in Quadrant A, there is the significant potential for fraud but only a small number of accounts. We believe the investigation cost is relatively low, as it is likely that there are few patterns of transactions to identify. Conversely, Quadrant B is more difficult to investigate as there are more accounts involved and, we believe, a larger set of patterns in the transactions.

To understand the large differences in the level of abnormal patterns in the journal entries we display in Table 7 above and to see if these journal entries could be seen as following the patterns we show in Fig. 2, we undertook additional analysis. We selected the last three digits (to the left of the decimal place) of each line item in each journal entry. In line with the previous discussion, we expected that these last three digits would be uniformly distributed. We viewed the line items in each journal from three perspectives: 1) all line items, 2) only line items greater than $1000 to eliminate review of minor journal entries and 3) journal entries totaling at least $1000. We counted the total number of line items within journal entries under each of the options. We then identified the number of line items with the five most common sets of last three digits and took this as a proportion of the total line items. We also considered the number of accounts

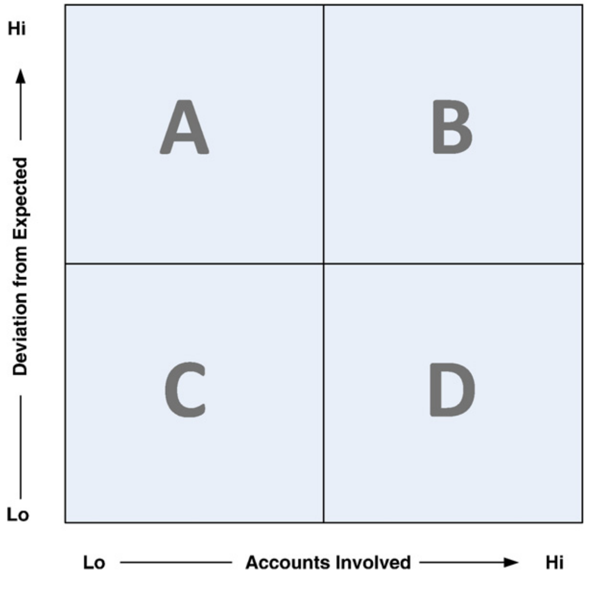

Fig. 2. Patterns of journal entry values vs. number of accounts involved.

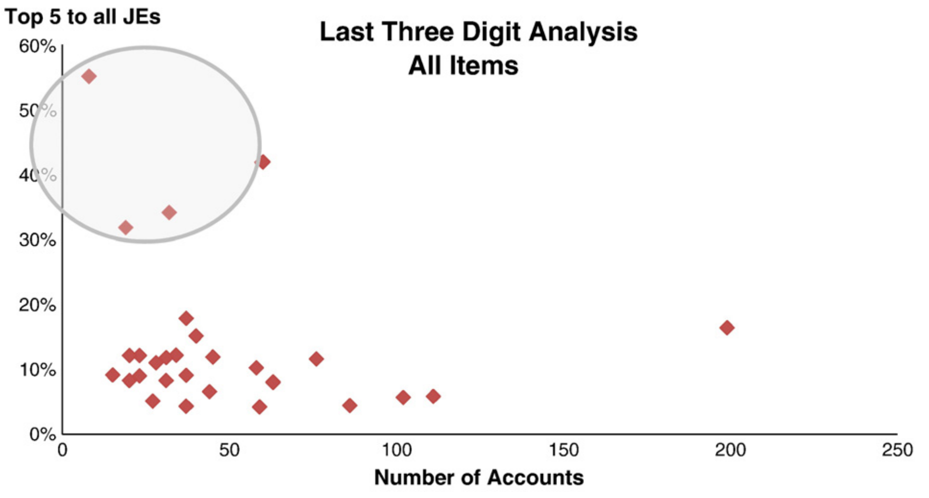

Fig. 3. Proportion of journal entries to number of accounts all items.

posted within the journal entries making up the line items with the five most common sets of last three digits. Figs. 3–5 show the three variants of journal entries as compared with the number of accounts.

There are interesting patterns in these three figures. There are several entities that have very high levels of abnormally frequent "last three digits." There are four of our 29 entities that have 30 to 60% of their transactions made up of just the top five of the last three digit patterns. With uniform distribution, we would expect any five three-digit numbers to represent only 0.5% of transactions. Interestingly, each of these entities employs only a maximum of 40 accounts within these "top five" transactions. Effectively, in our sample, all those entities that had strongly abnormal transaction patterns had only a relatively small number of accounts. These four entities could be placed in Quadrant A (high-small). No entities were in Quadrant B, one was in Quadrant D (low-large) and by far the most, 24, in Quadrant C (low-small). In general, all else being equal, the four firms in Quadrant A probably pose the highest risk of fraud for the auditors. These firms have a very high number of round number or consistent number transactions and

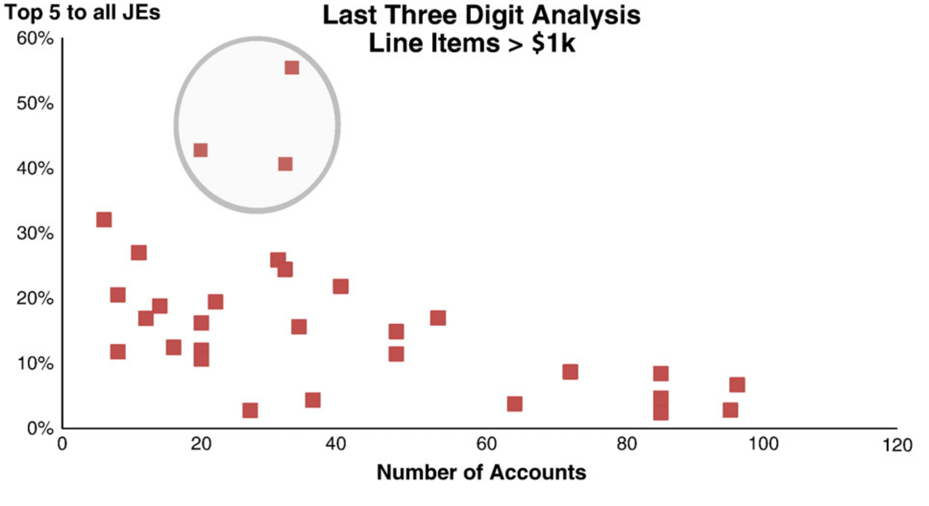

Fig. 4. Proportion of journal entries to number of accounts line items greater than $1000.

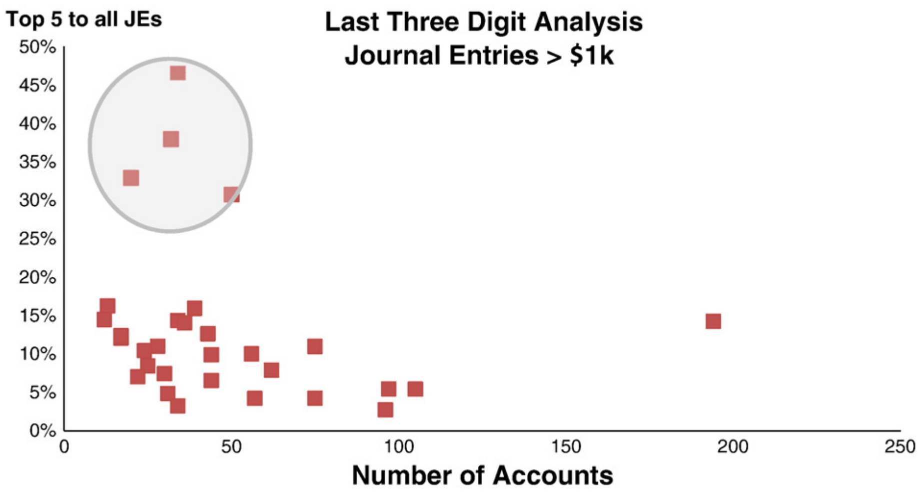

Fig. 5. Proportion of journal entries to number of accounts journal entries greater than $1000.

they are posted to just a few accounts which could indicate that the fraudster is covering up or falsifying a particular class of transactions such posting fictitious sales.

5.3. Unusual temporal patterns

The primary objective of data mining is finding unusual patterns (or outliers that do not fit a pattern) in the data. These unusual patterns would constitute the red flags that the auditor would subsequently investigate. One red flag that an auditor may investigate is unusual patterns in the journal entry activities. End-of-quarter and end-of-year journal activities are usually of particular interest. The auditor's concern is that management will make inappropriate journal entries to improve or manage their performance numbers (e.g., net income) prior to closing their books for their quarterly filings (form 10-Q for public companies) and the more closely followed annual report (form 10-K for public companies), which under goes a formal certified financial audit.

By far the most common form of financial fraud centers on revenue recognition. For example, a company may book a sale in one year that actually occurred in the following year. In the most egregious form of revenue recognition fraud, management books completely fictitious sales. These examples of revenue recognition fraud could result in a variety of journal entries to book those sales over and above the *normal* journal entries, which would therefore increase the overall number of journal entries posted during the period of fraud (e.g., the last month of the year).

As Fig. 6 illustrates, diagnosing *unusual* patterns can be challenging because defining *normal* is in itself a challenge. For example, increases in journal entry activities would be expected in the last month of the fiscal year as a variety of one-time *normal* closing journal entries and accruals would be posted. Graphs on the left side of Fig. 6 show journal entry line item volume for each month and graphs on the right side show the average dollar value of each journal entry line for each month. Visually comparing the organizations in the figure, it would be hard to define what is normal. The particular three organizations in Fig. 6 were selected to illustrate the wide differences in the 29 organizations in our database. For the 29 organizations for which we have 12 complete months of journal entries (including the ones shown in Fig. 6), only two organizations had the highest volume in the last month and only one of the 29 *other* organizations had the highest average dollar values in the last month. Does this mean that there was potential revenue recognition fraud for those one or two organizations and no potential for revenue recognition fraud at the other organizations? Of course not, on both counts. It does illustrate, however, that the auditor cannot visually cherry pick potential problem areas. It will take deeper data mining to isolate unusual patterns and transactions.

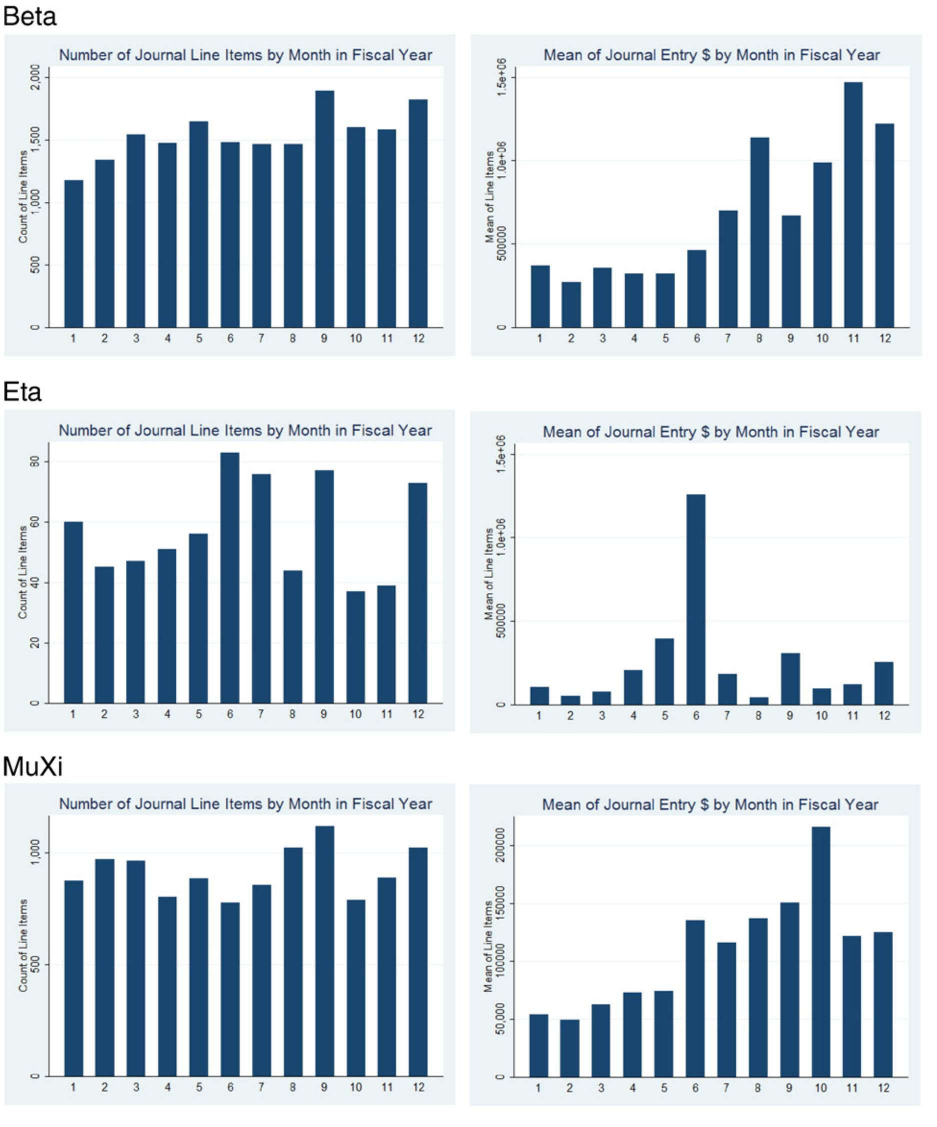

Fig. 6. Monthly distributions of monthly journal entry line volume and dollar volume for a sample of organizations.

6. Conclusion

Fraud detection has become an increasingly important element of the financial statement audit. There is clear evidence of the importance of journal entries in the conduct of financial statement frauds over the last decade, with one of the most egregious being WorldCom. It is hardly surprising, then, that a key element in recent professional developments in increasing the fraud detection requirements in the financial statement audit has been significantly heightened requirements to assess the controls on journal entries and to conduct substantive tests thereon. Unfortunately, research on data mining journal entries from a fraud detection perspective is essentially a null set. In this paper, we canvass a number of perspectives on such data mining.

The nature and form of the population of journal entries posted to the general ledger in computerized accounting information systems is a function of several technological and entity-level characteristics. In a modern ERP system, journal entries will be highly granular — even atomic. In more traditional accounting information systems, general ledger journal entries may be highly aggregated where the general ledger will receive summarized journals from subsidiary systems. These summarized journal entries will capture information with a very different profile than in an ERP system. Journal entries will flow from a variety of other systems and business processes. Journal entries may flow from consolidation systems, from automated or semi-automated general ledger and from manual entries. Data mining approaches must be sufficiently flexible to accommodate these different data structures and flows.

There is a clear and pressing need for research on a variety of interrelated areas in data mining journal entries. Data mining journal entries must bring together five characteristics, viz (a) amount, (b) chart of accounts code to establish impact on the general ledger, (c) source of the journal entry, (d) control characteristics surrounding the individual journal entry and (e) opening and, by extension, closing general ledger balances.

The biggest impediment to doing research in data mining of journal entries is getting access to one or more real-world journal entry data sets. For this project, we had access to 36 different data sets, of which 29 were appropriate for our initial analysis. The seven excluded data sets had less than 12 months of data. There are potentially many more data mining techniques that could be applied to this data set. However, our digital analysis techniques did bring up some interesting preliminary findings, including:

- For all 29 entities we tested, the Chi-square distribution indicates that the first digits of journal dollar amounts differs from that expected by Benford's Law. If, on one hand, we assume that Benford's law should apply to journal entries, these variations means the auditors would have a tremendous number of red flags to investigate. On the other hand, Benford's Law builds on certain assumptions about underlying data, so, further research is needed to explore whether or how journal entries violate one or more of those assumptions.
- Professional guidance recommends identifying journal entries that contain round numbers or a consistent ending number. Unlike first digits, which are expected to have a logarithmic distribution, the last digits would be expected to have a uniform distribution. Our test found that the distribution was definitely not uniform for many of the entities. Eight of the 29 entities had one of the fourth digits being three times more than expected. However, there could be situations in organizations that make some numbers appear more often, which would have to be identified by the auditors.
- Since investigating false positives could be expensive for the auditors, auditors will have to develop and select audit methodologies appropriate to the characteristics of the journal entries. We compared the number of accounts related to the top-five most-frequently occurring three last digit combinations. Of the 29 entities, four entities had a very high occurrences of the top-five three-digit combination that involved only a small set of accounts, one had a low occurrences of the top-five three-digit combination that involved a large set of accounts and 24 had a low occurrences of the top-five three-digit combination that involved a small set of accounts. In general, all else being equal, the first four firms probably pose the highest risk of fraud for the auditors since they had a very high number of rounded number or consistent number transactions and they are posted to just a few accounts which could indicate that the fraudster is covering up or falsifying a particular class of transactions.
- In term of general patterns of transaction volumes, there did not appear to be any. We expected to see increases at quarter end or year, but we did not find consistent examples of this in our 29 entities.

Our initial analysis of the 29 journal entry data sets just begins the potential analysis of these data sets. In the future, we expect to apply many more data mining techniques to discover other patterns and relationships in the data sets. We also want to start seeding the dataset with fraud indicators (e.g., pairs of accounts that would not be expected in a journal entry) and compare the sensitivity of the different data mining techniques to find these seeded indicators.

Acknowledgements

Our thanks for the important support of a software vendor in providing data, Neil Herman for data extraction and coding and the Shidler College of Business at the University of Hawai'i at Mānoa for financial support.

References

Arya A, Fellingham J, Schroeder DA. Accounting information, aggregation, and discriminant analysis. Manage Sci 2000;46(6): 790–806.
ASB. Statement on Auditing Standards no. 99 consideration of fraud in a financial statement audit, 26. New York: Auditing Standards Board, American Institute of CPAs; 2003.
ASB. Statement on Auditing Standards no. 109 understanding the entity and its environment and assessing the risks of material misstatement, 46. New York: Auditing Standards Board, American Institute of CPAs; 2006a.
ASB. Statement on Auditing Standards no. 110 performing audit procedures in response to assessed risks and evaiuating the audit evidence obtained, 26. New York: Auditing Standards Board, American Institute of CPAs; 2006b.
ASB. Exposure draft — proposed statement on auditing standards consideration of fraud in a financial statement audit (redrafted). New York: Auditing Standards Board, American Institute of CPAs; 2009.
Benford F. The law of anomalous numbers. Proc Am Philos Soc 1938;78:551–72.
Beresford, D. R., N. d. Katzenbach, and C. B. Rogers. 2003. Report of investigation by the Special Investigative Committee of the Board of Directors of WorldCom, Inc. Clinton, Miss.: WorldCom, Inc., 345.
CAQ. Practice aid for testing journal entries and other adjustments pursuant to AU section 316. Washington, DC: Center for Audit Quality; 2008.
Cheng M-Y, Hall P. Calibrating the excess mass and dip tests of modality. J R Stat Soc Ser B 1998;60(3):579–89.
Cleary R, Thibodeau JC. Applying digital analysis using Benford's Law to detect fraud: the dangers of type I errors. Auditing 2005;24 (1):77–81.
Curtis MB, Jenkins JG, Bedard JC, Deis DR. Auditors' training and proficiency in information systems: a research synthesis. J Inf Syst 2009;23(1):79–96.
DeVries DD, Kiger JE. Journal entries and adjustments — your biggest fraud danger. J Corporate Accounting Finance 2004;15(4): 57–62.
Diekmann A. Not the first digit! Using Benford's Law to detect fraudulent scientific data. J Appl Stat 2007;34(3):321–9.
Geyer CL, Williamson PP. Detecting fraud in data sets using Benford's Law. Commun Stat Simul Comput 2004;33(1):229–46.
Guan L, He D, Yang D. Auditing, integral approach to quarterly reporting, and cosmetic earnings management. Manage Auditing J 2006;21(6):569–81.
Hales DN, Sridharan V, Radhakrishnan A, Chakravorty SS, Siha SM. Testing the accuracy of employee-reported data: an inexpensive alternative approach to traditional methods. Eur J Oper Res 2008;189(3):583–93.
Hartigan JA, Hartigan PM. The dip test of unimodality. Ann Stat 1985;13(1):70–84.
Hill T. A statistical derivation of the significant-digit law. Stat Sci 1995;10(4):354–63.
IAASB. International standard on auditing 240 — the auditor's responsibility relating to fraud in an audit of financial statements. Handbook of international standards on auditing and quality control. New York: International Auditing and Assurance Standards Board, International Federation of Accountants; 2009a. p. 163–206.
IAASB. International standard on auditing 330 — the auditor's responses to assessed risks. Handbook of international standards on auditing and quality control. New York: International Auditing and Assurance Standards Board, International Federation of Accountants; 2009b. p. 331–54.
Ijiri Y. Theory of accounting measurement. Sarasota, Fla: American Accounting Association; 1975.
Ijiri Y, Kelly E. Multidimensional accounting and distributed databases: their implications for organizations and society. Acc Organ Soc 1980;5(1):115–23.
Janvrin D, Bierstaker J, Lowe DJ. An investigation of factors influencing the use of computer-related audit procedures. J Inf Syst 2009;23(1):97-118.
Janvrin D, Morrison J. Using a structured design approach to reduce risks in end user spreadsheet development. Inf Manage 2000;37 (1):1.
Leech SA. The theory and development of a matrix-based accounting system. Accounting Bus Res 1986;16(64):327–41.
Ley E. On the peculiar distribution of the U.S. stock indexes' digits. Am Stat 1996;50:311–3.
Mattessich R. Accounting and analytical methods. Homewood, Ill: Richard D Irwin; 1964.
Mattessich R. Accounting research and researchers of the nineteenth century and the beginning of the twentieth century: an international survey of authors, ideas and publications. Accounting Bus Financial Hist 2003;13(2):125–70.
Nigrini MJ. Digital analysis using Benford's Law: tests statistics for auditors. Second ed. Vancouver: Global Audit Publications; 2000.
Nigrini MJ, Mittermaier LJ. The use of Benford's Law as an aid in analytical procedures. Auditing A J Pract Theory 1997;16(2): 52–67.
Panko RR. Spreadsheets and Sarbanes–Oxley: regulations, risks, and control frameworks. Commun AIS 2006;2006(17):2-50.
PCAOB. AU section 316—consideration of fraud in a financial statement audit [URL]. Public Company Accounting Oversight Board [cited 20 October 2006]. Available from. http://pcaob.org/standards/interim_standards/auditing_standards/index_au.asp?series=300&section=300. 2002.
PCAOB. Proposed auditing standards related to the auditor's assessment of and response to risk and conforming amendments to PCAOB standards. Public Company Accounting Oversight Board, 21 October [cited 10 June 2009]. Available from. http://www.pcaobus.org/Rules/Docket_026/2008-10-21_Release_No_2008-006.pdf2008.
POB. Report and recommendations of the panel on audit effectiveness. New York: American Institute of Certified Public Accountants, Public Oversight Board; 2000.
PricewaterhouseCoopers. The use of spreadsheets: considerations for section 404 of the Sarbanes–Oxley Act. New York, NY: PricewaterhouseCoopers LLP; 2004. 9.
Sampson WC, Olan MJ. An innovation in auditing: matrix summaries of journal entries. Abacus 1992;28(2):133–41.
Skousen CJ, Guan L, Wetzel TS. Anomalies and unusual patterns in reported earnings: Japanese managers round earnings. J Int Financial Manage Accounting 2004;15(3):212–34.
Tackett JA. Digital analysis: a better way to detect fraud. J Corporate Accounting Finance 2007;18(4):27–36.
Wallace W. Reporting practices: potential lessons from Cendant Corporation. Eur Manage J 2000;18(3):328–33.

Weld LG, Bergevin PM, Magrath L. Anatomy of a financial fraud: a forensic examination of HealthSouth. CPA J 2004:44–9 (October).

Willkie Farr Gallagher, Arthur Andersen LLP. Report to the Audit Committee of the Board of Directors of Cendant Corporation, August 24 [cited 1 September 2007]. Available from. http://www.secinfo.com/dsvrn.71H3.htm. 1998.

Woods L. In confidence [Email], 26 April [cited 10 May 2007]. Available from. http://news.findlaw.com/hdocs/docs/worldcom/worldcom062602cstemail.pdf. 2002.