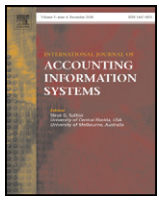
International Journal of Accounting Information Systems 11 (2010) 157–181

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).

1467-0895/$ – see front matter
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.