Association for Information Systems
AIS Electronic Library (AISeL)

ECIS 2013 Completed Research

ECIS 2013 Proceedings

7-1-2013

Auditing Journal Entries Using Extreme Value Theory

Argyris Argyrou

Hanken School of Economics, Helsinki, Finland, argyris.argyrou@hanken.fi

Follow this and additional works at: http://aisel.aisnet.org/ecis2013_cr

AUDITING JOURNAL ENTRIES USING EXTREME VALUE THEORY

Argyrou, Argyris, Hanken School of Economics, Arkadiankatu 22, 00101 Helsinki, Finland, argyris.argyrou@hanken.fi

Abstract

While a wealth of statutory and auditing pronouncements attest to the importance of the auditing of journal entries for preventing and detecting material misstatements to financial statements, existing literature has so far paid inadequate attention to this line of research. To explore this line of research further, this paper proposes a bipartite model that is based on extreme value theory and Bayesian analysis of Poisson distributions. The paper assesses the veracity of the model via a series of experiments on a dataset that contains the journal entries of an international shipping company for fiscal years 2006 and 2007. Empirical results suggest the model can detect journal entries that have a low probability of occurring and a monetary amount large enough to cause financial statements to be materially misstated. Further investigations reveal that the model can assist auditors to form expectations about the journal entries thus detected as well as update their expectations based on new data. The findings indicate that the model can be applied for the auditing of journal entries, and thus supplement existing procedures.

Keywords: auditing, journal entries, extreme value theory, Bayesian analysis

1 Introduction

The area of journal entries is deemed to pose a high risk of material misstatements to financial statements (Public Company Accounting Oversight Board (PCAOB), 2004); an egregious example is that of WorldCom who recorded false journal entries to artificially achieve expected revenue growth (Beresford et al., 2003). The example demonstrates that controls over the processing and recording of journal entries underpin the completeness and timeliness of financial reporting (Canadian Institute of Chartered Accountants, 2004). The importance of journal entries is attested by a wealth of auditing standards that require auditors to test the appropriateness of journal entries recorded in a general ledger (AICPA, 2002) and (IFAC, 2009).

However, absent a study by Debreceny and Gray (2010), existing literature has so far provided tenuous empirical evidence on how the auditing of journal entries can prevent and detect material misstatements to financial statements (Hogan et al., 2008) and (Grabski, 2010). Further, although literature suggests numerous procedures for detecting "anomalous" observations (Chandola et al., 2009), these procedures may not be sufficient for detecting those journal entries that can materially misstate financial statements.

Motivated by these issues, this paper proposes a bipartite model for detecting "suspicious" journal entries in order to assist auditors to identify and asses the risk of material misstatement to financial statements. The paper defines "suspicious" journal entries as having both a large monetary amount and a low probability of occurring; in other words, "suspicious" journal entries are rare and have a monetary amount that is large enough to materially misstate financial statements.

The first component of the model employs the peaks-over-threshold method (i.e. POT), a subset of extreme value theory, to estimate an optimum threshold that can differentiate the distribution of legitimate from that of "suspicious" journal entries. The second component models the number of monthly "suspicious" journal entries in terms of a univariate Poisson distribution and uses Bayesian analysis to draw inferences. The Bayesian analysis allows auditors to update their expectations concerning "suspicious" journal entries given new data as well as transfer their knowledge from an audit engagement to the next.

The paper applies the proposed model to a dataset provided by an international shipping company. The dataset contains the complete set of the company's journal entries for fiscal years 2006 and 2007; it was exported from the the company's database to a text file that consists of 55,350 lines and eight columns representing accounting transactions and variables, respectively. Exploratory data analysis has revealed that the transactions have different distributions between debit and credit balances as well as between one account category and another. For this reason, the paper combines the two variables, "Debit-Credit" and "Account Category", to partition the transactions into twelve experimental cells.

In the following section, the paper sets out the background and reviews procedures for anomaly detection. Section 3 provides the motivation behind the paper's using extreme value theory, describes the data, and introduces the model. Section 4 presents and analyses the results, and Section 5 discusses the main limitations of the paper. Section 6 draws conclusions and suggests possible directions for further research.

2 Background and procedures for anomaly detection

2.1 Background

The Statement on Auditing Standards 99: Consideration of Fraud in a Financial Statement Audit (AICPA, 2002) requires auditors, among other things, to test the appropriateness of journal entries recorded in a

general ledger and other adjustments made in preparing financial statements. The Standing Advisory Group of the PCAOB (2004) considers journal entries to be an area that poses a high risk of material misstatement to financial statements due to fraud. Further, controls over the recording and processing of journal entries underpin the completeness, accuracy, and timeliness of financial reporting (Canadian Institute of Chartered Accountants, 2004).

The fraud perpetrated at WorldCom exemplifies how journal entries can be manipulated to achieve, albeit artificially, expected revenue growth. False and unsupported journal entries were recorded to reduce operating line costs by capitalising these costs and by improperly releasing accruals; these journal entries had an estimated value of about US\$ 7.3 billions (Beresford et al., 2003, p.17 and 56).

Existing literature acknowledges the dearth of empirical evidence on how the reviewing of journal entries can detect and prevent financial statement fraud (Hogan et al., 2008) and (Grabski, 2010). A noteworthy exception is a study by Debreceny and Gray (2010) who used Benford's Law, or digit analysis, to detect fraudulent journal entries. In essence, the study compared the observed distribution of the first digit of USD amounts against that expected by Benford's Law; if the difference was statistically significant under a chi-square test, then the USD amount was deemed to be fraudulent.

Debreceny and Gray (2010) suggested that the observed distributions were significantly different from that expected by Benford's Law for all entities in the sample. Nonetheless, the results may have been an artefact of the chi-square test, as a large number of observations can induce statistically significant results (Grabski, 2010). An additional explanation is that either fraudulent journal entries were the norm in the sample, or Benford's Law is not applicable to journal entries (Grabski, 2010).

2.2 Procedures for anomaly detection

Although procedures for anomaly and novelty detection abound in the literature, they may not be sufficient for detecting "suspicious" journal entries, because they make restrictive assumptions concerning data. These procedures can be classified into three broad categories: (i) two-class classification or supervised, (ii) one-class classification or semi-supervised, and (iii) unsupervised. An extensive and thorough review of these procedures can be found in (Bolton and Hand, 2002) and (Chandola et al., 2009).

A two-class classification procedure assumes that a dataset contains observations labelled either as "legitimate" or as "anomalous". In this case, a model (e.g. neural networks based on supervised learning) is first trained on the dataset, and then used to determine the class (i.e. "legitimate" or "anomalous") to which a previously unseen observation belongs. Two issues arise: first, the prevalence, or prior probability, of "anomalous" observations occurring in the population may be orders of magnitude smaller than that of the "legitimate" observations; and second, it may be difficult to obtain accurate and representative class descriptions, especially for the "anomalous" class.

An one-class classification procedure first develops a reference model that can describe the behaviour of legitimate journal entries. It then estimates a similarity metric (e.g. Euclidean distance) between the reference model and novel journal entries; the similarity metric is monotonically related to the degree of suspiciousness. Finally, it considers a journal entry to be "suspicious", if the journal entry has a similarity metric in excess of an optimum threshold.

However, an one-class classification is prone to a high false-positive rate. The reason is that, unless it can encompass all possible instances of legitimate behaviour, it could classify a large number of legitimate journal entries as being "suspicious". Further, the status of legitimate behaviour is likely to change over time, and hence has to be updated as well. Additional shortcomings stem from the uneven sizes of legitimate and "suspicious" classes as well as the asymmetrical misclassification costs of Type I and Type II errors.

An unsupervised approach first estimates the probability density of data and then selects a threshold in such a way that the probability of a journal entry exceeding the threshold is very small (e.g. $P(X > u) = 10^{-4}$). A journal entry having such a small probability of occurring is deemed to be "suspicious".

However, this approach has three main limitations. First, it implicitly assumes that "suspicious" journal entries, observations occurring beyond the threshold, follow a uniform distribution; this assumption may be restrictive or invalid in practice (Lee and Roberts, 2008). Second, it has a valid probabilistic interpretation only for classification tasks whereby a single observation is being compared against a model describing legitimate behaviour (Clifton et al., 2010). Finally, it does not provide any guidance on how a threshold should be selected; instead, the threshold is selected in a heuristic manner based on past experience and knowledge.

The literature review has revealed a lack of knowledge on how the auditing of journal entries can prevent and detect material misstatements to financial statements. Further, it has indicated that procedures for anomaly detection may not suffice to detect "suspicious" journal entries. Motivated by these issues, the paper proposes an alternative model for detecting "suspicious" journal entries that is based on extreme value theory and Bayesian analysis of Poisson distributions.

A review of extreme value theory lies beyond the scope and confines of the paper; a thorough and comprehensive treatment of this subject can be found in Embrechts P. (1997), (Coles, 2001), and (Reiss and Thomas, 2007). At this juncture, it suffices to note that extreme value theory has been applied extensively in the discipline of Finance. For example, it has been applied to estimate Value-at-Risk (Longin, 2000) and expected shortfall (McNeil and Frey, 2000), to investigate contagion risk in the international banking sector (Ong et al., 2007), and to examine risk-based allocation of assets (Bensalah, 2002).

3 Research Design and Methodology

3.1 Motivation

The paper follows (Rohrbach, 1993) to conjecture that "suspicious" journal entries exhibit two distinguishing characteristics: they are rare, which means they have a low probability of occurring; and, they have a monetary amount that is sufficiently large to cause financial statements to be materially misstated. The corollary of this conjecture is that, if monetary amounts follow a unimodal distribution, then the amounts that are maxima in magnitude are also minima in probability values, and vice versa; in this case, these amounts would concentrate in the tail, or extreme-quantiles, of the unimodal distribution.

This insight motivates the paper to employ extreme value theory in order to model "suspicious" journal entries, as it is the appropriate statistical framework for studying observations that pertain to the tails, or extreme-quantiles, of a distribution.

3.2 Data description

The dataset has been provided by an international shipping company and consists of their journal entries for fiscal years 2006 and 2007. The dataset was exported from the database of the company to a text file that contains 55,350 lines and eight columns representing accounting transactions and variables, respectively; the variables are described in Table 1. For example, "Account Class" takes thirty values, such as: "Interest Received", "Office Expenses", "Trade Debtors" etc. In the present case, "Account Category" takes eight values: (i) "Non-Current Assets", (ii) "Cash and Cash Equivalents", (iii) "Trade and Other Receivables", (iv) "Income", (v) "Expenses", (vi) "Current Liabilities", (vii) "Non-Current Liabilities", and (viii) "Equity".

Name	Type	Values
Account Number	Alphanumeric	360
Account Description	Text	360
Posting Date	Date	24
Debit-Credit	Binary	2
USD Amount	Numerical	
Transaction Details	Text	
Account Class	Categorical-hierarchical	30
Account Category	Categorical-hierarchical	8

Table 1: *Description of variables*

The variables "Account Number", "Account Class", and "Account Category" group transactions in an ascending order of aggregation. For example, the account category "Trade and Other Receivables" consists of 23, 886 transactions; this number represents the aggregation of six account classes: "Sales Taxes Receivable" (1,410), "Trade Debtors" (6,217), "Other Debtors" (5,634), "Loans Receivable" (164), "Insurance Receivables" (9,361), and "Other Receivables" (1,100). The number of transactions is shown in parentheses.

The paper groups the transactions according to the "Account Category" variable, because there are not enough transactions at lowers levels of aggregation (i.e. "Account Number", "Account Class") to estimate the parameters of the proposed model. For the same reason, the paper excludes "Non-Current Assets" and "Equity" completely.

In addition, the paper excludes those transactions auditors would select as a standard procedure in the normal course of an audit; for example, transactions that record transfers to reserves, year-end consolidation, and closing Profit and Loss items to the Balance Sheet. Although the paper has not investigated the counter-factual, including these transactions would cause the model to estimate a higher threshold than otherwise; the reason is this type of transactions tends to have large monetary amounts and occur infrequently, often at the end of a fiscal year. As a result, the model would select transactions, which are selected anyway, but ignore transactions that may warrant further investigation.

Exploratory data analysis suggests the distributions of transactions are different between one account category and another as well as between debit and credit balances. As a result, the paper combines the two variables, "Debit-Credit" and "Account Category", to partition the transactions into twelve experimental cells, as shown in Table 2.

| Account Category	N	Mean	Mode	Median	Var
		USD:Credit			
Cash and Cash Equivalents	432	-2,449	-26	-215	13,424,043
Trade and Other Receivables	12,322	-2,879	-1,052	-1,053	14,819,048
Income	5,567	-1,081	0	-422	2,348,775
Expenses	860	-757	-11	-187	2,304,802
Current Liabilities	11,087	-1,270	-203	-272	5,604,960
Non-Current Liabilities	143	-1,228	-2,677	-624	2,974,177
	30,411				
		USD:Debit			
Cash and Cash Equivalents	2,655	6,154	1,274	4,168	29,997,396
Trade and Other Receivables	11,564	2,985	1,052	1,188	16,255,837
Income	1,663	1,225	0	441	4,944,962
Expenses	4,162	614	10	72	2,806,104
Current Liabilities	4,763	968	203	293	3,751,541
Non-Current Liabilities	89	2,000	109	735	9,538,657
	24,896				

Table 2: *Descriptive statistics for fiscal years 2006 and 2007*

3.3 Modelling "suspicious" transactions using the peaks-over-threshold method

In order to introduce the peaks-over-threshold method, the paper uses a concrete example that is based on the results and depicted in Fig. 1.

To elaborate, let variable $X = (x_1, \ldots, x_n)$ denote the monetary amounts of the transactions belonging to the "Debit" side of "Trade and Other Receivables", where $n = 11,564$ representing the number of transactions. Variable X can be assumed to be an independent and identically distributed (i.e. iid) random variable that follows an unknown distribution. The distribution is unimodal, and hence its probability density function decreases monotonically with increasing distance from the single mode, as shown in Fig. 1a.

Consequently, the further from the mode a monetary amount is, the larger its magnitude and the lower its probability of occurring would be. In other words, transactions that are extreme in monetary amounts are also minima in probability density; the converse is also true. This insight motivates the paper to estimate an optimum threshold, u, that can differentiate the distribution of legitimate from that of "suspicious" transactions. Transactions whose monetary amounts exceed the optimum threshold are considered to be "suspicious", because these amounts are both rare and large.
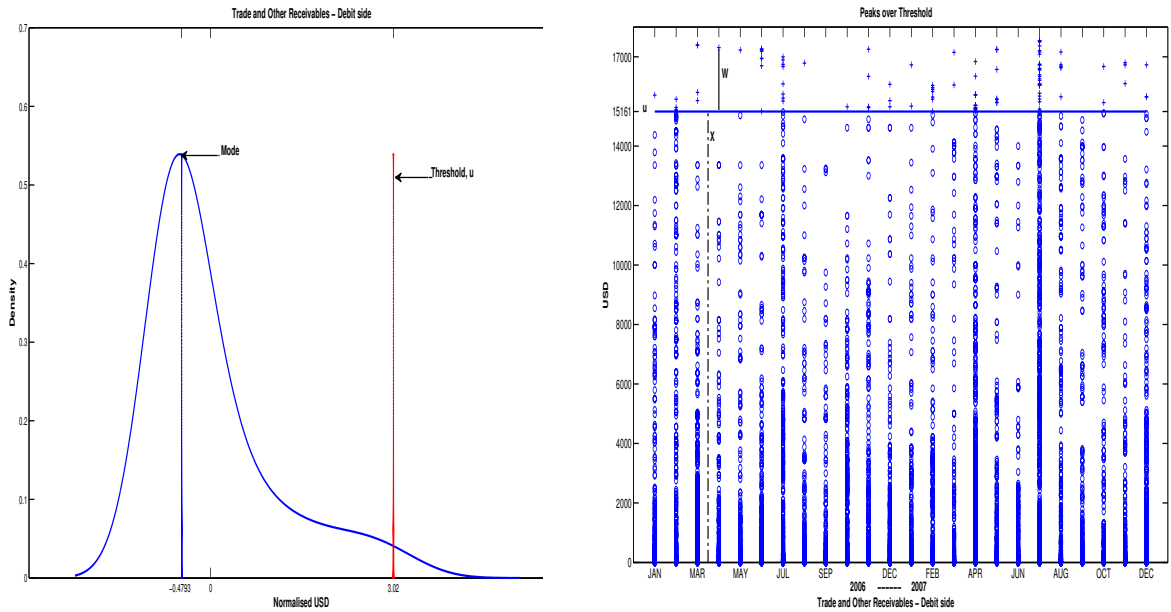
(a) *Probability density function showing the single mode and threshold, u=3.02*

(b) *Peaks over threshold*
The threshold, u=15,161 USD, corresponds to u=3.02

Figure 1: *Trade and Other Receivables - Debit side*

For a sufficiently high threshold, u, Pickand's theorem (Pickands, 1975) describes the distribution of excesses over threshold u conditional on the threshold being exceeded, i.e. $(X - u \mid X > u)$, in terms of a distribution within the Generalized Pareto (GP) family, as follows (Coles, 2001, p.75):

$$F(x - u; \tilde{\sigma}, \xi) = \begin{cases} 1 - \left[1 + \xi \left(\frac{x-u}{\tilde{\sigma}} \right) \right]_{+}^{\frac{-1}{\xi}} & \text{if } \xi \neq 0, \\ 1 - exp\left[-\frac{x-u}{\tilde{\sigma}} \right] & \text{if } \xi = 0. \end{cases} \tag{1}$$

6

The distribution, Eq. 1, has the following probability density function:

$$f(x-u;\tilde{\sigma},\xi) = \begin{cases} \left(\frac{1}{\tilde{\sigma}}\right)\left[1+\xi\left(\frac{x-u}{\tilde{\sigma}}\right)\right]_+^{\frac{-1-\xi}{\xi}} & \text{if } \xi \neq 0, \\ \left(\frac{1}{\tilde{\sigma}}\right)exp\left[-\frac{x-u}{\tilde{\sigma}}\right] & \text{if } \xi = 0. \end{cases} \tag{2}$$

Where $x - u > 0$, $\tilde{\sigma} > 0$, and $\tilde{\sigma}, \xi$ stand for the scale and the shape of a GP distribution, respectively. The shape parameter, ξ, describes the tail of a GP distribution; if $\xi < 0$ then the distribution of excesses, $x - u$, has an upper bound at $u - \frac{\tilde{\sigma}}{\xi}$, which means that the probability density for $x > u - \frac{\tilde{\sigma}}{\xi}$ is zero; if $\xi > 0$ then the distribution decreases polynomially and has a lower bound at $u - \frac{\tilde{\sigma}}{\xi}$; and if $\xi = 0$ then the distribution decreases exponentially and has no lower nor upper bound.

Selecting a threshold that is sufficiently high so that Pickand's theorem could apply is fraught with difficulties (Embrechts P., 1997, p. 355). Suffice it to note that a threshold strikes a trade-off between bias and variance in a model. A too low threshold could lead to sampling from the main body of a distribution, non-extremal values, and thus induce bias in estimating the parameters of a model. On the other hand, as a threshold increases the number of excesses with which parameters can be estimated decreases, and hence the standard errors of the parameters would increase.

In order to estimate an optimum threshold, the paper follows a three-step approach for each of the twelve experimental cells, described in Table 2. First, the paper initialises a set of candidate thresholds that take values between the 95% and 99% percentiles of the amounts. Second, at each candidate threshold, the paper fits a GP distribution, Eq. 1, and estimates the parameters, $\theta = (\tilde{\sigma}, \xi)$, by maximising the log-likelihood function (Coles, 2001, p.80):

$$LL(\tilde{\sigma},\xi;x-u) = \begin{cases} -Nlog(\tilde{\sigma}) - \left(\frac{1+\xi}{\xi}\right)\sum_{i=1}^{N} log\left[1+\xi\left(\frac{x_i-u}{\tilde{\sigma}}\right)\right]_+ & \text{if } \xi \neq 0, \\ -Nlog(\tilde{\sigma}) - \frac{1}{\tilde{\sigma}}\sum_{i=1}^{N}(x_i-u) & \text{if } \xi = 0. \end{cases} \tag{3}$$

Where N denotes the number of excesses over threshold u. Finally, the paper selects the threshold that corresponds to the GP distribution having the maximum log-likelihood.

This threshold becomes the decision boundary that distinguishes legitimate from "suspicious" transactions. For example, Fig.1b depicts the threshold, $u = 15,161$ USD, that separates legitimate from "suspicious" transactions; in this example, there are $11,348$ legitimate and 216 "suspicious" transactions.

3.4 Bayesian analysis of Poisson distributions

The paper models the number of monthly "suspicious" transactions in terms of a univariate Poisson distribution and draws inference via the Bayes' rule. Let the observed number of monthly "suspicious" transactions be denoted by the discrete variable V that follows a Poisson distribution having a probability mass function $\Pr(V = v) = \frac{\lambda^v e^{-\lambda}}{v!}$, where $\lambda > 0$ and $v \geqslant 0$.

Let λ be the unobserved and unknown average number of "suspicious" transactions over the 24-month period under investigation; and, let the prior distribution of λ, $p(\lambda)$, denote the degree of certainty, or inductive bias, about λ in the absence of any observed evidence.

The probability of a "suspicious" transaction occurring is assumed to depend on λ. This dependence can be formalised as $p(V|\lambda)$, which is the conditional probability of the observed number of "suspicious" transactions for each possible value of λ; it is also termed the likelihood function of λ, $L(\lambda)$. The Bayes' rule combines prior probability and likelihood function to estimate the conditional probability, $p(\lambda|V)$, for different values of λ taking into account observed evidence, V.

Formally, the Bayes' rule states:

$$p(\lambda|V) = \frac{p(V|\lambda)\,p(\lambda)}{\sum_\lambda p(V|\lambda)\,p(\lambda)}, \qquad \text{Posterior} = \frac{\text{Likelihood x Prior}}{\text{Marginalising factor}} \quad . \tag{4}$$

The likelihood function of λ is given by: $p(V|\lambda) = \prod_{i=1}^{m} \frac{\lambda^{v_i} e^{-\lambda}}{v_i!}$, where v_i represents the number of "suspicious" transactions for the ith month and $m = 24$ reflects the number of months.

The paper chooses the Gamma distribution as a conjugate prior, $\lambda \sim Gamma(\alpha, \beta)$, and sets the hyperparameters equal to one, $\alpha = \beta = 1$, for the prior to be non-informative. Because of the conjugacy property, the posterior distribution of the Poisson parameter follows a Gamma distribution as well: $\lambda|V \sim Gamma(\sum_{i=1}^{m} v_i + \alpha, \beta + m)$.

The posterior distribution has a closed-form solution, as follows:

$$p(\lambda|V) = \frac{\tilde{\beta}^{\tilde{\alpha}} \lambda^{\tilde{\alpha}-1} \exp(-\tilde{\beta}\lambda)}{\Gamma(\tilde{\alpha})} \quad , \tag{5}$$

where $\tilde{\alpha} = \sum_{i=1}^{m} v_i + \alpha$, $\tilde{\beta} = \beta + m$, and $\Gamma(\tilde{\alpha}) = (\tilde{\alpha} - 1)!$ is the Gamma function of $\tilde{\alpha}$.

4 Results presentation and discussion

4.1 Modelling "suspicious" transactions

The results, summarised in Table 3, identify the parameters of the best-fitted GP model, the thresholds, u, and the number of "suspicious" transactions, $N(u)$. For example, one hundred GP models at varying thresholds are fitted to the transactions that belong to the "Debit" side of "Trade and Other Receivables". The best-fitted GP model occurs when the threshold is USD 15,161, and hence 216 transactions are considered to be "suspicious", as depicted in Fig. 1b.

Account Category	N	$N(u)$	%	u	Gener. Pareto Distribution		
					Shape	Scale	N.Log.Likel.
USD:Credit							
Cash and Cash Equivalents	432	5	1.16%	-13,551	-1.421	0.300	-17.549
Trade and Other Receivables	12,322	554	4.5%	-12,695	-0.562	0.325	-379.392
Income	5,567	11	0.19%	-12,821	1.635	0.003	-34.214
Expenses	860	2	0.23%	-11,627	-4.399	6.151	-25.513
Current Liabilities	11,087	23	0.21%	-13,629	1.524	0.021	-31.134
Non-Current Liabilities	143	27	18.88%	-2,677	4.560	0.000	-100.799
	30,411	622	2.04%				
USD:Debit							
Cash and Cash Equivalents	2,655	225	8.47%	15,137	-0.98	0.43	-184.75
Trade and Other Receivables	11,564	216	1.86%	15,161	-0.51	0.34	-129.32
Income	1,663	23	1.38%	12,811	1.47	0.01	-51.24
Expenses	4,162	3	0.07%	15,657	-2.21	1.21	-20.42
Current Liabilities	4,763	8	0.17%	16,366	-1.75	0.64	-22.32
Non-Current Liabilities	89	2	2.25%	11,482	-3.65	4.44	-24.37
	24,896	477	1.91%				

Table 3: *Generalized Pareto models*

The results suggest the proposed model can perform more efficiently than simply selecting the largest $x\%$ (e.g. 5%) of the amounts. For example, the thresholds estimated for the debit and credit sides of "Trade and Other Receivables" select about 1.86% and 4.5% of the corresponding transactions, respectively. The reason for this efficiency is the model can estimate a threshold that is a function of three variables:

(i) "USD Amount", (ii) "Debit-Credit", and (iii) "Account Category". On the other hand, a uniform threshold is a function of only "USD Amount".

The proposed model can estimate a threshold that, at least in principle, can be interpreted in the context of extreme value theory, whereas a heuristic threshold lacks any interpretation. In this respect, the model can mitigate the subjectivity and bias that may occur when auditors select a threshold only on the basis of their past experience and knowledge (Tversky and Kahneman, 1974) and (Trotman et al., 2011).
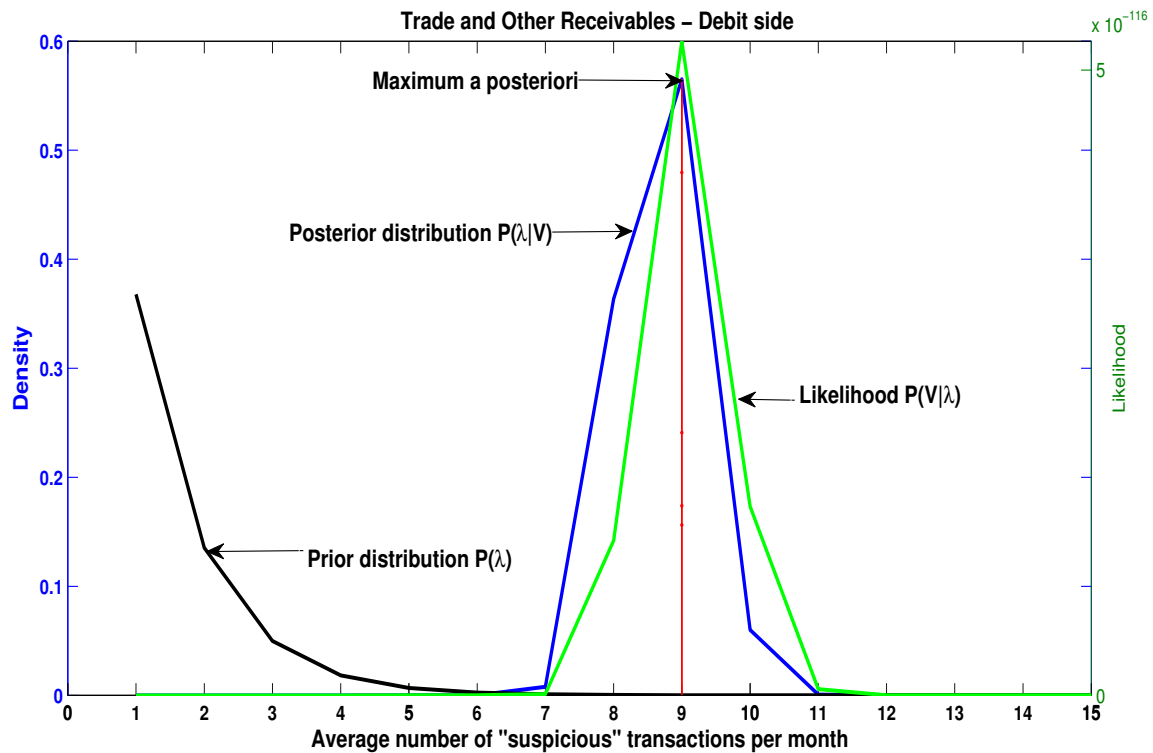
4.2 Bayesian analysis

Figure 2: Bayesian analysis for Trade and Other Receivables - Debit side

Figure 2 depicts the Bayesian analysis of the monthly "suspicious" transactions that belong to the "Debit" side of "Trade and Other Receivables". The maximum a posteriori (MAP), the mode of the posterior distribution, denotes the number of monthly "suspicious" transactions, $\lambda = 9$, that has the highest probability of occurring. Further, the MAP has a 95% credible interval of $8 - 10$, which means that, given the data, there is a 95% probability that the number of monthly "suspicious" transactions is either 8, or 9, or 10. In this case, the MAP is the same as the maximum likelihood estimate (MLE), which is the average calculated form the data (i.e. 216 "suspicious" transactions / 24 months). The reason is the posterior distribution is estimated only on the basis of the data, as the prior probability has been chosen to be non-informative.

Auditors can apply the Bayesian analysis in two ways. First, they can select those monthly transactions that have the $8 - 10$ largest monetary amounts, provided the monetary amounts follow a unimodal distribution. Second, they can update their evidence sequentially, as the current posterior distribution will become the prior distribution in next year's audit.

5   Caveats and limitations

In the context of this paper, the monetary amounts follow a unimodal distribution, and hence those transactions that are extreme in monetary amounts are also minima in probability density values. However, if the monetary amounts follow a multimodal distribution, then the correspondence between large amounts and low probability does not hold, because there is not a single mode from which distances can be defined. This possibility limits the applicability and veracity of the results. Nonetheless, this limitation is not irredeemable, as a future study could extend the paper to multimodal distributions by employing an appropriate methodology; for example, Clifton et al. (2010) describe such methodology.

The paper proposes a model that does not include two variables that are important in an audit engagement: time, and audit risk. In particular, the model is static in that it derives a time-invariant threshold, as shown in Fig. 1b. However, it may not be realistic to use a constant threshold throughout a fiscal year, because there is much more scope for manipulating journal entries at year-end than there is during the rest of the year. Ongoing research examines how a temporal covariate can incorporate information about cycles and trends that may be present in the distribution of journal entries.

Further, the model does not take into account the concept of audit risk, as it implicitly assumes that all audit engagements pose the same level of audit risk. An extension to the model could include auditors' assessment of audit risk and its components (i.e. inherent, detection, control). For example, an audit engagement having a high inherent risk would be assigned a much lower threshold than otherwise, other things being equal.

6   Conclusions and directions for future research

Existing literature, absent a study by Debreceny and Gray (2010), has so far provided little empirical evidence on how the auditing of journal entries can prevent and detect material misstatements to financial statements. The lack of evidence becomes more pronounced, given that a wealth of auditing standards require auditors to consider the complete set of journal entries in planning and performing audits. The auditing of journal entries becomes problematic considering that established procedures for anomaly detection may not suffice for detecting those journal entries that may cause financial statements to be materially misstated. Motivated by these issues, the paper proposes a bipartite model in order to assist auditors to detect such journal entries.

The results suggest the model can detect journal entries that are both rare and have a monetary amount large enough to materially misstate financial statements. Further, the Bayesian analysis indicate how auditors can form as well as update expectations about "suspicious" journal entries.

The paper has raised some questions that may support additional research. Ongoing research aims at incorporating a temporal covariate for the model to estimate a threshold that can capture potential cycles and trends existing in the distribution of journal entries; a further extension to the model could include a covariate for auditors' assessment of audit risk. An additional study could compare the performance of the model against that of unaided auditors who rely only on their past experience and knowledge.

References

AICPA, 2002. Statement On Auditing Standards 99 (SAS 99): Consideration of Fraud in a Financial Statement Audit. American Institute of Certified Public Accountant.

Bensalah, Y., 2002. Asset allocation using extreme value theory. Working Papers 2002-2, Bank of Canada. Available at: http://www.bankofcanada.ca/wp-content/uploads/2010/02/wp02-2.pdf.

Beresford, D. R., Katzenbach, N. d., Rogers, C. B. J., 2003. Report of Investigation by the Special Investigative Committee of the Board of Directors of WorldCom, INC. Available at: http://news.findlaw.com/wsj/docs/worldcom/bdspcomm60903rpt.pdf.

Bolton, R. J., Hand, D. J., Aug. 2002. Statistical fraud detection: A review. Statistical Science 17 (3), 235–249.

Canadian Institute of Chartered Accountants, 2004. IT Control Assessments in the context of CEO/CFO Certification. Toronto,Canada.

Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection: A survey. ACM Computing Surveys 41 (3), 1–58.

Clifton, D., Hugueny, S., Tarassenko, L., 2010. Novelty detection with multivariate extreme value statistics. Journal of Signal Processing Systems, 1–19.

Coles, S., 2001. An introduction to statistical modeling of extreme values. Springer Series in Statistics. Springer-Verlag, London, UK.

Debreceny, R. S., Gray, G. L., Sep. 2010. Data mining journal entries for fraud detection: An exploratory study. International Journal of Accounting Information Systems 11 (3), 157–181.

Embrechts P., Klüppelberg C., M. T., 1997. Modelling Extremal Events: for Insurance and Finance. Springer-Verlag.

Grabski, S., Sep. 2010. Discussion of 'Data mining journal entries for fraud detection: An exploratory study'. International Journal of Accounting Information Systems 11 (3), 182–185.

Hogan, C. E., Rezaee, Z., Riley, R. A., Velury, U. K., 2008. Financial statement fraud: Insights from the academic literature. Auditing: A journal of Practice and Theory 27 (2), 231.

IFAC, 2009. International Standards on Auditing 240 (ISA 240). The Auditor's Responsibilities Relating to Fraud in an Audit of Financial Statements. International Federation of Accountants.

Lee, H., Roberts, S., 2008. On-line novelty detection using the kalman filter and extreme value theory. In: Pattern Recognition 2008. Tampa, FL,USA, pp. 1–4.

Longin, F. M., Jul. 2000. From value at risk to stress testing: The extreme value approach. Journal of Banking & Finance 24 (7), 1097–1130.

McNeil, A. J., Frey, R., Nov. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance 7 (3–4), 271–300.

Ong, L. L., Mitra, S., Chan-Lau, J. A., 2007. Contagion Risk in the International Banking System and Implications for London as a Global Financial Center. IMF Working Papers WP-07-74, International Monetary Fund. Available at: http://www.imf.org/external/pubs/ft/wp/2007/wp0774.pdf.

Pickands, J., 1975. Statistical inference using extreme order statistics. The Annals of Statistics 3 (1), 119–131.

Public Company Accounting Oversight Board (PCAOB), 2004. Standing Advisory Group Meeting: Financial Fraud.
Available at : http://pcaobus.org/News/Events/Documents/09082004_SAGMeeting/Fraud.pdf.

Reiss, R., Thomas, M., 2007. Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields, 3rd Edition. Birkhäuser, Basel, Switzerland.

Rohrbach, K. J., 1993. Variance Augmentation to Achieve Nominal Coverage Probability in Sampling from Audit Populations. Auditing: A Journal of Practice and Theory 12 (2), 79–97.

Trotman, K. T., Tan, H. C., Ang, N., 2011. Fifty-year overview of judgment and decision-making research in accounting. Accounting & Finance 51 (1), 278–360.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. Science 185 (4157), 1124–1131.