

IBM Research Report

Audits and Business Controls Related to Receipt Rules: Benford's Law and Beyond

Vijay Iyengar

IBM Research Division

Thomas J. Watson Research Center

P.O. Box 704

Yorktown Heights, NY 10598



Research Division

Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

Audits and Business Controls Related to Receipt Rules: Benford's Law and Beyond

Vijay Iyengar
IBM Thomas J. Watson Research Center
19 Skyline Drive, Hawthorne, NY 10532
vsi@us.ibm.com

ABSTRACT

Computer Assisted Auditing Techniques (CAATs) are increasingly used in domains with vast amounts of data to extract information of real value in the audit process. Techniques based on applications of Benford's Law have been utilized in various domains. This paper goes beyond by introducing a class of focused tests (called Behavioral Shift Models) that can be tailored to the scenarios being investigated. The discussion is grounded by using scenarios related to policies around receipts. The new focused tests are compared with an adaptation of Benford's Law using carefully engineered synthetic data and real data from the business travel and entertainment expense management (T&E) domain. Results indicate that focused tests are robust with good detection power and significant reduction in false alarms when compared to either intuitive approaches or general tests based on Benford's Law. Application and validation with production data in the T&E domain suggests that these focused tests can leverage the deep knowledge of domain experts and play a valuable role in audit and business control processes.

Keywords: fraud detection; business rule violations; data analysis; false positives and negatives; Benford's Law, Computer Assisted Auditing Techniques (CAATs).

INTRODUCTION

Rules specifying the receipt requirements for various business expenses are an important part of an organization's expense management and control. These rules are formulated to conform to regulatory requirements and to implement the organization's business controls processes. For example, in the travel and business entertainment expense management domain (T&E) receipts might be required for expenses in a particular category (e.g., ground transportation) that exceed a specified amount threshold (e.g., US \$25). In most organizations, the threshold on the expense amount for a particular expense category is not to be considered as an entitlement and there are explicit business rules specifying that only actual expense amounts can be claimed. However, auditors do uncover suspicious "under-the-radar" behavior where a disproportionate number of expenses are claimed just under the corresponding amount thresholds above which receipts would be required. Beyond this behavior auditors and business controls personnel also monitor violations of business rules related to receipts (i.e., missing receipt exceptions). This sub-domain of scenarios related to receipts was chosen because it has important and interesting characteristics analyzed further in this paper. Also, the results can be extended to other sub-domains with similar characteristics (e.g., procurement limits based on item categories and the procurement process utilized).

Computer Assisted Auditing Techniques (CAATs) are increasingly being used as part of the audit process. The goal for such tools is to identify and prioritize cases for audit investigations and for taking business control actions. For example, digital analysis using Benford's Law identifies cases based on the non-conformance to the expected distribution by the target variable being analyzed (Nigrini 1996; Cleary and Thibodeau 2005). The business value of any technique depends on its Type I (false positives) and Type II (false negative) error rates. Earlier work has pointed out the need to apply Benford's Law considering all the digit possibilities in a single test to lower the Type I error rate (Cleary and Thibodeau 2005).

The task addressed in this paper is to analyze the expenses submitted by entities (e.g., employees in an organization) and to identify entities that exhibit suspicious or non-conforming behavior with respect to receipt rules. For the "under-the-radar" scenario, out-of-pocket expense claims that are below the receipt threshold are of particular interest. For this scenario, two methods will be compared to detect patterns in the expense claims data that are indicative of plausible alterations in the actual expenses incurred. The first method is an adaptation of the well known digit analysis based on Benford's Law (Nigrini 1996; Cleary and Thibodeau 2005). The second method (called Behavioral Shift Model: BSM) is a new application of hypothesis testing with the Likelihood Ratio Test (LRT) using an underlying distribution well suited to this domain (Iyengar, Boier, Kelley and Curatolo 2007). The comparison will be done first using synthetic data specifically engineered to include "under-the-radar" behaviors with respect to receipt limits. The comparison using synthetic data will allow quantification of the performance of both methods in terms of their false positive and false negative rates. The two methods will then be compared on real data from the T&E domain and conclusions will be drawn. Next, the problem of identifying excessive violations of business rules is considered. Analysis of exceptions triggered by missing receipts is a scenario of interest in the sub-domain related to receipt rules. A formulation of BSM tailored for analyzing event data is applied to this problem. Results of application to real data from the business travel domain are presented and discussed.

SYNTHETIC DATA

The synthetic data will be engineered in two steps. In the first step, data (referred to in this paper as Benford data (B) representing "non-fraudulent" behavior is created in a manner so that it conforms to Benford's Law. This data set B contains a list of expense data items, each item specifying an expense amount and the entity that incurred it. In the second step, a subset of entities is chosen randomly in each experiment for injecting fraudulent "under-the-radar" behavior for a specific receipt threshold. This engineered data allows quantitative assessment and comparison of detection capabilities of the two methods using the knowledge of which entities have the injected "under-the-radar" behavior.

There are many distributions that have been shown to obey Benford's Law (Leemis, Schmeiser and Evans 2000). The expense amounts T in the Benford data B are generated using

$$T = 2 \times 10^W, \text{ where}$$

$$W = \text{Triangular}(0, 1, 3).$$

The receipt threshold for this data set is at a value of \$25 (i.e., expense amounts above \$25 will trigger the receipt requirement). The triangular distribution (with minimum value 0, maximum value 3 and a mode of 1) was chosen to create data with some of the characteristics seen in real life data sets in the T&E domain. Histograms showing the distribution of the expenses in the Benford data B are provided in Figures 1 and 2. Figure 2 shows only the subset of these expenses that are below the receipt threshold of \$25.

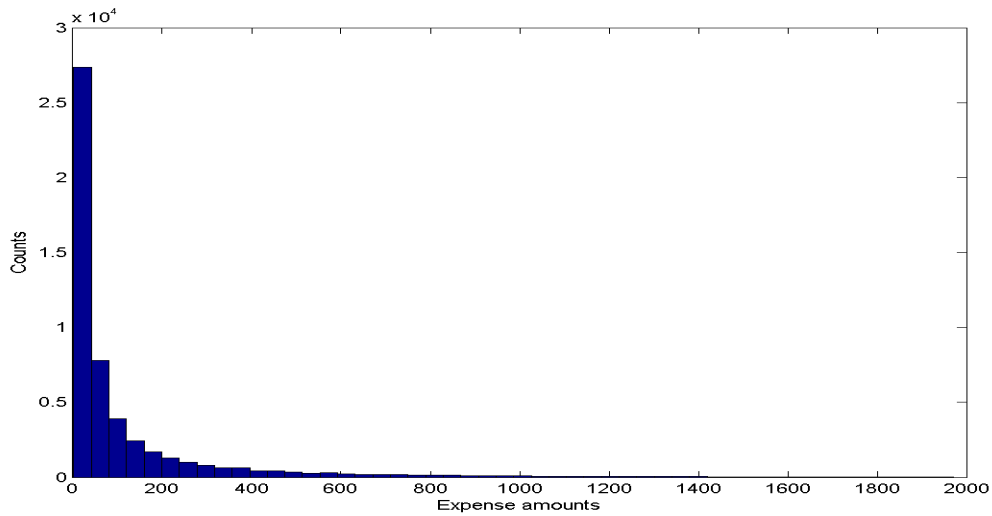


Figure 1. Histogram of expense amounts in the Benford data B before injection of "under-the-radar" behavior.

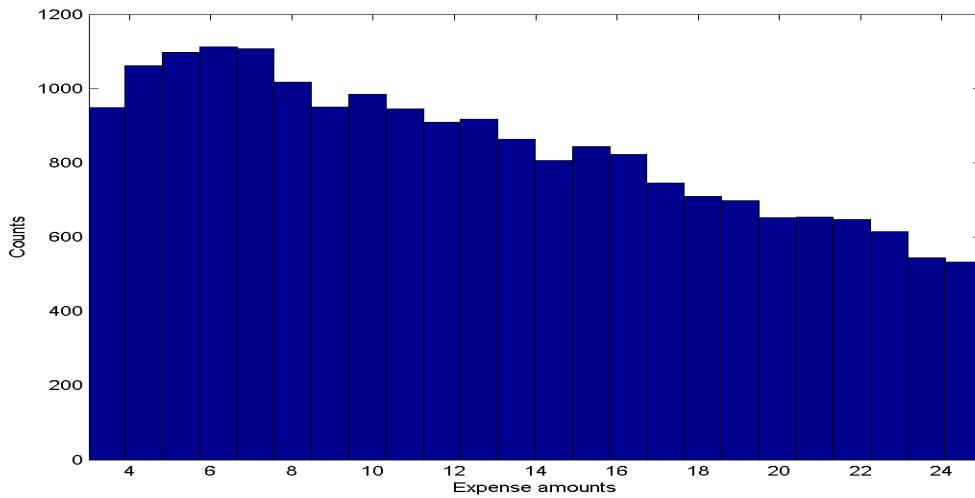


Figure 2. Histogram of expense amounts in the Benford data B before injection of "under-the-radar" behavior (expenses restricted to those not requiring receipts).

The synthetic data models the expense amounts for 1000 entities. We use an exponential distribution with a mean of 50 to model the number of expense items for each entity (based on analysis of real T&E data sets). The expense amounts in each of these expense items are chosen by random sampling from T without replacement.

This initial data set B was used to generate independent data sets with injected “under-the-radar” behavior. Each generated data set (D_i) is used in an independent random experiment (i). In each generated data set (D_i), 20 entities (representing 2% of the total number of entities) were selected randomly for injection of “under-the-radar” behavior. For each chosen entity, expenses that are below the receipt threshold are replaced by an amount that is chosen from a uniform distribution ranging from the original amount to the receipt threshold. In essence, the amounts below the receipt threshold are adjusted upwards without triggering the receipt requirement. This is illustrated in Figure 3 which shows the expense amounts for a chosen entity under the receipt threshold both before and after the injection of the “under-the-radar” behavior. The two histograms clearly show the shift towards the receipt threshold created by the injection. The next section describes and compares the detection of such behavior using two analysis methods.

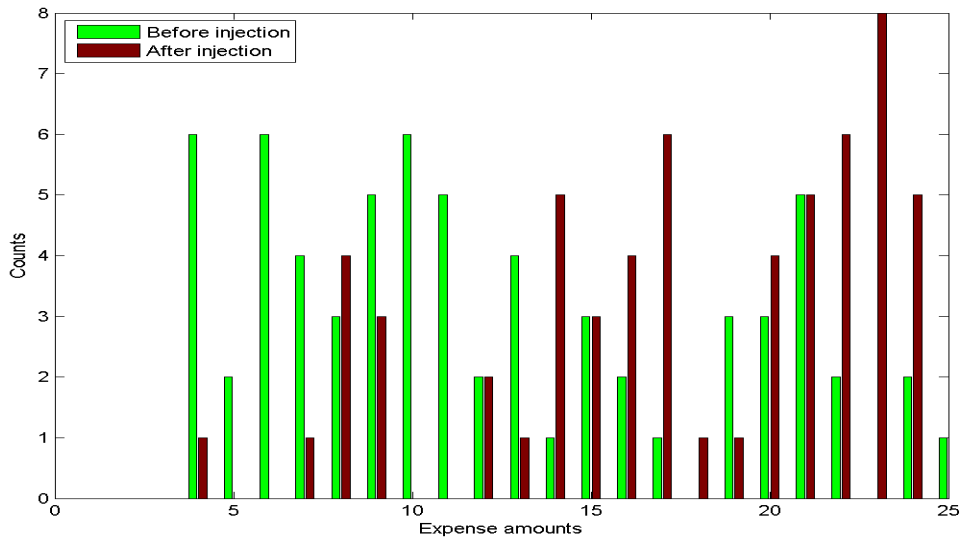


Figure 3. Example of expenses below the receipt threshold for an entity before and after injection of "under-the-radar" behavior (shown using two histograms).

ANALYSES OF THE SYNTHETIC DATA

Analysis using Benford’s Law

Entities suspected of “under-the-radar” behavior could be identified by a direct application of Benford’s Law. The expense data for each entity can be tested by using a single Chi-Square test to assess the conformance of the first digit’s distribution to

Benford's Law (Cleary and Thibodeau 2005). Entities with significant deviation from the expected distribution would be candidates for audit. Intuitively, this approach has applicability since entities without the injected "under-the-radar" behavior have expenses randomly chosen from data (B) that conforms to Benford's Law. The performance of this method can be assessed using the knowledge of entities with injected behavior with metrics based on counts of true positives, false positives and false negatives.

However, this analysis method can be tuned further using the knowledge of the receipt threshold and the nature of the "under-the-radar" behavior that is being targeted for detection. The scale invariance property of data conforming to Benford's Law is well known (Pinkham 1961). Hence, multiplying expense data conforming to Benford's Law by a non-zero constant would not impact the conformance. However, scaling does impact the performance of the method in detecting non-conforming expense data. Therefore, the analysis method can be tuned by utilizing a scaling factor for expenses that improves the detection power for the "under-the-radar" behavior and the specific receipt threshold (\$25 in our experiments). The expected first digit frequencies conforming to Benford's Law are given in Table 1. These frequencies decrease monotonically for first digit values from one to nine. The "under-the-radar" behavior is expected to increase the proportion of expenses closer to the receipt threshold of \$25. The impact of this excess will be magnified if the expenses are scaled such that digit values with lower expected frequencies have the excess occurrences. For example, analyzing the expenses after multiplying them by the scaling factor of four would cause any excessive proportion of expenses in the range (22.5, 25) to result in higher frequencies for the first digit value being nine (which has the lowest expected frequencies if conforming to Benford's Law).

Digit value	Expected frequencies for the first digit position
1	0.3010
2	0.1761
3	0.1249
4	0.0969
5	0.0792
6	0.0669
7	0.05780
8	0.0512
9	0.0458

Table 1. Expected frequencies for the first digit position conforming to Benford's Law

The impact of scaling was assessed experimentally for a set of scaling factors. The results are presented in Table 2 and discussed next in detail. Table 2 presents the average performance over twenty five independent experiments, where in each experiment a randomly chosen set of twenty entities was injected with the "below-the-radar" behavior. The scaled values for each entity were tested using a single Chi Square test and p-values computed (Cleary and Thibodeau 2005). In each experiment, entities with p-values less than the threshold of 0.05 are tagged "positive" for audit and further investigation. Each

of these tagged entities can be classified as either “true positive” or “false positive” based on whether it was injected or not with a “below-the-radar” behavior in that experiment. The false positive rate expresses the false positives as a fraction of the number of negative instances in each experiment (980 in our experiments). The false negative rate expresses the false negatives as a fraction of the total number of positive instances with injected behavior (20 in our experiments). The intuitive choice of four for the scaling factor indeed has the highest power with an average of 11.2 true positives. However, the analysis using Benford’s law results in a very high number of false positives (ranging from 41.1 to 66.64). At the intuitive choice of four for the scaling factor, for every entity correctly identified with the “under-the-radar” behavior there would almost 5 other incorrectly identified entities. This performance would be deemed very expensive from an audit cost perspective.

Scaling factor	Number of true positives	Number of false positives	Number of false negatives	False positive rate	False negative rate
1	5.32	47.04	14.68	0.048	0.734
2	9.80	41.16	10.2	0.042	0.510
3	9.68	47.04	10.32	0.048	0.516
4	11.2	54.88	8.8	0.056	0.440
5	3.68	47.04	16.32	0.048	0.816
6	4.72	66.64	15.28	0.068	0.764
7	4.44	58.8	15.56	0.06	0.778
8	4.76	59.78	15.24	0.061	0.762
9	2.96	55.86	17.04	0.057	0.852

Table 2. Results of applying the Benford's Law based method in 25 experiments using synthetic data

Analysis using Behavioral Shift Model (BSM)

Behavioral Shift Model (BSM) is new application of hypothesis testing using Likelihood Ratio Tests (LRT) that is adapted for this specific task (Iyengar et al. 2007). Detection of the “under-the-radar” behavior is done with BSM by analyzing the subset of the data with only those expenses that are under the receipt threshold. This subset of expenses over all the entities is used as a “normal” baseline and the likelihood ratio test considers the two hypotheses: a particular entity’s mean expense is either the same as (null hypothesis) or greater than (alternate hypothesis) the “normal” mean from the baseline. The likelihood ratio test is formulated using the exponential distribution and this formulation has been shown to have wider applicability to other distributions and can be applied to censored data like the subset below the receipt threshold (Huang et al. 2007, Iyengar et al. 2007). Consider an entity E with M expenses in the filtered set F that sum up to Q. Let N denote the total number of expenses in the subset (considering all the entities) and P denote their sum. The score for entity E using BSM is given by:

$$Score(E) = M \times \log\left(\frac{M}{Q}\right) + (N - M) \times \log\left(\frac{N - M}{P - Q}\right) - N \times \log\left(\frac{N}{P}\right).$$

The p-value is computed directly by Monte Carlo experiments to empirically determine the distribution for the scores. In each of these experiments, each entity is assigned expenses by sampling at random from the entire set of expenses in the subset data. Entities with estimated p-values below 0.05 are tagged as “positive” for audit and further investigation. The performance of the BSM analysis (averaged over the same twenty five data sets used in the earlier Benford’s Law based analysis) is given in Table 3. The results from the earlier analysis using Benford’s Law (with the best scaling factor of four) are also repeated in Table 3 for ease of comparison. The results indicate that BSM analysis correctly detects fewer entities with “under-the-radar” behavior when compared to the earlier analysis using Benford’s Law (scaling factor of four). The average number of entities correctly detected is around 78% of the number correctly detected by the use of Benford’s Law. But, this relatively small degradation in detection by BSM comes with a remarkable improvement in the false positive rate (for these twenty five experiments there were no false positives). In most audit applications, the BSM performance would represent the preferred tradeoff between detecting suspicious behavior worthy of investigation and false alarms leading to wasteful investigations.

Method	Number of true positives	Number of false positives	Number of false negatives	False positive rate	False negative rate
BSM	8.68	0	11.32	0.0	0.566
Benford’s Law (Scaling factor = 4)	11.2	54.88	8.8	0.056	0.440

Table 3. Results of applying the BSM method in 25 experiments using synthetic data (earlier results from Benford’s Law based analysis repeated for easy comparison).

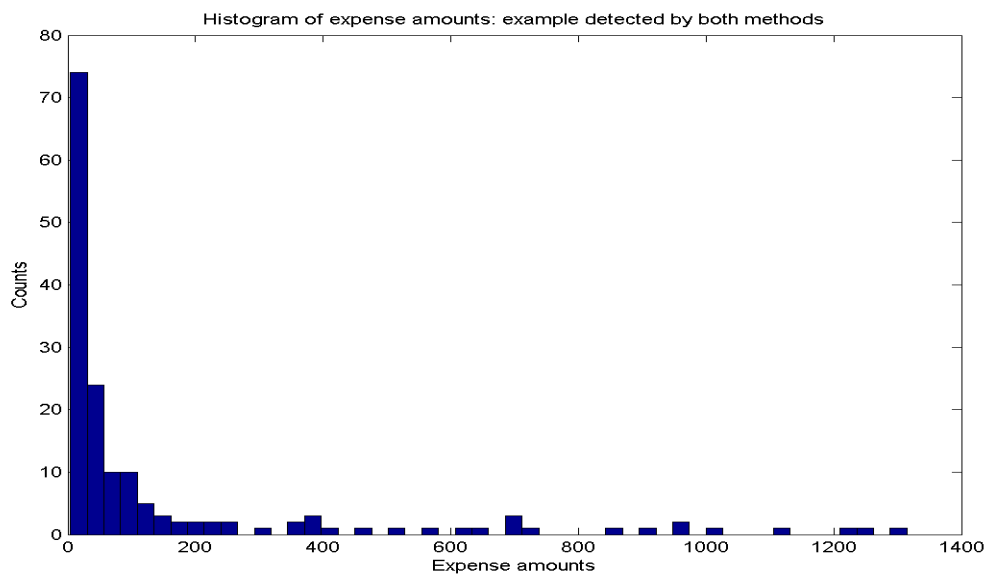


Figure 4. Histogram of expenses for an entity detected by both methods

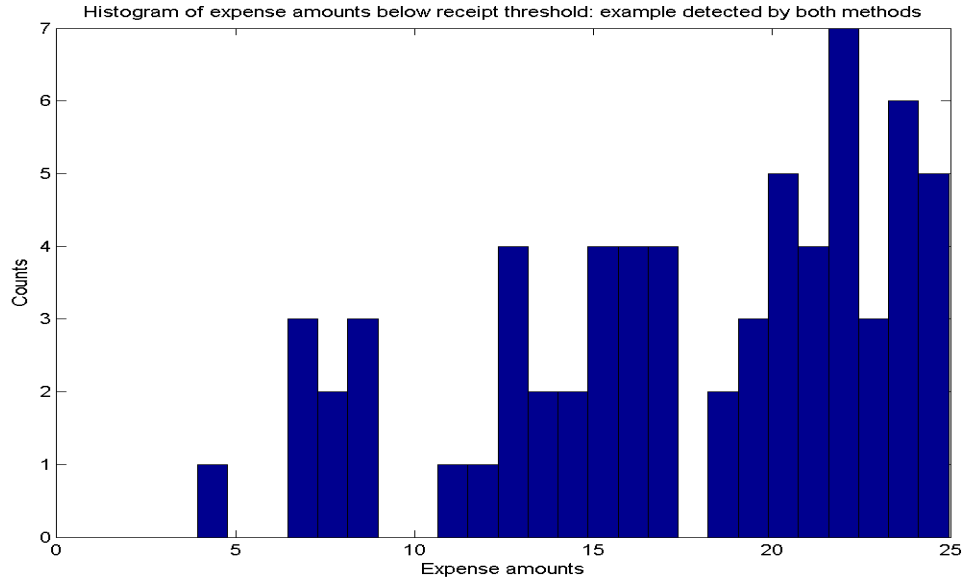


Figure 5. Histogram of expenses below the receipt threshold for an entity detected by both methods

An intuitive understanding of the detection of the “under-the-radar” behavior can be gotten by considering the distribution of expenses for an entity that was correctly detected by both methods. Figure 4 displays the histogram of all the expenses for such an entity while Figure 5 displays the histogram of the subset of its expenses below the receipt threshold. The disproportionate fraction of expenses below the \$25 receipt threshold is quite apparent in Figure 5.

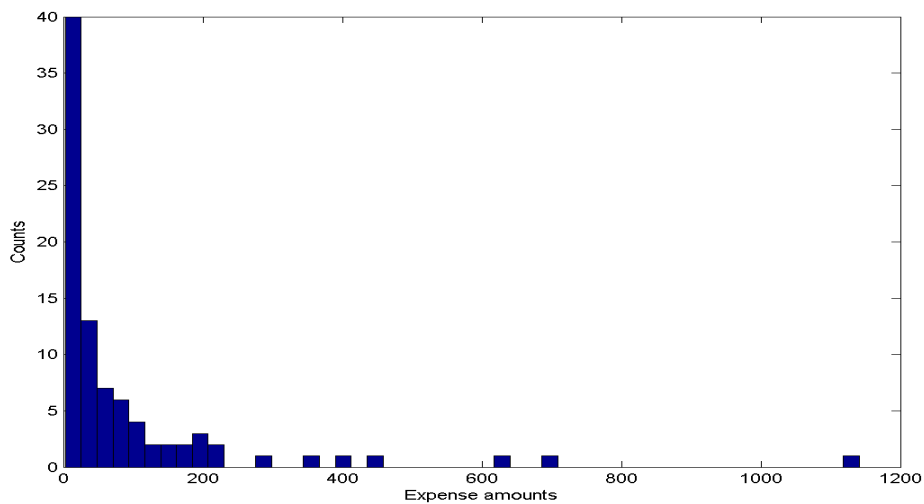


Figure 6. Histogram of all expenses for an entity incorrectly detected only by Benford's Law based analysis

In addition, examining the expenses corresponding to an entity that was incorrectly detected by the Benford’s law based analysis but not by BSM can shed some further insight into their performances. Figure 6 shows the histogram of all expenses for such an entity while Figure 7 shows the histogram of the subset of its expenses below the receipt threshold. Unlike the earlier case, there is no apparent disproportionate concentration of expenses just below the receipt threshold for this entity. However, a general test like the one based on Benford’s Law detects non-conformance and flags this entity for further investigation

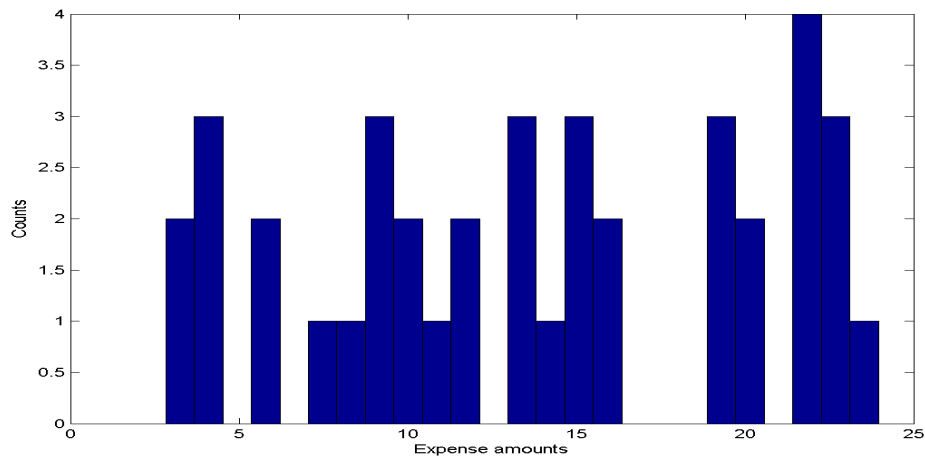


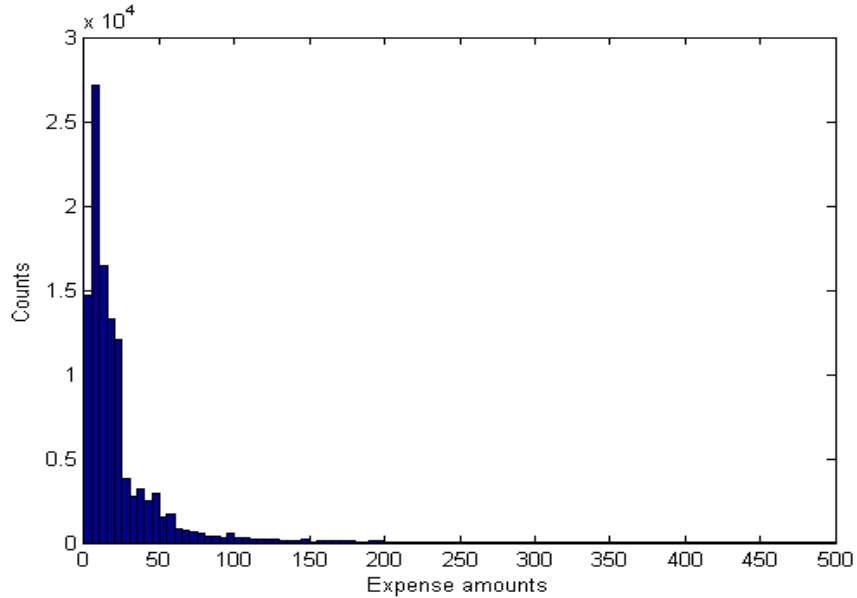
Figure 7. Histogram of expenses below the receipt threshold for an entity incorrectly detected only by Benford's Law based analysis

ANALYSES OF THE REAL DATA

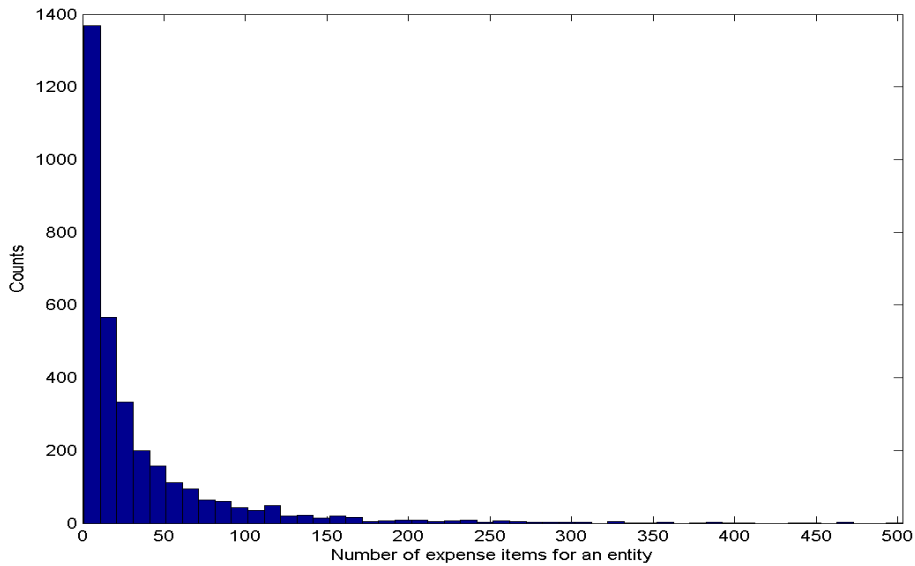
The earlier analysis of carefully engineered synthetic data provided some insights into the performance of the two methods (based on Benford’s Law and Behavioral Shift Models) in detecting “under-the-radar” behavior. In this section, the two methods will be compared by applying them to real data in the T&E domain extracted from an enterprise expense reporting system (GERS). Unlike the case with synthetic data the results are not clearly quantified in terms of true positives, false positives, and false negatives. Many of the entities identified by BSM were validated by audit and business control professionals. However, entities identified only by the Benford’s Law based analysis were too numerous and were not validated. The comparison between the two methods will be done in this section by illustrating examples of cases categorized by the outcomes (positive or negative) from the methods. This deep dive complements the earlier analysis of the synthetic data to provide a comprehensive comparison of the two methods.

The T&E data analyzed contains 112116 expense claims by a sample of 3265 employees in an organization from a one year time period. Only those expenses that were paid in cash and with a receipt threshold of \$25 were included in this data. The distribution of expenses is shown in Figure 8. The expense data has a long tail ranging up to a maximum expense of \$25000. For clarity, the histogram in Figure 8 shows only the distribution of expenses up to \$500, even though the analyzed data included the tail

beyond. The distribution of the number of expense items claimed by a single entity is shown in Figure 9. The histogram shows the typical profile with many employees having just a few expense claims and a decreasing number of frequent travelers with many claims. The average number of expense items claimed by an employee in this data set is thirty four.



**Figure 8. Histogram of expenses in real data set
(Tail of distribution for expenses beyond \$500 is not shown)**



**Figure 9. Histogram showing the distribution of the number of expenses claimed by an employee
(one year period)**

This data set was analyzed using BSM and the Benford’s Law based method. The scaling factor of four was used for the Benford’s Law method. A p-value threshold of 0.05 was used for both methods to determine which entities tagged as positive. The comparative results are shown using the 2-by-2 matrix in Table 4 that shows the counts of entities categorized by the outcomes from both methods. A pattern similar to that seen with the synthetic data emerges in Table 4. The number of entities tagged positive by the BSM method is small (137) compared to those identified by the method based on Benford’s Law (1357). Almost all (132 out of 137) entities tagged as positive by BSM are also positively identified by the Benford’s Law based method. Next, examples will be examined based on the categories indicated by this 2-by-2 matrix.

	BSM method: negative	BSM method: positive
Benford’s Law method: negative	1903	5
Benford’s Law method: positive	1225	132

Table 4. Comparative results on the real data for BSM and Benford's Law based methods

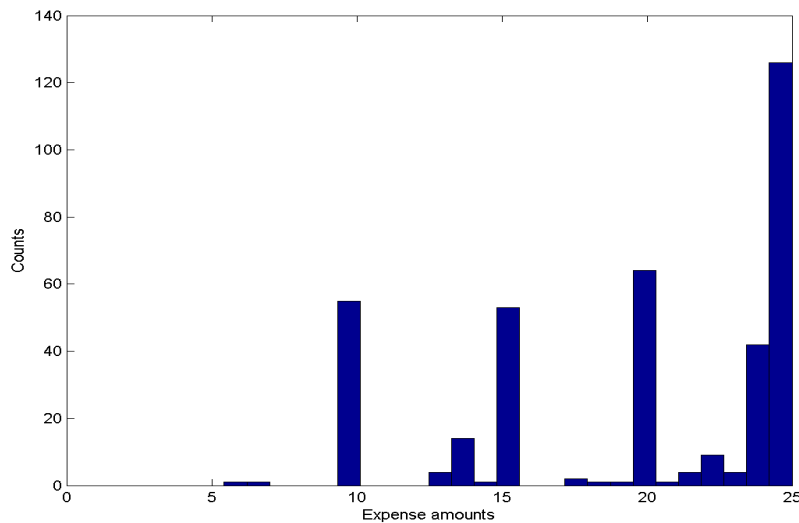


Figure 10. Histogram showing distribution of expenses for an employee tagged as positive by both methods (excludes two expenses of \$135 and \$137)

Consider an employee tagged as positive by both methods. The histogram of expenses for this employee is shown in Figure 10. Two expenses claimed for the amounts \$135 and \$137 were excluded from this histogram to increase the clarity in the range below the receipt threshold. The disproportionate spikes just below the \$25 threshold correspond to expense claims for individual meal items (breakfast, lunch or dinner) and for ground

transportation expenses. In addition, the large number of claims at the rounded expense amounts of \$10, \$15 and \$20 add to the evidence suggesting further investigation of these expense claims. This example also illustrates how a simple heuristic that considers counts and proportions of expenses in a fixed window below the receipt threshold would not factor in more complex patterns of disproportionate claims.

Next, consider an employee tagged as positive by BSM and negative by the Benford's Law based method. The histogram of the expenses for this employee is shown in Figure 11. The histogram clearly shows the spike in number of claims just below the receipt threshold that would have contributed to the positive outcome from BSM. The scaled Benford's Law based analysis computes the digit frequencies shown in Table 5 which do not pass the significance threshold for being tagged positive.

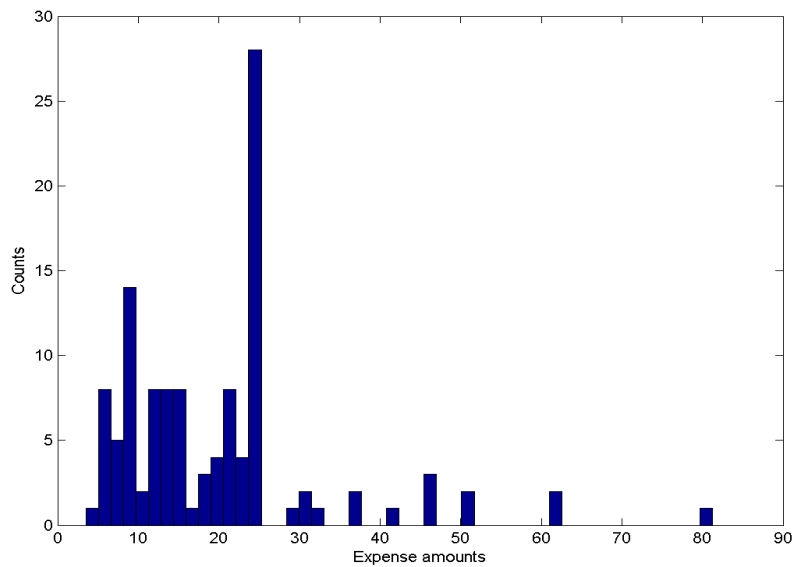


Figure 11. Histogram showing distribution of expenses for an employee tagged positive by BSM and negative by the Benford's Law based analysis

Digit value	Actual frequencies for the first digit position	Expected frequencies for the first digit position
1	0.3077	0.3010
2	0.1197	0.1761
3	0.1538	0.1249
4	0.0684	0.0969
5	0.0940	0.0792
6	0.0684	0.0669
7	0.0598	0.0580
8	0.0684	0.0512
9	0.0598	0.0458

Table 5. Actual and expected digit value frequencies from the scaled Benford's Law based analysis that tagged the employee negative (BSM tagged this employee as positive)

Lastly, consider an employee tagged positive by Benford’s Law based analysis but not by BSM. Figure 12 shows the distribution of the employee’s expenses and Table 6 has the actual and expected frequencies for the first digit values from the Benford’s Law based analysis of the scaled expenses (scaling factor was four). There is no clear pattern of “under-the-radar” behavior while the digit analysis shows clear non-conformance to Benford’s Law.

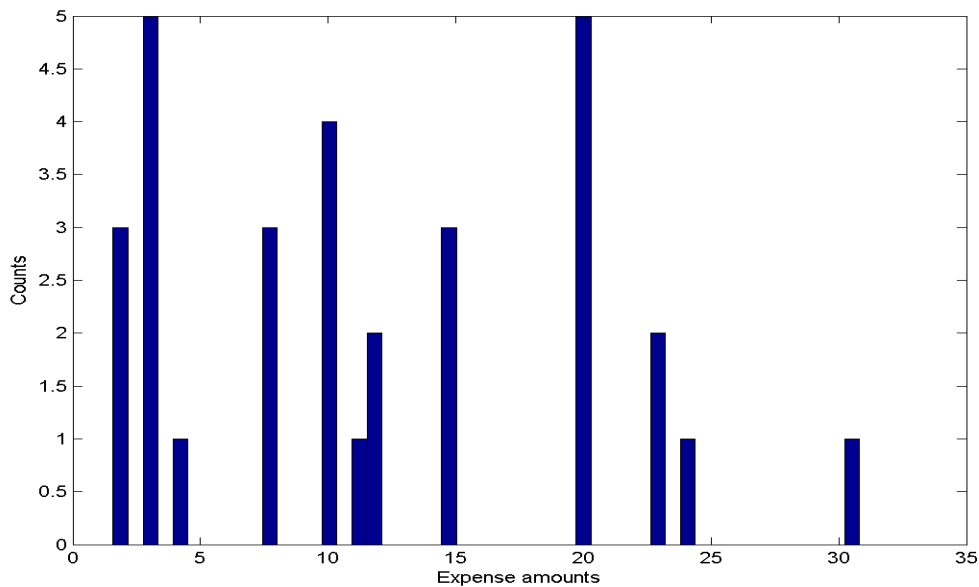


Figure 12. Histogram showing distribution of expenses for an employee tagged as positive by Benford’s Law based analysis and negative by BSM

Digit value	Actual frequencies for the first digit position	Expected frequencies for the first digit position
1	0.2258	0.3010
2	0	0.1761
3	0.0968	0.1249
4	0.2258	0.0969
5	0	0.0792
6	0.1290	0.0669
7	0	0.0580
8	0.2258	0.0512
9	0.0968	0.0458

Table 6. Actual and expected digit value frequencies from the scaled Benford’s Law based analysis that tagged the employee positive (BSM tagged this employee as negative)

The examples with differing outcomes from the two analyses methods reinforce some of the issues with applying Benford’s Law to this problem scenario. The use of scaling to tune the Benford’s Law based method to the specifics of the scenario (e.g., the specific receipt threshold) does not go far enough in making the test focused on the problem scenario, namely detecting “under-the-radar” behavior. The BSM approach, on the other

hand, is more focused by identifying entities with shifted distributions of expenses when compared to the distribution over all entities. The use of the likelihood ratio test adds to the robustness by making the scoring depend on both the amount of deviation from the expected behavior and its repetitiveness. The use of Monte Carlo based p-value estimation takes into account the actual distribution of expenses (typically with significant tails) and reduces the bias in the estimated p-value.

ANALYZING RECEIPT RELATED EXCEPTIONS

The discussion on scenarios related to receipts is not complete without considering the analysis of policy exceptions related to receipts. Specifically, this section will consider the handling of missing receipts exception. From a business controls perspective, excessive submissions of expense claims with the required receipts missing (triggering an exception in each case) is important to monitor and control. Other control points include excessive approval of the missing receipt exceptions by individual approvers or after aggregation within larger organizational units.

Consider the analysis of excessive missing receipt exceptions by employees. A simple heuristic would be to rank employees by the total number of these exceptions and consider employees at the top of the list for further investigation and control action. However, this simple heuristic would penalize the frequent travelers since they have higher numbers of expenses and consequently a higher number of opportunities to incur this exception. A possible fix to this bias might be to consider exception rates for employees that normalize the number of exceptions triggered by the number of exception opportunities. This simple fix still has issues since it could highlight employees who have very few expense claims but happen to generate exceptions on a significant fraction of them. Usually, these cases are not good candidates for further investigation or business control action. There is also a subtle problem with these simple heuristics. The missing receipt exception rate is not uniform across all expense types. For example, missing receipts exception rates are typically higher for ground transportation expenses than for hotel room rate expenses. This non-homogeneity occurs across the organization and should be taken into account if the goal is to identify employees with significant deviation in policy regarding receipt submissions when compared to the entire organization.

Behavioral Shift Models (BSM) can be formulated for event count scenarios and this formulation can, by choice, factor in non-homogeneity of the kind discussed above due to different expense types (Iyengar et al. 2007). Consider an entity E with V occurrences of the event of interest (e.g., missing receipt exception) and W opportunities. For example, an entity might have 10 missing receipt exceptions in a set of expense claims that had 100 expenses requiring receipts. The exception rate for this entity is 10% based on this data. Similarly, the total number of exceptions T and the actual exception rate Z achieved for the entire organization can be calculated. If the exception rate Z for the entire organization is excessive then the control action has to be applied organization-wide. The more interesting situation is when the organization's rate is well controlled and the goal is to identify entities (e.g., employees) with excessive rates. This formulation of

BSM uses the likelihood ratio test (using an underlying Poisson distribution) to score entities with excessive exceptions (Kulldorff 1997). The simplified score for an entity E is given by:

$$Score(E) = V \times \log\left(\frac{V}{W \times Z}\right) + (T - V) \times \log\left(\frac{T - V}{T - (W \times Z)}\right)$$

As mentioned earlier, the BSM formulation can factor in non-homogeneity in event rates due to other factors (e.g., varying missing receipt exceptions for different expense types). In the homogeneous case, the expected rate of exceptions for an entity E is simply $(W \times Z)$. For the non-homogenous case, the expected rate is computed for each homogeneous segment and then aggregated (Iyengar et al. 2007, Kulldorff 1997). The p-values are estimated by Monte Carlo methods using the Poisson distribution.

This BSM formulation was applied to the real T&E enterprise data set considered earlier (same set of employees, same time period and restricted to cash payments) to identify employees with excessive missing receipt exceptions. The number of employees with one or more missing receipt exceptions in this data set was 1098. At the significance threshold of 0.05, there were 78 employees identified by BSM as having excessive missing receipt submissions. Table 7 summarizes the results for the top three employees identified by BSM. For each identified employee the table lists the actual number of missing receipt exceptions triggered in the time period of the analysis, the expected number of exceptions and the number of exception opportunities (i.e., expenses claimed in the same time period that require receipts). The ranking is based on the scores from the likelihood ratio test. The expected number of exceptions is computed by considering each expense type as a homogeneous segment for missing receipt exception rates. The expected numbers of exceptions in Table 7 are not proportional to the opportunities since the distribution of expense types varies from one employee to another. For example, the second ranked employee has many hotel room rate expenses leading to a reduction in the employee's expected number of exceptions. The employees identified in Table 7 clearly have excessive missing receipt exceptions using the expected number as reference. The likelihood ratio test formulation using in BSM factors in the repetitiveness of the behavior (i.e., exception triggering) allowing auditors to focus on the entities with higher numbers of violation taking into account the expected numbers of violations.

BSM Rank	Actual number of exceptions	Expected number of exceptions	Number of exception opportunities
1	128	9	263
2	129	11	519
3	86	9.5	418

Table 7. Summary results for the top three employees identified by BSM as having excessive missing receipt exceptions

CONCLUSIONS

Computer aided audit techniques (CAAT) have an important role especially in domains with large volumes of data. They can provide real value to a domain if the analysis methods are capable of sifting through the data and identify cases for further investigation with high accuracy. Analysis based on Benford's Law has been used in various scenarios to identify suspicious cases. This paper focused on scenarios related to receipt rules in an organization that included attempts to pass under the threshold of detection (called "under-the-radar" behavior) and incurring excessive policy violations. Two analysis methods were discussed for the "under-the-radar" detection. The first method is based on an adaptation of Benford's Law to this specific problem. The second method is based on a new application of likelihood ratio tests (called Behavioral Shift Models) focused on these specific problem scenarios. Experimental results were presented from application of both methods to carefully engineered synthetic data and to real data from the travel and business entertainment expense management domain. Results indicate that a focused method like BSM can achieve almost the same detection power as the Benford's Law based analysis but with significantly fewer false alarms. A different formulation in BSM also addresses the analysis of business rule violations to identify entities with excessive exceptions. Our experience in applying analyses targeted at specific scenarios (e.g., BSM for receipt scenarios) in the business travel and entertainment expense management domain suggests that focused tests can effectively leverage the domain knowledge of experts and provide value to audit and business control actions.

ACKNOWLEDGMENTS

I would like to acknowledge the contributions of my colleagues Karen Kelley, Ioana Boier and Ray Curatolo to the preceding project that developed Behavioral Shift Models and applied them to the business travel and entertainment expense management domain.

REFERENCES

Cleary, R., and J.C. Thibodeau. 2005. Applying digital analysis using Benford's law to detect fraud: The dangers of type I errors. *Auditing: A Journal of Practice & Theory* 24 (1): 77-81.

GERS: IBM Global Expense Reporting Solutions,
<http://www-935.ibm.com/services/us/index.wss/offering/igs/a1009035>.

Huang, L. Kulldorff, M., and Gregorio, D. 2007. A Spatial Scan Statistic for Survival Data. *Biometrics* 63(1): 109-118.

Iyengar, V.S., I. Boier, K. Kelley, and R. Curatolo. 2007. Analytics for audit and business controls in corporate travel and entertainment. Proceedings of the Sixth Australasian Data Mining Conference, Conferences in Research and Practice in Information Technology Series 70: 3-12.

Kulldorff, M. 1997. A Spatial Scan Statistic, Communications in Statistics: Theory and Methods, 26:1481-1496.

Leemis, L.M., B.W. Schmeiser, and D.L. Evans. 2000. Survival Distributions Satisfying Benford's Law. The American Statistician, 54(4): 236-241.

Nigrini, M. 1996. A taxpayer compliance application of Benford's law. The Journal of the American Tax Association 18 (Spring): 72-91.

Pinkham, R. S. 1961. On the distribution of first significant digits. Annals of Mathematical Statistics 32: 1223-1230.