DOI: 10.32602/jafas.2019.25

JOURNAL OF ACCOUNTING, FINANCE AND AUDITING STUDIES
http://www.jafas.org

Estimating the Proportion of Misstated Records in an Audit Data set using Benford's Law

Carlos Gomes da Silva[a] Pedro Manuel Rodrigues Carreira[b]

[a] Corresponding Author, School of Technology and Management, Polytechnic Institute of Leiria, and *INESC Coimbra*, Portugal, cgsilva@ipleiria.pt
[b] School of Technology and Management, Polytechnic Institute of Leiria, Portugal, pedro.carreira@ipleiria.pt

Keywords
Accounting and Finance, Auditing, Fraud Detection, Sampling, Benford's Law.

Jel Classification
M42, H26, C83.

Abstract
Auditors are required to provide high levels of assurance that financial statements are free of material misstatements. This paper contributes to the literature on the field of audit sampling, by proposing a procedure to estimate the proportion of misstated records in a numerical audit data set based on stratified sampling, which can also be of assistance in financial fraud detection. Stratification rules based on the expected profile of misstated records and on Benford's law are evaluated and compared through an empirical experiment. The results show that: 1) the examined stratification rules perform significantly better than a simple random sampling approach; 2) when using Benford's law, combining it with other methods does not seem to improve the performance of the estimation. The proposed procedure can be embedded in an audit software and contribute to enhance the effectiveness of audits and fraud detection.

1. Introduction

Organizations produce and use information to support their own decisions and the decisions of their stakeholders. In the absence of misstatements in the data, the quality of these decisions improves. Auditing procedures are thus important and requested to be in constant development in order to work efficiently and effectively in a complex interconnected world that is continuously producing and processing huge amounts of data. This challenge is being pursued by both professionals and academics.

The present paper contributes to the auditing field by focusing on an attribute of the numerical records in the financial statements of organizations, which is their dichotomic nature of being either misstated or non-misstated, with misstatements being either unintentional (errors) or deliberate (fraud), such as the manipulation of revenues, sales, receivables, inventory, debts, allowances and expenses, for example. In detail, we aim to answer the following research question: from a set of numerical records (that can be the ones of a particular account or class of accounts), which is the proportion of misstated records?

Given that it is frequently too costly to audit all records from an account due to the oversize of the data population, even when large computing capabilities are available, audit sampling is useful, particularly stratified sampling. Audit sampling is a widespread technique used by auditors with increasing requirements and challenges (Danuescu and Anca-Oanab, 2012; Elder *et al*., 2013; Lombardi *et al.*, 2014; Christensen *et al.*, 2015). Of course, when using sampling, audit risk (i.e. the risk of forming an incorrect audit conclusion) is naturally present, as the number of errors can be either overestimated or underestimated given that only part of the records are examined. The auditor must thus make efforts so that the sampling method minimizes this risk and provides a reasonable basis for drawing conclusions about the population, which justifies the pursuit for more effective and efficient sampling procedures. Indeed, as referred in Christensen *et al.* (2015), the Public Company Accounting Oversight Board identified sampling as an area needing more emphasis.

In our approach, the proportion of misstated records in a data set of numerical records is estimated through an inference procedure based on stratified sampling. Logically, we aim to find an unbiased estimator with high precision (low variance). While the advantages of stratified sampling over simple random sampling in terms of precision of the underlying

estimators are well known (Lohr, 2010), they vary with the quality of the stratification rule employed to define the stratums, which justifies the need to evaluate and compare the performance of alternative stratification rules. This is done through an empirical experiment in a subsequent section of the paper.

Given the dichotomic nature of the records, two stratums are proposed, one constituted by the records signaled by the stratification rule and the other by the remaining ones. In detail, a stratification rule operates here by classifying each record as either suspicious (potentially misstated) or non-suspicious (potentially non-misstated) and by forming the two stratums accordingly.

In the literature, statistical classification methods that can be used in auditing to signal suspicious records are either supervised or unsupervised. Supervised methods (for example linear discriminant analysis, logistic discrimination, neural networks and genetic algorithms) require the collection of information about occurrences of both misstated and non-misstated records in order to derive a model capable of predicting the nature of new records. However, the required information may not always be available or can be costly to obtain, thus making of interest unsupervised methods, which do not require such information. Generally, unsupervised methods compare the behavior of the data with some expected pattern and are usually based on cluster analysis or profiling and outlier detection. Applications of several supervised and unsupervised methods to the auditing field can be found for instance in the review by Bolton and Hand (2002). It is however important to note that it is not possible, using statistical analysis alone, to conclude that fraud has been perpetrated. It only alerts the auditor for suspicious data that needs further examination.

In the present paper, we define the stratums in an unsupervised way, by profiling the suspicious records and by using Benford's law (Newcomb, 1881; Benford, 1938).

Within a data analysis framework, as it is the case of our study, Nigrini and Mittermaier (1997) considered as audit targets (suspicious records) rounded numbers, numbers below psychological thresholds or internal authorization limits and numbers occurring with a relatively high frequency. This defines a profile for the suspicious records.

Moreover, by allowing to analyze and detect distortions in the pattern of the digits of the numbers in a data set, Benford's law has been proved to be useful in audit contexts (Carslaw, 1988; Nigrini, 1994; Hill, 1995; Nigrini and Mittermaier, 1997; Nigrini, 1999; Durstchi *et al.*,

2004; Johnson and Weggenmann, 2013). Indeed, Benford's law predicts a specific pattern for the digits of numbers and is expected to be followed by a large range of variables, namely by many accounting and financial variables, such as, for example, transaction amounts, corporate net incomes, individual taxable incomes and stock prices (Hill, 1995; Nigrini and Mittermaier, 1997; Durstchi *et al.*, 2004). Also, Nigrini (1999) refers some practical applications of the law in the audit context, as for example to analyze accounts payable data, estimations in the general ledger, the relative size of inventory unit prices among locations, duplicate payments, computer system conversion of accounts, new combination of selling prices and customer refunds.

When using Benford's law to identify the set of suspicious records, the usual procedure is to compare the expected frequencies of the digits, according to the law, with the observed ones in the data set under audit. For example, Nigrini and Mittermaier (1997) also consider as suspicious the records with first-two digits that register significant positive spikes, i.e. for which the observed relative frequency is significantly higher than the expected according to the law. Also, in a more recent approach by Gomes da Silva and Carreira (2013), the set of suspicious records is the solution of a mathematical programming model that uses multiple conformity tests and test statistics simultaneously to evaluate the statistical divergence between the pattern of the digits in the observed data and the expected pattern according to Benford's law. In detail, the model works by identifying the smallest subset of records (the ones that are responsible for the nonconformity and labeled as suspicious) from the initial data set of records, so that the set of the remaining records becomes conforming. Despite more demanding, this last approach is more likely to detect subtle data manipulations, such as number invention (which occurs for example when an employee invents the number of units produced in a given day instead of executing a true inspection). In the present paper, we embed Benford's law in an inference process, which, despite being straightforward, is new to the literature.

The remaining of the paper is as follows. In section 2, we present the proposed approach to estimate the proportion of misstated records in a numerical data set. The empirical experiment is conducted in section 3, including the discussion of the results. Finally, in section 4, the main conclusions are presented.

2. Estimating the Proportion of Misstated Records

In this section, we develop the procedure to estimate the proportion of misstated records in a data set under stratified sampling, considering an audit context where the auditor has a set of records that are numerical occurrences of an audit variable, such as, for example, sales revenues, payments or inventory, and wants to assess the proportion of misstated records in it, with sampling being convenient.

In order to formally describe the procedure, we introduce the following notation:

N - number of records in the audit population;

N_i - number of records in stratum i (i=1,2);

C - number of misstated records in the audit population;

C_i - number of misstated records in stratum i (i=1,2);

p - proportion of misstated records in the audit population;

p_i - proportion of misstated records in stratum i (i=1,2);

n_i - number of records in the sample from stratum i (i=1,2);

c_i - number of misstated records in the sample from stratum i (i=1,2);

p_{strat} - estimator for p.

As the proportions of misstated records in stratums 1 (assumed to be the one constituted by the suspicious records signaled by the stratification rule) and 2 (constituted by the non-suspicious records) are $p_1=C_1/N_1$ and $p_2=C_2/N_2$, the proportion of misstated records in the population is given by $p=(N_1/N)p_1+(N_2/N)p_2$. If p_1 and p_2 are too costly to obtain, they must be estimated. These estimations require sampling within each stratum.

By sampling n_1 records in a random manner from the set of N_1 records in stratum 1, and by examining them (i.e. by verifying their true nature - misstated or non-misstated), an unbiased estimator for p_1 is $\hat{p}_1=c_1/n_1$. Similarly, by examining n_2 records selected in a random manner from the set of N_2 records of stratum 2, an unbiased estimator for p_2 is $\hat{p}_2=c_2/n_2$. Hence, an unbiased estimator for p is $p_{strat}=(N_1/N)\hat{p}_1+(N_2/N)\hat{p}_2$.

Thus, we suggest the following procedure to estimate p:

Procedure: Estimation of p

Step 1. Stratify the population in two stratums according to a given stratification rule (N_1 and N_2 are defined).

Step 2. Select a random sample of $n_i=\pi_i N_i$ records from stratum i, where i=1,2 and $\pi_i \in]0,1]$, and examine the sampled records to determine which of them are misstated (determine c_1 and c_2).

Step 3. Compute p_{strat}.

With respect to the property of unbiasedness, note that the natural alternative estimator N_1/N (the proportion of records classified as suspicious), which does not require inspection of the records of any sample, is, in general, biased for p. This is because the composition of the stratums is likely to be imperfect in the sense stratum 1 contains non-misstated records (false positive errors) and stratum 2 contains misstated ones (false negative errors). This estimator would be unbiased only in the unlikely event that all and only misstated records are signaled as such, i.e. in the case where the composition of the stratums is free of errors. In the general case where errors in the composition of stratums are present, p can only be estimated in an unbiased way by examining random samples of both stratums. Table 1 summarizes the situations that can occur regarding the types of errors that can be made in stratification process.

Table 1: Stratification errors

Stratification result	p_1	p_2	Errors
$C_1=N_1, C_2=0$	1	0	-
$C_1<N_1, C_2=0$	C_1/N_1	0	False positives
$C_1=N_1, C_2>0$	1	C_2/N_2	False negatives
$C_1<N_1, C_2>0$	C_1/N_1	C_2/N_2	False positives and false negatives

Beyond the unbiasedness, the precision is also important to assess the quality of p_{strat} as an estimator for p. As referred previously, stratified sampling increases the precision of estimators relatively to simple random sampling. This is indeed the case whenever the stratums are built so that the variance within each stratum is lower than the variance in the total population. Moreover, the reduction in the variance is larger the better the stratification

rule works in allocating data with different characteristics to different stratums (Lohr, 2010).

In our context, this means that the variance of p_{strat}, given by $Var(p_{strat})=(N_1/N)^2(p_1(1-p_1))/n_1)(N-n_1)/(N-1)+(N_2/N)^2(p_2(1-p_2)/n_2)(N-n_2)/(N-1)$ (Lohr, 2010), depends on the number of errors made by the stratification rule. Clearly, for given values of N_1, N_2, n_1 and n_2, the closer p_1 is to 1 and p_2 to 0, respectively, the lower the total number of errors and the lower the variance of p_{strat}. Hence, the lower the total number of errors made by the stratification rule, the higher the quality of p_{strat} as an estimator for p.

In addition, note that while the minimization of false positive errors is most probable with rules that produce small values for N_1 (i.e. rules that have stricter requirements to classify a record as suspicious), the minimization of false negative errors is most likely with stratification rules that produce high values for N_1 (i.e. rules that have stricter requirements to classify a particular record as non-suspicious). Hence, making a good balance in terms of the size of the stratums seems to be important so as to minimize the total number of errors.

Given the prior discussion, we now define three possible stratification rules that auditors can employ to define the stratums within the current approach. In practice, as most manipulations of financial records are made by rounding numbers, placing numbers just below psychological thresholds or internal authorization limits, duplicating records or by inventing numbers, the suggested rules are based on the profile of suspicious records defined earlier and on Benford's law, as follows:

Rule AP1: stratum 1 is the reunion of the following subsets of records:

(a) multiples of 100 (targeting rounded numbers);

(b) numbers with last two digits 99 (targeting numbers just below psychological thresholds or internal authorization limits);

(c) numbers that appear 5 times or more (targeting duplications);

(d) numbers with first-two digits with significant positive spikes at a 5% level (targeting rounded numbers, numbers just below psychological thresholds or internal authorization limits, duplications and invented numbers);

(e) solution of Model 1 of Gomes da Silva and Carreira, 2013 (targeting rounded numbers, numbers just below psychological thresholds or internal authorization limits, duplications, invented numbers and other distortions).

The logic for this rule is to "catch-it-all", i.e. to apply all the main tools offered by the literature to signal suspicious records when the audit data set is solely a numerical set of records from an account (or class of accounts) of a given firm over one specific time period. These are the profile of suspicious records, the simplified use of Benford's law (first-two digits test only) and the more exhaustive use of Benford's law (several conformity tests and test statistics simultaneously).

As this rule is relatively undemanding to classify a record as suspicious, there is some risk that it produces too high values for N_1, with the correspondent negative consequences for the number of classifying errors and for the precision of p_{strat}. Hence, the other two suggested stratification rules (AP2 and AP3) are subsets of AP1 so that N_1 is lower. In particular, it is likely to exist some redundancy between Model 1 of Gomes da Silva and Carreira (2013), which has the ability to capture all typical data manipulations, with the reunion of the other four subsets in AP1. We thus define rules AP2 and AP3 accordingly.

Rule AP2: stratum 1 is the reunion of subsets (a) to (d) in AP1.

Rule AP3: stratum 1 is subset (e) in AP1.

3. Empirical Experiments to Compute P

In this section, we evaluate the performance of the procedure proposed in the previous section, achieved under each of the stratification rules AP1, AP2 and AP3, concerning the effectiveness in identifying correctly the misstated and non-misstated records and the precision of the resulting estimator for p, for different intensity levels of misstatements present in the data. Also, we compare these performances with the one where simple random sampling is used to estimate p.

Data and experiment

In this empirical study, we use thirty simulated Benford compatible data sets (equivalent to 30 variables with no misstatements), each consisting of 5000 records with four digits each. Each of those records was simulated using the integer part of the result of the operation 1000×10^{r}, where r is a uniformly distributed random number between zero and one (Hill, 1995).

Each data set was afterwards contaminated under five different levels of intensity: modification of 2% of its records (100 records), 5% (250 records), 10% (500 records), 20% (1000 records) and 40% (2000 records). This resulted in a total of 150 data populations.

The records that were modified reflect four types of real-world data misstatements: rounding, psychological thresholds or internal authorization limits, number invention and duplications. In detail, in each data population, the records that were modified are divided in four equally sized groups as follows:

Group 1: records rounded to the nearest multiple of 100 (for example, the records 1534 and 1457 were modified to 1500);

Group 2: records modified to reflect psychological thresholds or internal authorization limits (the last two digits were replaced by 99 and the second digit was decreased by one unit, so that, for example, the record 1831 was modified to 1799; of course, other digit combinations could be defined to reflect psychological thresholds or internal authorization limits);

Group 3: records modified by number invention (substituting the original records by four-digit numbers simulated from the uniform distribution);

Group 4: records modified by replacing the original records by the multiple use of one of them, taken randomly.

In the application of AP3, we considered the conformity tests, test statistics and significance levels as in the experiment in Gomes da Silva and Carreira (2013), considering as audit targets the records identified in the solutions of the model, extending however their experiment with an additional contamination level (5%).

The main reason for using simulated data is to know which records are indeed misstated, i.e. to know the true status of each record, which allows to evaluate and compare the effectiveness of the stratification rules in identifying misstated data (with real data, knowing the true status of each record would be more difficult) and, consequently, to assess the precision of the proposed estimator for p by computing $Var(p_{strat})$.

Finally, for all populations, proportional allocation ($\pi_1=\pi_2=\pi$) was considered to define the sample size in each of the stratums, arbitrarily defining $\pi=0.1$ in all cases. Concerning sample sizes, other more sophisticated allocation methods, such as optimal allocation, could have been considered. Nevertheless, in general they require, as input, information about the true variances within the stratums that is not known in practice. Moreover, proportional allocation is probably the best allocation method for increasing precision if the variances are more or less equal across all the stratums (Lohr, 2010).

Results

The obtained results are presented in Tables 2 and 3 and in Figures 1, 2, 3 and 4. Tables 2 and 3 contain, for each stratification rule, the average values of N_1, N_2, C_1, C_2, percentage of false positive errors (% FP=$100(N_1-C_1)/N_1$), percentage of false negative errors (% FN=$100C_2/N_2$), percentage of errors (% Errors=$100(N_1-C_1+C_2)/N$), p_1, p_2 and Var(p_{strat}), computed for the thirty data sets within each contamination level. Due to the unbiasedness of the underlying estimator and to the size of the experiment, the value of p_{strat} is 0.02, 0.05, 0.1, 0.2 and 0.4 (the true value of p) in each contamination level, respectively, for any stratification rule. We thus omit it in the tables below.

Table 2: Results for low contamination levels

	Contamination 2%			Contamination 5%			Contamination 10%		
	AP1	AP2	AP3	AP1	AP2	AP3	AP1	AP2	AP3
N_1	615.8	609.4	33.6	722.2	715.7	130.9	929.5	926.7	323.2
N_2	4384.2	4390.6	4966.4	4277.8	4284.3	4869.1	4070.5	4073.3	4676.8
C_1	78.5	77.8	15.1	194.1	193.9	91.1	387.7	387.2	286.5
C_2	21.5	22.2	84.9	55.9	56.1	158.9	112.3	112.8	213.5
% FP	86.6	86.5	55.5	72.4	72.1	30.2	56.6	56.5	11.4
% FN	0.5	0.5	1.7	1.3	1.3	3.3	2.8	2.8	4.6
%	11.2	11.1	2.1	11.7	11.6	4.0	13.1	13.0	5.0
p_1	0.134	0.135	0.445	0.276	0.279	0.698	0.434	0.435	0.886
p_2	0.005	0.005	0.017	0.013	0.013	0.033	0.028	0.028	0.046
Var(p_{stra}	0.00003	0.00003	0.00003	0.00007	0.00007	0.00006	0.00012	0.00012	0.00008

Table 3: Results for large contamination levels

	Contamination 20%			Contamination 40%		
	AP1	AP2	AP3	AP1	AP2	AP3
N_1	1358.0	1355.0	703.7	2439.6	2413.1	1481.6
N_2	3642.0	3645.0	4296.3	2560.4	2586.9	3518.4
C_1	780.1	779.1	640.9	1593.1	1584.4	1404.2
C_2	219.9	220.9	359.1	406.9	415.6	595.8
% FP	41.6	41.6	8.9	34.6	34.2	5.2
% FN	6.0	6.1	8.4	15.9	16.1	16.7
% Errors	16.0	15.9	8.4	25.1	24.9	13.5
p_1	0.584	0.584	0.911	0.654	0.658	0.948
p_2	0.060	0.061	0.084	0.159	0.161	0.169
Var(p_{strat})	0.000203	0.000203	0.000143	0.000339	0.000339	0.000212

The average results show that AP3 behaves very differently than AP1 and AP2, which seem to produce similar results. First, N_1 is much smaller for AP3 in all contamination levels. This reveals that AP3 is more cautious when signaling a record as suspicious. Second, the performance of AP3 also differs significantly from AP1 and AP2 with respect to the correct identification of misstated and non-misstated records. Indeed, in all contamination levels, even though AP1 and AP2 have a lower percentage of false negatives, AP3 has a significantly lower percentage of false positives and a significantly lower overall percentage of errors.

In a more detailed analysis, Figures 1 and 2 display the boxplots of the distributions of the percentage of false positive and false negative errors, respectively, across the thirty data sets within each contamination level (the symbol circle corresponds to an outlier and the symbol star to a severe outlier).

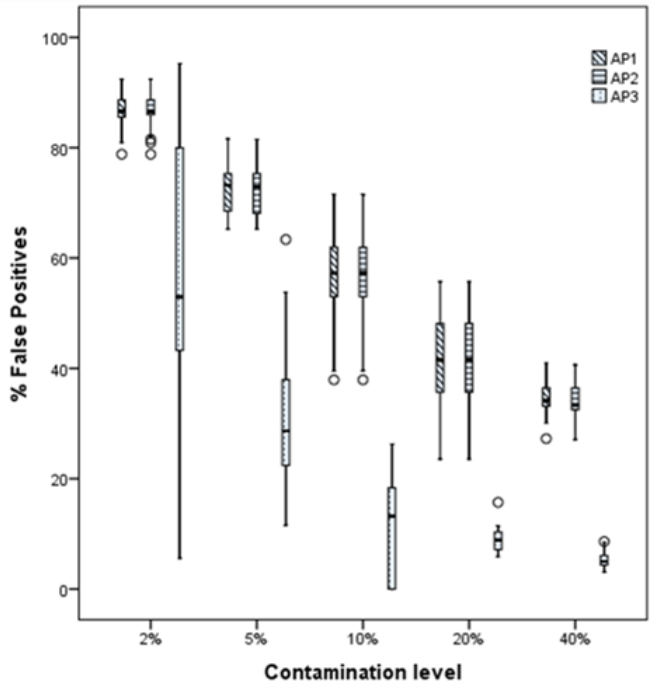

Figure 1: Distributions of the percentage of false positive errors, by stratification rule and contamination level.

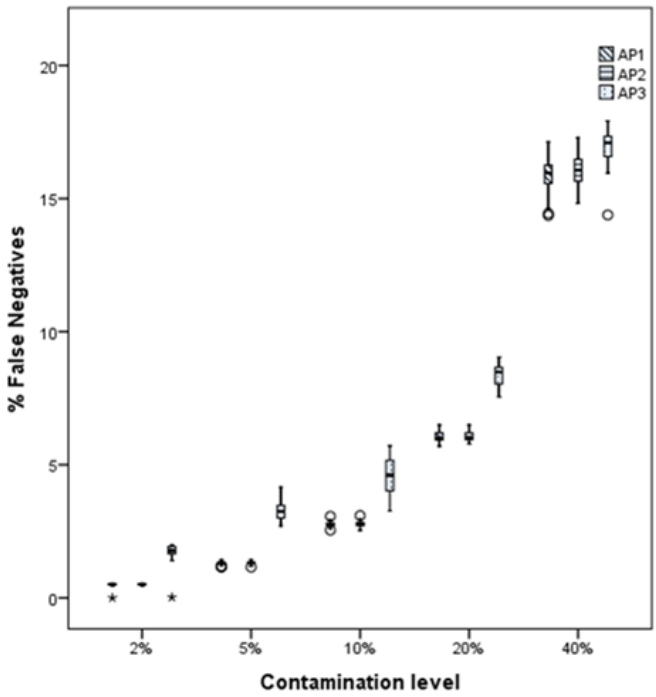

Figure 2: Distributions of the percentage of false negative errors, by stratification rule and contamination level.

In Figure 1, it can be observed that AP3 performs well better than the alternative rules in terms of percentage of false positive errors. Indeed, the worst performance of AP3 in terms of the percentage of false positive errors is in general better than the best performance of the alternative rules across all contamination levels. Furthermore, the dispersion of the values of the percentage of false positive errors under AP3 seems to decrease with the contamination level.

Concerning the percentage of false negative errors, Figure 2 shows that AP1 performs slightly better than AP2, and that both rules perform better than AP3, but, in this last case, with differences of less magnitude than the ones observed in Figure 1.

Globally, it can be concluded that AP3 performs better than the alternative rules, given that it leads to a lower percentage of errors in all contamination levels, which reflects that the lower percentage of false positive errors of AP3 more than compensates the fact that it captures a lower number of misstated records than the alternative rules. Figure 3 displays the distributions of the percentage of errors obtained under the three stratification rules

across the thirty data sets within each contamination level. The percentage of errors is almost always lower for AP3, revealing its increased effectiveness in signaling correctly the true nature of the records. Moreover, the relative gain of AP3 seems to be higher for low contamination levels, which are more frequent in real financial data. Note however that, by capturing a lower number of misstated records than AP1 and AP2 in its set of suspicious records, using AP3 may harm audit conclusions and the auditor's reputation if he indeed inspects only the records signaled as suspicious.

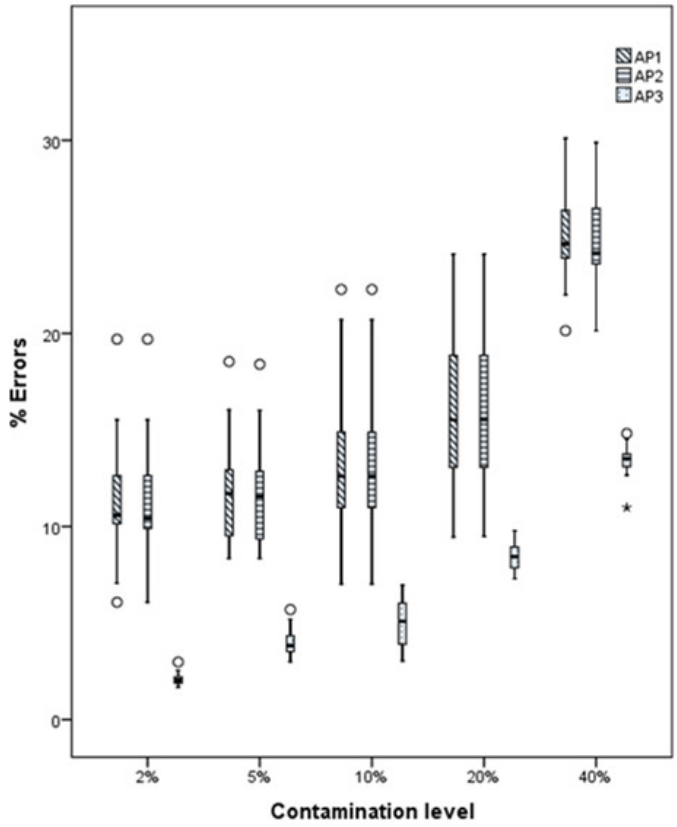

Figure 3: Distributions of the percentage of errors, by stratification rule and contamination level.

With respect to the main purpose of the paper, the results suggest that estimating the proportion of misstated records using AP3 as a stratification rule allows for increased precision in the estimation, which can be observed by the lower values of $Var(p_{strat})$ achieved under AP3 in Tables 2 and 3. Figure 4 displays the average variances of p_{strat} achieved within each contamination level, under each stratification rule (denoted in the figure as SS AP1, SS

AP2 and SS AP3, respectively). To have a baseline value of the variance for reference, we also compute it under simple random sampling (denoted in the figure as RS). In this case, an unbiased estimator for p is $\hat{p}$=c/n, where n is the number of randomly sampled records (assumed to be 0.1N here), and c is the number of misstated records found in the sample, with variance given by Var($\hat{p}$)=(p(1-p)/n)(N-n)/(N-1).

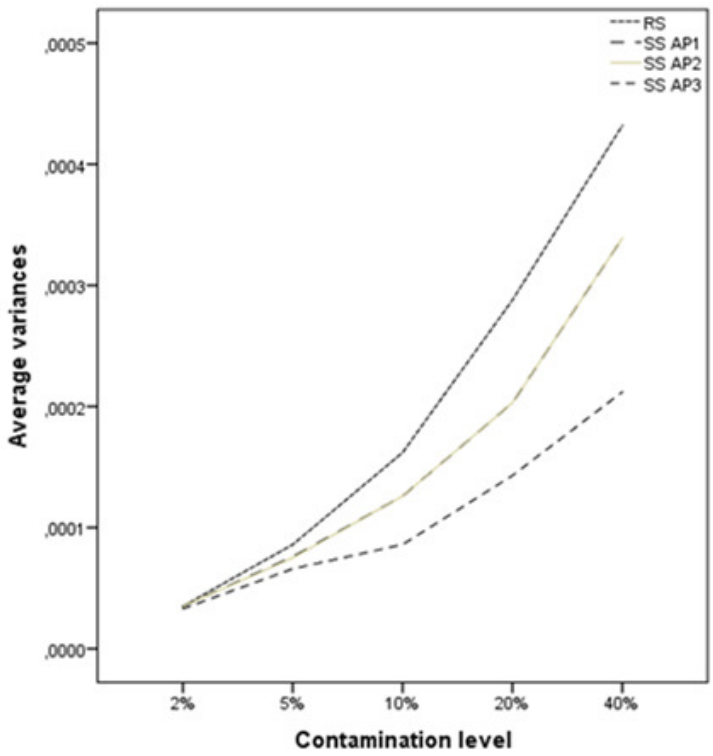

Figure 4: Average variances of the estimator for p by stratification rule and contamination level.

According to the figure, none of the stratification rules generates lower precision for p_{strat} than the simple random sampling approach, which means that, in general, the suggested rules are indeed able to capture useful information about the misstatements present in the data and, consequently, that stratified sampling based on them indeed allows to increase the precision of the estimators for the proportion of misstated records in a data set, as compared to simple random sampling. Additionally, it can be confirmed that stratified sampling under AP3 indeed leads to lower variances (higher precisions) for p_{strat} than under AP1 or AP2.

4. Conclusions

In this paper, we contributed to the issue of estimating the proportion of misstated records in a data set of numerical records, by suggesting a procedure based on stratified sampling and three possible stratification rules that can be employed, and by assessing the quality of the procedure under each of the rules. This study may assist auditors to form their overall conclusion about whether or not the financial statements of an audited entity are absent from material misstatements when it is convenient to use sampling.

As Benford's law is proved to be helpful in identifying misstated records in a numerical data set, we investigated it as a basis for stratifying the data. The empirical experiment showed that AP3, which uses uniquely but exhaustively the law, increases the precision of the proposed estimator relatively to AP1, AP2 and simple random sampling. Moreover, AP3 seems to generate higher precision when used in isolation than when combined with other rules. This is because AP3, when used separately, allows for a lower number of total classification errors, by making a better balance between false positive and false negative errors. Combining it with other rules appears to disturb that balance.

The proposed procedure can be embedded in audit software, contributing to enhance the effectiveness of audits and fraud detection, to provide useful information to the stakeholders of the audited entity and to enlarge the scope of Benford's law in such software.

The proposed procedure must however be employed only to audit the accounts or variables that are expected to follow Benford's law. Otherwise, the precision of the underlying estimator for p is expected to diminish. Naturally, even though difficult to execute in practice, the precision of the estimator for p is also expected to decrease if the data manipulations are done with the knowledge of Benford's law (so that the pattern of the digits is not affected).

As for future research, other stratification rules can be investigated so as to try to reduce even further the expected number of false positive and false negative errors. Also, the set of conformity tests and test statistics can be extended to prevent some limitations on the use of Benford's law in auditing (Cho and Gaines, 2007; Barney and Schulzke, 2016; Goodman, 2016), which could improve even further the performance of the stratification rule suggested in the present paper. Indeed, there is a growing concern with the excess of false positives that results from the commonly used Chi-square and Z-statistics and Mean Absolute Deviation (MAD), and with the consequent waste of time and resources required to

inspect the respective records, and some new test statistics have been suggested (Cho and Gaines, 2007; Barney and Schulzke, 2016). Moreover, the estimation of p could be made using also information about other available attributes of the records in a data set (other than their digits), leading to hybrid procedures that combine both supervised and unsupervised approaches.

References

Barney, B. and Schulzke, K. (2016), Moderating "Cry Wolf" Events with Excess MAD in Benford's Law, *Journal of Forensic Accounting Research*, 1, 1, pp. A66-A90.

Benford, F. (1938), The law of anomalous numbers, *Proceedings of the American Philosophical Society*, 78, 4, pp. 551-572.

Bolton, R. and D. Hand (2002), Statistical fraud detection: a review, *Statistical Science*, 17, 3, pp. 235-255.

Carslaw, C. (1988), Anomalies in income numbers: Evidence of goal oriented behavior, *The Accounting Review*, 63, 2, pp. 321-327.

Cho, W. and Gaines, B. (2007), Breaking the (Benford) law, *The American Statistician*, 61, 3, pp. 218-223.

Christensen, B., Elder, R. and Glover, S. (2015), Behind the numbers: insigths into large audit firm sampling policies, *Accounting Horizons*, 29, 1, pp. 61-81 (doi:10.2308/acch-50921).

Danuescu, T., Anca-Oanab, C. (2012), Opportunity and necessity in audit sampling: non-statistical sampling method, *Procedia Economics and Finance*, 3, pp. 1128 -1133.

Durtschi, C., Hillison, W. and Pacini, C. (2004), The effective use of Benford's law to assist in detecting fraud in accounting data, *Journal of Forensic Accounting*, 5, pp. 17-34.

Elder, R., Akresh, A., Glover, S., Higgs, J. and Liljegren, J. (2013), Audit sampling research: a synthesis and implications for future research, *Auditing: A Journal of Practice & Theory*, 32, 1, pp. 99-129.

Goodman, W. (2016), The promises and pitfalls of Benford's law, *Significance*, pp. 38-41 (doi:10.1111/j.1740-9713.2016.00919.x).

Gomes da Silva, C. and Carreira, P. (2013), Selecting audit samples using Benford's law, *Auditing: A Journal of Practice & Theory*, 32, 2, pp. 53-65 (doi:10.2308/ajpt-50340).

Hill, T. (1995), A statistical derivation of the significant-digit law, *Statistical Science*, 10, 4, pp. 354-363.

Johnson, G. and Weggenmann, J. (2013), Exploratory research applying Benford's law to selected balances in the financial statements of state governments, *Academy of Accounting and Financial Studies*, 17, 3, pp. 31-43.

Lombardi, D., Bloch, R., Vasarhelyi, M. (2014), The future of auditing, *Journal of Information Systems and Technology Management*, 11, 1, pp. 21-32 (doi:10.4301/S1807-17752014000100002).

Lohr, S. (2010), Sampling: Design and Analysis (2nd edition), Brooks/Cole, Cengage Learning, Boston.

Newcomb, S. (1881), Note of the frequency of use of the different digits in natural numbers, *American Journal of Mathematics*, 4, pp. 39-40.

Nigrini, M. (1994), Using digital frequencies to detect fraud, *The White Paper*, April, pp. 3-6.

Nigrini, M. (1999), I've got your number, *Journal of Accountancy*, 187, 5, pp. 79-83.

Nigrini, M. and Mittermaier, L. (1997), The use of Benford's law as an aid in analytical procedures, *Auditing: A Journal of Practice & Theory*, 16, 2, pp. 52-67.