Instituto de Engenharia de Sistemas e Computadores de Coimbra

Institute of Systems Engineering and Computers

INESC - Coimbra

Carlos Gomes da Silva and Pedro M.R. Carreira

Selecting Audit Targets Using Benford's Law

No. 8 2011

ISSN: 1645-2631

Instituto de Engenharia de Sistemas e Computadores de Coimbra

INESC - Coimbra

Rua Antero de Quental, 199; 3000-033 Coimbra; Portugal

www.inescc.pt

Selecting Audit Targets Using Benford's Law

Carlos Gomes da Silva* Pedro M. R. Carreira†

May 2011

Abstract

We provide a contribution to digital auditing and financial fraud detection by developing two general mathematical programming models that can help auditors selecting audit targets more promisingly, using Benford's Law. One model highlights the k most suspicious records in a data set and the other identifies the subset of nonconforming records. The models take into account several conformity tests and test statistics in simultaneous. Also, we solve some particular cases of such models for a set of simulated data and provide some insights about the relation between the required computational time and the initial characteristics of the data set, its number of records, conformity tests considered, and test statistics used.

Keywords: Digital Auditing, Benford's Law, Fraud Detection, Mathematical Programming.

*School of Technology and Management - Polytechnic Institute of Leiria, Campus 2, Morro do Lena, Alto do Vieiro, Apartado 4163, 2411-901 Leiria, Portugal, email: cgsilva@ipleiria.pt, and INESC Coimbra, R. Antero de Quental, 3000-033 Coimbra, Portugal.

†School of Technology and Management and CIGS - Polytechnic Institute of Leiria, Campus 2, Morro do Lena, Alto do Vieiro, Apartado 4163, 2411-901 Leiria, Portugal, email: pedro.carreira@ipleiria.pt.

1 Introduction

Over the last years, auditing standards have been recommending in an increasing trend the use of technology-driven or analytical procedures during the planning stage of an audit, commonly known and aggregated under the term of Digital Analysis. Current auditing standards suggest that Digital Analysis may allow an auditor to identify irregularities in accounting data sets, and to detect fraud symptoms more easily. In particular, it can be used to highlight suspicious transactions, accounts, events or trends, over which to address direct auditing procedures (Durtschi et al., 2004).

Digital Analysis is not abnormally new, though. Many basic digital analysis techniques are in use for some time. For example, Coderre (1999) suggests techniques such as the search for duplicate transactions, the search for even amounts (that may have been rounded up), and ratio analysis, comprising the ratio of the highest value to the lowest (maximum/minimum), the ratio of the highest value to the next highest (maximum/second highest) and the ratio of the current year to the previous year. Naturally, the accounts with larger ratio values are the ones for which direct auditing procedures are recommended.

Coderre (2009) presents the main basic digital analysis techniques and provides a good guide for the implementation of such techniques under the auditing software ACL. Beyond the ones referred in the last paragraph, techniques as the search for blank/absent records, cross-tabulations, the search for duplicates (e.g. same direct deposit number, but different employee number; same purchase order number, but different vendor number; same invoice number, amount, and date, but different vendor number; same employee number, but different work department), the search for gaps (e.g. missing accounts receivable payments, purchase orders not recorded, branch offices not reporting revenues, receipts missing for a given day, missing cash register tapes, water or electricity meters readings not recorded), and the search for exceptions to normal values (e.g. large-dollar transactions, negative quantities in inventories, payments to unusual vendors, purchases on weekends, purchases for exact dollar amounts, abnormally high frequencies of some numbers), are often used by many auditors.

Even though extensively adopted with respect to its basic tools, more complex and powerful Digital Auditing techniques are rarely used, especially in a consistent manner (Akresh et al., 1998). One of these more powerful and less often adopted Digital Analysis technique comes from Benford's Law (Benford, 1938), also known as the law of natural numbers or the law of significant digits. Having also in mind that recent auditing standards have been chalenging professionals and academics to develop new and more effective Digital Auditing techniques (Nigrini, 2000), our commitment in this paper is to enhance the use of Benford's Law by auditors in the planning stage of an audit, by proposing a new and more productive way of using it.

Indeed, we argue that there is still much to be done in using Digital Analysis and Benford's Law to detect irregularities in the data and to fight financial fraud. In particular, there is still much to be done on how to use Benford's Law to select audit targets. In its current use, even when adopting an integrative approach that uses multiple conformity tests and selects the set of records that fail all those tests, auditors often remain with unmanageable large audit target sets that are not compatible with the typical constraints they face (budget and time). Consequently, in order to shrink such target sets, auditors usually end up selecting randomly, from the initial (large) target set, the records over which to address direct auditing procedures. Somehow, this implies both a return to the traditional approach to audit planning, and the existence of some room for further increases in the contribution of analytical methods to identify audit targets. More problematic, the current integrative approach does not consider the interdependence between conformity tests, by not taking into account the impact of removing a certain record in the different tests in simultaneous, assuming instead the impact in each test in isolation. In the end, more narrow and consistent criteria seem to be required so that auditors can use Benford's Law in a non-arbitrary and effective way.

We contribute to this by developing two mathematical programming models that help auditors selecting, in a data set, a subset of audit targets, i.e. the records for which direct auditing procedures are addressed. The first model reveals the subset of the k most nonconforming records, where k is chosen by the auditor given his restrictions, while the second model identifies the smallest subset of nonconforming records. This specific and valuable information is new to literature and constitutes the main purpose of our paper.

The organization of the paper is as follows. In section 2, we give a brief overview of Benford's Law and its most common conformity tests and test statistics. In section 3, we formulate two general mathematical programming models that allow to identify each of the target subsets, and we solve them for two particular cases. In section 4, we run an experiment for a set of simulated data and provide some insights about the computational time required to solve such problems depending on the characteristics of the initial data set, its number of records, conformity tests considered and test statistics used. Section 5 concludes the paper.

2 Benford's Law and Conformity Tests

Broadly speaking, Benford's Law states that there are more numbers starting with the numeral one than with the numeral two, more numbers starting with the numeral two than with the numeral three, and so on. The theory was first described by Newcomb (1881), who observed that books of logarithms were considerably more worn in the first pages, which correspond to low numbers, than in the last pages, which correspond to high numbers, in a continuous way. He concluded that researchers looked more frequently for the logarithms of numbers with first digit 1 than with first digit 2, with first digit 2 than with first digit 3, and so on. Hence, there would exist more numbers starting with numeral 1 than with larger numbers.

The theory became known as Benford's Law more than 50 years later due to the seminal work of Benford (1938), in which he gathered more than 20 000 observations of data from many different areas and showed that the observed first digits offer a remarkably good fit to the logarithmic distribution.

Let D_1 be the first digit of a number. According to Benford's Law, the probability P that a number has first digit $i_1 = 1, ..., 9$ is given by

$$P\left(D_1 = i_1\right) = \log_{10}\left(1 + \frac{1}{i_1}\right) \tag{1}$$

Likewise, let $D_1 D_2$ be the first two digits of a number. The probability P that a number has first two digits $i_1 i_2 = 10, ..., 99$ is given by

$$P\left(D_1 D_2 = i_1 i_2\right) = \log_{10}\left(1 + \frac{1}{i_1 i_2}\right) \tag{2}$$

From this, we can obtain the probability for the second digit of a number. Let D_2 be the second digit of a number. The probability P that a number has second digit $i_2 = 0, ..., 9$ is given by

$$P\left(D_2 = i_2\right) = \sum_{i_1=1}^{9} \log_{10}\left(1 + \frac{1}{i_1 i_2}\right) \tag{3}$$

Using the formulas, one can observe that approximately 30.1% of the numbers have first digit one, while only 4.6% of the numbers have first digit nine. For the second digit, this gap is not so evident, though. Indeed, approximately 12% of the numbers have second digit zero, 11.4% have second digit one and 8.5% have second digit nine. Actually, once we increase the position of the digit in a number, the gap converges to zero. In the limit, it is usual to assume uniform distributions for the last and last-two digits of a number.

Of course, the Law does not apply to all lists of numbers. According to Nigrini and Mittermaier (1997), while the Law applies well to lists of numbers that describe the relative sizes of similar phenomena, such as market values, net incomes, daily stock trading volumes, transaction amounts, populations, etc., the Law is not likely to apply to data sets featuring minimum or maximum cut-off points (such as a top 100 revenue ranking of firms), numbers influenced by human thought (such as ATM cash withdrawals' amounts) or assigned numbers (such as purchase orders, personal identification numbers, telephone numbers, or car license plate numbers).

In some of the theoretical approaches to Benford's Law, Hill (1995) showed that "if distributions are selected at random (in any "unbiased" way) and random samples are then taken from each of these distributions, the significant digits of the combined sample will converge to the logarithmic distribution" (Benford distribution), which means that Benford's Law can be viewed as a law of true randomness of numbers. Also, Boyle (1994) showed that a list of numbers obtained by multiplying, dividing, or raising to integer powers numbers taken from random variables, converges to the Benford distribution.

It is unquestionable that most accounting and financial numbers are indeed result of multiplying or dividing mathematical operations, such as transaction amounts, corporate net incomes, individual taxable incomes, stock prices or quantity volumes. Applications of Benford's Law to auditing and accounting began in the late 1980s. Carlslaw (1988) found that, for a sample of New Zealand companies, reported net incomes revealed excess of second digits 0 and few second digits 9, which supported his theory that managers tend to round up values near psychological boundaries. Nigrini (1994) was apparently the first researcher proposing Benford's Law to assist in fraud detection. He admitted that, if individuals invent numbers, the numbers would not conform to Benford distribution. This assumption was supported later on by many experimental studies, such as the ones from Diekmann (2007), who found that auditors should focus more on the last than on the first digits though, and from Watrin et al. (2008), who found that individuals cannot adapt sufficiently to Benford's Law when inventing numbers, even when they are educated to do so. However, they also alert auditors for the need to engage in the critical step of ensuring that unmanipulated data that respects to the audit target indeed follows Benford's Law. In this particular, Durtschi et al. (2004) provide a good summary of conforming and nonconforming accounting-related data sets.

The natural question that arises next is which auditing tests to perform.

2.1 Conformity Tests

The most common conformity tests that can be used to verify whether or not a particular data set of records conforms with Benford's Law were mainly suggested by Nigrini and Mittermaier (1997).

Let $f_i(T, N)$ be the observed relative frequency of a particular digit (or digits) i in the context of conformity test T in a set of N records, and let $e_i(T)$ be the respective (expected) Benford probability as presented in the previous section.

Conformity Test 1 ($T = 1$)**: First Digits Test** Compares f_i with e_i for the first digit of numbers, i.e. for $i = i_1$, $i_1 = 1, ..., 9$.

Conformity Test 2 ($T = 2$)**: Second Digits Test** Compares f_i with e_i for the second digit of numbers, i.e. for $i = i_2$, $i_2 = 0, 1, ..., 9$.

Conformity Test 3 ($T = 3$)**: First-Two Digits Test** Compares f_i with e_i for the first two digits of numbers, i.e. for $i = i_1 i_2$, $i_1 i_2 = 10, 11, ..., 99$.

Conformity Test 4 ($T = 4$)**: Last Digit Test** Compares f_i with e_i for the last digit of numbers, i.e. for $i = i_k$, $i_k = 0, 1, ..., 9$, with k denoting the last digit.

Conformity Test 5 ($T = 5$)**: Last-Two Digits Test** Compares f_i with e_i for the last two digits of numbers, i.e. for $i = i_{k-1} i_k$, $i_{k-1} i_k = 00, 01, ..., 99$.

2.2 Test Statistics

Each of these tests can be operated under a collective statistic in the sense that all the relevant frequency deviations are joined in a single statistic, or under individual statistics. The most widely adopted collective statistics are the Chi-Square statistic and the Mean Absolute Deviation (MAD), while the most common individual statistic is the Z-statistic.

2.2.1 Collective Statistics

Chi-Square Statistic The Chi-Square statistic is given by

$$S_1(T, N) = N \sum_i \frac{[f_i(T, N) - e_i(T)]^2}{e_i(T)} \sim \chi^2_{[n(T)-1]} \tag{4}$$

where $n(T)$ is the number of feasible digits (classes) i in the context of conformity test T, i.e. $n(1) = 9$, $n(2) = n(4) = 10$, $n(3) = 90$ and $n(5) = 100$.

The null hypothesis corresponds to conformity with Benford's Law.

Mean Absolute Deviation (MAD) The Mean Absolute Deviation (MAD) statistic is given by

$$S_2(T, N) = \frac{\sum_i |f_i(T, N) - e_i(T)|}{n(T)} \quad (5)$$

Unlike the Chi-Square, this statistic does not follow any well-known distribution. Drake and Nigrini (2000) defined however some critical values that can be used to conclude about conformity for the first digits, first-two digits and second digits tests, resumed as follows.

MAD	*Nonconformity*
$T = 1$	> 0.012
$T = 2$	> 0.016
$T = 3$	> 0.0018

2.2.2 Individual Statistic

Z-Statistic The Z-statistic is an individual test statistic in the sense it allows to test whether or not the deviation of a particular feasible digit(s) i from the respective Benford probability, say $i = i_{12} = 99$ in the first-two digits test for example, is significant. The Z-statistic can be written as

$$Z_i(T, N) = \frac{f_i(T, N) - e_i(T)}{\sqrt{\frac{e_i(T)[1-e_i(T)]}{N}}} \sim N(0, 1) \quad (6)$$

The null hypothesis here corresponds to the conformity of digit(s) i with Benford's Law.

In any conformity test T operated under the Z-statistic, the auditor may be interested in calculating up to $n(T)$ Z-statistic values, one per each feasible digit(s) i.

This statistic could be written in an adjusted form with a continuity correction term in the numerator, as was presented by Nigrini and Mittermaier (1997).

3 Models for Selecting Audit Targets

We introduce now the mathematical programming models that should help auditors to identify audit targets. As we have seen before, we will formulate two different models. In the first model, the problem is to identify the k most suspicious records in a data set. In the second model, the problem is to determine the smallest subset of records that must be removed from the data set in order to achieve conformity, or to improve the level of conformity.

3.1 The K Most Suspicious Records

Suppose that we want to use conformity test T and collective statistic S_1 in order to evaluate the conformity with Benfords's Law of a certain set of records (there is *nonconformity* if the value of S_1 is greater than a critical value or pre-defined threshold $S_1^*(T)$). Here, we want to identify k records that, when removed from the data set, generate the highest improvement in statistic S_1. The test statistic to use is thus

$$S_1(T, N-k) = (N-k)\sum_i \frac{\left[\frac{a_i(T)-l_i(T)}{N-k} - e_i(T)\right]^2}{e_i(T)} \sim \chi^2_{[n(T)-1]} \tag{7}$$

where $a_i(T)$ is the initial (before the removal) number of records with digit(s) i in the context of conformity test T, and $l_i(T)$ is the number of records with digit(s) i removed from the data set. The ratio $[a_i(T) - l_i(T)]$ / $(N-k)$ is thus the relative frequency of digit(s) i after the removal of the k records in the context of conformity test T.

Now, the k records that affect statistic S_1 by the most, which we name the k most suspicious records, can be identified by solving the following mathematical integer and nonlinear programming model:

$$\begin{aligned}
&Min\ z = S_1(T, N-k) \\
&s.t. \\
&k = \sum_i l_i(T) \\
&l_i(T) \leq a_i(T),\ \forall i \\
&l_i(T) \geq 0 \text{ and integer},\ \forall i
\end{aligned} \tag{8}$$

The objective function in problem (8) consists on the minimization of the value of statistic S_1, whereas the first constraint limits the number of records to be removed (audited), the second set of constraints is of coherence since no more than the initial number of records can be removed, and the last set of constraints imposes integrality on the solution.

Note that if $k \leq a_i(T), \forall i$, then the set of constraints $l_i(T) \leq a_i(T), \forall i$ can be omited.

Example 1 The 3 500 weekly sales records of a company have the following distribution for the first digit:

i	1	2	3	4	5	6	7	8	9
a_i	946	583	437	352	297	258	230	207	190
f_i	0.27	0.167	0.125	0.101	0.085	0.074	0.066	0.059	0.054

The initial value for the statistic S_1 is 30.9, which means that the conformity of the data set with Benford's Law for a significance level of 5% is rejected.

Suppose now that the auditor wants to select the 20 most suspicious records to audit in detail. These records can be found by solving problem (8), which simplifies as follows:

$$Min\ z = (3500 - 20) \left[\begin{array}{c} \frac{\left(\frac{946 - l_1(1)}{3480} - 0.301\right)^2}{0.301} + \frac{\left(\frac{538 - l_2(1)}{3480} - 0.176\right)^2}{0.176} + \\ \frac{\left(\frac{437 - l_3(1)}{3480} - 0.125\right)^2}{0.125} + \frac{\left(\frac{352 - l_4(1)}{3480} - 0.097\right)^2}{0.097} + \\ \frac{\left(\frac{297 - l_5(1)}{3480} - 0.079\right)^2}{0.079} + \\ \frac{\left(\frac{258 - l_6(1)}{3480} - 0.067\right)^2}{0.067} + \frac{\left(\frac{230 - l_7(1)}{3480} - 0.058\right)^2}{0.058} + \\ \frac{\left(\frac{207 - l_8(1)}{3480} - 0.051\right)^2}{0.051} + \frac{\left(\frac{190 - l_9(1)}{3480} - 0.046\right)^2}{0.046} \end{array} \right]$$

$$s.t.$$
$$\sum_i l_i(1) = 20,\ i = 1, ..., 9$$
$$l_i(1) \geq 0 \text{ and integer},\ i = 1, ..., 9$$

Solving the problem, the following optimal values $l_i^*(1)$ are obtained:

i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$l_i^*(1)$	0	0	0	0	0	0	3	6	11

The minimum value found for statistic S_1 is 24.88, lower than the original value, but still revealing the presence of nonconformity. ◄

The problem of verifying conformity can be conducted by analyzing more than the first digit alone, though. As discussed in section 2, the distributions of the second, first-two, last and last-two digits, for example, are also relevant. Moreover, we can use either collective or individual test statistics, or both in simultaneous.

Naturally, this generalization increases the complexity of problem (8). Indeed, when a record is removed, more than one absolute frequency is affected. For instance, if the record 1256 is removed from a given data set, the frequencies of first digit 1, second digit 2, first-two digits 12, and so on, are affected. Hence, each particular record is likely to have a different impact in the problem.

We generalize problem (8) in order to take into account:
1) the consideration of multiple conformity tests;
2) the use of multiple collective statistics for each conformity test;
3) the use of multiple individual statistics for each conformity test.

The main issue is to determine whether or not a specific record t, from the initial set of N records, is to be removed. For this, we introduce the following binary variables:

$$y_t = \begin{cases} 1, \text{ if record } t \text{ is removed} \\ \qquad otherwise \end{cases}, \; t = 1, ..., N$$

Note that, for all the previously introduced test statistics, a commom term is $f_i(T, N) - e_i(T)$. When k records are removed from a data set, thus remaining $N-k$ records, this term becomes given by $\frac{a_i(T)-l_i(T)}{N-k} - e_i(T)$, with $l_i(T) = \sum_{t=1}^{N} h_{it}(T)\, y_t$, $a_i(T) = \sum_{t=1}^{N} h_{it}(T)$, $k = \sum_{t=1}^{N} y_t$ and $h_{it}(T) = 1$ if record t has digit(s) i in the context of conformity test T, and $h_{it}(T) = 0$ otherwise.

For example, consider conformity tests $T = 1$, $T = 2$, $T = 3$ and $T = 5$, and the set of records $\{1256, 4567, 1457\}$. Then, $h_{11}(1) = h_{42}(1) = h_{13}(1) = h_{21}(2) = h_{52}(2) = h_{43}(2) = h_{12,1}(3) = h_{45,2}(3) = h_{14,3}(3) = h_{56,1}(4) = h_{67,2}(4) = h_{57,3}(4) = 1$, while all other $h_{it}(T)$ are 0.

Moreover, when considering multiple test statistics and removing a set of records from the data set, one can be faced with a trade-off between the values of the statistics, in the sense that improving the value of one statistic may imply deteriorating the value of others.

A reasonable strategy to deal with this consists on minimizing the maximum relative deviation between the value of each test statistic and its respective defined threshold or critical value ($S_j^*(T)$ or Z^*), i.e. minimizing the worst-case relative deviation. With this strategy, the general mathematical problem to identify the k most suspicious records in a data set can be formulated as follows (Model 1).

$$\begin{array}{l} Min\ \theta \\ s.t. \\ \theta \geq \frac{S_j(T,N-k)-S_j^*(T)}{S_j^*(T)};\ j=1,2;\ T=1,...,5 \\ \theta \geq \frac{|Z_i(T,N-k)|-Z^*}{Z^*};\ \forall i\ ;\ T=1,...,5 \\ \sum_{t=1}^{N} y_t = k \\ y_t \in \{0,1\},\ t=1,...,N \end{array} \tag{9}$$

Note that the notation in the model is very compact for convenience of presentation, but in order to define $S_j(T, N-k)$ and $Z_i(T, N-k)$ the use of parameters $h_{it}(T)$ is required. Also, note that one is obviously not obliged to use all possible test statistics, neither all possible conformity tests and all feasible digit(s) i in the individual test statistic Z. In such cases, the number of constraints becomes smaller.

An alternative strategy to deal with the trade-off between the multiple test statistics could be minimizing the value of one statistic, imposing conformity constraints on all other. However, in this case, a leading test statistic must be chosen, which can be subjective.

Of course, the k most suspicious records to audit in detail are the ones for which $y_t = 1$ in the solution.

3.2 The Smallest Nonconformity Set

Another problem consists on determining the smallest nonconformity set of records in the sense that, if it is removed from the initial data set, conformity in the data set is achieved (or improved).

This problem can also be formulated as a mathematical programming model, where the objective function is the size of the set of records to be removed from the data set, and there are additional constraints for the desirable value (usually, the critical value) of the statistics used.

As in the beginning of subsection 3.1, assume conformity test T, test statistic S_1, and critical value $S_1^*(T)$. The mathematical programming problem can thus be written as follows.

$$
\begin{aligned}
&Min\ z = k \\
&s.t. \\
&(N-k)\sum_i \frac{\left[\frac{a_i(T)-l_i(T)}{N-k}-e_i(T)\right]^2}{e_i(T)} \leq S_1^*(T) \\
&k = \sum_i l_i(T) \\
&l_i(T) \geq 0 \text{ and integer, } \forall i
\end{aligned}
\tag{10}
$$

Example 2 Using the data from Example 1, assuming $T = 1$ and test statistic S_1, imposing $S_1^*(1) = 15.51$, and replacing the values $a_i(1)$, $e_i(1)$, N and $S_1^*(1)$, the simplified program is

$$
\begin{aligned}
&Min\ z = k \\
&s.t. \\
&(3500-20)\left[
\begin{array}{c}
\frac{\left(\frac{946-l_1(1)}{3480}-0.301\right)^2}{0.301} + \frac{\left(\frac{538-l_2(1)}{3480}-0.176\right)^2}{0.176} + \\
\frac{\left(\frac{437-l_3(1)}{3480}-0.125\right)^2}{0.125} + \frac{\left(\frac{352-l_4(1)}{3480}-0.097\right)^2}{0.097} + \\
\frac{\left(\frac{297-l_5(1)}{3480}-0.079\right)^2}{0.079} + \\
\frac{\left(\frac{258-l_6(1)}{3480}-0.067\right)^2}{0.067} + \frac{\left(\frac{230-l_7(1)}{3480}-0.058\right)^2}{0.058} + \\
\frac{\left(\frac{207-l_8(1)}{3480}-0.051\right)^2}{0.051} + \frac{\left(\frac{190-l_9(1)}{3480}-0.046\right)^2}{0.046}
\end{array}
\right] \leq 15.51 \\
&k = \sum_i l_i(1) \\
&l_i(1) \geq 0 \text{ and integer, } i = 1, ..., 9
\end{aligned}
$$

with solution given by

i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$l_i^*(1)$	0	0	0	0	4	11	16	18	21

When this set of records is removed from the data set, the value of S_1 becomes 15.42 and conformity is achieved. A total of 70 records must be removed in order to satisfy the desirable value for test statistic S_1. Interpreting the solution, the auditor should thus select 4 records starting with digit 5, 11 starting with digit 6, 16 starting with digit 7, 18 starting with digit 8 and 21 starting with digit 9, and audit them in detail. ◀

Similarly to problem (8), problem (10) can be generalized in order to consider multiple conformity tests and multiple collective and individual statistics for each conformity test. The generalization can be written as follows (Model 2).

$$
\begin{array}{l}
Min\ z = k \\
s.t. \\
S_j\left(T, N-k\right) \leq S_j^*\left(T\right);\ j = 1,2;\ T = 1,...,5 \\
\left|Z_i\left(T, N-k\right)\right| \leq Z^*;\ \forall i\ ;\ T = 1,...,5 \\
k = \sum_{t=1}^{N} y_t \\
y_t \in \{0,1\},\ t = 1,...,N
\end{array}
\quad (11)
$$

Again, parameters $h_{it}\left(T\right)$ are required to define $S_j\left(T, N-k\right)$ and Z_i $\left(T, N-k\right)$, and one is not obliged to use all possible test statistics, i.e. one may be interested in imposing only some of the constraints. Also, the records that should be audited in detail are again the the ones for which $y_t = 1$ in the solution.

4 Solving the Models

Once Models 1 and 2 are built, the remaining issues relate with solving them. In this section, we run an experiment over simulated data, and we illustrate some particular features of the resolution process for both Models.

As presented in equations (9) and (11), the models have some mathematical difficulties. They depend on the number of binary $\{0,1\}$ variables and the objective function and the constraints can be nonlinear.

The models were tested using simulated data sets of 1 000, 3 000, 5 000, 7 000 and 10 000 records of five digits numbers. In each simulated data set, 50% of the numbers were generated according to the Benford's distribution and the other 50% were randomly picked from a uniform distribution between 10000 and 99999.

In order to specify the models, we defined the following cases:

Case 1: Consideration of single conformity tests, using single test statistics;

Case 2: Consideration of conformity tests $T = 1, 2, 3, 4, 5$ in simultaneous, using a single collective test statistic (Chi-square) for each test;

Case 3: Consideration of conformity tests $T = 1, 2, 3, 4, 5$ in simultaneous, using the collective test statistic Chi-square for all tests, the collective test statistic MAD for $T = 1, 2, 3$, and the individual test statistic Z for $i = 1, 9$ in $T = 1$, $i = 0$ in $T = 4$, and $i = 00$ for $T = 5$. With respect to the test statistics Chi-square and Z, the confidence level was set to 95%.

Additionally, in Model 1, k was set to 10% of N, i.e. we assume that the auditor wants to identify the 10% most suspicious records for further examination.

Table 1 presents the performance of the initial (simulated) data set, concerning conformity tests and test statistics considered in Cases 1, 2 and 3. It also contains the critical value for each test statistic for each conformity test. One can observe that there is a mix of conformity and nonconformity in the initial data set (values above critical values reveal nonconformity), with the most severe nonconformity occurring in conformity tests for the first and first-two digits. Furthermore, one can also observe that, for $T = 2$, there is a conflict in the conclusions that come from collective statistics S_1 and S_2.

In order to solve the integer nonlinear mathematical programming models, we used the NEOS platform (www.neos-server.org) and the *solver* BARON /GAMS. This *solver* has showed to have some advantages over other *solvers* (Neumaier, 2005). This platform/solver has a predefined limit of 1000 seconds for solving this type of models. Also, there is no guarantee by the *solver* that the solution found is a global optimum. For such cases, it is useful to have a quality measure of the solution. Therefore, we present the gap between the best possible solution (not necessarily feasible) and the obtained solution, together with the CPU time (in seconds) and the values achieved for the objective function and test statistics, in the results that follow. This information is obtained from the output of the *solver*.

N	Critical Value	**1000**	**3000**	**5000**	**7000**	**10000**
$S_1(1,N)$	15.507	72.155	310.549	470.924	833.300	986.900
$S_2(1,N)$	0.012	0.026	0.029	0.029	0.033	0.030
$Z_1(1,N)$	1.96	-5.448	-11.029	-15.636	-18.245	-20.281
$Z_9(1,N)$	1.96	3.517	10.374	9.489	12.567	15.622
$S_1(2,N)$	16.919	13.966	13.218	18.134	47.654	41.015
$S_2(2,N)$	0.016	0.011	0.006	0.004	0.007	0.005
$S_1(3,N)$	113.145	152.936	401.183	589.075	974.192	11106.001
$S_2(3,N)$	0.002	0.003	0.003	0.003	0.003	0.003
$S_1(4,N)$	16.919	4.760	8.887	3.924	14.386	15.528
$Z_0(4,N)$	1.96	1.160	1.217	0.519	-1.633	-1.167
$S_1(5,N)$	123.225	91.800	93.867	65.320	97.571	101.600
$Z_{00}(5,N)$	1.96	-0.636	-0.183	1.279	-1.081	-0.603

Table 1: Characteristics of the simulated data

4.1 Results for Model 1

Relatively to Case 1, we solved 12 different problems for each value of N. For $T = 1$, we used the Chi-square statistic, the MAD, and Z_1 and Z_9 statistics. For $T = 2$, we considered the Chi-square statistic and the MAD. For $T = 3$, we used the Chi-square statistic and the MAD. For $T = 4$, we considered the Chi-square statistic and Z_0. Finally, for $T = 5$, we assumed the Chi-square statistic and Z_{00}.

Table 2 presents a summary of the results for Case 1.

For most of the solved problems, the *solver* returned a solution within 1000 seconds. The unique exception is for conformity test $T = 5$ under the Chi-square test statistic.

Apparently, the use of Z statistics instead of collective statistics seem to decrease the CPU and the gap to the global optimum. By construction, Z statistics are directed to a single digit, shrinking the number of binary variables that are relevant for the objective function. For instance, if $N = 1000$, the 1000 binary variables considered by the MAD statistic are reduced to 222 relevant for $Z_1(1, 0.9N)$, i.e. there are 222 records with first digit 1.

Moreover, when applied to a conformity test with initial nonconformity (see Table 1), the MAD statistic is very time consuming, which suggests that it can be a challenge for a *solver* to stop the search for a solution when MAD is used in this case.

Tests	N	**1000**	**3000**	**5000**	**7000**	**10000**
$S_1\,(1, 0.9N)$	$CPU\,(s)$	1.060	8.600	20.620	35.710	72.880
	θ	0.210	4.910	9.690	18.720	19.926
	Gap	0.019	0.446	0.881	1.702	1.811
$S_2\,(1, 0.9N)$	$CPU\,(s)$	1000*	1000*	1000*	1000*	1000*
	θ	0.206	0.607	0.618	0.866	0.607
	Gap	0.495	0.646	0.654	0.694	0.629
$Z_1\,(1, 0.9N)$	$CPU\,(s)$	0.040	0.160	0.450	0.860	1.770
	θ	0.814	2.998	4.913	5.859	6.378
	Gap	0.000	0.000	0.000	0.000	0.000
$Z_9\,(1, 0.9N)$	$CPU\,(s)$	0.040	0.230	0.430	0.840	1.730
	θ	-0.985	-0.979	-0.997	-0.992	-0.995
	Gap	0.015	0.021	0.003	0.008	0.005
$S_1\,(2, 0.9N)$	$CPU\,(s)$	0.310	22.920	10.250	44.930	87.010
	θ	-0.912	-1.000	-0.991	-0.964	-1.000
	Gap	0.013	0.000	0.001	0.004	0.000
$S_2\,(2, 0.9N)$	$CPU\,(s)$	0.790	18.930	25.140	22.340	100.190
	θ	-0.8417	-0.991	-0.975	-0.959	-0.996
	Gap	0.079	0.009	0.014	0.021	0.004
$S_1\,(3, 0.9N)$	$CPU\,(s)$	18.730	1000*	1000*	473.590	1000*
	θ	-0.471	0.132	0.841	2.146	2.275
	Gap	0.027	0.076	0.089	0.184	0.241
$S_2\,(3, 0.9N)$	$CPU\,(s)$	1000*	1000*	1000*	1000*	1000*
	θ	0.353	0.170	0.188	0.355	NA
	Gap	0.537	0.473	0.497	0.526	NA
$S_1\,(4, 0.9N)$	$CPU\,(s)$	2.370	521.520	4.370	9.400	22.000
	θ	-0.963	-0.993	-1.000	-1.000	-1.000
	Gap	0.003	0.001	0.000	0.000	0.000
$S_1\,(5, 0.9N)$	$CPU\,(s)$	1000	1000	1000	221.660	1000
	θ	NA	NA	NA	-0.896	NA
	Gap	NA	NA	NA	0.012	NA
$Z_0\,(4, 0.9N)$	$CPU\,(s)$	0.040	0.170	0.420	0.790	-1.600
	θ	-1.000	-1.000	-1.000	-1.000	-1.000
	Gap	0.000	0.000	0.000	0.000	0.000
$Z_{00}\,(5, 0.9N)$	$CPU\,(s)$	0.030	0.140	0.380	0.730	1.520
	θ	-0.829	-1.000	-1.000	-0.871	-1.000
	Gap	0.000	0.000	0.000	0.000	0.000
NA - the solver did not return a solution in 1000 seconds; * - the solver was interrupted but a solution was returned.						

Table 2: Model 1 - Results for Case 1

N	**1000**	**3000**	**5000**	**7000**	**10000**
$CPU\,(s)$	13.180	52.690	161.680	366.580	401.460
θ	0.230	4.913	9.695	18.724	19.927
Gap	0.021	0.447	0.014	0.021	1.812
$S_1\,(1, 0.9N)$	19.071	91.692	165.842	305.869	324.519
$S_1\,(2, 0.9N)$	15.129	11.861	18.749	34.791	37.425
$S_1\,(3, 0.9N)$	105.335	188.168	280.184	420.268	440.317
$S_1\,(4, 0.9N)$	2.933	13.444	4.093	10.327	17.820
$S_1\,(5, 0.9N)$	80.000	99.481	64.133	89.365	117.911

Table 3: Model 1 - Results for Case 2

In its hand, the Chi-square statistic, while performing well for conformity tests with a low $n\,(T)$ (for $T = 1$, $T = 2$ and $T = 4$), the CPUs and gaps become larger (or even undetermined) for conformity tests with a high $n\,(T)$ ($T = 3$ and $T = 5$).

Note further that θ is greater than 0 in many problems. This means that removing 10% of the records from the data set is not sufficient to achieve conformity.

Concerning Case 2, for which the results are summarized in Table 3, one can see that the *solver* remains acceptable for practical usage. Indeed, all CPUs are small and all the problems were solved within the time limit. When comparing to Case 1, this can be seen as somehow counter intuitive, since there are more constraints here than in Case 1, where some problems were not solved. This may suggest that the *solver* benefits from the existence of certain constraints, reducing the search space and accelerating the achievement of its stopping conditions. Furthermore, the CPU is always increasing with N, but more than linearly.

Note also that conformity is never achieved when removing 10% of the records from the data set. The simultaneous consideration of several constraints implies a compromise where the worst relative deviation remains positive.

As the number of records N increases, the value of θ also increases. This reflects the increase of nonconformity with N, presented in the initial data (Table 1).

Finally, Table 4 presents results for Case 3.

N	**1000**	**3000**	**5000**	**7000**	**10000**
$CPU\,(s)$	6.150	80.450	180.700	311.950	660.040
θ	0.814	4.923	9.694	18.720	19.926
Gap	0.814	0.448	0.881	1.702	1.811
$S_1\,(1, 0.9N)$	23.380	91.846	165.830	305.803	324.505
$S_2\,(1, 0.9N)$	0.016	0.019	0.019	0.022	0.019
$Z_1\,(1, 0.9N)$	−3.555	−7.836	−11.590	−13.444	−14.460
$Z_9\,(1, 0.9N)$	1.407	2.252	3.303	4.626	4.549
$S_1\,(2, 0.9N)$	15.809	11.835	17.243	38.112	39.021
$S_2\,(2, 0.9N)$	0.012	0.006	0.004	0.006	0.006
$S_1\,(3, 0.9N)$	110.822	186.320	277.121	462.331	445.250
$S_2\,(3, 0.9N)$	0.003	0.002	0.002	0.003	0.002
$S_1\,(4, 0.9N)$	5.844	12.044	440.000	11.683	16.718
$Z_0\,(4, 0.9N)$	−1.444	2.181	0.596	0.630	2.003
$S_1\,(5, 0.9N)$	86.000	102.963	64.267	98.952	103.067
$Z_{00}\,(5, 0.9N)$	−0.670	0.387	1.049	−0.253	0.424

Table 4: Model 1 - Results for Case 3

Case 3 considers several conformity tests and several test statistics in simultaneous. However, results are very similar to the ones of Case 2. Indeed, all problems were solved in time, all θ are positive, and both the gaps and the θ increase with N. More precisely, the values of θ are of the same magnitude than in Case 2. This reveals that the relevant constraints to define θ are the same in both cases, which can be perceived by looking at the structure of the initial data (Table 1), and observing that the most severe sources of conformity are the ones that were already present in the test statistics used in Case 2 (Chi-square statistics).

4.2 Results for Model 2

We now solve Model 2 for the same Cases and data sets than in the previous section.

Table 5 presents a summary of the results for Case 1.

The main difference in Case 1 as opposed to Model 1 is that now the Chi-square statistic becomes more manageable even for the conformity tests with high $n\,(T)$ ($T = 3$ and $T = 5$). For the rest of the statistics, Z statistics

Tests	N	**1000**	**3000**	**5000**	**7000**	**10000**
$S_1\,(1, 0.9N)$	$CPU\,(s)$	1.520	9.360	23.290	51.400	93.810
	k	112	633	1325	1900	2576
	Gap	10.182	57.545	120.455	172.727	234.182
$S_2\,(1, 0.9N)$	$CPU\,(s)$	1000*	1000*	1000*	1000*	1000*
	k	124	509	960	1396	1697
	Gap	51.209	153.875	436.310	352.470	495.779
$Z_1\,(1, 0.9N)$	$CPU\,(s)$	0.360	1.440	1.770	3.510	12.330
	k	177	780	1509	2118	2838
	Gap	16.091	70.909	137.182	192.545	258.000
$Z_9\,(1, 0.9N)$	$CPU\,(s)$	0.190	0.630	0.790	1.680	6.080
	k	11	102	117	195	300
	Gap	1.000	9.273	10.636	17.727	27.273
$S_1\,(2, 0.9N)$	$CPU\,(s)$	0.490	1.960	26.020	40.970	102.880
	k	0	0	7	185	177
	Gap	0.000	0.000	0.636	16.818	16.091
$S_2\,(2, 0.9N)$	$CPU\,(s)$	0.500	1.410	8.690	8.050	30.680
	k	0	0	0	0	0
	Gap	0.000	0.000	0.000	0.000	0.000
$S_1\,(3, 0.9N)$	$CPU\,(s)$	2.620	15.930	64.280	80.290	188.850
	k	26	341	840	1389	1924
	Gap	2.364	31.000	76.364	126.273	174.909
$S_2\,(3, 0.9N)$	$CPU\,(s)$	1000	1000*	1000*	1000*	1000*
	k	NA	396	684	1139	1265
	Gap	NA	360.390	574.094	817.123	1163.270
$S_1\,(4, 0.9N)$	$CPU\,(s)$	0.300	2.970	4.030	7.110	44.860
	k	0	0	0	0	0
	Gap	0.000	0.000	0.000	0.000	0.000
$S_1\,(5, 0.9N)$	$CPU\,(s)$	2.060	13.160	20.070	37.170	74.040
	k	0	0	0	0	0
	Gap	0.000	0.000	0.000	0.000	0.000
$Z_0\,(4, 0.9N)$	$CPU\,(s)$	0.100	0.600	0.820	1.320	6.320
	k	0	0	0	0	0
	Gap	0.000	0.000	0.000	0.000	0.000
$Z_{00}\,(5, 0.9N)$	$CPU\,(s)$	0.070	0.340	0.780	1.310	3.410
	k	0	0	0	0	0
	Gap	0.000	0.000	0.000	0.000	0.000
NA - the solver did not return a solution in 1000 seconds; * - the solver was interrupted but a solution was returned.						

Table 5: Model 2 - Results for Case 1

N	**1000**	**3000**	**5000**	**7000**	**10000**
$CPU\,(s)$	15.570	57.430	196.330	362.630	528.140
k	117	634	1323	1910	2575
Gap	10.636	57.636	120.273	1910	234.091
$S_1\,(1, 0.9N)$	14.132	15.311	15.344	14.833	15.412
$S_1\,(2, 0.9N)$	8.475	7.033	12.204	15.277	4.827
$S_1\,(3, 0.9N)$	86.115	111.419	112.636	112.182	113.121
$S_1\,(4, 0.9N)$	5.301	5.834	12.489	3.026	6.145
$S_1\,(5, 0.9N)$	71.803	77.702	121.722	49.804	91.970

Table 6: Model 2 - Results for Case 2

are again very tractable in the sense they consume a small amount of time, and MAD statistics without initial conformity are revealed to be more problematic for the adopted *solver*. The CPU always achieves the time limit in such cases and, in one of the cases, no solution was returned by the solver.

Furthermore, the values of k are in general increasing with N, as the values of the gaps to the global optimums. Obviously, the cases for which $k = 0$ are the cases where there was conformity in the initial state of the data set (see Table 1). Of course, in these cases, in order to use Model 2, auditors are invited to impose more demanding critical values so that there is never initial conformity.

Table 6 summarizes the results for Case 2.

The main observation here is that the problems that consider several conformity tests in simultaneous can still be solved with Model 2 in short time, which supports its practical use. Note also that k, CPU and the gap always increase with N. Furthermore, k is always greater than 10% of N, which is expectable from the result obtained in Model 1 where θ was always greater than zero.

Finally, Table 7 presents results of Case 3.

Despite the large number of constraints imposed in this Case, the problem has not became infeasible, for all N. There exists a compromise among the test statistics.

N	**1000**	**3000**	**5000**	**7000**	**10000**
$CPU\,(s)$	122.630	138.630	247.230	441.890	1000*
k	181	783	1511	2122	2842
Gap	181	71.182	137.364	192.909	2842
$S_1\,(1, 0.9N)$	11.298	11.294	10.982	14.706	14.240
$S_2\,(1, 0.9N)$	0.011	0.005	0.005	0.006	0.003
$Z_1\,(1, 0.9N)$	−1.870	−1.916	−1.930	−1.917	−1.927
$Z_9\,(1, 0.9N)$	0.422	0.768	0.839	0.877	0.083
$S_1\,(2, 0.9N)$	9.936	9.096	13.040	11.255	15.562
$S_2\,(2, 0.9N)$	0.010	0.005	0.005	0.004	0.004
$S_1\,(3, 0.9N)$	51.133	112.249	112.900	112.477	110.737
$S_2\,(3, 0.9N)$	0.002	0.001	0.002	0.001	0.001
$S_1\,(4, 0.9N)$	6.849	13.117	5.987	7.301	7.074
$Z_0\,(4, 0.9N)$	−1.153	−0.474	−0.446	0.153	1.466
$S_1\,(5, 0.9N)$	98.582	118.363	92.685	121.795	118.502
$Z_{00}\,(5, 0.9N)$	−0.067	−0.463	1.040	0.176	−0.425
* - the solver was interrupted but a solution was returned.					

Table 7: Model 2 - Results for Case 3

Also, as in Case 2, for all N, a solution is found by the solver in reasonable time. The values of k are always greater than in Case 2, which is expectable since new nonconformity situations are introduced when more test statistics are considered (in particular some MAD and Z statistics, as can be seen in Table 1).

At last, note that after $N = 1000$, the percentage of k is very homogeneous around 30% of N.

5 Conclusions

We contributed to Digital Analysis by formulating two general mathematical programming models that can help auditors to select a specific target set of records from a data set when planning an audit. The models make use of Benford's Law in a more consistent and effective way than in the current approach, by considering multiple conformity tests and test statistics in simultaneous, which allows for example to account for the interdependencies between conformity tests, and by allowing the auditor either to select a specific number of records over which to address direct auditing procedures or to identify the subset of nonconforming records in a data set.

The models are general for the most common conformity tests and test statistics. Nevertheless, without changing their main goal, i.e. to identify potential problematic records, they can be easily extended to incorporate other conformity tests, test statistics and objective functions.

Of course, before using these models to plan an audit, auditors should certify that unmanipulated data from the accounts to be audited should conform with Benford's Law. This is a critical step, assumed throughout the paper, that can never be neglected, and which is only vaguely discussed in Section 2.

There are several *solvers* that can be used by auditors to solve the suggested problems. Each one has its own computational strategy and its own limitations, namely in terms of decision variables. Importantly, note that in order to solve the generalized versions of the proposed models, one needs a *solver* capable of processing up to N binary $\{0,1\}$ decision variables, and some sort of quality measure is desirable when a global optimum is not assured.

A small experiment was conducted in order to try to identify some regularities in the relation between the computational time required to solve the proposed problems and the characteristics of the initial data set, its number of records, conformity tests considered, and test statistics used. The computational time required to solve the problems appears to increase with the number of records N, but not properly with the number of conformity tests and test statistics considered. Anyway, the computational time is very reasonable for most cases, which offers feasibility to the implementation of the proposed analytical audit planning procedures, at least with the used *solver*.

The characteristics of the initial data set seem also to be significantly related with computational time, especially with respect to the MAD statistic, for which the resolution becomes difficult when there is initial nonconformity.

The *solver* used in the paper faced difficulties in some problems, which can be due to its technical assumptions in the search for solutions and to the nature of the constraints. It would be important to develop and test more flexible tools, capable of dealing with a greater variety of constraint types and objective functions.

Another suggestion for future research is testing the proposed approach to detect irregularities and fraud symptoms using real data sets, and comparing its effectiveness with the one obtained with traditional audit planning approaches. Considering the fact that the critical values can be subjective, it could also be of interest to perform sensitivity analysis over their values in the proposed models.

References

[1] Akresh, A., J. Loebbecke and W. Scott (1998), *Research opportunities in auditing: The second decade*, in *Audit approaches and techniques*, edited by A. Abdel-khalik and I. Solomon, Sarasota, FL: American Accounting Association.

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

[3] Boyle, J. (1994), *An application of Fourier series to the most significant digit problem*, American Mathematical Monthly (November), pp 879-886.

[4] Carlslaw, C. (1988), *Anomalies in income numbers: Evidence of goal oriented behavior*, The Accounting Review, LXIII (2), pp 321-327.

[5] Coderre, D. (2009), *Fraud Analysis Techniques Using ACL*, Book, John Wiley and Sons, Inc.

[6] Coderre, D. (1999), *Fraud Detection Using Digital Analysis*, EDPACS, 27:3, págs. 1-8.

[7] Diekmann, A. (2007), *Not the first digit! Using Benford's Law to detect fraudulent scientific data*, Journal of Applied Statistics, Vol 34, N^o 3, pp 321-329.

[8] Drake, P. and M. Nigrini (2000), *Computer assisted analytical procedures using Benford's Law*, Journal of Accounting Education, N^o 18, págs. 127-146.

[9] Durtschi, C., W. Hillison and C. Pacini (2004), *The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data*, Journal of Forensic Accounting, Vol. V, pp 17-34.

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

[11] Neumaier, A., O. Shcherbina, W. Huyer and T. Vinkó (2005), *A comparison of complete global optimization solvers*, Mathematical Programming: Serie B, Vol. 103, N^o 2, pp 335-356.

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

[13] Nigrini, M. (2000), *Continuous Auditing*, Ernst & Young Center for Auditing Research and Advanced Technology, University of Kansas.

[14] Nigrini, M. and L. Mittermaier (1997), *The Use of Benford's Law as an Aid in Analytical Procedures*, Auditing: A Journal of Practice & Theory, Vol. 16, N^o 2, págs. 52-67.

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

[16] Watrin, C., R. Struffert and R. Ullmann (2008), *Benford's Law: an instrument for selecting tax audit targets?*, Review of Management Science, 2, pp 219-237.