Mateusz Baryła
Uniwersytet Ekonomiczny w Krakowie

O PEWNYM MODELU POZWALAJĄCYM IDENTYFIKOWAĆ K NAJBARDZIEJ PODEJRZANYCH REKORDÓW W ZBIORZE DANYCH KSIĘGOWYCH W PROCESIE WYKRYWANIA OSZUSTW FINANSOWYCH

Wprowadzenie

W polskiej ustawie o rachunkowości znajduje się zapis mówiący, że „(…) celem badania sprawozdania finansowego jest wyrażanie przez biegłego rewidenta pisemnej opinii wraz z raportem o tym, czy sprawozdanie finansowe jest prawidłowe oraz rzetelnie i jasno przedstawia sytuację majątkową i finansową, jak też wynik finansowy badanej jednostki" (*Ustawa o rachunkowości z dnia 29 września 1994 r.*, art. 65, ust. 1)[1]. Należy wyraźnie zaznaczyć, że przeprowadzane przez biegłego rewidenta badanie sprawozdania finansowego dotyczy również wykrywania oszustw finansowych. Wskazują na to zapisy zawarte zarówno w przepisach krajowych, jak i międzynarodowych[2]. Zgodnie z nimi to właśnie na biegłym rewidencie spoczywa obowiązek wykrywania oszustw, jeżeli tylko wywierają one istotny wpływ na informacje prezentowane w sprawozdaniach finansowych.

W procesie wykrywania oszustw finansowych pomocne okazują się procedury oparte na prawie Benforda. Problematyka dotycząca identyfikowania nadużyć finansowych z wykorzystaniem wspomnianego prawa została podjęta między innymi w następujących pracach: Nigrini (1994, 2012); Durtschi, Hillison, Pacini (2004); Saville (2006).

[1] W zasadzie taki sam cel badania sprawozdania finansowego został sformułowany w regulacjach międzynarodowych (por. *Międzynarodowy Standard Rewizji Finansowej (MSRF) nr 200*, ust. 3).

[2] Zob.: *Międzynarodowy Standard Rewizji Finansowej nr 240* zatytułowany „Odpowiedzialność biegłego rewidenta podczas badania sprawozdań finansowych dotycząca oszustw", ust. 5, 3; *Krajowy standard rewizji finansowej nr 1*, ust. 54. Treści tych standardów można znaleźć na stronie internetowej Krajowej Izby Biegłych Rewidentów: http://www.kibr.org.pl/.

1. Analizy oparte na badaniu rozkładu cyfr

Najogólniej rzecz ujmując, prawo Benforda dotyczy częstości występowania cyfr na określonych pozycjach znaczących w liczbie. Można tutaj rozważać zarówno pojedyncze pozycje znaczące, jak i brać pod uwagę większą ilość pozycji znaczących liczby jednocześnie. Wynikający ze wspomnianego prawa rozkład cyfr na określonych pozycjach znaczących jest znany w literaturze pod nazwą rozkładu Benforda.

Prawdopodobieństwo tego, że na pierwszej, drugiej oraz − uogólniając problem − k-tej pozycji znaczącej liczby (co symbolicznie będziemy oznaczać: D_1, D_2 i D_k) pojawi się cyfra odpowiednio i_1, i_2 oraz i_k, obliczamy następująco:

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

$$P(D_2 = i_2) = \sum_{h=1}^{9} \log_{10}\left(1 + \frac{1}{10h + i_2}\right), \tag{2}$$

$$P(D_k = i_k) = \sum_{h_1=1}^{9} \sum_{h_2=0}^{9} \dots \sum_{h_{k-1}=0}^{9} \log_{10}[1 + (h_1 \cdot 10^{k-1} + \dots + h_{k-1} \cdot 10 + i_k)^{-1}], \tag{3}$$

gdzie: $i_1 \in \{1, 2, ..., 9\}$, $i_j \in \{0, 1, ..., 9\}$, $j = 2, 3, ..., k$.

Chcąc przykładowo obliczyć prawdopodobieństwo tego, że pierwszą i drugą znaczącą cyfrą liczby będą odpowiednio cyfry i_1 oraz i_2, należy to uczynić zgodnie z następującym wzorem:

$$P(D_1 D_2 = i_1 i_2) = \log_{10}\left(1 + \frac{1}{i_1 i_2}\right). \tag{4}$$

Zastosowanie prawa Benforda w wykrywaniu oszustw sprowadza się do dokonywania porównań częstości występowania cyfr (obliczonych np. dla zbioru danych księgowych) na określonych pozycjach znaczących z prawdopodobieństwami wynikającymi z rozkładu Benforda. W tym zakresie Nigrini i Mittermaier (1997) zaproponowali sześć testów wykorzystujących analizę cyfr (ang. digital analysis tests), wśród których znalazły się między innymi następujące: test pierwszych cyfr, test drugich cyfr, test dwóch pierwszych cyfr oraz test dwóch ostatnich cyfr.

W celu przeprowadzenia dalszych rozważań poczyńmy kilka uwag do przyjętych oznaczeń. Niech $w_i(t)$ oznacza zaobserwowaną w zbiorze liczącym n elementów częstość względną cyfry (lub cyfr) i w teście o numerze t, czyli

$w_i(t) = n_i(t)/n,$ gdzie $n_i(t)$ reprezentuje ilość wystąpień cyfry (lub cyfr) i w t-tym teście. Niech $p_i(t)$ oznacza wynikające z rozkładu Benforda prawdopodobieństwo wystąpienia cyfry (lub cyfr) i w teście o numerze t. Wówczas, posługując się wprowadzonymi oznaczeniami, wymienione testy oparte na analizie cyfr będzie można scharakteryzować tak, jak to uczyniono w tabeli 1.

Tabela 1

Charakterystyka wybranych testów opartych na analizie cyfr

Test	Charakterystyka testu
$t = 1$	Polega na porównaniu częstości względnych w_i z prawdopodobieństwami p_i dla pierwszej cyfry znaczącej; $i = i_1$, przy czym $i_1 = 1, 2, ..., 9$
$t = 2$	Polega na porównaniu częstości względnych w_i z prawdopodobieństwami p_i dla drugiej cyfry znaczącej; $i = i_2$, przy czym $i_2 = 0, 1, ..., 9$
$t = 3$	Polega na porównaniu częstości względnych w_i z prawdopodobieństwami p_i dla dwóch pierwszych cyfr znaczących; $i = i_1i_2$, przy czym $i_1i_2 = 10, 11, ..., 99$
$t = 4$	Polega na porównaniu częstości względnych w_i z prawdopodobieństwami p_i dla dwóch ostatnich cyfr znaczących; $i = i_{s-1}i_s$, przy czym $i_{s-1}i_s = 00, 01, ..., 99$, natomiast s oznacza ostatnią cyfrę

Źródło: Opracowanie własne na podstawie: Nigrini, Mittermaier (1997, s. 57); Silva, Carreira (2011, s. 5).

W przypadku każdego z przeprowadzanych testów pojawia się konieczność dokonania oceny badanego zbioru danych pod kątem jego zgodności z prawem Benforda. W tym zakresie proponuje się wykorzystanie różnych miar. Niektóre z nich zestawiono w tabeli 2.

Tabela 2

Wybrane miary służące do oceny zgodności danych z prawem Benforda

Miara	Równanie	Miara	Równanie
M_1	$\chi^2(t) = n\sum_i \frac{[w_i(t) - p_i(t)]^2}{p_i(t)}$	M_3	$d(t) = \frac{1}{n(t)}\sqrt{\sum_i [w_i(t) - p_i(t)]^2}$
M_2	$MAD(t) = \frac{1}{n(t)}\sum_i \lvert w_i(t) - p_i(t)\rvert$	M_4	$u_i(t) = \frac{w_i(t) - p_i(t)}{\sqrt{\frac{p_i(t)[1 - p_i(t)]}{n}}}$

Uwaga: W przypadku miary M_2 i M_3 za wartość wyrażenia $n(t)$, w zależności od wybranego testu t, podstawiamy: $n(t = 1) = 9$, $n(t = 2) = 10$, $n(t = 3) = 90$, $n(t = 4) = 100$.

Źródło: Opracowanie własne.

Pojawiające się w literaturze metody oceny zgodności danych ze wspomnianym prawem można w zasadzie podzielić na dwie grupy. Jedna z nich obejmuje te metody, które opierają się na teorii weryfikacji hipotez statystycznych (np. miara M_1 i M_4), natomiast drugą grupę stanowią metody, które nie wykorzystują takiego podejścia (np. miara M_2 i M_3). Podstawowy problem związany z drugą wymienioną z kolei grupą miar dotyczy braku jakichkolwiek wartości granicznych, na podstawie których można by jednoznacznie stwierdzić, czy zbiór danych podlega prawu Benforda, czy też nie. Jedynie w przypadku miary M_2 można znaleźć w literaturze pewne sugestie, które zestawiono w formie tabeli 3.

Tabela 3

Zaproponowane wartości graniczne dla miary M_2

Stopień zgodności	$t = 1$	$t = 2$	$t = 3$
Duża zgodność	0,000 - 0,004	0,000 - 0,008	0,0000 - 0,0006
Akceptowalna zgodność	0,004 - 0,008	0,008 - 0,012	0,0006 - 0,0012
Skrajnie akceptowalna zgodność	0,008 - 0,012	0,012 - 0,016	0,0012 - 0,0018
Brak zgodności	> 0,012	> 0,016	> 0,0018

Źródło: Opracowanie własne na podstawie: Drake, Nigrini (2000, s. 133-134).

W dalszej części niniejszego artykułu zostanie przedstawiony model (w swojej najprostszej postaci), zaproponowany w pracy Silva, Carreira (2011), służący do identyfikacji k najbardziej podejrzanych rekordów w zbiorze danych księgowych.

2. Konstrukcja modelu

Przyjmijmy, że biegły rewident dysponuje zbiorem danych składającym się z n zapisów księgowych. Chce on zidentyfikować w tym zbiorze zadaną z góry liczbę k zapisów, jaką należy poddać szczegółowemu badaniu celem wykrycia nieprawidłowości spowodowanych oszustwami. Aby przeprowadzić analizę zgodności danych z prawem Benforda, biegły rewident wybiera test t oraz jedną z miar przedstawionych w tabeli 2, oznaczoną symbolem M z odpowiednim subskryptem. Biegły chce zidentyfikować k najbardziej podejrzanych zapisów księgowych, tj. takich rekordów, które − gdy zostaną usunięte z wyjściowego zbioru danych − spowodują największą poprawę wartości wybranej miary M. W dalszej części, spośród miar zaprezentowanych w tabeli 2, weźmiemy pod uwagę jedynie te, które uwzględniają łączny rozkład cyfr na określonej pozycji w liczbach.

W celu rozwiązania postawionego problemu decyzyjnego, w zapisie miar służących do oceny zgodności danych z prawem Benforda trzeba uwzględnić owe k liczb, które należy usunąć z pierwotnego zbioru danych. Mamy zatem:

$$M_1(t, n-k) = (n-k)\sum_i \frac{\left[\frac{n_i(t) - k_i(t)}{n-k} - p_i(t)\right]^2}{p_i(t)}, \tag{5}$$

$$M_2(t,n-k)=\frac{1}{n(t)}\sum_i\left|\frac{n_i(t)-k_i(t)}{n-k}-p_i(t)\right|, \tag{6}$$

$$M_3(t,n-k)=\frac{1}{n(t)}\sqrt{\sum_i\left[\frac{n_i(t)-k_i(t)}{n-k}-p_i(t)\right]^2}\,, \tag{7}$$

gdzie $n_i(t)$ oraz $k_i(t)$ oznaczają ilość liczb mających cyfrę (cyfry) i na określonej pozycji znaczącej odpowiednio dla wyjściowego zbioru danych (liczącego n elementów) i zredukowanego zbioru danych (tj. powstałego po usunięciu ze zbioru wyjściowego k liczb). Pojawiające się w powyższych miarach wyrażenie $[n_i(t)-k_i(t)]/(n-k)$ oznacza częstość względną występowania cyfry (cyfr) i w wyniku zastosowania wybranego testu t, obliczoną na podstawie zbioru danych, z którego usunięto k liczb.

Przejdźmy do zapisania optymalizacyjnego modelu programowania matematycznego dla omawianego problemu decyzyjnego. Jako funkcję kryterium (celu) przyjmiemy jedną z uprzednio zdefiniowanych miar, której wartość będziemy minimalizować:

$$M(t,n-k)\to \min,$$

przy następujących ograniczeniach:

$$k=\sum_i k_i(t),$$

$$\forall i\ \ k_i(t)\le n_i(t),$$

$$\forall i\ \ k_i(t)\ge 0 \text{ i całkowite.}$$

Zauważmy, że spośród wyżej zapisanych warunków ograniczających, pierwszy z nich dotyczy łącznej liczby rekordów, które mają zostać usunięte, czyli poddane przez audytora szczegółowemu badaniu. Drugi z zapisanych warunków odnosi się do ilości usuniętych liczb zawierających cyfrę (cyfry) i na określonej pozycji znaczącej; ilość ta nie może być większa niż ilość takich liczb znajdujących się w wyjściowym zbiorze. Z kolei trzeci zapisany warunek mówi o tym, iż ilość liczb, jakie należy usunąć z cyfrą (cyframi) i na danej pozycji znaczącej, musi się wyrażać liczbą całkowitą nieujemną. Zauważmy ponadto, że drugi z zapisanych warunków okazuje się być zbędny w sytuacji, gdy dla każdego i jest spełniona nierówność $k\le n_i(t)$.

3. Przykład empiryczny

W celu zilustrowania omawianego problemu posłużmy się następującym przykładem. Załóżmy, że biegły rewident dysponuje zbiorem danych księgowych liczącym n = 3200 rekordów. Do jego analizy wybiera przykładowo test t = 1, a więc decyduje się na analizę opartą na rozkładzie pierwszej cyfry znaczącej. Przyjmijmy, że dla rozważanego zbioru uzyskano rozkład pierwszej cyfry znaczącej taki, jak przedstawia to tabela 4.

Tabela 4

Rozkład pierwszej cyfry znaczącej dla analizowanego zbioru danych

i	1	2	3	4	5	6	7	8	9
$n_i(1)$	1132	518	409	302	233	172	158	168	108
$w_i(1)$	0,354	0,162	0,128	0,094	0,073	0,054	0,049	0,052	0,034

Źródło: Opracowanie własne.

Załóżmy dodatkowo, że do oceny zgodności danych z prawem Benforda audytor wybiera odchylenie przeciętne, czyli miarę oznaczoną symbolem M_2. Dla rozpatrywanego przypadku otrzymujemy wartość wspomnianej miary wynoszącą 0,0127, co wskazuje na brak zgodności danych z prawem Benforda.

Przyjmijmy także, że biegły rewident chce zidentyfikować k = 30 zapisów, na które należy zwrócić szczególną uwagę, aby wykryć nieprawidłowości spowodowane oszustwami. Wówczas optymalizacyjny model matematyczny przyjmie następującą postać:

$$M(t=1, n-k=3170) = \frac{1}{9}\left(\left|\frac{1132-k_1(1)}{3170}-0{,}3010\right| + \left|\frac{518-k_2(1)}{3170}-0{,}1761\right| +\right.$$
$$+\left|\frac{409-k_3(1)}{3170}-0{,}1249\right| + \left|\frac{302-k_4(1)}{3170}-0{,}0969\right| + \left|\frac{233-k_5(1)}{3170}-0{,}0792\right| +$$
$$+\left|\frac{172-k_6(1)}{3170}-0{,}0670\right| + \left|\frac{158-k_7(1)}{3170}-0{,}0580\right| + \left|\frac{168-k_8(1)}{3170}-0{,}0512\right| +$$
$$\left.+\left|\frac{108-k_9(1)}{3170}-0{,}0458\right|\right) \to \min,$$

$$\sum_i k_i(1) = 30, \quad i = 1, 2, ..., 9,$$

$$k_i(1) \geq 0 \text{ i całkowite, } i = 1, 2, ..., 9.$$

Rozwiązując powyższy model, uzyskano następujące optymalne wartości $k_i^*(1)$, które zestawiono w tabeli 5.

Tabela 5

Rozwiązanie optymalne rozważanego problemu decyzyjnego

i	1	2	3	4	5	6	7	8	9
$k_i^*(1)$	23	0	4	0	0	0	0	3	0

Źródło: Opracowanie własne.

Analizując otrzymane wyniki, należy stwierdzić, że w celu wykrycia zadanej liczby najbardziej podejrzanych księgowań, biegły rewident powinien wybrać i poddać szczegółowemu badaniu 23 zapisy księgowe, które mają cyfrę 1 na pierwszej pozycji znaczącej, 4 zapisy z cyfrą 3 jako pierwszą cyfrą znaczącą oraz 3 zapisy mające na pierwszej pozycji znaczącej cyfrę 8. Na uwagę zasługuje fakt, że po usunięciu ze zbioru owych 30 rekordów, wartość odchylenia przeciętnego wynosi około 0,0117, co wskazuje już na skrajnie akceptowalną zgodność danych z prawem Benforda.

Podsumowanie

Opisany w niniejszym artykule model programowania matematycznego może posłużyć jako użyteczne narzędzie w trakcie identyfikowania najbardziej podejrzanych zapisów księgowych w procesie wykrywania oszustw finansowych przez biegłych rewidentów. Należy jednak wyraźnie zaznaczyć, że zaprezentowany model posiada co najmniej dwa zasadnicze ograniczenia.

Podczas konstrukcji modelu przyjęto, że audytor wybiera jedną z analiz opartych na rozkładzie Benforda. W rzeczywistości może on być zainteresowany jednoczesnym przeprowadzeniem większej liczby tego typu analiz. Prezentując model założono także, iż do oceny zgodności danych z prawem Benforda jest wykorzystywana miara uwzględniająca łączny rozkład cyfr na określonej pozycji w liczbach. Rozbudowując ten problem, można wziąć pod uwagę możliwość wykorzystania wielu miar jednocześnie, także tych, które w swojej konstrukcji nie opierają się na łącznym rozkładzie cyfr. Nie ulega wątpliwości, że chęć uwzględnienia przedstawionych uwag przyczyni się do wzrostu złożoności omówionego modelu.

Literatura

Drake P.D., Nigrini M.J. (2000): *Computer Assisted Analytical Procedures Using Benford's Law*. „Journal of Accounting Education", No. 18.

Durtschi C., Hillison W., Pacini C. (2004): *The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data*. „Journal of Forensic Accounting", Vol. V.

http://www.inescc.pt/documentos/Researchreport8.pdf (dostęp: 08.04.2013).

Krajowy standard rewizji finansowej nr 1. Krajowa Rada Biegłych Rewidentów.

Międzynarodowy Standard Rewizji Finansowej (MSRF) 200: Ogólne cele niezależnego biegłego rewidenta oraz przeprowadzanie badania zgodnie z Międzynarodowymi Standardami Rewizji Finansowej. Międzynarodowa Federacja Księgowych.

Międzynarodowy Standard Rewizji Finansowej (MSRF) 240: Odpowiedzialność biegłego rewidenta podczas badania sprawozdań finansowych dotycząca oszustw. Międzynarodowa Federacja Księgowych.

Nigrini M.J. (1994): *Using Digital Frequencies to Detect Fraud*. „The White Paper", April/May.

Nigrini M.J. (2012): *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. Wiley, New Jersey.

Nigrini M.J., Mittermaier L.J. (1997): *The Use of Benford's Law as an Aid in Analytical Procedures*. „Auditing: A Journal of Practice & Theory", Vol. 16, No. 2.

Saville A. (2006): *Using Benford's Law to Detect Data Error and Fraud: An Examination of Companies Listed on the Johannesburg Stock Exchange*. „South African Journal of Economic and Management Sciences", Vol. 9, No. 3.

Silva C.G., Carreira P.M.R. (2011): *Selecting Audit Targets Using Benford's Law*. Institute of Systems Engineering and Computers, INESC, Coimbra.

Ustawa o rachunkowości z dnia 29 września 1994 r. Dz.U. 1994, nr 121, poz. 591 z późn. zm.

ABOUT A MODEL IDENTIFYING THE K MOST SUSPICIOUS RECORDS IN AN ACCOUNTING DATA SET IN THE PROCESS OF FINANCIAL FRAUD DETECTION

Summary

Financial frauds lead to the disturbance of a normal development of stock markets. When they appear, the funds are not properly allocated, which has a negative impact on economic growth. In most cases, investors make decisions taking into consideration economic information presented by companies. Financial frauds substantially affect data included in financial statements. For this reason, it seems important to undertake steps aiming at fraud detection. A key role in this issue may be assumed by auditors who are responsible for identifying significant financial irregularities.

In the paper, a certain mathematical programming model which can be useful for auditors during detecting irregularities caused by financial frauds is discussed. In the case of this model, the decision problem consists in finding a given number of the *k* most suspicious records in a data set that should be thoroughly audited. Benford's Law is used as a base while constructing the model.