基于本福德定律发现企业舞弊的适用性

——针对我国上市公司财务数据的实证测试

陈 曦　万宇飞　李 璐

（山东工商学院会计学院　山东烟台　264005）

【摘要】本福德定律描述了在一系列数据中首位数字从数字1到数字9的数据在整体中的分布概率。根据该定律，统计数据的有效数字分布概率符合对数规律，人为的舞弊活动将破坏这种规律。本文将这一原理应用到审计工作中，研究这一种基于数字分布规律的舞弊检测方法，并提供了专用分析工具。

本文对11年来中国上市公司的财务数据情况进行统计分析，并与本福德定律标准值进行比较，对其在中国上市公司的适用性作出检测，验证了运用本福德定律识别企业舞弊的有效性，提出了根据本福德定律来查找企业舞弊迹象的新方法，并进一步为其在实际工作的使用提出了建议。

【关键词】本福德定律　报表审计　实证测试　标准意见　非标准意见

一、本福德定律介绍

本福德定律是个古老而有趣的定律。近年来，随着信息技术在审计领域的应用，借助于计算机进行数值分析，人们发现了本福德定律在识别财务舞弊迹象方面所具有的特殊功效。

本福德定律是由美国人塞蒙·纽卡姆（Simon Newcomb）在1881年首次发现的。在1881年的某一天，塞蒙在使用对数表时，意外注意到一件事：对数表的首页明显比其后的纸页使用次数更多。塞蒙对这一奇怪的现象产生了浓厚的兴趣，在研究之后，当时他所找到的合理的解释是相比较于对大数字的运算量，人们对于小数字的运算量更大。

在大量的统计分析后，他发现了更多的古怪现象：以1为第一位数的随机数的出现次数要比以2为第一位数的随机数的出现次数要多，而以2为第一位数的随机数的出现次数又比以3为第一位数的随机数的出现次数多，以此类推。当时塞蒙只是出于好奇才对这一数学现象进行关注和研究，并没有往其实际应用的方向进行考虑，这也导致了当时的人对这一数学现象缺乏兴趣，很快这个现象就被人们忘记了。

到了1938年，弗瑞克·本福德（FrankBenford）也注意到了类似的现象。弗瑞克采集了总数为20 229的数字并对其进行统计证明，发现在所有的这些数字中，整数1在数字首位出现的概率约为30%，整数2在数字首位出现的概率约为17%，整数3在数字首位出现的概率约为12%，而8和9在数字首位出现的概率分别约为5%和4%。因此人们也称之为“第一位数分布规律”（见表1）。

表1　　数字1～9在数字首位出现的期望概率

N	1	2	3	4	5	6	7	8	9
E`[digit(n)]	0.301 0	0.176 1	0.124 9	0.096 9	0.079 2	0.066 9	0.058 0	0.051 2	0.045 8

数学表达式为：

$$E`[digit(n)]=\log10(1+1/n)$$

在该数学表达式中，n为整数1，2，3，4，5，6，7，8，9……E`[digit(n)]为各整数在数字首位出现的对应期望概率（即未受人为影响的理论值）。本福德定律揭示了数字1~9在自然状态下的数据的首位的概率分布。

Hill（1938）证明，本福德定律对任何整数的分析对象都有效，因此在审计领域，本福德定律经常被用来检测人类对数据的影响程度，比如对心理价格屏障、数据篡改、公司欺诈等行为的检测。本福德定律测试方法隐含的结果是，如果一组数据所统计分析出的首位数字分布与本福德定律理论分布结果不一致，则该数据有可能受到某种人为影响。

二、论文结构与研究思路

本文通过本福德定律对中国上市公司从2000年到2010年的主要财务数据的首位数字分布进行统计分析，验证其是否符合本福德定律的首位数字分布理论值，以期获得本福德定律在中国上市公司的适用情况。

本文对本福德定律在中国上市公司的实用情况提出假设，以中国上市公司2000年到2010年的实际财务数据为样本数据，对其进行统计分析，以验证这些假设。同时参考前人的研究方法，除整体分析之外，也以不同的标准对数据进行分类，并对其分析结果进行比较，从多个角度验证本福德定律在中国上市公司实际应用的合理性。

三、研究方法

本文选取的数据来自国泰安数据库，包括沪深两市2000年至2010年2 087家上市公司的年度、季度财务数据，剔除了少量数据为0的无效样本。

针对2010年的各公司年报，会计师事务所共出具了2 087份审计报告，其中标准审计意见报告1 984份，带强调事项段的无保留意见审计报告80份，保留意见加事项段审计报告6

份,保留意见审计报告12份,无法表示意见审计报告5份。

(一)测试方法

1. 分析工具。由于从数据库获得的数据库格式为Excel所支持的xls,即电子表格格式,且Excel软件本身有丰富的数据筛选和运算功能,因此本文使用微软办公软件的Microsoft Office Excel的2003版作为数据处理软件。

为方便比较,本文以本福德定律的首位数字为1至9的数据的理论出现概率为基准,以线性相关回归分析来计算实际数值与理论值的相关系数,该相关系数代表二者的相关程度,其数据在0与1之间,数值越大则相关程度越高,数值越小则相关程度越低。

2. 分析手段。假设从A1到A8000是分析对象的所有数据,那么:

(1)求各项绝对值。在B1输入"=ABS(A1)"(不包括双引号,下同),用单元格填充工具将该函数拖至B8000。

(2)求各绝对值首位数字。在C1输入"=LEFT(B1,1)",用单元格填充工具将该函数拖至C8000。

(3)统计首位数字的出现次数。在D1输入"=COUNTIF (C1:C8000,1)",在D2输入"=COUNTIF(C1:C8000,2)",以此类推,直至在D9输入"=COUNTIF(C1:C8000,9)",并在D10输入"=SUM(D1:D9)"来求和。

(4)计算首位数字的出现概率。在E1输入"=D1/XX"(XX为D10中的数字),用单元格填充工具将该函数拖至E9,并在E10输入"=SUM(E1:E9)"来求和。

(二)提出假设

中国证监会要求,截至每年4月30日,所有上市公司要公布其上一年度财务报告,因此本文提出了以下假设:

H1:上市公司年度报告在总体上来说应符合本福德定律。

H2:标准意见公司和非标准意见公司同一年的财务报表数据的首位数字分布与本福德定律的相关系数有差异,且非标准审计意见的财务报表数据的首位数字分布与本福德定律的相关系数低于标准审计意见的财务报表数据的首位数字分布与本福德定律的相关系数。

H3:对中国上市公司的不同财务指标来说,其总体的数据表现在与本福德定律理论值的吻合程度上应有所差异。

H4:因违规被公开处罚的上市公司财务数据分布与本福德定律的理论数值分布存在着差异。

H5:上市公司不同年度财务数据分布与本福德定律的理论数值分布存在一定差异。

(三)测试结果

1. 各年主要财务数据比较。截至2010年12月31日,我们对沪深两市发行A股的2 087家上市公司公布的主要财务数据的首位数分布情况,及其与本福德定律理论分布值进行比较。测试结果如表2:

表2 中国上市公司各年度主要财务数据本福德定律测试相关系数结果

年 份	资产总额	资本公积	所有者权益合计	营业收入	营业成本	利润总额	净利润
2000	0.962658	0.965912	0.972068	0.994884	0.997086	0.988762	0.994844
2001	0.962595	0.976897	0.967618	0.998218	0.998044	0.998509	0.997373
2002	0.973309	0.966981	0.967080	0.998797	0.998659	0.999776	0.998224
2003	0.985585	0.966146	0.969857	0.999559	0.999148	0.998678	0.998143
2004	0.990322	0.976839	0.977594	0.998454	0.999489	0.997912	0.999270
2005	0.994676	0.976831	0.987366	0.999108	0.999130	0.998450	0.998466
2006	0.996448	0.980664	0.986456	0.999154	0.999355	0.998919	0.999100
2007	0.994953	0.990374	0.985416	0.999562	0.999471	0.998111	0.998774
2008	0.993280	0.996233	0.989041	0.999435	0.999654	0.997790	0.999588
2009	0.990332	0.995249	0.993312	0.998719	0.998914	0.999603	0.999535
2010	0.990826	0.992485	0.981097	0.995230	0.999659	0.999394	0.999420
BenfordLaw	1	1	1	1	1	1	1

从表2所列的概率分布数值以及相关系数可以看出:总体上,中国沪深两市上市公司的主要财务数据首位数的分布表现出了本福德定律所描述的规律,与本福德定律所描述的理论数值分布高度相关,其相关系数达到了0.990。

不同财务数据的数值与本福德定律理论分布值之间存在一定程度的差异,其中营业成本与本福德定律的理论值有较好的一致性,资本公积和所有者权益合计与本福德定律的符合程度较差。

2. 各年整体财务数据比较。

(1)资产负债表。表3给出了在不同年份中国上市公司的各年资产负债表整体财务数据的首位分布,及其与本福德定律理论值的比较结果。

表3 不同年份资产负债表数据首位数分布情况比较

年 份	首位数									相关系数
	1	2	3	4	5	6	7	8	9	
2000	0.2840	0.1624	0.1186	0.1032	0.0894	0.0701	0.0673	0.0549	0.0500	0.997960
2001	0.2782	0.1520	0.1229	0.0992	0.0849	0.0725	0.0702	0.0550	0.0493	0.994494
2002	0.2898	0.1557	0.1251	0.0984	0.0853	0.0682	0.0681	0.0595	0.0499	0.996634
2003	0.2907	0.1617	0.1256	0.0988	0.0871	0.0698	0.0626	0.0559	0.0478	0.998415
2004	0.2893	0.1663	0.1217	0.1001	0.0853	0.0710	0.0623	0.0519	0.0467	0.999105
2005	0.2876	0.1743	0.1243	0.1001	0.0863	0.0647	0.0648	0.0532	0.0446	0.998900
2006	0.2870	0.1719	0.1285	0.0949	0.0876	0.0655	0.0624	0.0563	0.0458	0.998848
2007	0.2905	0.1741	0.1336	0.0920	0.0866	0.0672	0.0537	0.0607	0.0461	0.997051
2008	0.3007	0.1692	0.1254	0.0998	0.0826	0.0661	0.0553	0.0536	0.0474	0.999291
2009	0.3045	0.1699	0.1197	0.0920	0.0889	0.0667	0.0589	0.0527	0.0466	0.998154
2010	0.3083	0.1596	0.1149	0.0892	0.0880	0.0705	0.0636	0.0568	0.0493	0.994012
BenfordLaw	0.3010	0.1761	0.1249	0.0969	0.0792	0.0669	0.0580	0.0512	0.0458	1

从表3中可看出,从2001年至今,相关系数的变化大体可分为三个阶段:

第一阶段:从2000年到2002年。在2000年、2001年和2002年的相关系数都比较低,这3年正是中国上市公司报表披露的混乱时期,大型报表舞弊案件在这3年出现多起,比如银广夏舞弊案等,这段时期中国上市公司的报表舞弊情况普遍比较严重,因此出现了相关系数较低的情况。

第二阶段:从2003年到2007年。这段时期的相关系数突然回升,兼有逐年降低的趋势。在2003年和2004年,审计署对报表舞弊严重的银广夏、西安达尔曼、天津磁卡、蓝田股份、大庆联谊做出公开处罚,纠正了当时粉饰报表的不正之风。在现在来看,2003年和2004年的相关系数突然升高,与当时审计署的相关工作是分不开的。

同时,这段时期实行了新会计准则,很多企业利用会计准则进行舞弊,比如在效益好的年份将部分利润隐藏在坏账准备里,并在业绩不好的年份中将其提出以粉饰利润。

另外,在2007年1月1日之后,中国再次实行新会计准则,使得很多在之前会计准则中可以使用的舞弊手段不能再用,因此在2007年集中了大量非正常与正常的调整,造成当年相关系数成为这段时期低谷的情况。

第三阶段:从2008年至今。该阶段,相关系数又表现出一个回升之后逐年降低的趋势。2008年是中国实行新会计准则的第二年,大部分调整(包括正常和非正常调整)已集中体现在2007年,因此当年相关系数相对于2007年来说出现显著回升。

而2008年又是全球金融危机出现的一年,在随即到来的2009年和2010年里,金融危机的影响尚未结束,造成企业普遍出现效益不好的情况,企业可能会出于对融资、股价等方面的考虑对财务数据进行人为操纵,使得相关系数逐年降低。

(2)利润表。表4给出了在不同年份,沪深两市上市公司的各年整体财务数据的首位分布,及其与本福德定律理论值的比较结果。

表4　不同年份利润表数据首位数分布情况比较

年份	首位数									相关系数
	1	2	3	4	5	6	7	8	9	
2000	0.2829	0.1827	0.1300	0.1009	0.0790	0.0728	0.0567	0.0507	0.0443	0.997126
2001	0.2903	0.1788	0.1292	0.0975	0.0760	0.0697	0.0594	0.0544	0.0448	0.999050
2002	0.3007	0.1732	0.1286	0.0963	0.0803	0.0696	0.0572	0.0503	0.0437	0.999662
2003	0.2957	0.1735	0.1299	0.0986	0.0797	0.0677	0.0587	0.0500	0.0461	0.999619
2004	0.3028	0.1723	0.1266	0.0962	0.0821	0.0673	0.0562	0.0515	0.0450	0.999697
2005	0.2977	0.1780	0.1233	0.0970	0.0785	0.0674	0.0617	0.0504	0.0460	0.999780
2006	0.3008	0.1735	0.1289	0.0988	0.0776	0.0658	0.0573	0.0529	0.0444	0.999675
2007	0.3002	0.1739	0.1226	0.1010	0.0802	0.0675	0.0565	0.0607	0.0480	0.999685
2008	0.2998	0.1727	0.1259	0.0992	0.0808	0.0662	0.0593	0.0524	0.0437	0.999762
2009	0.2988	0.1769	0.1238	0.0991	0.0822	0.0646	0.0577	0.0502	0.0468	0.999763
2010	0.2970	0.1737	0.1263	0.0964	0.0812	0.0693	0.0583	0.0514	0.0464	0.999904
BenfordLaw	0.3010	0.1761	0.1249	0.0969	0.0792	0.0669	0.0580	0.0512	0.0458	1

从表4中可以看出,利润表的整体财务数据与本福德定律理论值的相关系数普遍很高,绝大部分都在0.999以上,与资产负债表的情况有很大不同,造成这种情况的原因可能包括:公司普遍对本文研究时选取的利润表样本(营业收入、营业成本、利润总额、净利润)进行的数据操纵行为较少;整体上,中国上市公司的利润表相比较资产负债表而言,的确受相关公司人为操纵的因素较少。

3. 不同审计意见的上市公司对比。表5到表7分别给出了在不同年份、不同审计意见的中国上市公司的财务报表整体财务数据的首位分布,及其与本福德定律理论值的比较结果。

表5　不同审计意见的中国上市公司在2008年的主要财务数据首位分布

首位数	审计意见		BenfordLaw
	非标准意见	标准意见	
1	0.299 1	0.299 9	0.301 0
2	0.166 0	0.173 1	0.176 1
3	0.130 3	0.126 1	0.124 9
4	0.100 3	0.098 3	0.096 9
5	0.083 9	0.080 4	0.079 2
6	0.066 8	0.066 3	0.066 9
7	0.065 7	0.058 6	0.058 0
8	0.048 2	0.052 8	0.051 2
9	0.039 6	0.044 4	0.045 8
相关系数	0.997 555	0.999 845	1

表6　不同审计意见的中国上市公司在2009年的主要财务数据首位分布

首位数	审计意见		BenfordLaw
	非标准意见	标准意见	
1	0.299 7	0.299 4	0.301 0
2	0.299 7	0.176 6	0.176 1
3	0.119 3	0.124 2	0.124 9
4	0.108 7	0.098 5	0.096 9
5	0.081 8	0.082 0	0.079 2
6	0.067 2	0.064 3	0.066 9
7	0.063 9	0.057 3	0.058 0
8	0.031 9	0.051 3	0.051 2
9	0.056 0	0.046 4	0.045 8
相关系数	0.993 443	0.999 808	1

表7　不同审计意见的中国上市公司在2010年的主要财务数据首位分布

首位数	审计意见		BenfordLaw
	非标准意见	标准意见	
1	0.312 9	0.295 8	0.301 0
2	0.173 5	0.174 3	0.176 1
3	0.108 2	0.126 9	0.124 9
4	0.093 2	0.096 9	0.096 9
5	0.082 3	0.081 6	0.079 2
6	0.080 9	0.068 5	0.066 9
7	0.066 0	0.057 8	0.058 0
8	0.040 1	0.051 9	0.051 2
9	0.042 9	0.046 2	0.045 8
相关系数	0.993 182	0.999 882	1

表5到表7给出了2008～2010年非标准意见公司和标准意见公司的主要财务数据首位分布情况以及与本福德定律的分布值的比较。从表中可以看出,非标准意见公司的首位分布和标准意见公司的首位分布均呈现出本福德定律描述的首位出现概率递减规律。

但是非标准意见公司与本福德定律相关系数分别为0.997 555、0.993 443、0.993 182，标准意见公司与本福德定律相关系数分别为0.999 845、0.999 808、0.999 882，两者存在一定的差异,标准意见公司的首位分布更接近于本福德定律,真实性更高。

4. 财务报表舞弊公司与上市公司整体对比。表8列出了对5家已经被证实曾经实施了财务报告舞弊的上市公司进行财务数据首位分布的测试结果，以及与中国上市公司整体的对比：

表8　已经被证实的5家实施了财务报告舞弊的上市公司测试数据

舞弊公司	首位数						平均值
	2000	2001	2002	2003	2004	2005	
银广夏	0.983 3	0.977 0	0.975 2	0.981 8	0.994 0	0.979 0	0.981 7
西安达尔曼	0.948 0	0.917 7	0.972 3	0.990 8			0.957 2
天津磁卡	0.979 6	0.968 9	0.988 0	0.941 5	0.911 0	0.936 5	0.954 3
蓝田股份	0.981 5	0.930 2	0.938 6				0.950 1
大庆联谊	0.984 9	0.943 7	0.937 9	0.938 4	0.949 3	0.883 1	0.939 6
中国上市公司	0.997 5	0.996 8	0.998 1	0.999 0	0.999 4	0.999 3	0.998 3

从表8可看出，这5家上市公司的相关系数明显低于正常值，由此可为本福德定律在中国上市公司审计工作中应用的可行性做出进一步证实。

四、结论

本文在对本福德定律进行简单介绍的基础上，对中国上市公司自2000年来的主要财务数据进行统计分析并将结果与本福德定律理论值进行比对,得出了一系列结论：

(一)实用性

从表2、表3、表4中十余年来上市公司的整体财务数据来看，本福德定律在中国上市公司的实证测试中表现出良好的契合,即使是在实行新会计准则之后的2008年至今,资产负债表与利润表的平均相关系数也有0.996的水平，这足以说明,本福德定律在我国上市公司的报表审计工作中是可以作为一项审计工具使用的。

(二)辅助审计工作

从表3和表4来看，在资产负债表和利润表的主要财务数据中,营业成本与本福德定律理论值的相关性较高,资本公积与所有者权益合计的相关性相对较低;比较两表的整体数据,资产负债表的相关性较低,利润表的相关系数较高。这对中国审计工作者的工作方向提供了帮助：在对中国上市公司进行审计时,若资源有限,审计人员应为资本公积和所有者权益合计设置更低的重要性水平,以进行更谨慎的审计。

(三)相关系数应用方法

从表5至表7来看，标准审计意见的财务报告的数据与本福德定律理论值的相关系数高于非标准审计意见财务报告;从表8来看，在对已被证实有报表舞弊的5家中国上市公司的实证测试以及与中国上市公司整体的比较中发现,5家上市公司的报表数据与本福德定律的相关系数都低于整体水平。由此我们可得出以下结论:“相关系数与报表舞弊的可能性之间存在负相关关系,即:相关系数越小,那么报表舞弊的可能性越高。”

(四)局限性

本福德定律有着使用方便、可操作性强、对使用者要求的相关工作经验少的优点。但是在实际应用过程中应注意,财务数据与本福德定律理论值的相关系数低并非财务数据舞弊的充要条件。事实上,相关系数较低并不一定意味着财务数据舞弊。应用本福德定律应作为初步审计过程,其相关系数只用于辅助审计人员找出报表可疑点，应在进一步的审计之后再下结论。

主要参考文献

1. 张苏彤，康智慧. 信息时代舞弊审计工具. 审计研究,2007;3

2. 刘福清,陈义文.浅谈奔福德定律在审计中的应用.会计之友,2010;12

3. 李岩.浅析奔福德定律在舞弊审计中的应用.会计与审计,2010;6

4. 狄为,施鹏仙.基于Benford定律的会计舞弊发现研究.会计之友,2010;9

5. Cindy Durtsch. The Effectice Use of Benford´s Law to Assist in Detecting Fraud in Accounting Data. Journal of Forensic Accounting,2004;6

6. 王福胜,李勋,孙逊.奔福德定律及其在审计中的应用研究.财会通讯,2007;3

7. 张苏彤.奔福德定律:一种舞弊审计的数值分析方法.中国注册会计师,2005;11

8. 朱文明,王昊,陈伟.基于Benford法则的舞弊检测方法研究.数理统计与管理,2007;1

9. 许存兴,张芙蓉.上市公司财务舞弊实证分析.天津商业大学学报,2010;3