Detecting Fraud in Foreign Financial Assistance Transactions using Benford’s Law

Benjamin Arancibia

**Abstract**

Transparency of foreign financial assistance is being recognized to be one of the key areas whereby aid effectiveness can be improved. Increased aid transparency allows for outside organizations to begin study and analyze foreign financial assistance transactions. In this paper, the author will use foreign aid transactions that have been published to the International Aid Transparency Initiative to identify transactions that are suspicious and possibly fraudulent. These transactions will be identified using a numerical law about the frequency distribution of leading digits that occur in nature. This paper will also discuss data collection processes and the lack of important fields in the dataset that could result in higher susceptibility of corruption.

**Keywords**

Foreign aid, Benford’s Law, fraud, data collection

**1. Introduction**

*1.1 Background*

Foreign financial assistance also known as Official Development Assistance (ODA) is one of the most important instruments of collaboration between rich nations/organizations and poor nations. In 2009, net disbursements of ODA were $140.2 billion and over $3.76 trillion between 1960 and 2009. For 31 recipient countries ODA was greater than 10% of their GDP in 2008 (Ghosh & Kharas, 2011). There is a growing consensus that aid transparency must be improved to get better aid effectiveness. Part of this improved aid effectiveness is transparency so that data collection and project implementation processes can be improved also so that advanced analytics can be performed transactional information to better under project funded by ODA.

Fraud and error detection in accounting dates back to the 13th century when Europe instituted the practice of bookkeeping. Numerous methods have been created to discover and diagnose fraud with an aim of decreasing negative economic impact. This has resulted in the creation of forensic accounting. Internal auditing, as a sub-genre of forensic accounting, relies on systematic approaches to evaluate and understand a company’s financial statements and corresponding data. One particular methodology that can be applied to understand financial statements and interpret deviations is Benford’s Law which relies on the underlying use of digit distribution across figures in financial statements (Boronico, Harris, & Teplitsky, 2014). This law and methodology can be used and applied to ODA financial transactions.

*1.2 Problem Definition*

This paper will use an analytic methodology to address the following problem:

*Official Development Assistance is often characterized as not effective because it goes to corrupt nations, and as a result, ODA transactions are often described as being corrupt and going to waste. However, no statistical analysis has been performed to find suspicious transactions. This paper will also investigate data quality and how different organization processes can result in ODA transactions being susceptible to fraudulent activity.*

*1.3 Literature Review*

There are currently no papers detailing an analytical approach to identify foreign projects linked to corruption. Most papers deal with the subject in anecdotal manner. Most analytical papers dealing with the subject of foreign aid and international development focus on the transparency of foreign aid amounts and the effect of foreign aid on corruption in the recipient country. Ghosh and Kharas detail in their paper, *The Money Trail: Ranking Donor Transparency in Foreign Aid*, one of the main themes of analytics in foreign aid.

In their paper, Ghosh and Kharas detail how transparency of aid activity is recognized as one of the key areas so that aid effectiveness can be improved. In the paper they propose an index to measure and rank donors on transparency of their aid activities. The authors used the Transparency Index and rates 31 bilateral and multilateral donor agencies on six measure of transparency. They found that being a member of the International Aid Transparency Initiative (IATI) is a powerful signal of a donor being more transparency across other dimensions not measured in the Transparency Index. The authors also did not find any relationship between transparency and donor aid volume values. Overall, the World Bank’s International Development Association (IDA) and Australia are identified as the most transparency donors, while Korea and Inter-American Development Bank Special Fund are the least transparent (Ghosh & Kharas, The Money Trail: Ranking Donor Transparency in Foreign Aid, 2011).

One of the major topics that Ghosh and Kharas describe in their paper is the fragmentation of aid activities. Figure 1 below from Ghosh and Kharas shows the increase in the number of commitments and the decrease in the mean of the commitment size (Ghosh & Kharas, The Money Trail: Ranking Donor Transparency in Foreign Aid, 2011). This might be an advantage in trying to discern possible commitments (foreign aid projects) that are meant for corruption. Collins, Zubairi, Nielson, and Barder estimate that at least $18 billion of aid a year is susceptible to corruption (Collins, Zubairi, Nielson, & Barder, 2009).



Okada and Samreth in their paper investigate the effect of foreign on corruption using a quantile regression method. The authors in their paper describe that foreign aid reduces corruption and its reduction effect is greater in less corrupt countries. Though the authors acknowledge that this effect is different by different donor countries. What distinguishes their paper from other papers in the international development field is that they focus on the effect of foreign aid on corruption in recipient countries. Their conclusion is that foreign aid generally decreases corruption level. Their methodology and variables that they include in their study are not robust enough to make this claim. Okada and Samreth look at Aid (Total), GPP per Capita, Democracy, and English Legal origin as three variables that are tied to corruption. The authors do not account for other indicators which have been shown to have a large impact on corruption in countries (Okada & Samreth, 2011). These indicators include, but are not limited to health levels, education levels, and government institution strength (Kimura & Todo, 2010). Even though Okada and Samreth do not take these variables into account, their paper is important because of analytical approach they take to answering a problem in the international development field.

Benford’s law is a law of mathematics that describes a numerical regularity in real-world numbers expressed in the decimal system. According to Benford, various digits do not occur with the same frequencies. For example, the formula for the first digit is P(d1) = log(1+(1/d1)) with d1 being one of the numerals 1,…,9. The joint distribution of the first and all later digits adhere to the following probabilities: P(D1 = d1,…,Dk = dk) = log[1+ with D1, D2 signifying the first, second, etc. significant digit and di the numerals 0,1,…,9 (j=2,…,k) (Benford, 1938). Benford’s law was then further investigated to study two-digit distribution and Mebane in 2006 argues that the frequencies of the numerals of election counts at precinct level approximate a Benford distribution of the second digit (Mebane, 2006). Breunig and Goerres then used Mebane’s methodology to look at electoral irregularities in Germany. The authors were able to replicate Mebane’s methodology and found irregularities in different levels of the Bundestag elections in Unified Germany (Breunig & Goerres, 2011).

Apart from elections, Benford’s Law has been used extensively in detecting fraud in financial records. In Durtshi, Hillison, and Pacini, the authors describe Benford’s Law and how it can be used to in auditing. The authors show that a digital analysis of records can be effectively used and show where auditors should exercise caution when using Benford’s Law. Finally, the authors identify data sets that can be expected to follow Benford’s distribution, discuss the power of statistical tests, types of frauds that would be detected and not be detected by such analysis, the potential problems that arise when account contains too few observations, as well as issues related to base rate of fraud. Overall, this paper has an in-depth and practical explanation of the analytical technique and explains how to be successful when using it (Durtschi, Hillison, & Pacini, 2004).

Benford’s law has also been applied to drug discovery data. Orita, Moritomo, Niimi, and Ohno demonstrate that several data sets in the field of drug discovery follow Benford’s distribution, whereas ‘doctored’ data do not. Their findings indicate the applicability of Benford’s law in assessing data quality in the field of drug discovery. We also propose a useful index of evaluating data quality based on Benford’s law (Orita, Moritomo, Niimi, & Ohno, 2010). Finally, Kraus and Valverde describe in their paper, *A Data Warehouse Design for the Detection of Fraud in the Supply Chain by Using the Benford’s Law*, how to develop a data warehouse solution that supports forensic analytics to use Benford’s Law to detect fraud. Their application primarily focuses on supply chain management processes such as procurement and inventory management. They used parameterized stored procedures with Dynamic SQL to analyze their supply chain data (Kraus & Valverde, 2014).

Based on the literature review, Benford’s Law and its variation of examining second digit distributions will work for attempting to detect fraud in foreign aid transactions. The importance of this paper is that no other authors have used this analytical technique, nor any other analytical techniques, to attempt to pinpoint fraud in foreign aid.

*1.4 Approach*

The remainder of the work is organized as follows. Section 2 explains the methods used, particularly the fraud detection methodologies used to flag suspicious transaction. Section 3 discusses the implementation of the methodologies and results. Section contains conclusions, processes overviews and their role in suspicious transaction, and areas for further analysis.

**2. Methods**

2.1 Data Collection and Clean Up

Transaction level data from the International Aid Transparency Initiative (IATI) was used for the analysis foreign aid transactions and this data can be easily accessible via IATI’s API or IATI’s datastore, which allows users to query the desired data. When the data is queried and downloaded there are several different qualitative and quantitative fields within the dataset. The full IATI dataset has 74 variables and 471,395 transactions, but many columns are missing values. The percentages of missing values for columns range from 0.07% (transaction value) to 100% (Transaction Recipient Region). Since there are a large amount of variables with missing data, the raw dataset was filtered to only contain columns that are relevant to the investigation. The relevant data columns with description can be seen in Table 1.

|  |  |
| --- | --- |
| Variable | Description |
| Transaction Type | Type of transaction |
| Default Currency | Currency value |
| Transaction Value | Numeric Amount |
| Transaction Value Date | Date of transaction |
| Transaction Provider Organization | Organization providing funding |
| Transaction Receiver Organization | Organization receiving funding |
| Reporting Organization | Organization reporting funding |
| Title | Title of Project that transaction is part of |
| Description | Description of project that transaction is part of |
| Start Planned Date | Planned start date |
| End Planned Date | Planned end date |
| Start Actual Date | Actual start date |
| End Actual Date | Actual end date |
| Recipient Country | Recipient country of transaction |
| Sector Vocabulary | Sector of transaction |

Table 1: Filtered Data Columns

After filtering the data set the transaction amounts were transformed to one currency. There are 18 different currencies, including no currency identified, within the dataset and these were all transformed to 2015 United States Dollars (USD). Those transactions without identified currencies were assumed to be in USD. After performing this the data was then removed of all transactions that have no values, which there are 339. One aspect of the dataset that must be discussed are when multiple countries have been entered into the recipient country variable. As an example, a transaction might have Kenya, Ethiopia, and Uganda as the listed recipient countries. There are two ways to deal with these sort of cases, divide into three equal transaction values or leave it as it is. It was decided to leave these transactions alone because it is not a safe assumption that the transaction values are split evenly.

*2.2 Benford’s Law*

Benford’s Law is the mathematical tenet that says that low digits occur more frequently in large number than high digits. In 1995 Hill, a mathematician, provided a proof for Benford’s law as well as demonstrating how it applied to stock market data, census statistics, and certain accounting data. He noted that Benford’s distribution, like the normal distribution, is an empirically observable phenomenon. Hill’s proof relies on the fact that the numbers in sets that conform to the Benford distribution are second generation distributions, that is, combinations of other distributions. If distributions are selected at ran- dom and random samples are taken from each of these distributions, then the significant- digit frequencies of the combined samplings will converge to Benford’s distribution, even though the individual distributions may not closely follow the law ( (Durtschi, Hillison, & Pacini, 2004). Based on Benford’s Law the probability that a number has any particular non-zero first digit is:

Where d is a number 1,2 … 9, and P is the probability

Using this formula, the probability that the first digit of a number is one is estimated to be 30 percent while the probability the first digit is a nine is 4.6 percent. Formulas for expected digital frequencies are:

First digit of a number:

For second digit of a number:

To calculate the two digit combinations:

D1 represents the first digit of a number and D2 represents the second digit of a number, etc.

**3.0 Results**

Once the data has been transformed it can be analyzed using Benford’s Law via the Benford Analysis package available on CRAN (Cinelli, 2015). This package provides a robust tool for using Benford’s Law to investigate a specific dataset. Its main purposes are to find out where the dataset deviates from Benford’s Law and to identify suspicious data that need further verification.

*3.1 Summary Statistics*

The summary statistics of USD transaction amounts are seen below in Table 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Mean (SD) | Median | Min | Max |
| USD Transaction Amount | 8,811,817 USD (1108042891 USD) | 30,000 USD | - 2,147,483,648 USD | 166,298,000,000 USD |

Table 2: Summary Statistics

Using the Befords Analysis R package and the described filtering, the outputs of this tool are a Chi Square test and the Mantissa Arc Test. The Mantissa Arc Test is a commonly used digit forensic tool looks to see if mantissas are uniformly distributed and if they are uniformly distributed then the result is a perfect circle with a radius of 1 and a center of gravity of (0,0). The following is the output of the Benford Law Analysis Package:

Benford object:

Data: benfords$usd.conversion

Number of observations used = 456111

Number of obs. for second order = 292817

First digits analysed = 2

Mantissa:

Statistic Value

Mean 0.4968

Var 0.0848

Ex.Kurtosis -1.1999

Skewness -0.0087

The 5 largest deviations:

digits absolute.diff

1 10 4461.34

2 50 2738.37

3 20 1950.33

4 30 1872.78

5 15 1774.79

Stats:

Pearson's Chi-squared test

data: benfords$usd.conversion

X-squared = 7442.799, df = 89, p-value < 2.2e-16

Mantissa Arc Test

data: benfords$usd.conversion

L2 = 1e-04, df = 2, p-value < 2.2e-16

Mean Absolute Deviation: 0.001038434

Distortion Factor: -0.6463139

What the output is showing is that the results appear to be significant based on the resulting p-values, which are highlighted. Plotting the results also give a good overview of what is occurring in the dataset, which can be seen in Figure 2 below.

Figure 2: Benford’s Law Analysis

Based on this initial results, the digits by decreasing order of discrepancy are in Table 3.

|  |  |  |
| --- | --- | --- |
| Rank | Digits | Absolute Difference |
| 1 | 10 | 4461.3410 |
| 2 | 50 | 2738.3671 |
| 3 | 20 | 1950.3276 |
| 4 | 30 | 1872.7791 |
| 5 | 15 | 1774.7908 |
| 6 | 17 | 1201.3096 |
| 7 | 40 | 1073.7270 |
| 8 | 13 | 1037.7881 |
| 9 | 11 | 940.7783 |
| 10 | 25 | 804.9066 |

Table 3: Digits by Decreasing Order of Discrepancies

These results show that the first two digits of 10 in that combination appear the most often together and occur more often than they should. Table 4 shows the largest number of duplicate values in the dataset.

|  |  |  |
| --- | --- | --- |
| Rank | Numbers | Duplicate Counts |
| 1 | 100000 | 1517 |
| 2 | 500000 | 959 |
| 3 | 50000 | 884 |
| 4 | 1000000 | 840 |
| 5 | 1 | 752 |
| 6 | 200000 | 739 |
| 7 | 300000 | 646 |
| 8 | 10000 | 598 |
| 9 | 250000 | 519 |
| 10 | 150000 | 503 |

Table 4: Duplicates by Decreasing Order

These results show that the value 100000 appears the most often in the dataset, exactly 1,517 times. This makes sense in comparison to the leading digits that were flagged for discrepancies. The package also is able to get data that is suspicious based on the digits groupings by employing the following command:

suspects <- getSuspects(benford.data, benfords, how.many=2)

suspects

The function results in creating a dataframe with 30,002 observations that are suspicious. When that dataframe is investigated the most frequent suspicious countries can be seen in Table 5.

|  |  |  |
| --- | --- | --- |
| Rank | Country | Counts |
| 1 | Blank | 8182 |
| 2 | Tanzania | 643 |
| 3 | India | 607 |
| 4 | Kenya | 590 |
| 5 | Mozambique | 575 |
| 6 | Uganda | 538 |
| 7 | Ethiopia | 463 |
| 8 | Sudan | 432 |
| 9 | Afghanistan | 427 |
| 10 | Indonesia | 423 |

Table 5: Suspicious Countries by Decreasing Order

By far the most suspicious transactions in the dataset that is flagged by Benford’s Law analysis are transactions with no recipient country. The following analysis is performed for transaction providing organizations and the results can be seen in Table 6.

|  |  |  |
| --- | --- | --- |
| Rank | Organization | Counts |
| 1 | Blank | 20161 |
| 2 | Department for International Development | 1059 |
| 3 | Bill and Melinda Gates Foundation | 1032 |
| 4 | Ministry of Foreign Affairs (DGIS) | 838 |
| 5 | International Development Association | 500 |
| 6 | International Bank for Reconstruction and Development | 441 |
| 7 | Oxfam Novib | 282 |
| 8 | GlobalGiving | 273 |
| 9 | UNICEF (FOR GR Allocations Only) | 265 |
| 10 | United National Development Programme | 246 |

Table 6: Suspicious Organizations by Decreasing Order

As can be seen by the suspicious recipient countries and suspicious organizations, transactions that are missing these key fields have the highest number of counts. To investigate how blank fields effect suspicious transactions, all transactions that had a blank recipient country and blank transaction providing organization were removed. This decreases the total number of transactions studied from 471,056 to 179,786. Benford’s Law analysis provides the below output:

Benford object:

Data: benfords2$usd.conversion

Number of observations used = 172627

Number of obs. for second order = 109627

First digits analysed = 2

Mantissa:

Statistic Value

Mean 0.4931

Var 0.0845

Ex.Kurtosis -1.1861

Skewness -0.0046

The 5 largest deviations:

digits absolute.diff

1 10 2369.50

2 50 1081.38

3 20 883.15

4 30 737.72

5 15 688.49

Stats:

Pearson's Chi-squared test

data: benfords2$usd.conversion

X-squared = 3985.72, df = 89, p-value < 2.2e-16

Mantissa Arc Test

data: benfords2$usd.conversion

L2 = 0, df = 2, p-value = 0.04314

Mean Absolute Deviation: 0.001273255

Distortion Factor: -1.364675

What the output is showing is that the results appear to be significant based on the resulting p-values, which are highlighted. Plotting the results also give a good overview of what is occurring in the dataset, which can be seen in Figure 3 below:



Figure 3: Benford’s Law Analysis Filter Transactions

The complete code and data is available upon request.

**4.0 Conclusions**

References