Benjamin Arancibia

CUNY IS 698

Data Collection and Analysis

**Using Benford’s Law to Detect Fraud in Foreign Financial Assistance Transactions**

**Project Summary**

The main focus of this research is to apply Benford’s law to foreign financial aid transactions and answer the following questions:

* What foreign financial aid transactions have a high likelihood of corruption during project implementation?
* What type of organizations, multilateral or bilateral, have more foreign financial aid transactions flagged as possibly fraudulent?
* What is the distribution of type, private or government expenditure, of transactions that are flagged as fraudulent?

**Methods**

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 United States Dollars (USD). The transformation was done by filtering the transactions by years and then transforming the transaction amount based on the currency exchange rate for that year into 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.