Newcomb-Benford's law helps customs officers to detect fraud in international trade

Lucas Lacasa
School of Mathematical Sciences, Queen Mary University of London, Mile end Road E14NS London (UK)

Email: l.lacasa@qmul.ac.uk
Acknowledgments:
The author's research is supported by an ESPRC Early Career Fellowship EP/P01660X/1
Author contributions:
L.L. wrote the paper. The author declares no conflict of interest.

The leading digit of a number represents its non-zero leftmost digit. For instance, the leading digits of 19 and 0.072 are 1 and 7, respectively. Benford's law (BL) was originally discovered in the late nineteenth century [1,2] as an anecdotal pattern emerging in seemingly as disparate datasets as streets addresses, freezing points of chemical compounds, house prices or physical constants, with the leading digit *d* in those datasets following a logarithmically decaying distribution $P(d)=\log_{10}(1+1/d)$ instead of being uniformly distributed as one may naïvely assume. Later this pattern was shown to be a consequence of a central limit-type mechanism [3,4,5] emerging not only empirically but also in mathematical sequences of several garments. A few years ago, some authors devised a way to leverage BL as an 'anti-fraud' tool [6,7], based on a simple idea: assuming this law is expected to naturally emerge in a certain dataset, then if the dataset has been manipulated or data has been fabricated, the statistics would deviate from the law in a way which could be quantitatively measured. Accordingly, BL and variants have been proposed to assess fraud in contexts ranging from election data [8-11] to financial accounting in external, internal, and governmental auditing [12]. In PNAS, Cerioli et al [13] take this strategy to the next level, and propose a sophisticated statistical modelling framework which can be used to monitor and detect hints of individual fraudulent behavior in the context of international trade (i.e. imports and exports that are declared by national traders and shipping agents). The authors of this work develop a mathematical model which provides the correct statistical tests to assess the conformance to BLs of individual traders, and validate the whole framework in realistic scenarios enabled by high resolution transaction data from the customs of various EU member states.

A naïve approach to flag potential misuse or manipulation of data using BL is the following: under the premise that a particular dataset conforms to a specific theoretical distribution (BL or other variants of the law for e.g. the second leading digit), the strategy is to compare the empirical leading digit distribution found in the actual data against the expected BL. Applying goodness of fit and a contrast of hypothesis one can conclude whether the null hypothesis ('the empirical one conforms to the expected BL') can be rejected up to a certain confidence level. This is typically the approach taken to assess vote count statistics in the context of election forensics [8-11]. Note, however, that methodologies based on BL and variants are not totally free of healthy controversy [15,16], and it's also true that potential fraudsters aware of these regularities can try to make up data in such a way that specific patterns which emerge in fair data also hold even in the manipulated data. Even more dramatically, despite some common misperception that the origin of BL is reasonably well understood, this is actually not the case [17]. In other words, there is no general theory fully guaranteeing that BL should systematically emerge in a particular un-manipulated dataset.

In [13], Cerioli et al are able to circumvent these issues by pre-assessing the conditions under which BL should emerge in the context of international trade, i.e. the conditions under which the subsequent inferences can be trusted. These authors actually pioneer the application of BL to the context of international trade, but why would this be an important area of application? If only, because of the money. Big money. Just to give a sense of scale, the European Union (EU) accounts for around 15% of world's trade in goods, translating in over 1800 billion euros on imports and a similar quantity of exports only in 2017, with a positive balance between the two of over 20 billions [19]. Being able to detect fraudulent behavior --including under-reported goods-- is therefore of paramount importance [14]. Devising a robust and accurate statistical method which could be easily embedded online to automatically monitor transaction activities and flag suspicious ones would tick all the boxes (reliability, flexibility, efficiency, cost-effectiveness). This also aligns with the overall modernization strategy currently pursued by the World Customs Organization [18].

Cerioli et al [13] successfully address the two main challenges that might otherwise preclude a realistic detection of fraud via BL analysis of individual traders:
First, they establish solid conditions for the validity of the BL in the field of international trade data –this being an essential for step for the implementation of large-scale, automatic monitoring processes--. Second, they find approximations and corrections to the adequate test statistics needed to scrutinize fraud in those instance where BL is not actually expected to hold. Importantly, the authors are therefore able to discriminate two cases for non-conformance to BL: those that are related to data fabrication and fraud from the so-called 'false positives': legitimate deviations which emerge for instance for traders who operate on a limited number of products (so that there is not enough variability in their transactions for the BL to emerge in the first place). As a result, traders can be classified into three groups: legitimate traders whose activity conforms to BL, those whose activity is still legitimate but does not conform to BL due to controlled factors, and those that do not conform to BL even if they should, probably due to data fabrication (see Fig.1 for an illustration).

On a technical level, Cerioli et al [13] initially argue that the sequence of transaction values for a particular trader complies with Hill's central limit theorem [3] hypothesis. By generating synthetic transactions of 'idealized traders', the authors can statistically assess the conditions where BL holds, and in what circumstances classical chi-squared tests can therefore be applied. An important result is that BL breaks down as the theoretical expected distribution when the number of different traded goods m is much smaller than the total number of transactions n $(m<<n)$, so in those instances a blind chi-squared test is not recommended.
One of the important novelties of [13] is that they model transaction data via a trader-specific contamination model composed by linearly interpolating a mixture of two distributions for the leading digit: a 'legitimate' one (usually BL) which is

parametrically dependent on *n* and *m*, and a contaminant distribution that models the effect of data manipulation. The null hypothesis (legit trader) is equivalent to having a null coefficient on the contaminant distribution, however this scenario is only equivalent to conformance to BL when the transaction data fulfil the criteria discussed above. In those cases where the expected distribution for non-fraudsters deviates from BL for legitimate reasons, the authors can plug in the actual expected distribution as the legitimate one, and can deduce –among other things-- what is the correct distribution of the test statistic which shall be used to reject the null hypothesis of a legit transaction pattern.

Validation and calibration of the theoretical framework was possible thanks to the access to real-data of import/exports declared by national traders and shipping agents using the form called Single Administrative Document (SAD) (data was provided by Italian customs and by the Customs Office of another EU member state not disclosed for its specific confidentiality policy). Results overall show that Cerioli et al's methodology [13] can flag fraudsters and 'de-flag' traders whose activity deviates from expected BLs due to legitimate reasons. The authors discuss a particularly illuminating case of a trader extracted from an archive of fraudulent declarations provided by the Italian Customs, whose fraudulent behavior was only discovered after substantial investigation on two of the declarations concluded they were fraudulent. In such case, a standard protocol based on robust regression techniques aiming at the automatic detection of value frauds in customs data did not provide clear evidence of substantial undervaluation or of other major anomalies. Cerioli et al's analysis on the other hand produced a strong signal of contamination of the digit distribution for this trader, and their statistical analysis safely concluded the presence of fraudulent manipulation.

In conclusion, Cerioli et al's [13] provides a principled framework for goodness-of-fit testing of BL for anti-fraud purposes, with a focus on customs data. This methodology has the potential to be embedded –probably in combination with more standard and model-free approaches [6,12]-- in real international trade anti-fraud protocols and audits in the near future. On this respect, a web application [13] developed with the purpose to assist customs officers and auditors in the screening task has already been set in place.

References

[1] Newcomb, S. (1881) Note on the frequency of use of the different digits in natural numbers. *Am. J. Math.* 4, 39–40.

[2] Benford, F. (1938) The law of anomalous numbers. *Proc. Am. Philos. Soc.* 78, 551–572.

[3] Hill, T. P. (1995) A statistical derivation of the significant-digit law. *Stat. Sci.* 10, 354–363.

[4] Berger A, Hill TP (2015) *An Introduction to Benford's Law.* (Princeton Univ. Press, Princeton).

[5] Miller SJ, ed. (2015) *Benford's Law: Theory and Applications.* (Princeton Univ. Press, Princeton).

[6] Kossovsky AE (2015) *Benford's Law: Theory, The General Law Of Relative Quantities, And Forensic Fraud Detection Applications.* (World Scientific, Singapore).

[7] Lacasa L, Fernández-Gracia J (2018) Election Forensics: quantitative methods for electoral fraud detection, *Forensic Science International* (in press, arXiv:1811.08502)

[8] Mebane, WR (2006) Election Forensics: Vote Counts and Benford's Law (Political Methodology Society, University of California, Davis, CA)

[9] Pericchi L., Torres D. (2011) Quick anomaly detection by the Newcomb-Benford Law, with applications to electoral processes data from the USA, Puerto Rico, and Venezuela. *Stat. Sci.* 26, 502-516 (2011).

[10] Fernández-Gracia J, Lacasa L. (2018) Bipartisanship Breakdown, Functional Networks, and Forensic Analysis in Spanish 2015 and 2016 National Elections, *Complexity* 9684749
[11] Breunig C., Goerres A. (2011) Searching for electoral irregularities in an established democracy: Applying Benford's law tests to Bundestag elections in Unified Germany, *Electoral Studies* 30 pp.534-545.
[12] Nigrini MJ (2012) *Benford's Law.* (Wiley, Hoboken).
[13] Cerioli A., Barabesi L., Cerasac A., Menegatti M., and Perrotta D. (2018) The Newcomb-Benford Law and the detection of frauds in international trade, *Proc. Natl. Acad. Sci. USA* XXX
[14] European Commission (2014) Operation SNAKE: EU and Chinese customs join forces to target undervaluation of goods at customs (Press release IP-14-1001. URL:http://europa.eu/rapid/).
[15] Deckert JD, Myagkov M, Ordeshook PC (2011) Benford's Law and the detection of election fraud. *Polit Anal* 19 pp.245-268.
[16] Mebane WR (2011), *Comment on "Benford's Law and the detection of election fraud*", *Polit Anal* 19 pp.269-272
[17] Berger A, Hill TP (2011) Benford's law strikes back: no simple explanation in sight for mathematical gem. *Mathematical Intelligencer* 33:85–91.
[18] World Customs Organization (2017) Message from the WCO Secretary General (URL: http://www.wcoomd.org/en/about- us/international-customs-day/icd-2017.aspx).
[19] International trade in goods - a statistical picture (URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International_trade_in_goods_-_a_statistical_picture).

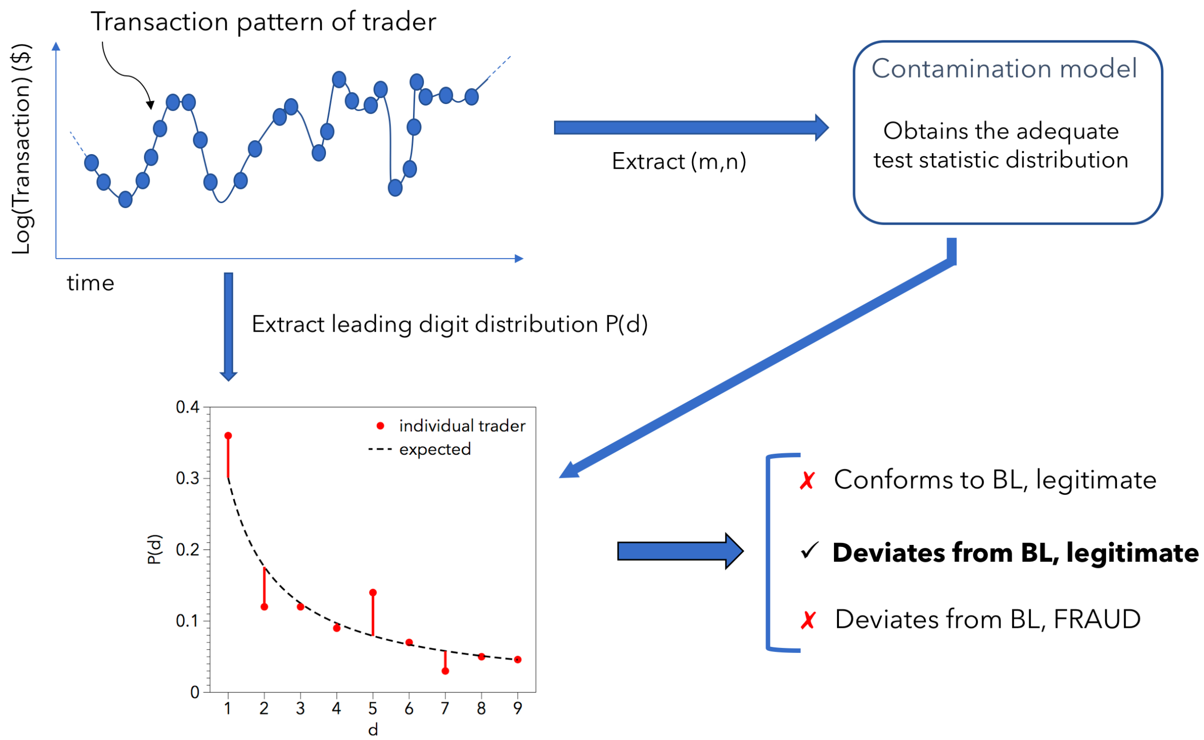

Fig.1 An individual trader leaves a trace of transaction activities. The method computes the leading digit distribution P(d) and (i) evaluates which is the expected law if no manipulation has occurred, and (ii) obtains the adequate test statistic distribution. Statistical comparison between the empirical and expected distribution concludes whether the data conform or deviate from BL, and in the latter case, whether this deviation is due to legitimate reasons or not.