European Commission

Benford's Law Conference

10-12 July 2019 - Stresa, Italy

Book of Abstracts

Joint Research Centre

Book of Abstracts

Table of contents

Wednesday 10 July 2019

10:00 – 10:30

Position Talk — Chair: W. Kleinegris

Scientists and Practitioners: the JRC and Italy's Customs on Fraud Detection

Marco De Andreis
Italian Customs Agency, Italy

ABSTRACT: This talk takes stock of a collaboration between the JRC scientists and Italy's Customs Agency practitioners on innovative methods for fraud detection – a collaboration which has been going on for nearly a decade. To exploit the great potential offered by these methods, EU customs organizations should avoid internal compartmentalization or, even better, the EU should overcome the current fragmentation of its customs policy among 28 different national actors.

Wednesday 10 July 2019

11:00 – 12:00

Keynote Talk — Chair: A. Berger

Benford's Law and Detection of Anomalies in Data

Theodor P. Hill

School of Mathematics, Georgia Institute of Technology, USA
Email: hill@math.gatech.edu

ABSTRACT: This talk will consist of three parts. First, a very brief review of the history of Benford's Law (BL) and both classical and recent empirical evidence of it will be presented. Second will be an overview of recent applications of BL to detect fraud in both financial and non-financial data, and then recent applications to detect non-fraudulent anomalies in data from processes ranging from the physical sciences to music, medicine, and data encryption. (This broad collection of the use of BL to detect abnormal patterns in non-financial data hopefully will suggest new ideas for additional BL applications for detection of financial fraud.) The last and main part of the talk will focus on several theoretical aspects of BL, with focus on several main probability theorems as well as common probability errors in applications of BL theory

KEYWORDS: Benford's Law, fraud detection, data anomalies

Wednesday 10 July 2019

12:10 – 13:10

Keynote Talk — Chair: S. J. Miller

A Review of the Forensic Accounting Benford Applications and Future Directions

Mark Nigrini

West Virginia University, Morgantown, WV, USA
Email: ForensicAnalytics@gmail.com

ABSTRACT: The stream of accounting and finance related applications of Benford's Law started by evaluating whether managers tend to round up their income numbers when they are just below a psychological threshold. The current forensic applications are summarized with some reservations for certain applications and a green light for selected others

KEYWORDS: Benford's Law, forensic accounting, financial statement fraud.

Wednesday 10 July 2019

15:30 – 16:30

Position Talk — Chair: D. Perrotta

Big data, Benford & the Allegory of the Cave: Strengths and Weaknesses of Profiling for the purpose of preventing and detecting fraud

Winfried Kleinegris

European Commission, European Anti-Fraud Office (OLAF), Belgium

ABSTRACT: The collection and processing of data in research for policy purposes have always been impacted by reliability and validity issues: data are collected following the elaboration of constructs or concepts (proxies) that were previously defined in the framework of a model or theory that simplifies reality in an attempt to explain it. The quality of a model (or theory) depends on the extent to which that model can satisfactorily explain past events or successfully predict future ones. Increasing the amount of data for analyses without due examination of the reliability (or accuracy) of the data or the validity of the concepts and constructs that underpin the data collection and input, does not necessarily improve the quality of the predictions or explanations of past behaviour. Biases in the results of data analyses can be traced back to several types of shortcomings: Coding errors account for a high number of often embarrassing mistakes (accuracy or reliability). However, a serious problem arises when there is a mismatch between the phenomenon under investigation and the definition of the constructs elaborated for its measurement or quantification (validity). This mismatch is even exacerbated if the amount of data used for testing hypotheses is increased or if intermediate results are used as inputs to feed subsequent analyses that turn into an 'analyses loop', without further human checks to assess whether or not the results 'make sense'. Experiments with profiling in law enforcement, such as predictive policing, have shown –sometimes spectacular- short-term results but have equally led to a substantial increase of false positives which eventually cancelled out the initial savings in manpower, not to mention the breach of certain legal principles. Benford's law is an example of a proxy that should only be used in combination with a larger set of red flags: descriptions of symptoms of irregularities and fraud. Any findings from the application of Benford's law should be the basis for further investigative work: it is a shadow on the wall but does not replace a complex, multifaceted and multidimensional reality

Wednesday 10 July 2019

15:30 – 16:30

Keynote Talk — Chair: T. Verdonck

Robust Modern Multivariate Data Analysis

Marco Riani
University of Parma, Italy

ABSTRACT: Data rarely follow the simple models of mathematical statistics. Often, there will be distinct subsets of observations so that more than one model may be appropriate. Further, parameters may gradually change over time. In addition, there are often dispersed or grouped outliers which, in the context of international trade data, may correspond to fraudulent behavior. All these issues are present in the datasets that are analyzed on a daily basis by the Joint Research Centre of the European Commission and can only tackled by using methods which are robust to deviations to model assumptions.
This distance between mathematical theory and data reality has led, over the last sixty years, to the development of a large body of work on robust statistics. In the seventies of last century it was expected that in the near future "any author of an applied article who did not use the robust alterative would be asked by the referee for an explanation". Now, a further forty years on, there does not seem to have been the foreseen breakthrough into the wider scientific universe. In this talk, we initially sketch what we see as some of the reasons for this failure, suggest a system of interrogating robust analyses, which we call "monitoring" and describe the robust and efficient methods which are currently used by the Joint Research Centre of the European Commission to detect model deviations, groups of homogeneous observations, multiple outliers and/or sudden level shifts in time series

Wednesday 10 July 2019

17:10 – 17:40

Invited Talk — Chair: C. Macmillan

Benford's Law and the Detection of Frauds in International Trade

Andrea Cerioli[1], Lucio Barabesi[2], Andrea Cerasa[3], Mario Menegatti[1], Domenico Perrotta[3]

1 Department of Economics and Management, University of Parma, Italy

2 Department of Economics and Statistics, University of Siena, Italy

3 European Commission, Joint Research Centre (JRC), Italy

ABSTRACT: We develop a principled framework for goodness-of-fit testing of Benford's law in customs data. Our approach relies on a trader-specific contamination model, under which fraud detection has close connections with outlier testing.

KEYWORDS: Anomaly detection, Benford's law, anti-fraud, customs fraud

Wednesday 10 July 2019

17:40 – 18:10

Invited Talk — Chair: M. De Andreis

Detecting Manipulations in Accounting Data of EU Organizations

Caroline Gasparro[1], Francesca Torti[2]

1 Fincons, Italy

2 European Commission, Joint Research Centre (JRC), Italy

ABSTRACT: Benford's Law has long been a tool used by forensic accountants to detect and uncover fraud, errors or biases in reported financial statements and accounting data. Typically, the Benford analysis is applied to financial statement data of individual companies to uncover whether that specific company might have altered its finances. In this work, we follow a different approach: we examine nearly 60,000 Italian companies and apply the Benford analysis to all individual voices of the financial statements to identify which voices are Benford or not Benford compliant. The final goal is to indicate to auditors which items in financial statements are most sensible to manipulations. In a second step, we extend this work and apply the Benford analysis to the financial statements of a number of European political parties and related organizations. The analysis was carried out with a SAS code that we have developed, which will subsequently be made publicly available through an interactive and practical web application for the Benford analysis of data

Wednesday 10 July 2019

18:10 – 18:30

Contributed Talk A — Chair: D. Lemons

Several Common Probability Distributions Obey Benford's Law

Guojun Fang[1], Qihong Chen[2]

1 Institute of Scientic Computation and Financial Data Analysis, China

2 School of Mathematics, Shanghai University of Finance and Economics, China

ABSTRACT: Benford's law states that the frequency of lower rst signicant digits(FSD) is higher than that of upper FSD in many naturally produced numbers. About 30% are 1's and only less than 5% are 9's.The law can be applied to many dierent elds,so it's important to know which common probability distributions obey Benford's law.In this paper. We revisited whether the Log-normal probability distribution obeys the law by using fourier analysis and numerical simulation. Moreover,we also used simulation method to judge whether the Weibull distribution and Inverse gamma distribution are almost close to Benford's law under some conditions. Because these distributions are widely seen in natural phenomena,so our work give some reasons why Benford's law is universe in the real world.

KEYWORDS: Benford's law, Log-normal distribution, Weibull distribution, Inverse gamma distribution

Wednesday 10 July 2019

18:10 – 18:30

Contributed Talk B — Chair: E. Sordini

Property Appraisals and Loan Performance

Hong Lee[1], James E. Larsen[1], Joseph R. Mason[2]

1 Wright State University, United States

2 Louisiana State University and Wharton Financial Institutions, United States

ABSTRACT: The present paper analyzes whether Bedford's Law and cluster analysis can be used to isolate appraisal manipulation related to the mortgage crisis. We find that the frequency of the leading (left-most) digit of a appraisals varies significantly from Benford's distribution. Cluster analysis using the three digits provides a crucial separation that is associated with riskier mortgages. One-digit Benford analysis and two- and three-digit appraisal clusters are associated with increased rates of Serious Delinquency and Early Payment Default. The effects are robust to controlling for loan characteristics like credit scores, high loan to value ratios, high debt to income ratios, junior liens, interest only payments, negative amortization, balloon payments, cash out refinancing or exotic interest-only, negative amortization, and balloon mortgages originated in 2006 and 2007., as well as time-varying risks and feature bias.

KEYWORDS: mortgage, defaults, appraisal, Benford's Law, clustering, feature bias, mortgage crisis

Thursday 11 July 2019

09:30 – 10:30

Keynote Talk — Chair: M. J. Nigrini

Benford's law, or, Why the IRS and Others Care about Number Theory

Steven J. Miller
Williams College, United States

ABSTRACT: Benford's Law of Digit Bias is observed in many diverse systems, and is a powerful tool to detect data fraud. For such applications, though, it is necessary to know that the data should follow Benford's Law. We discuss some useful techniques for proving a system is Benford, especially Fourier Analysis (especially the Poisson Summation Formula), Irrationality Exponents, and the Central Limit Theorem. We apply these methods to products and compositions of random variables, decomposition processes, and recurrence relations, to name a few. We also discuss generalizations to certain dependent random variables.

KEYWORDS: Benford's law, Fourier Analysis, Poisson Summation, Irrationality Exponents, Central Limit Theorem, Dependent Random Variables

Thursday 11 July 2019

10:35 – 11:35

Keynote Talk — Chair: D. Perrotta

The General Law of Relative Quantities and Digital Development Forensics

Alex Ely Kossovsky
United States

ABSTRACT: The Benford phenomenon is found to be essentially quantitative and physical, not digital or numerical, constituting a scientific reality independent of our arbitrarily invented positional number system. This is termed 'The General Law of Relative Quantities', and it is expressed algebraically as **ln((D + d(F – 1))/(D + (d – 1)(F – 1)))/ln(F)**. Classic Benford's Law, namely **LOG(1 + 1/d),** is then demonstrated to be merely a consequence and a special case of this more general law. In the field of data fraud detection, a new forensic algorithm is presented, enabling us to detect fraud even when the sophisticated cheater is aware of the law and invents numbers accordingly. The algorithm employs a subtle inner digital pattern of development within the Benford's pattern itself, and which is found to be nearly universal, being more prevalent than the Benford phenomenon itself, applicable even to non-Benford types.

KEYWORDS: Benford's law, digit distribution, quantitative configuration, data skewness, data symmetry, relative quantities, positional number system, data fraud, data forensics, digital development, localized digit configuration, sophisticated data fraud

Thursday 11 July 2019

12:00 – 12:40

Invited Talk — Chair: P. Filzmoser

Instance-Dependent Cost-Sensitive Logistic Model for Detecting Transfer Fraud

Höppner Sebastiaan[1] , Baesens Bart[2], Petrides George[3], Verbeke Wouter[3], Verdonck Tim [1,4]

1 KU Leuven, Department of Mathematics, Belgium

2 KU Leuven, Faculty of Economics and Business, Belgium

3 Vrije Universiteit Brussel, Departement of Business Informatics and Data Analytics, Belgium

4 University of Antwerp, Department of Mathematics, Belgium

ABSTRACT: Credit card fraud is a growing problem that affects card holders around the world. Every year billions of Euros are lost worldwide due to credit card fraud. Financial institutions are, therefore, forced to continuously improve their fraud detection systems and they do so by increasingly relying on predictive machine learning models. The aim of detecting transfer fraud is to identify transactions with a high probability of being fraudulent. From the perspective of machine learning, the task of predicting the fraudulent nature of transactions can be presented as a binary classification problem. Different solutions for detecting fraud are then commonly evaluated based on some sort of misclassifiction measure, and do not take into account the actual financial costs associated with the fraud detection process. Fraud detection, however, is a typical example of cost-sensitive classification, where the costs due to misclassification vary between instances. Nevertheless, current transfer fraud detection algorithms often miss to include the real costs associated with credit card fraud. Based on an instance-dependent cost matrix for transfer fraud detection, a cost measure is introduced that represents the monetary gains and losses due to the classification of credit transfers. We present a classifier that minimizes this instance-dependent cost measure directly into the model construction during the training step, where the classifier's interior model structure resembles a lasso-regularized logistic regression. Moreover, using the cost matrix, an instance-dependent threshold can be derived that allows for making the optimal cost-based decision for each transfer. As an illustration, we apply the technique on two publicly available binary imbalanced datasets on credit card fraud and compare our proposed method against existing credit card fraud detection models.

KEYWORDS: Cost-sensitive classification, fraud detection, lasso-regularized logistic regression, cost-based model evaluation

Thursday 11 July 2019

12:40 – 13:10

Invited Talk — Chair: A. Cerioli

A Robust Clustering Approach to Fraud Detection

Luis A. García-Escudero, Alfonso Gordaliza, Agustín Mayo-Iscar
Departamento de Estadística e I.O. e IMUVA, Universidad de Valladolid, Spain.

ABSTRACT: An overview of different robust clustering procedures, which allow for simultaneous detection of clusters and outliers, will be given. This approach could be useful when data is clustered in groups due to unknown latent factors. We will focus on a trimming-based methodology, where trimmed units are the ones estimated most likely to correspond with fraud transactions. Some extensions to clustering around linear subspaces and to robust functional clustering will be shown.

KEYWORDS: Cluster Analysis, robustness, outliers, trimming

Thursday 11 July 2019

15:00 – 15:30

Invited Talk — Chair: A. E. Kossovsky

Thermodynamics of Benford's Law

Don S. Lemons
Bethel College, United States

Abstract: I derive Benford's first digit from an equlibrium model of a conserved quantity that is randomly fragmented and combined into pieces of different sizes. Each set of pieces of one size are in equilibrium with other sets of pieces of other sizes. An equilibrium parameter that is structurally identical to a temperature regulates the equilibrium between pieces of different sizes and produces Benford's law. The key result of the derivation is an entropy function $S = \ln(\frac{e^2 N}{I^2})^I$ where I is the number of piece sizes, N the number of smallest fragments, and e the base of the natural logarithm. The argument $(\frac{e^2 N}{I^2})^I$ is the number of points on or near the surface that defines the conserved quantity.

KEYWORDS: Benford's law, thermodynamics, statistical mechanics

Thursday 11 July 2019

15:30 – 16:00

Invited Talk — Chair: M. Riani

Statistical (Local) Data Depths and Segmentation of International Trade

Claudio Agostinelli, Giacomo Fracisci

Department of Mathematics, University of Trento. Italy

ABSTRACT: We introduce a new class of statistical (local) data depths which are designed for distributions in an Euclidean space which are symmetric with respect to subspaces. We apply these new concept on the segmentation of international trade data and we illustrate the procedure by looking for substructure in the relation between weight and value.

KEYWORDS: Clustering, Outliers detection, Statistical Data Depths

Thursday 11 July 2019

16:00 – 16:40

Contributed Talks C — Chair: A. Cerasa

An Approach for Dealing with Imbalanced Data in Fraud Detection

Höppner Sebastiaan[1], Ortner Iréne[2], Verdonck Tim[1], Baesens Bart[3]

1 KU Leuven, Department of Mathematics, Belgium

2 Applied Statistics, Austria

3 KU Leuven, Faculty of Economics and Business, Belgium

ABSTRACT: We introduce a new class of statistical (local) data depths which are designed for distributions in an Euclidean space which are symmetric with respect to subspaces. We apply these new concept on the segmentation of international trade data and we illustrate the procedure by looking for substructure in the relation between weight and value. A major challenge when trying to detect fraud is that the fraudulent activities form a minority class which make up a very small proportion of the dataset. Detecting fraud in an imbalanced dataset typically leads to predictions that favor the majority group, causing fraud to remain undetected. Popular over-sampling techniques like SMOTE [1] and ROSE [2] solve the problem of imbalanced data by creating synthetic samples that mimick the minority class. However, when outliers are present in the data these techniques are prone to introduce synthetic samples which distort the detection algorithm. We present a new sampling technique, called rob-ROSE, which combines several promising approaches to elevate the problems of imbalanced data and outliers in fraud detection. The proposed method achieves to enhance the presence of the fraud cases while ignoring contaminated observations, leading to superior fraud detection

KEYWORDS: Classification problem, over-sampling, misclassification cost

REFERENCES:

[1] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[2] Giovanna Menardi and Nicola Torelli. Rose: random over-sampling examples. *Data Mining and Knowledge Discovery*, 28(1):92–122, 2014.

Average, Minimal and Maximal Significands

Zoran Jasak
NLB Banka d.d., Bosnia Herzegovina

ABSTRACT: Calculation of theoretical values of average signi cands is presented by use of simple property of logarithmic function, with narrow connection to sum invariance property. Another topic is recurence formula for minimal and maximal significands for interval [d; d + 1).

KEYWORDS: Clustering, Outliers detection, Statistical Data Depths

Thursday 11 July 2019

16:00 – 16:40

Contributed Talks D — Chair: R. Stadler

The TC Data 360 Dataset and the Benford's Law

Francesco Porro

Università degli Studi di Milano-Bicocca, Italy,

ABSTRACT: The Benford's law is a very well-known distribution observed for the first time by Simon Newcomb [3] and some decades later, popularized by Frank Albert Benford [1]. They discovered that in many datasets that satisfy some particular features, the first significant digits are not uniformly distributed as one can imagine, but they follow a different law (the same occurs for the second, the third and the fourth significant digits). The theoretical framework of these distributions has been studied in many books and in many papers, for example in [2]. Also the applications are numerous and related to different fields. In this presentation, the TC data 360 dataset is considered (dataset available at the URL: https://tcdata360.worldbank.org/). This is a very huge amount of data that is collected and directly provided by The World Bank. It contains 2400 indicators evaluated on more than 200 countries from all over the world. Among many others, there are variables related to the trade (both the "classical" one and the more recent e-trade), the competitiveness, the economic investments, the innovation, the economic indicators. The aim of this presentation is to consider some of these variables, to analyze them by exploring their relationships with the Benford's law. The fitting of the distributions of the selected variables to the Benford's law is assessed by the most current statistical tools, used in goodness-of-fit procedures

REFERENCES:

[1] Frank Benford. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4):551–572, 1938.

[2] Arno Berger and Theodore P. Hill. A basic theory of benfords law. *Probability Surveys*, 8:1–126, 2011.

[3] Simon Newcomb. Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematic*, 4(1):39–40, 1881.

Application of Benford's Law in the e-Audit Tool„ SRP – Summary Risk Profiling" of the German Tax Administration – More than 11 Years of experience in the Field

Tobias scholz, Björn Golnick

Tax Authority of Schleswig-Holstein, Germany

ABSTRACT: For more than 11 years German Tax Auditors have been using Benford's law assessing bookkeeping data and data from electronic cash registers or other sub-systems. The NBL (Newcomb-Benford-Law) is inter alia used to examine daily cash revenue. Sums thought up by the taxpayer can be detected. To determine the degree of trustworthiness the chi²-test is applied. The tests are standardized and embedded in an excel template called "SRP – Summary Risk Profiling". Once the data is stated reliable (no complete manipulation is detected by Benford's law) other analyzes on patterns are applied. The SRP System combines different tests also on economical data. Today this anti-fraud tool is also used by customs (customs fraud) and by other countries.

KEYWORDS: Benford's law, SRP, chi²-test, structure analysis, application

Empirical Evidence of Financial Statement Manipulation in Going Concern and Bankrupted Companies

Barney Bradley[1], Kurt Schulzke[2], Pier Luigi Marchini[3]

1 University of Utah, School of Medicine, United States

2 University of North Georgia, Mike Cottrell College of Business, United States

3 University of Parma, Italy

ABSTRACT: This study uses comparative archival data on going concern and bankrupt companies to discover whether and to what extent Benford's Law metrics signal financial distress in advance of bankruptcy.

KEYWORDS: Benford's law, Financial statement, Bankruptcy

Friday 12 July 2019

09:30 – 10:30

Keynote Talk — Chair: T. P. Hill

Digits and Dynamics – Exploring Benford's Law Through Dynamical Systems

Arno Berger
University of Alberta, Edmonton, Canada

ABSTRACT: Dynamical systems often arise from difference or differential equations that provide basic models for many important processes in science, engineering, and beyond. The numerical data generated from such models pervade present-day society. Do these data conform to Benford's law in one way or another? And if they do, does this in turn say anything interesting about the underlying dynamical system? Proceeding leisurely through a panorama of intriguing yet elementary examples, this lecture will answer both questions in the affirmative. Challenging questions in probability, number theory, and dynamics, some open, some recently resolved, are going to emerge naturally along the way. Though mathematical in nature, the lecture is aimed at the non-specialist.

KEYWORDS: Benford's law, first-digit law, linear system, resonance

Friday 12 July 2019

10:30 – 11:00

Invited Talk — Chair: L. Barabesi

Benford's Law Between Number Theory and Probability

Rita Giuliano

Dept. of Mathematics, University of Pisa, Italy. The EC's Support is acknowledged.

ABSTRACT: In this talk we propose two interpretations and generalizations of Benford's Law pertaining to two different fields of Mathematics: Number Theory and Probability (so indicating that these two fields are not so far apart as they are usually felt). Let P is the set of prime numbers. In terms of conditional density with respect to P, Benford's Law can be stated as follows. If Aq the set of integers the first digit of which is q, then the conditional logarithmic density of Aq, given P, is equal to its (non-conditional) logarithmic density, i.e. to the number log (1+1/q). In the paper [1] we discuss an extension of the previous result in two directions, i.e. we prove that Aq can be replaced by any set A belonging to a rather large class of sets and show that, in conditioning, P can be replaced by any "regular"set H (during the talk we shall explain what "regular"means). From the point of view of Probability Theory, we can say that Benford's Law predicts the probability distribution of all digits in everyday-life numbers. As Benford himself noticed, the greater the number of sources of the data, the better the mantissae of these data fit his law. So imagine that whenever we have a huge amount of data coming from numerous origins, their mantissae are distributed according to a fixed distribution μ. Assume now that the data come from a random variable X distributed on some interval [0;A]. Since the maxima themselves come from various origins, we expect that the mantissae of the maxima, too, conform to μ. In the paper [2] we observe that a random variable X on some interval [0;A] can be seen as the product of A with a random variable Y on [0;1]. Regarding as a random variable, we translate into mathematical terms the requirement that the mantissae of X and of A follow the same law, obtaining that this law has to be Benford's Law. This approach can be linked to the characterization of Benford's Law by scale-invariance, as well as to the convergence of a product of independent random variables to Benford's Law.

KEYWORDS: Benford's law,conditional density, mantissa

REFERENCES:

[1] Giuliano Antonini, R., Grekos, G., (2005), Regular sets and conditional density: an extension of Benford's Law, Coll. Math, 103, 2, 173–192. https://www.impan.pl/en/publishing-house/journals-andseries/colloquium-mathematicum/all/103/2/87871/regular-sets-andconditional-density-an-extension-of-benford-s-law

[2] Giuliano, R., Janvresse, ´ E., (2010), A unifying probabilistic interpretation of Benford's Law, Unif. Distrib. Theory 5, 2, 169–182.. https://math.boku.ac.at/udt/vol05/no2/91GiulJan10-2.pdf

Friday 12 July 2019

11:30 – 12:10

Invited Talk —
Chair: L. A. García-Escudero

Fraud Detection in the Digital Music Industry

Peter Filzmoser, Nermina Mumic, Radostina Kostadinova

TU Wien (Vienna University of Technology), Austria

ABSTRACT: The digital music industry is getting more and more important, and this implies that data from music sales also get bigger. It is practically impossible for the artists to check for the correctness of their sales figures, and thus there is an urgent need for tools to detect eventual frauds. Such frauds could originate from different platforms, but it would be unlikely that fraud happens at the same time on all platforms. Thus the idea is to design a fraud detection method which is making use of the (log-)ratios between the values of the different platforms. Potential fraud would happen if the log-ratios are large.

KEYWORDS: Compositional data, robust filtering, outlier detection

Friday 12 July 2019

12:10 – 12:40

Invited Talk — Chair: F. Greselin

Denoising and Trimming for Improved Cluster Solutions with Applications to Customs Frauds

Andrea Cerioli[1], Luis A. García-Escudero[2], Alfonso Gordaliza[2], Agustín Mayo-Iscar[2], Domenico Perrotta[3], Marco Riani[1], Francesca Torti[3]

1 Department of Economics and Management, University of Parma,, Italy

2 Departamento de Estad´ıstica e I.O. e IMUVA, Universidad de Valladolid, Spain

3 European Commission, Joint Research Centre (JRC), Italy

ABSTRACT: The ComExt Extra-European trade database gives statistics on merchandise trade. It is interesting to identify anomalous patterns in datasets included in this database, which could be useful for detecting fraud. Robust clustering procedures can be applied for addressing this aim. TCLUST methodology provides robust clustering procedures in a variety of statistical settings. There are versions of TCLUST available for the identification of clusters of data around linear subspaces. These proposals work when the sample contains pointwise contamination linked to high leverage points, which are the most harmful. An additional non-standard issue appears when applying clustering procedures to data sets which include a very high proportion of concentrated contamination, possibly greater than 50%, without containing any information related to the assumed existing groups. TCLUST proposals, modified by the inclusion of thinning, are able to downweight successfully the influence of these observations in the estimation process. Statistical procedures around TCLUST methodologies have well known statistical properties including robustness. These procedures have proven to be useful in the analysis of ComExt data.

KEYWORDS: Cluster Analysis, robustness, outliers, trimming

Friday 12 July 2019

12:40 – 13:10

Invited Talk — Chair: F. Torti

Use of Benford's law in the Case Selection of Ex-Post Clearance Customs Audits in Austria

Robert Stadler[1], Andrea Cerasa[2]

1 Predictive Analytics Competence Center, Austrian Ministry of Finance, Austria

2 European Commission, Joint Research Centre (JRC), Italy

ABSTRACT: Customs duties are collected in the interest of both the EU and Member States. The protection of the EU budget is one of the major responsibilities of the national customs administrations. Analyses using Benfords Law (BL) are a good tool to identify fake customs values. But selection methods based only on anomalies detected on the basis of a single statistical criteria, including BL, are risky because they can generate many false positives if the deviations are motivated by normal market conditions. For this reason, Austrian customs used the results of Benford analyses as an input variable in predictive models among many others. The advantage of this approach is that it does not give too much weight on anomalies according to BL when there is a good reason apart from fraud. Our contribution will illustrate the approach and show concrete findings in a real operational context.

KEYWORDS: Benford's law, anti-fraud, customs fraud

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub