Chaos, Solitons and Fractals 90 (2016) 8–17

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Regularities and discrepancies of credit default swaps: a data science approach through Benford's law

Marcel Ausloos [a,c,*], Rosella Castellano [b], Roy Cerqueti [b]

[a] School of Management, University of Leicester, University Road, Leicester LE1 7RH, UK
[b] Department of Economics and Law, University of Macerata, via Crescimbeni 20, I-62100, Macerata, Italy
[c] GRAPES, Beauvallon Res., rue de la Belle Jardiniere, 483/0021, B-4031, Liege Angleur, Euroland

ARTICLE INFO

Article history:
Received 29 September 2015
Revised 18 January 2016
Accepted 2 March 2016
Available online 29 March 2016

Keywords:
Benford's law
Credit default swaps
Complex systems
Data science

ABSTRACT

In this paper, we search whether the Benford's law is applicable to monitor daily changes in sovereign credit default swaps (CDS) quotes, which are acknowledged to be complex systems of economic content. This test is of paramount importance since the CDS of a country proxy its health and probability to default, being associated to an insurance against the event of its default. We fit the Benford's law to the daily changes in sovereign CDS spreads for 13 European countries – both inside and outside the European Union and European Monetary Union. Two different tenors for the sovereign CDS contracts are considered: 5 years and 10 years – the former being the reference and most liquid one. The time period under investigation is 2008–2015 which includes the period of distress caused by the European sovereign debt crisis. Moreover, (i) an analysis over relevant sub-periods is carried out; (ii) several insights are also provided by implementing the tracking of the Benford's law over moving windows. The main test for checking the conformance to Benford's law is – as usual – the χ^2 test, whose values are presented and discussed for all cases. The analysis is further completed by elaborations based on Chebyshev's distance and Kullback and Leibler's divergence. The results highlight differences by countries and tenors. In particular, these results suggest that liquidity seems to be associated to higher levels of distortion. Greece – representing a peculiar case – shows a very different path with respect to the other European countries.

1. Introduction

Quantitative phenomena are characterized by numbers, whose values exhibit specific features. For instance, it is clear that the distribution of individuals' age in the world is not uniform, implying that it is more probable to randomly extract a twenty-year old human than a hundred-year old one. However, deviations from the uniform law have been observed for large dataset also in not very intuitive contexts. In this respect, the so-called Benford's law (BL), introduced by Newcomb [26] and formalized by Benford [7], plays a prominent role. BL describes the regularity of the frequency distribution of leading digits in some important large data set. In particular, it states that the distribution of first digit is more concentrated on smaller values, and the digit "one" has the highest frequency. The story of BL is interestingly traced in Torres et al. [33], where both the theoretical advancements from the original formulation and a list of its applications are provided. In brief, the assertion of the BL is that, once taken a large set of numbers, the first digit distribution follows a logarithmic law:

$$P(d) = log_{10}\left(1+\frac{1}{d}\right), \qquad d = 1, 2, \ldots, 9, \tag{1}$$

where $P(d)$ is the probability that some numbers from the data set have their first digit equal to d; log_{10} being the logarithm in base 10. For the reader's convenience, in Table 1 we show the frequencies of the first digit as given by the BL.

The applications of BL range in a wide set of fields. A list of recent contributions (published after the list of Torres et al.) should e.g. include Tödter [32], Bazzani et al. [6], Rauch et al. [30], De and Sen [14], Mir [23,24], Mir et al. [25], Ausloos and Clippe [3], Ausloos et al. [4]. In the specific context of financial applications, BL violations might be meaningfully associated to data misalignments (as much demonstrated by Nigrini [27,28]) and toxicity of the markets, but also to other cases (Ausloos et al. [4]). Some papers are particularly relevant here. Varian [34] tested – through BL – the presence of anomalies over a set of land planning data for 777 tracts, roughly corresponding to census tracts in the area of San Francisco Bay: input data (in year 1967) and forecasts (for years

* Corresponding author. Tel./fax: +3243714340.
E-mail addresses: marcel.ausloos@ulg.ac.be, ma683@le.ac.uk (M. Ausloos), rosella.castellano@unimc.it (R. Castellano), roy.cerqueti@unimc.it (R. Cerqueti).

http://dx.doi.org/10.1016/j.chaos.2016.03.002
0960-0779/

Table 1
Frequencies of the first digit in a set of data – values ranging from 1 to 9, see the first row – according to BL.

First digit	1	2	3	4	5	6	7	8	9
Frequency	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

between 1970 and 1990). He found surprisingly, yet somewhat expected, agreement (not his own words) with BL.

Nigrini [27,28] assessed mistakes over a large collection of accounting data, and explained them by invoking frauds or, simply, misprints in the data reporting process (see also Nigrini and Mittermeir, [29]). It is also worth noting the contributions of Ley [19] who studied the stock market by using Benford's law; Realdon [31] who applied the BL and linked the CDS market to the stock market, and Carrera [9] who used BL to analyze numerical patterns in exchange rates in order to verify whether they appear to have been subject to some degree of policy management.

In line with Abrantes et al. [1], Judge and Schechter [17] and Varian [34], we use the BL to verify the *quality* and *credibility* of CDS data. This is a key issue since CDS are traded in decentralized over the counter (OTC) markets which are often pictured as opaque and with little information about pricing mechanisms, price settlement and trading volumes. Moreover, we scientifically move from the perspective of CDS as complex systems (see e.g. Markose et al. [15] and Kim and Jung, [18], where the CDS volatility is studied by adopting a network approach; supporting arguments on the complex nature of the CDS can also be found in Burns et al., [8]). In so doing, we are also in line with the literature stating that BL provides an informative content mainly when the data generating process is a complex system (see e.g. Li et al., [20]).

Here, we adopt a data science perspective, aiming at deciphering the presence of anomalies in the daily changes of sovereign CDS quotes. There are three main reasons why we chose the sovereign CDS: firstly, before the recent crisis, CDS were often lauded as derivative instruments that could stabilize the financial system as a whole because of their portentous risk transferring and risk signaling abilities. Secondly, the CDS market has grown tremendously over the last decade and is currently an integral part of the financial system. Thirdly, there is some anecdotal evidence that major banks manipulated CDS prices. For example, it is worth citing the case of Reuters reporting that big US banks face CDS manipulation lawsuit[1] and the one of US Senator Carl Levin that has accused Goldman Sachs to manipulate CDS quotes.[2]

Thus, the up-to-date relevance of the present investigation can be found in the growing importance of the role played by sovereign CDS in the globalized financial market (Longstaff et al., [21]).

Indeed, after the financial crisis started in 2007 in the United States, financial markets have found a new concern since the beginning of 2010: the levels of deficits and public debts. All developed countries, even the largest, are suspected of being able to default on their debt. Rating agencies, bankers and investment funds have begun to worry about the sustainability of public finances and required countries to reduce their debt by cutting government spending, especially social spending. In this context, it can be observed that the recent development of CDS on the debt of developed countries has launched CDS as tools for estimating the probability of default of the states, although their quotes are determined on an opaque and poorly regulated market (Jarrow, [16]). As a consequence, the scientific interest for sovereign CDS mainly focuses on a time span which includes the key dates characterizing the financial crisis which is still experienced. Indeed, a considerable number of sovereign CDS contracts were signed in the early stages of the crisis (2008–2010), particularly on the wake of the turbulence following the failure of Lehman Brothers. Pre-crisis periods are generally associated to low volumes of these products. Consequently, the limited meaningfulness of the analysis of their paths allows us to shorten such considerations in this paper.

A broad range of literature analyzes the CDS market (Bao et al., [5]; Castellano and D'Ecclesia, [10]; Castellano and Scaccia, [11], only to mention the most recent ones) and, especially in the last six years, due to the European sovereign debt crisis, the scientific literature has placed more emphasis on sovereign CDS. Many of the recent studies have found that CDS quotes on debt securities increase substantially before financial crises become full-blown. However, it is worth to mention that, beyond its signaling power, the CDS market still poses many questions in relation to its full transparency and its capacity to spread of market disturbances – mostly on the downside – from one country to the other. For instance, Afonso et al. [2] use EU sovereign CDS spreads to carry out an analysis on the reaction of spreads to announcements from rating agencies. They show that quotes are sensitive to announcements, and that spillover effects from lower to higher rated countries among EMU countries, together with persistence effects, can be observed. Longstaff et al. [21] show that sovereign CDS quotes are driven more by global market factors, risk premiums, and investment flows than by country-specific macroeconomic fundamentals. In other words, their analysis supports a view of the sovereign CDS market in which investors play a predominant role.

Some other studies highlight that the price discovery mechanism and the knowledge of actual net positions of financial institutions are necessary to ensure the transparency of the CDS market (Cont and Kokholm, [12]).

2. Paper content

The dataset considered in this paper is given by Thomson Reuters Composite Sovereign CDS spreads for 13 European countries on a daily basis. Countries are distributed in four main groups: a) the *core economies* – France Germany, United Kingdom; b) the *most worrying economies* – Ireland, Italy, Portugal and Spain; c) the *Eastern economies* – Croatia, Czech Republic, Poland, Romania – and Turkey; d) Greece. Spreads are provided for two different tenors: 5 and 10 years. The period under investigation is August 2008 to April 2015. Such a period includes the Lehman Brother's bankruptcy and the sovereign debt crisis of developed countries, which are associated to increasing level of signed sovereign CDS contracts and to the interest in the analysis of their paths.The resulting time series gives an amount of 42,000 spread observations.

The conformance between daily changes in CDS spreads and BL has been checked by performing two different typologies of procedures. First, for both tenors (5 and 10 years) a χ^2 test has been implemented over the entire period and over four noticeable sub-periods : (i) August 8, 2008 to April 25, 2015; (ii) August 8, 2008 to January 1, 2010; (iii) January 1, 2010 to October 31, 2013; (iv) November 1, 2013 to April 25, 2015; (v) January 1, 2010 to April 25, 2015. Such sub-periods were chosen to highlight the time span of the European sovereign debt crisis and the one that preceded it. Secondly, for the reference contract (5 years), a BL conformity track has been implemented by computing three statistical indicators over moving windows of 120 days length, with 60 days overlap: a χ^2 test, the Chebyshev's distance and the Kullback and Leibler's divergence.

It will be shown that several insights can be derived from the analysis: first of all, there is much evidence of conformity with BL, but not in all cases. In particular, the BL validity hypothesis can

[1] http://mobile.reuters.com/article/idUSL1N0R51TS20140904?irpc=932
[2] http://blogs.reuters.com/felix-salmon/2010/12/10/annals-of-cds-manipulation-goldman-sachs-edition/