



Statistical methods for decision support systems in finance: how Benford's law predicts financial risk

Roy Cerqueti^{1,2,3} · Mario Maggi⁴ · Jessica Riccioni¹

Accepted: 11 April 2022
© The Author(s) 2022

Abstract

This paper merges the statistical analysis of data regularities and decision support systems for investors. Specifically, it discusses the Benford's law as a decision support device for financial investments. In particular, we illustrate the role of such a property of financial data as risk predictor for financial markets. First of all, we show empirical evidence of accordance between data on market index daily returns and Benford's law. Then, we highlight that on short time period (1 year) the deviations from Benford's law are related to low risk and positive trend periods; the p value of the χ^2 test against the Benford's distribution displays some predicting power for the market average return and risk level.

Keywords Data regularity · χ^2 test · Financial risk prediction · Statistical analysis

1 Introduction

Financial scientific community is paying growing attention to statistical methods for data analysis. Such a scientific approach is grounded on the explosion of information technology and the growing availability of large datasets. In general, financial market data exhibit various kinds of structures and regularities. The assessment of the interrelations among such structures allows facing the classical theme of market prediction by using peculiar properties of the considered empirical samples as a decision support system. Thus, the assessment of data

✉ Roy Cerqueti
roy.cerqueti@uniroma1.it

✉ Mario Maggi
mario.maggi@unipv.it

✉ Jessica Riccioni
jessica.riccioni@hotmail.it

¹ Department of Social and Economic Sciences, Sapienza University of Rome, Piazzale Aldo Moro, 5, 00185 Rome, Italy

² School of Business, London South Bank University, London SE1 0AA, UK

³ GRANEM, Université d'Angers, SFR Confluences, 49000 Angers, France

⁴ Department of Economics and Management, Università di Pavia, via S. Felice, 5, 27100 Pavia, Italy

properties and regularities and their usefulness in explaining financial paths appear to be of paramount relevance.

In this respect, du Jardin and Severin (2012) took into full consideration the relevant theme of the inhomogeneous periodicity of the available data used for making financial prediction. With this aim, the authors introduced the Kohonen map as a decision support system for investors. In Fischer and Krauss (2018), the authors discussed how deep learning could be applied to explore the persistence properties in the long run of the financial series, and from there, the forecast of financial markets. In the same line and by following a different approach, Oztekin et al. (2016) provided a general statistical framework for predicting daily stock returns. The authors integrated different data analytical models to pursue their scopes and tested their proposal on a wide range of daily data taken from the BIST 100 Index. Noakes and Rajaratnam (2016) examined the efficiency of the stock market on the Johannesburg Stock Exchange (JSE). The authors considered the unique characteristics of this market by making adjustments for thin trading, which occurs during the sample period through a random number generator test. Avdoulas et al. (2018) dealt with stock return predictability by applying various modifications of a nonlinear model estimation and forecasting optimization algorithm in the context of the Eurozone southern periphery stock markets; the considered study has fundamental implications for the predictability of “PIIGS” markets and—more generally—for market efficiency. Guerard et al. (2021) used data mining tests and several modern regression techniques for modelling expected returns in global markets. Akyildirim et al. (2021) tested the predictive power of intraday returns of the twelve major cryptocurrencies using different methodologies, including logistic regression, random forest classification algorithms, support vector machines and artificial neural networks. The authors have identified the most performing and robust method for predicting future daily returns through numerical experiments. Very recently, Jana et al. (2021) proposed a model for predicting the one-day ahead price of Bitcoin by integrating appropriate alternate components. The intent of Jana et al. (2021) is to support investors in making good financial choices, hence obtaining high earnings. Of particular originality is Venkatesh et al. (2014), where the authors applied data clustering and neural networks to provide a forecast of the cash demand in the ATMs. Neural networks are exploited even earlier by Desai and Bharati (1998) to study the predictive power of economic and financial variables by replacing the linear regression method with neural network models (nonlinear regression technique). The authors used within-sample and out-of-sample data to make and validate the predictions. A fully nonparametric smoother with the covariates (a machine learning technique) was applied by Kyriakou et al. (2021) to evaluate the performance of benchmarks of long-term stock returns, to obtain future forecasts useful for investors in determining the values of financial instruments. They used earnings-by-price ratio, inflation, short interest rate, long interest rate, dividend-by-price ratio, and term spread as predictors. More generally, it is evident that the tools for predicting the returns and prices of financial instruments are fundamental for investors. The deepening of the characteristics and properties of the historical series and of the available information are crucial in the analytical process that supports the financial decision system. A good forecast is a decision-making tool and is a key component to successful investing and risk management.

In the field of data mining and decision analytics, the compliance of financial data with specific functional laws might be viewed as a decision support system for investors and institutions. Some relevant contributions are worth to be mentioned in this field. In Huang et al. (2008), the authors presented Zipf’s law as a device to be used by auditors for detecting frauds. The ability of Zipf’s law as a valuable tool to detect the presence of frauds has also been confirmed by Pietronero et al. (2001). Detection systems have been studied by many

researchers who have sought innovative solutions based on different data mining techniques such as machine learning, neural networks and clustering analysis (see, e.g., Bernard et al., 2019; Boros et al., 2011; Duan et al., 2009; Jiang et al., 2018; Ngai et al., 2011).

This paper presents a decision support system contextualization of data regularity for the case of financial investments decisions.

As a premise, we point out that how investors and institutions take a position in a financial market strongly depends on the considered assets' risk profiles and the overall market. This evidence explains why the theme of financial risk assessment is a classical issue in finance, and a special focus is clearly pointed to risk forecasting (see, e.g. the excellent monograph Alexander, 2008 but also the recent contributions in Borges, 2010; Castellano et al., 2020; Cooper et al., 2014; Cooper et al., 2021; Gu & Peng, 2019; Kürüm et al., 2018; Meng & Taylor, 2020). We move from this premise and treat here the peculiar issue of explaining how data regularity can be helpful in predicting the market risk. In so doing, we are in line with some important contributions dealing with the forecast of market risk as a relevant target in the field of decision support systems (see, e.g. D'Ecclesia & Clementi, 2021; Feldman & Xu, 2018; Groth & Muntermann, 2011; Huang & Kou, 2014; Al Janabi, 2013).

As a data regularity, we are here interested in Benford's law (BL, hereafter). As we will show in the literature review presented in the next section, such a law is quite popular in the environment of fraud detection but quite neglected as a device for market data analysis. Indeed, to the best of our knowledge, no attention has been paid to the informative content of BL for risk assessment. This paper moves some steps to fill this gap. Specifically, we aim at exploring the predictive properties of this law when applied to the daily returns of a stock index. In so doing, we advance the proposal that BL can also be exploited for long-term forecasting of financial risk.

The reasons behind the usefulness of BL for risk management can be found in the informative content of BL. Specifically, in the context of financial markets, the violation of BL for daily prices and volumes might be associated with exogenous shocks or socio-economic-political instability, which can modify the evolution of such financial quantities. In this respect, it is worth mentioning the exhaustive discussion in Riccioni and Cerqueti (2018). The authors analyze long series of daily volumes and prices of 4166 stocks listed in important international stock markets.

This said—and in line with the literature on BL, see the next section—the analysis of the risk on the basis of the validity of BL points the attention to the possible connection between the occurrence of exogenous events having an impact on the financial markets and risk management. More in detail, we here aim to present a framework for stating how the effect of such events, observed through possible deviations of the financial time series from the BL, can provide information on the future evolution of the risk level of the considered financial quantities. It is essential to notice that a full explanation of how events impact the validity of BL is out of the scope of the present paper—such a challenging theme deserves more targeted research. However, the methodology proposed in this paper may represent a preliminary step for focusing more attention on the causality effect between external events and financial risk management.

To sum up, our method is useful for risk forecasting and can provide information on how exogeneous shocks—which can modify the compliance with the BL of a given financial time series—can be used for forecasting financial risk.

Our approach can be considered an alternative to those presented in the existing literature in financial risk management. Perhaps the studies that are closer to our perspective are those evaluating the impact of events like news announcements on prices' volatility (see, e.g. Neely, 2011 and references therein). However, the quoted papers adopt an event study approach,

where the event is isolated and its role is fully identified. Our perspective is radically different in that we do not need to identify the specific external events affecting a financial time series. Therefore, the proposed method appears to be more general in that it does not require a preliminary knowledge of the external events affecting the financial time series for risk management.

For our purpose, we consider the series of the daily returns of S&P 500, for about 30 years (from Mar-21-1988 to Mar-21-2018).

Our analysis is carried out under three different perspectives: first, we consider the overall sample and check the validity of the BL for the leading digit. After this step, in-sample and out-of-sample experiments in a moving-window framework have been implemented. The investigation target lies in assessing the risk level of the considered returns conditioned on compliance with the BL.

The obtained findings can be synthesized in saying that risk grows as compliance with BL becomes more evident. This outcome has the noteworthy consequence that BL can be interpreted as a risk assessment instrument for financial returns.

The rest of the paper is organized as follows. Section 2 provides a literature review on BL and its applications. Section 3 describes the used dataset, along with its main statistical features, and the methodologies employed for the analysis. Section 4 is devoted to the description of the empirical results and to their discussion. This section also contains some further elaborations in the light of performing robustness checks of the results. Last section offers some conclusive remarks.

2 Literature review on the Benford's Law

BL—which has been introduced by Newcomb (1881) and formalized by Benford (1938)—is one of the most interesting properties of a large set of data. It provides the expected frequencies of the digits in tabulated data and highlights that the lowest digits appear the largest part of time. In fact, BL asserts that the frequency of the leading digit of the values of a dataset decreases with the value of the digit and reaches its maximum when the digit is “one”. We recall that the first leading or significant digit—in brief, the first digit—is nothing but its first digit, by excluding zero. For example, the first significant digit of 7899 is 7, while that of 0.0329 is 3. For the first digit, BL states that:

$$P(\text{first digit} = n) = \log_{10} \left(1 + \frac{1}{n} \right), \quad n = 1, \dots, 9, \quad (1)$$

where $P(\text{first digit} = n)$ is the probability that a number has the first digit equals to n , \log_{10} being the logarithm in base 10. The property (1) can also be generalized to digits beyond the first (see Hill, 1995c for more details).

The pioneering work of Benford shows that the first significant digits of a wide set of randomly collected data satisfy the property in formula (1). Thus, an immediate question can be stated: what about the meaningfulness of the randomly collected data which deviate from BL? The response is questionable (refer to Hill, 1995a for some theoretical suggestions). A paper that created debate in the scientific world was the recent work by Mir and Ausloos (2018) who compared the first two papers, in Refs. Newcomb (1881) and Benford (1938), where the BL was discovered, to the story of “A sleeping beauty,” in the sense that they have been forgotten in a sort of deep sleep for more than 100 years the former one and almost 50 years the latter one.

This said, one can conjecture that non compliance with BL means that data have suffered from some sort of manipulation. Such a property explains why BL has been used in the contexts of economics and finance to investigate whether data have been manipulated and result, therefore, unreliable.

In the context of accounting, Hill (1995a,b,c, 1998) devoted many of his studies to investigate the particular phenomenon of BL and was the first who applied it to assist in detecting fraud in accounting data. Nigrini, in Nigrini (1996, 1999, 2012), was inspired by Hill (1995a,b,c, 1998) and focused his research on the possibility of turning BL into a real tool able to detect the falsification and the frauds in accounting and auditing (whether these are arithmetic errors and errors in calculation or misapplication of the appropriate accounting standards, or frauds, such as the alteration of records or documents, the lack of enforcement of accounting standards, the omission of some results or finally recording non existent transactions). This type of study has evolved over the years, and nowadays, the BL is a standard tool to support the identification of tax fraud in the US. In the same environment of financial audit, Bhattacharya et al. (2011) discussed BL as a decision support system for detecting frauds and the presence of tax evasion.

Another management application concerns the validation of self-reported data from the employees of a company. As shown by Hales et al. (2008, 2009), the use of BL can provide a low-cost method to detect internal data manipulation, and so it allows to improve the operating performances of a firm.

In macroeconomics, it is worth mentioning some relevant papers dealing with the application of BL. Nye and Moul (2007) studied the GDP data, in particular those contained on the Penn World Tables. Tödter 2009 applied BL to regression coefficients and standard errors in empirical economics, hence constructing an indicator of fraud in economic research. Günnel and Tödter (2009) have dealt with the study of the forecast of GDP growth and inflation of German consumer prices. Rauch et al. (2011) examined the abnormal data of national and financial accounts of the EU countries from 1999 to 2009. Rauch et al. (2011), detected the quality of macroeconomic data relevant to the deficit criteria reported to Eurostat by the EU member states. Michalski and Stoltz (2013) tested the hypothesis that “a country may want to hide its true state of the world to prevent capital outflows or attract inflows” examining the balance of payments data for 103 countries between 1989 and 2007. Mir (2016) applied BL on the illicit financial outflows from developing countries. Holz (2014), instead, deepened data from the National Bureau of Statistics of China to measure the quality of the Chinese GDP. Rauch et al. (2014) have compared government social security statistics with deficit related data reported by the EU member states to Eurostat. Some contributions have tested the law on the aggregate of income taxes of municipalities and Italian regions for the period between 2007 and 2011 (see Mir et al., 2014; Cerqueti & Ausloos, 2015, Ausloos et al., 2017). Deleanu (2017) analyzed a 2003–2007 dataset of indicators of compliance and efficiency in combatting money laundering collected by Eurostat.

In finance, BL represents a useful tool for verifying the efficiency of financial indexes. Ley (1996) checked daily returns of two American stock indexes (i.e. the S&P for the period from 1926 to 1993 and the Dow Jones for the period from 1900 to 1993), observing that BL holds for both stock indexes. Clippe and Ausloos (2012) proposed the analysis of the validity of BL on a set of financial data. Corazza et al. (2010) investigated the trend of S&P 500 stock quotations. De Ceuster et al. (1998) analyzed any psychological barriers at the Dow Jones 30 Industrial Average, the Financial Times—Stock Exchange 100 and the Nikkei Stock Average 225. Other studies tested the validity of the law on the sovereign credit default swap markets (see e.g. Realdon, 2008; Ausloos et al., 2016). Patton et al. (2015) provided a deep exploration of the reliability of voluntary disclosures of performance in the context of hedge funds by

applying several instruments, including BL. Juergens and Lindsey (2009) studied trading volume for Nasdaq market makers around analyst recommendation changes issued by an analyst at the same firm. Alali and Romero (2013) started to study 10-years data concerning financial accounting data using a large sample of US public companies. Karavardar (2014) applied BL to investigate the Istanbul stock exchange. Nigrini (2015) dealt with daily returns, daily volumes, expected returns and abnormal returns, discussing the approach of dividing a population into a subset and analyzing the compliance of BL on these subsets. Carrera (2015), instead, in the context of policy management, analyzed exchange rates. Shi et al. (2018) decided to apply BL on ten industrial sectors of the main developing countries over a period of fourteen years, focusing their attention on reported financial data. Riccioni and Cerqueti (2018) have tried to interpret the international financial markets (it has been the first time for a global context) through the analysis of volumes and adjusted closing prices of all stock indexes listed on the stock exchange of several countries from listing day to November 2014. Abrantes-Metz et al. (2012) highlighted the usefulness of BL for comparing the Libor with other short-term borrowing rates to investigate potential anticompetitive behaviour in real markets.

So, we are witnessing remarkable popularity of the BL among financial data scientists and economists so that, in commenting on the important paper by Sudjianto et al. (2010) on financial frauds assessment, Hand (2010) states: *I was surprised that Sudjianto et al. (2010) did not mention Benford's law.*

3 Materials and methods

3.1 Dataset

The investigated dataset collects the daily returns of the S&P 500 index from Mar-21-1988 to Mar-21-2018, for a total amount of 7561 observations. The data downloaded from the Bloomberg data provider concern the S&P 500 Total Return index, which includes dividends.

As a premise, we state that we have complied with the terms of service for the Bloomberg platform.

Let P_t be the closing price of the index on day t . To preserve the time scale, returns are computed taking into account the number of calendar days between observations, so the return of trading day t is

$$r_t = \frac{\ln(P_t) - \ln(P_{t-1})}{d_t},$$

where d_t is the number of calendar days between the trading days $t - 1$ and t .

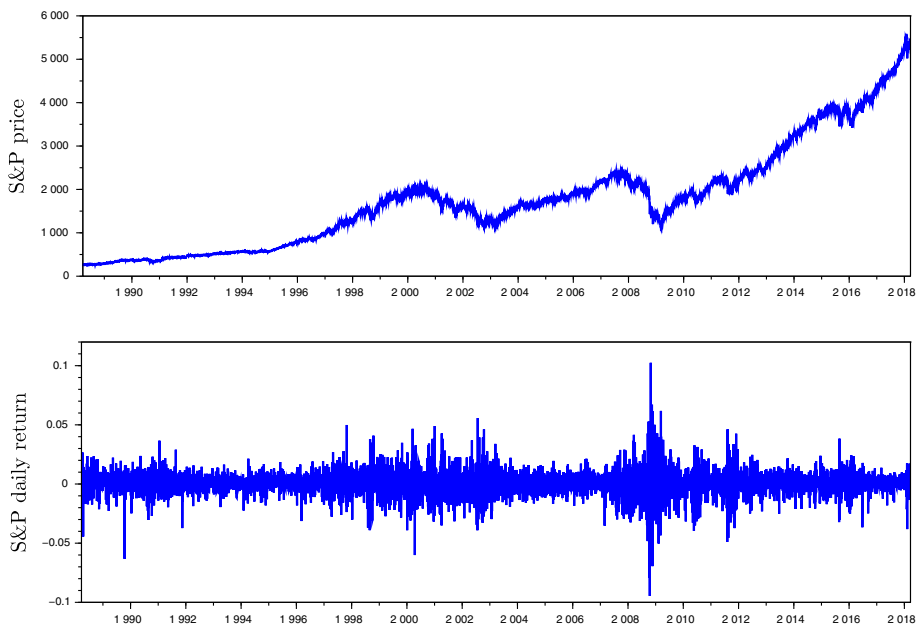
Table 1 shows some statistical characteristics of the considered sample. Prices are also shown, for the sake of completeness. Figure 1 describes the time-evolutions of the daily prices and returns over the considered period.

Table 1 confirms the stylized facts on the index prices and returns, which are broadly stable with a small number of very relevant outliers. The mentioned outliers are referred to the distribution of the whole sample. The number of observations beyond three standard deviations from the average is around 1.68%, on the whole sample, and 1.25% on the annual sliding windows used in the paper. We have carefully checked those data: generally, they are unexpected returns, often located at the beginning of a rise in volatility; moreover, we can exclude they result from error in the data sampling. We do not eliminate those data from our analysis, because “anomalous” returns are the subject of our study, as elements

Table 1 Statistical characteristics of daily returns and prices

Statistical characteristics	Daily prices	Daily returns
Minimum	254.51	−0.094595
Maximum	5585.55	0.1024541
Mean	1765.25	0.0003307
Median	1670.62	0.0004237
Standard deviation	1167.98	0.0096108
Skewness	0.9155747	−0.2107351
Kurtosis	3.4290946	11.96854

Data of the S&P 500 Total Return index. The considered period ranges from Mar-21-1988 (prices) and Mar-22-1988 (returns) to Mar-21-2018

**Fig. 1** Time series of the daily prices and returns of the S&P 500

of the financial risk. Back to Table 1, we observe a wide range in prices that reflects the overall growth of the index value along the considered time period. Prices also exhibit a positive skewness with tails tending towards smaller data and low concentration of values. All these features are consistent with the non-stationarity of the price time series, and the fairly long time window. Differently, for returns there is a negative skewness, with a tail that tends towards higher values which indicates a propensity to positive returns and a leptokurtic distribution very concentrated in the central values. Daily returns exhibit high kurtosis with a much more peaked distribution. So, small fluctuations are less frequent, since returns are rather clustered around the mean.

3.2 Methodology

The analysis is carried out in several consecutive steps.

First of all, we detect whenever the considered data on the returns satisfy the requirements of the BL, according to Eq. (1). With this aim, a χ^2 goodness-of-fit test is implemented over the first significant digit of the return series, to verify whether the empirical frequencies are statistically different from the theoretical ones, as described by BL. Specifically, we have

$$\chi^2_{\text{stat}} = \sum_{i=1}^9 \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(8)},$$

where O_i is the empirical frequency detected for digit i from the original sample, whilst E_i is the theoretical frequency of digit i , according to the BL in Eq. (1).

The p values correspond to various significance levels for 8 degrees of freedom. Indeed, 8 degrees of freedom are needed for verifying the conformity of the first significant digit—being $n = 9$ the number of possible significant first digits.

After this preliminary step, we have carried out an extensive in-sample and out-of-sample forecasting experiment using moving-windows. To this end, we evaluate the compliance with the BL at the individual windows level, to assess the forecasting power of such a rule. The degree of such a compliance is captured by the p -value of χ^2 test, with null hypothesis the BL distribution (1).

To carry out the in- and out-of-sample analysis, we have denoted by $w \in \mathbb{N}$ the sliding window length and $w_f \in \mathbb{N}$ the forecasting horizon. For each window ending on time t , some indicators have been computed on the data: p_t is the p -value of the χ^2 test against (1); m_t , s_t , v_t are the average, the standard deviation and the VaR at 5% of the return, respectively; R_t is the total return on the w_f days ahead (from $t + 1$ to $t + w_f$), converted on a daily basis (i.e. the average daily return).

For the out-of-sample analysis, we computed the distribution of R_t , conditioned on the values of p_t in the window ending in t . For conditioning, the range $[0, 1]$ of the values of p_t has been divided into 40 equally spaced bins. Furthermore, the average, the standard deviation, the Value at Risk of the w_f -days ahead returns' conditional distributions are considered.

3.3 The contribution of the Benford's law to the exploration of financial data

First of all, we have wide evidence of the compliance of our dataset with the BL. Figure 2 shows that the returns of the S&P index seems to not substantially deviate from the BL. It is a visual inspection, but rather satisfactory in terms of rendering the compliance of the data with BL for the first leading digit.

Such a suggestion seems not formally supported by the χ^2 test over the entire sample (see Table 2). Indeed, the p -value close to zero and the large χ^2 -value state that there is not statistical compliance of the original sample with the BL. Deviations between visual appealing and statistical tests may occur in a large sample of data, as in this specific case (see e.g. Ley, 1996; Nigrini, 2012, and Cerqueti & Magg, 2021).

Compliance improves when time periods are shortened. In this respect, in Table 2 we present also the cases of four consecutive subperiods with the same time length and which divide equally the whole time period under investigation. In all subperiods, BL is accepted at a significance level of at least 5%. The motivation behind such a result may be that data are more homogeneous in short periods, while the sample—which covers a period of 30 years—may have remarkable discrepancies and exhibit several regimes.

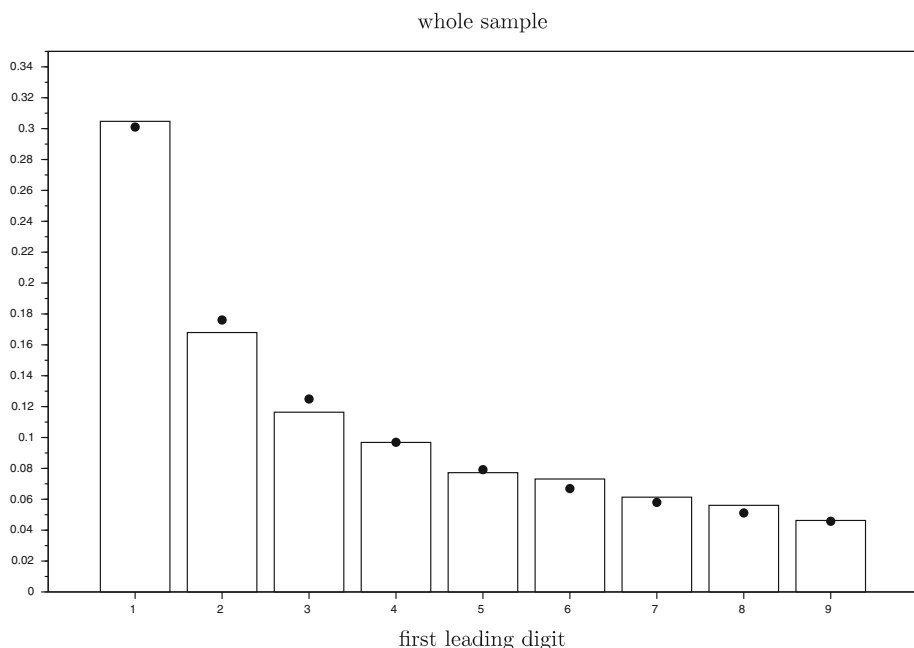


Fig. 2 First leading digit of the S&P 500 frequency distribution (bars), compared to the Benford's law probabilities (dots)

Table 2 Output of the χ^2 test with 8 degrees of freedom (null hypothesis BL), applied on the leading digit distribution of the S&P 500 returns

	Whole sample	1st	2nd	3rd	4th
χ^2 stat	20.32	9.32	14.23	6.42	9.848
<i>p</i> Value	0.0092	0.3156	0.0760	0.6008	0.2758

The sample has been also divided into 4 equal sub-periods (columns 1st, 2nd, 3rd and 4th)

Moreover, we can show that the S&P returns display the typical scale invariance which characterize BL data. Figure 3 presents the first digit distribution of the S&P returns, rescaled to match different volatility scenarios. The selected volatility levels are consistent with the variation range of the standard deviation of S&P 500 in the considered period. Remark that the distribution is very similar and close to the BL in all cases.

A further discussion stems from the consideration that the simplest model we can use to describe the financial returns assumes that the log-returns are independent and identically Gaussian. As a premise, purely analytical arguments suggest that there is not a convincing reason to expect the validity of BL when data have a mean close to zero and they are normally distributed. This said, we can observe the statistical discrepancy between the considered data and the normal distribution, in accord to well-established stylized facts in finance. To support this outcome, we here present some graphical representations of the resampled data along with the BL distribution plot, for comparison purposes (see Fig. 4). We also provide some normality tests to formalize the non compliance of the data with the normal distribution. Specifically, we use the classical statistics of Doornik-Hansen, Shapiro-Wilk, Lilliefors and Jarque-Bera; see e.g. Yap and Sim (2011) for a comparison of normality tests. We observe

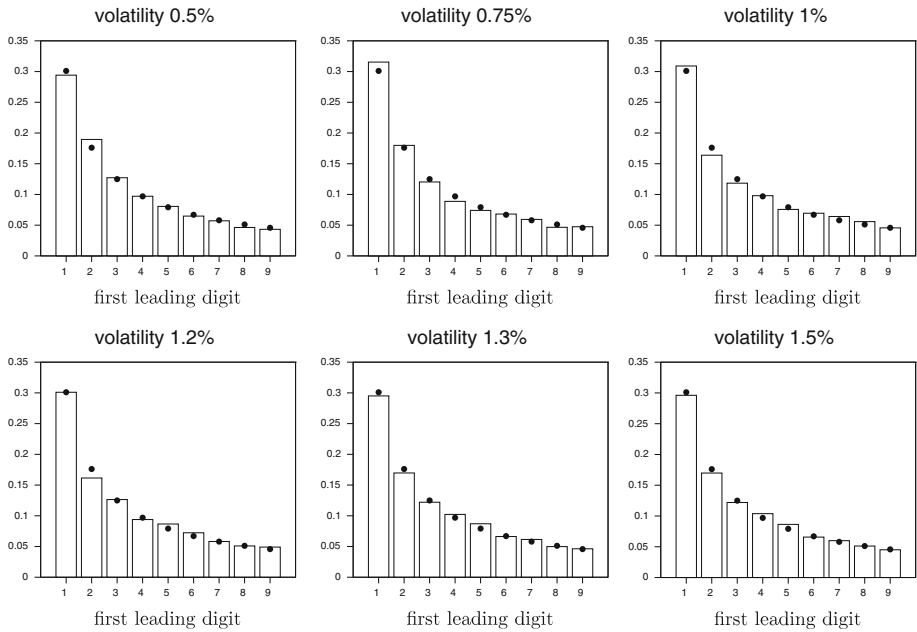


Fig. 3 First digit frequency distribution (bars) of the returns of the S&P 500, rescaled to produce six different volatility scenarios, compared to the Benford's law probabilities (dots)

Table 3 Tests for normality of the considered series of returns

Statistics	Value	<i>p</i> value
Doornik-Hansen	12453.3	$< 10^{-6}$
Shapiro-Wilk	0.530843	$< 10^{-6}$
Lilliefors	0.368234	$< 10^{-6}$
Jarque-Bera	1.06278e+06	$< 10^{-6}$

that all the considered tests have p values close to zero, and so the normality null hypothesis is always strongly rejected (see Table 3).

The main insights about a substantial difference between Gaussian and real data can be obtained by comparing Figs. 3 and 4. The first one displays the first digit distribution of the S&P 500 data, rescaled to match the considered six different volatility scenarios; the second one uses Gaussian data drawn from the same different volatility scenarios. It is noticeable that Gaussian data do not display the scale invariance that appears instead in the financial returns.

Now, we can observe two levels of information provided by the compliance of the financial returns with the BL: firstly, our approach gives insights on the statistical description of the financial returns; secondly, BL can be effectively used to make forecasting exercises.

We enter the details.

To investigate whether p_t , i.e., the p values of the χ^2 statistics with respect to the BL at the t -th time window, is able to add some insights on the understanding of the return distribution, we compute the main distribution indicators of the considered time series on an annual rolling window basis. So for each rolling time interval, beside p_t , m_t and s_t , we also compute

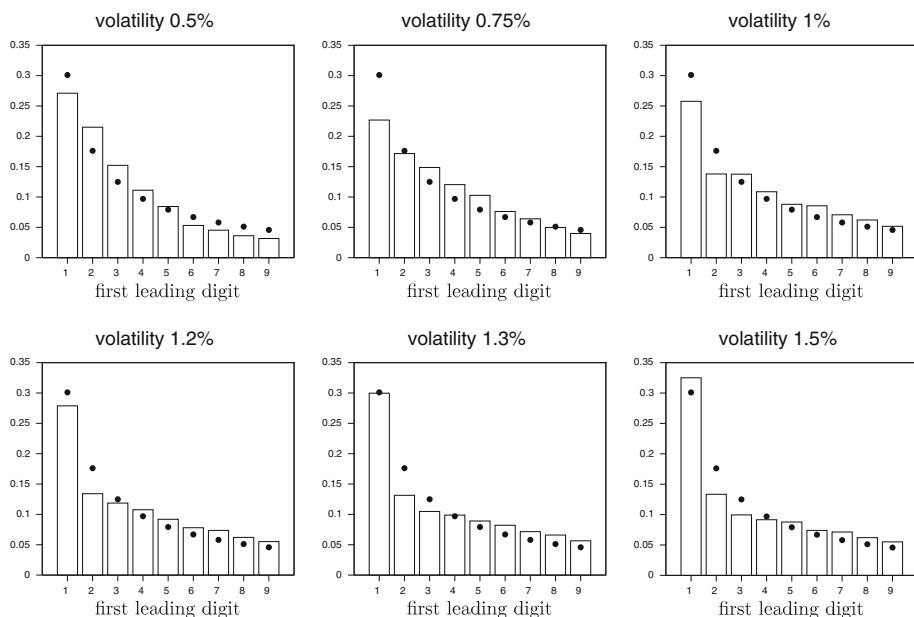


Fig. 4 Here we show the comparison of frequency distribution between Gaussian data (bars) and Benford's law data (dots)

the skewness, kurtosis coefficients to better describe the shape of the return distribution. Moreover, to take into consideration also some dynamical features of the return process, we also compute the first order autocorrelation of the returns and the first order autocorrelation of the squared return in excess to the mean. We then measure the correlation between all these indicators and the p_t . We obtain that the correlations between the p_t and the mean, the standard deviation, the skewness, the kurtosis, the first-order auto-correlation of the return, the first-order auto-correlation of the square return in excess to the mean computed on an annual sliding window is -0.1647 , 0.1486 , 0.0411 , 0.0059 , 0.0333 and 0.0738 , respectively. Such low correlations indicate that the p_t are able to describe a feature of the return process which is not significantly related to the common statistical properties. Therefore, we can claim that the p_t provide additional information to the description of the return process.

Moreover, it is well known that the other models, like GARCH and heterogeneous autoregression (HAR) (see Engle, 1982; Bollerslev, 1986; Bollerslev et al., 1992; 2018; Nelson 1991, and Corsi et al., 2012; Santos & Ziegelmann, 2014; Vortelinos, 2017), and the whole family born from their variants, can be used to forecast the volatility, and their performances are more satisfactory for shorter time horizons (e.g., see Ding & Granger, 1996; Baillie, 1996; Zumbach, 2004). Differently, in this work we are interested in measuring effects on risk over a longer time horizon. For sake of comparison, we also applied a GARCH(1,1) and a HAR models to forecast the risk. A more detailed discussion is provided in Sect. 4.1.1. Here we notice that the correlation between the volatility forecast obtained with the GARCH and HAR models and the p_t are 0.11 and 0.04, respectively, confirming the fact that the p_t provides another kind of information about the return process, than other time-series models.

In addition, our proposal is based on a property of the considered dataset as a whole, without requiring any time-order and the consequent specification of the time-evolution of the considered phenomenon. This grounding assumption allows to implement forecasting

exercises also when the time-dimension is lost—as it may happen in the case of missing data or for data which are heterogeneous in terms of their periodicity. Differently, for example, a GARCH is a time-series model; its calibration requires a dataset composed of consecutive observations. In this respect, we provide an additional tool in risk forecasting, to be used alongside other predictors. Hence, the risk prediction procedure presented in this paper is more general and allows a high degree of tractability in several real-data situations.

4 Results and discussion

As a premise on the discussion of the results, we point the attention of the readers to an interpretation of our study, which can be implemented by looking at Ley and Varian (1994). The quoted paper is quite close to our perspective. Indeed, Ley and Varian discuss the behavior of the investors in presence of some peculiar values of the Dow-Jones index. The authors explore the prediction power of such so-called *resistance levels* or *psychological barriers*. They claim that their analysis is of purely empirical nature, as the one we present in our paper. We observe that by one side, the quoted paper is similar to us in that it aims at discussing the existence of a relationship between financial markets digits data and forecasting exercises; by the other side, we here deal with statistical conformity with a well-known law and forecast while Ley and Varian deal with an empirically-obtained law and forecast.

We now present the results by dealing with the local analysis, along the moving windows.

We use p_t to assess the compliance of the data over the related window $(t - w + 1, t]$ with the BL. For what concerns the in-sample analysis, we have taken $w = 250$ trading days, i.e. about one year.

The frequency distribution of p_t on the moving windows is represented in Fig. 5. Figure 6 shows the series of the four quantities m_t , s_t , v_t , p_t . In Fig. 7 one can find mean and standard deviation of the returns conditioned to the p values of the χ^2 test. In Fig. 8 the VaRs—whose values are conditioned to the values of p_t —are displayed.

Some insights can be obtained by looking at those figures.

First of all, we notice that the most frequent case is the one with weak deviations from BL (i.e. p values larger than 0.05 or any other common significance level). This means that there is a substantial compliance with BL for the first significant digits over the moving windows. This is in contradiction with what happens for the entire original sample, and is in agreement with the already claimed improvement of the compliance level as the sample becomes smaller.

Notice from Figs. 7 and 8 that VaRs at 5% have a behavior similar to the standard deviations, even if an amplified effect can be registered. It is also interesting to note the specular behaviors of the means and standard deviations of the returns, so that one raises once the other decreases. In presence of weak compliance with the BL (low p values) there is a low expected value with high variance, hence leading to a high level of riskiness.

The analysis of financial risk is further stressed when comparing the VaRs at different levels conditioned to the compliance with the BL. One can argue that the VaRs are below their unconditional values when p values are low, hence supporting the finding that high compliance with the BL leads to high risk. We can explain the reduction of the VaR at 1% for values of p_t close to 1, because the classes with high p_t contain fewer observation.

In the context of the out-of-sample analysis, we have taken $w = 250$ and different forecasting horizons $w_f \in \{20, 60, 120, 180, 250, 375\}$, corresponding to about 1, 3, 6, 9, 12, 18 months, respectively.

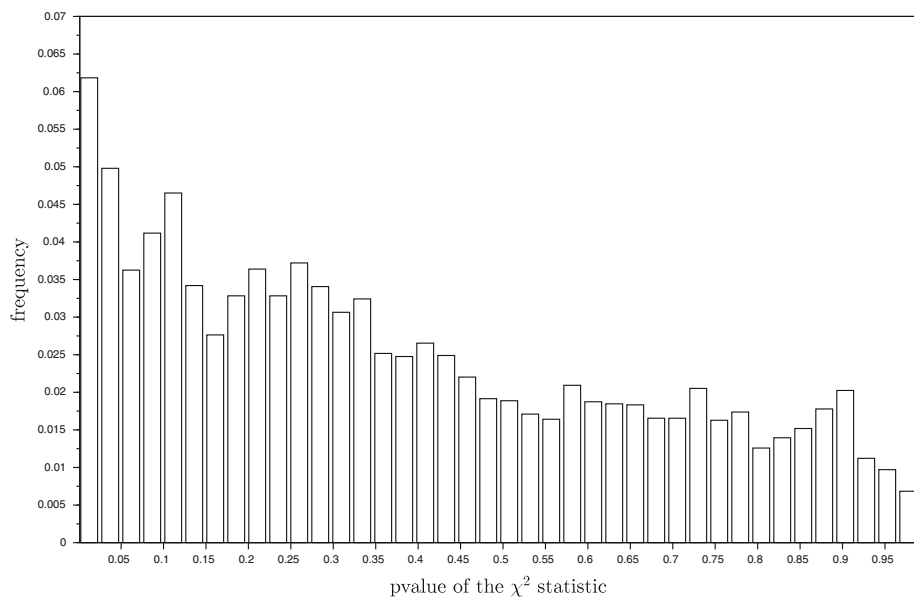


Fig. 5 Frequency distribution of p_t obtained on the sliding windows

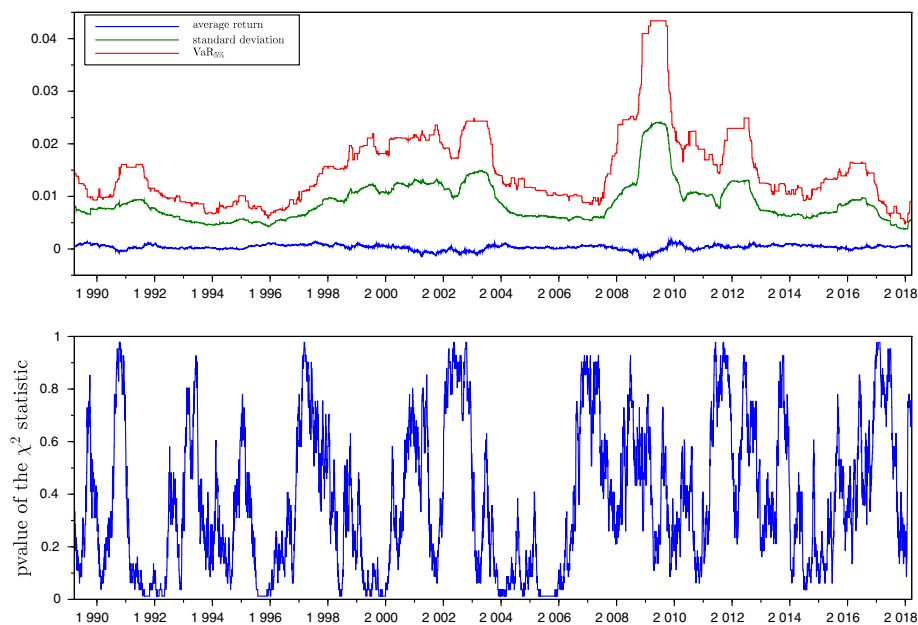


Fig. 6 Time series of m_t , s_t , v_t (top) and of p_t (bottom) for S&P 500

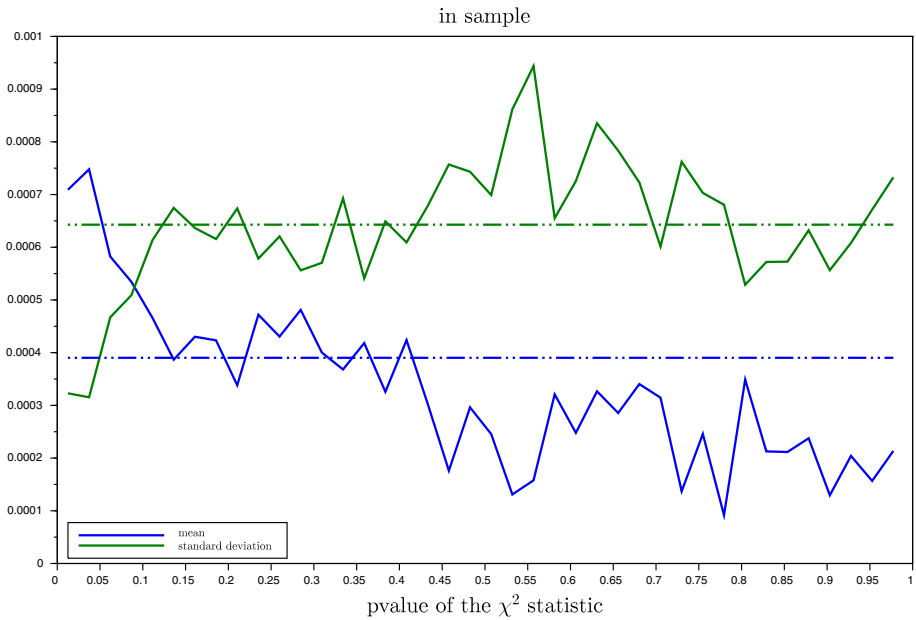


Fig. 7 Return mean and standard deviation, conditional with respect to the p value of the χ^2 statistic. The continuous lines represent the conditional mean and standard deviation; the dashed lines show the corresponding unconditional values

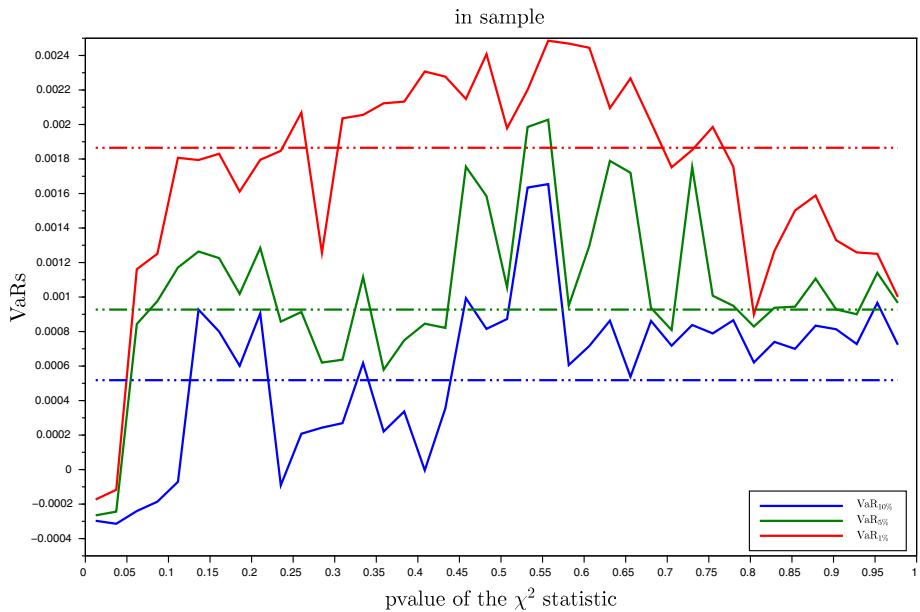


Fig. 8 Value at Risk, conditional with respect to the p value of the χ^2 statistic. The continuous lines represent the conditional VaR at levels 10% (blue), 5% (green), 1% (red); the dashed lines show the corresponding unconditional values. (Color figure online)

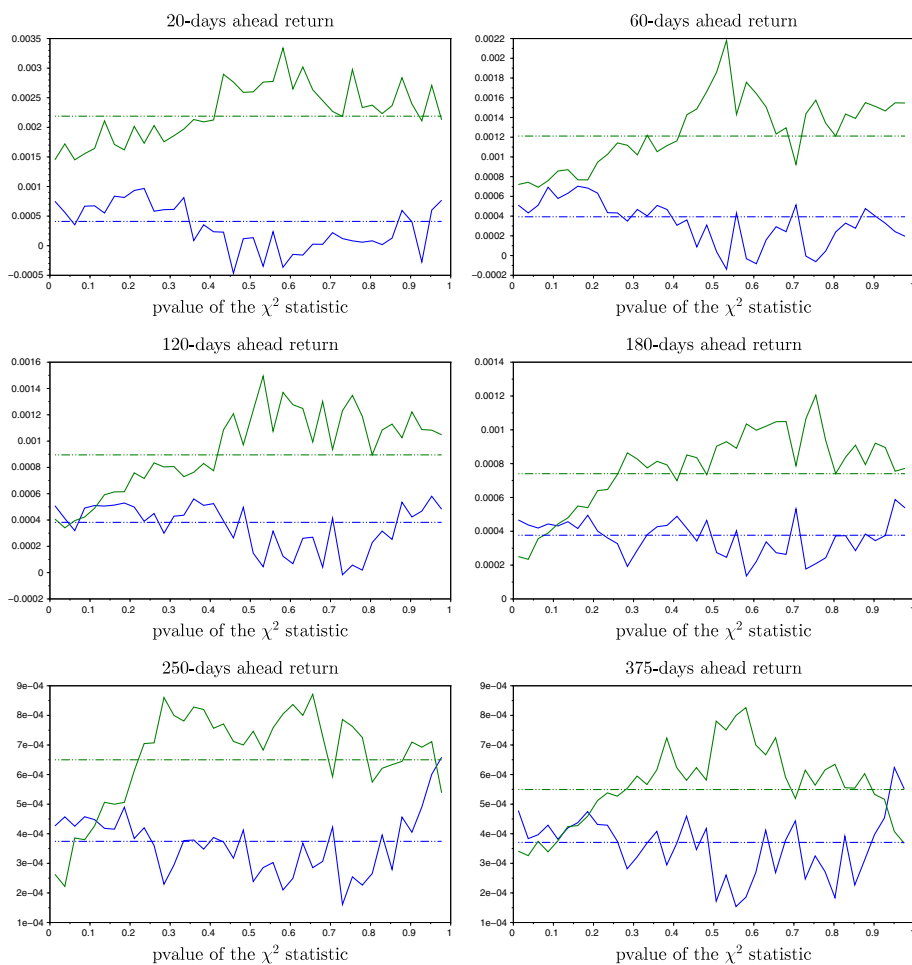


Fig. 9 Out-of-sample. Mean and standard deviation for the forecast return, conditional with respect to the p value of the χ^2 statistic. Different forecasting horizon are considered, corresponding to about 1, 3, 6, 9, 12, 18 months, respectively. The continuous lines represent the conditional mean (blue) and standard deviation (green); the dashed lines show the corresponding unconditional values. (Color figure online)

The panels in Figs. 9 and 10 illustrate the means, standard deviations and VaRs at 1%, 5%, 10% levels for the windows at the considered forecasting horizons.

For the cases of mean and standard deviations, one can broadly confirm the specular behaviors observed in the in-sample analysis. However, several facts emerge. First, the horizons of 60 days ahead and longer lead to similar shapes. The high compliance with BL leads to low expected returns with high standard deviation, thus suggesting financial distress. As the horizon increases, the ranges of variation of conditioned means and standard deviations decrease.

Interesting insights are provided by the analysis of the conditioned VaRs at different levels. Notice that VaRs are generally low when p -values are small, for all the considered horizons and levels. In this case, the high risk is associated to the compliance with BL, especially for short horizons. Quite surprisingly even for the longer horizons of 250 and 375 days, low p

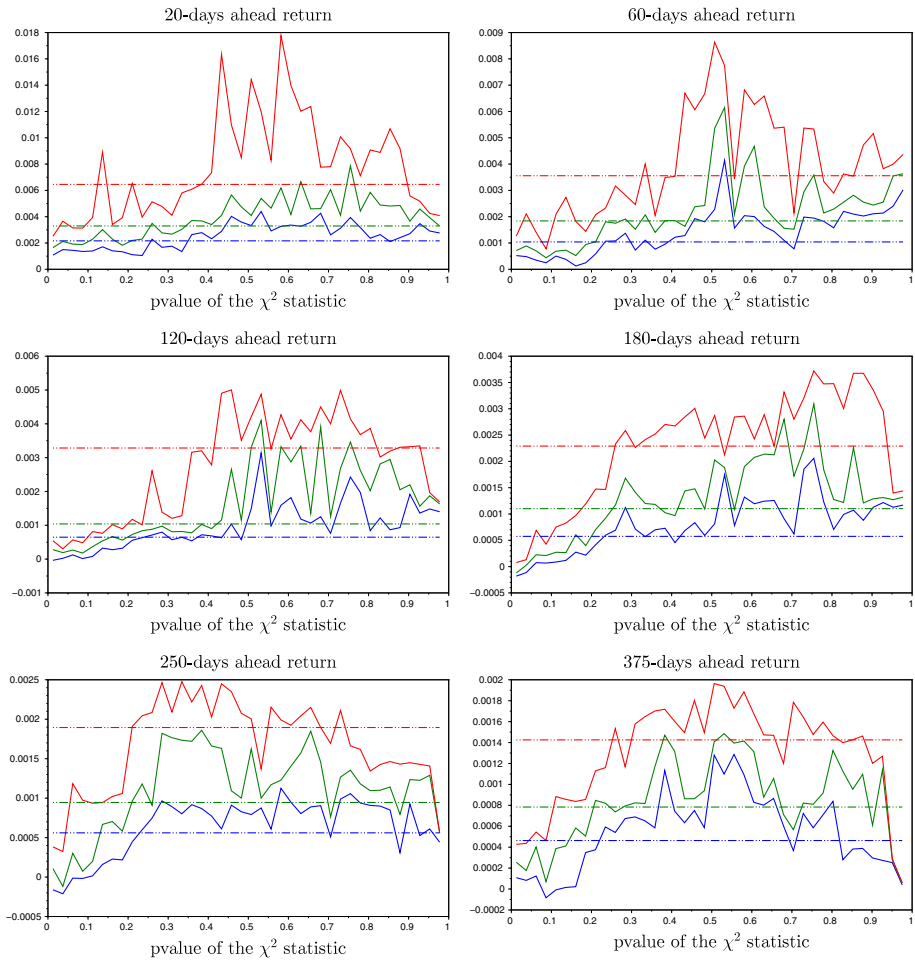


Fig. 10 Out-of-sample. Case of Value at Risk for the forecast return. Refer to the caption of Fig. 9

values are still associated to VaRs below their unconditional values. However, we notice a reduction of the VaRs even for p values close to 1. The case of p values around one concerns classes with fewer observations, therefore the statistical meaning may be weaker. In general, these results can be hardly appreciated for the 1% VaR and for the 20-days horizon. This can be due to the fact that 1% is a rather severe level for the VaR. Moreover, a 20-days horizon is short, so the results may result quite noisy.

4.1 Consistency and robustness checks

In order to relate the results with existing risk predictors, and give the results a more robust basis, some checks have been performed.

First of all, Sect. 4.1.1 compares the performances of the proposed indicator with the predictions obtained by the GARCH(1,1) and HAR models. Then, in order to check whether the inclusion of dividends into stock values has significant consequences, in Sect. 4.1.2 we also applied the analysis to the S&P 500 index, that does not include dividends. To ascertain

that the described phenomenon is stable trough time, in Sect. 4.1.3 we analyzed separately the first and the second part of our sample. We remark that both subsamples include the occurrence of large financial distresses. Finally, Sect. 4.1.4 shows that the highlighted patterns do not depend on the fact that the distribution of p_t produces few observations for large values of p_t .

For each presented check, we have considered the different forecasting horizons presented above, i.e. $w_f \in \{20, 60, 120, 180, 250, 375\}$. However, for sake of space and interest, in most cases only the forecasting horizon $w_f = 60$ days is shown; the full set of results concerning all the forecasting horizons discussed above are available upon request.

4.1.1 Comparison with other variance predictors

First of all, we show a comparison with other risk prediction models. The financial risk forecasting is a relevant matter and different methods and techniques have been proposed in the literature. These proposals span from time series analysis to machine learning (see, e.g., Bollerslev et al., 2018; Gavrishchaka & Banerjee, 2006; Liu, 2019; Satchell & Knight, 2011), with a recent interest in high-frequency data. However, there is a vast consensus about the usefulness of GARCH and heterogeneous auto-regression (HAR) in the volatility prediction, at least for short time horizons (see, e.g., Bollerslev et al., 2018; Corsi et al., 2012, Santos & Ziegelmann, 2014; Vortelinos, 2017, and references therein). For this reason, we propose a comparison between the compliance indicator to the BL and the variance forecasting obtained by the GARCH(1,1) and the HAR (the HAR settings are with lags 22, 5, and 1, as usual). Figures 11, and 12 show the results obtained with the GARCH(1,1) and the HAR. Comparing Figs. 11 and 12 with Fig. 10, we can notice that the three methods have an overall ability to predict large VaR, with some differences. On short forecasting horizons, all the predictors have good performances, with a neater behavior of the GARCH and HAR models. Instead, on longer forecasting horizons the outcomes appear comparable across the three methods, with more consistent results in favor of the p value of the χ^2 statistics.

Besides, we found that the predicted variance obtained from the GARCH and the HAR has a small correlation with respect to the $\chi^2 p$ value. This means that we are using a predictor containing different information with respect to the common variance predictors. For instance, the BL $\chi^2 p$ value, and the 60-days variance forecasting obtained from the GARCH and the HAR have the following correlation matrix

	$\chi^2 p - \text{val}$	GARCH(1, 1)	HAR
$\chi^2 p - \text{val}$	1.	0.1078538	0.0401525
GARCH(1, 1)	0.1078538	1.	0.9568413
HAR	0.0401525	0.9568413	1.

In light of this, our proposal offers an additional indicator to be considered for financial risk prediction. The overall consistency with other variance predictors, together with their weak correlation, makes the proposed tool a useful additional instrument to be considered in variance forecasting, alongside other possible predictors. In addition—as discussed in Sect. 3.3—our proposal is not based on a time-series model, making it more affordable, and, in case of missing data and short data series, more tractable.

4.1.2 Dividend inclusion

It is important to stress that in the presented analysis we used the S&P 500 Total Returns index, which includes dividends to describe the effective return of an investor. As a device for

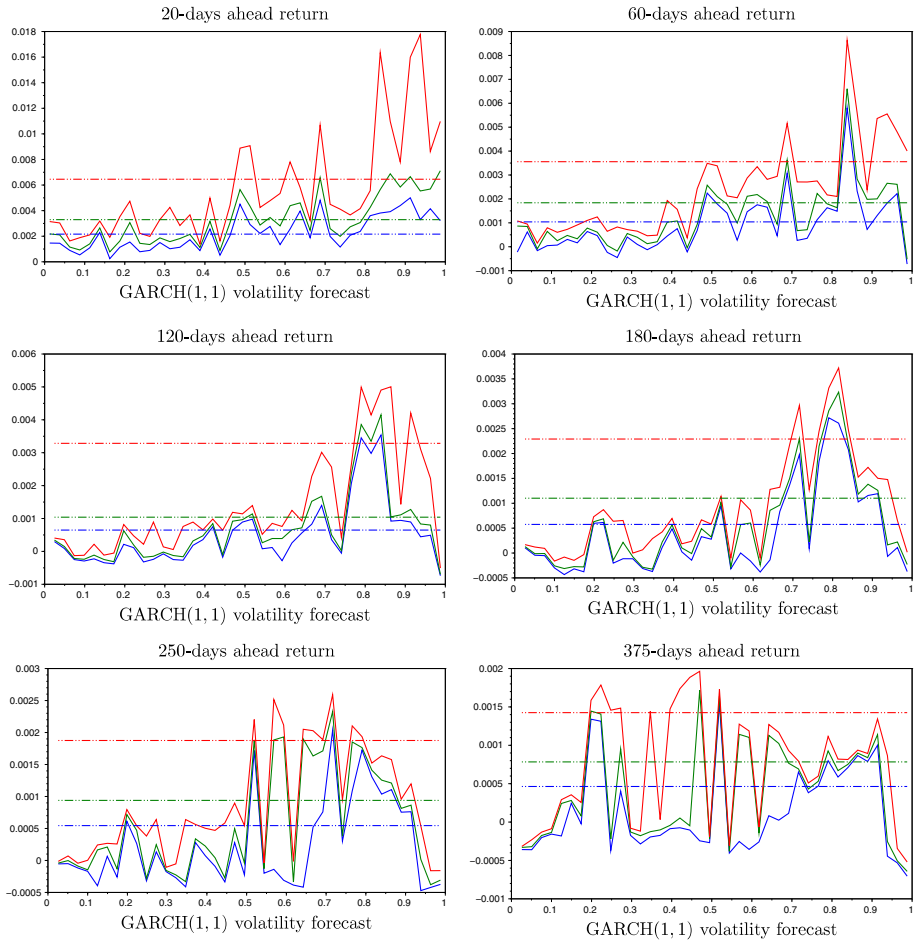


Fig. 11 Out-of-sample. Value at Risk for the forecast return, conditional with respect to the GARCH(1,1) variance forecast. For the complete description of the graphs for all the considered horizons and levels see Fig. 9

robustness check, we here consider also the S&P 500 index without dividends and present two different ways of using it. First, we simply replace the data with the S&P 500 index; then we propose to use the S&P 500 to measure the indicator (the p value of χ^2 test against the BL) and the S&P 500 Total Returns index to compute the returns (as in an actual investment). The results are summarized in Fig. 13. In these cases, the phenomenon is confirmed and it is even more pronounced. We can even conclude that the index without dividend is able to provide a better risk signal, than the Total Return index.

4.1.3 Subperiods

To ascertain that the phenomenon is persistent, another check we present is about the time period analyzed. We consider two time intervals: from Jan 4, 1988 to Jul 18, 2003, and from Jul 21, 2003 to Feb 15, 2019. Both intervals are shorter than the analyzed full sample

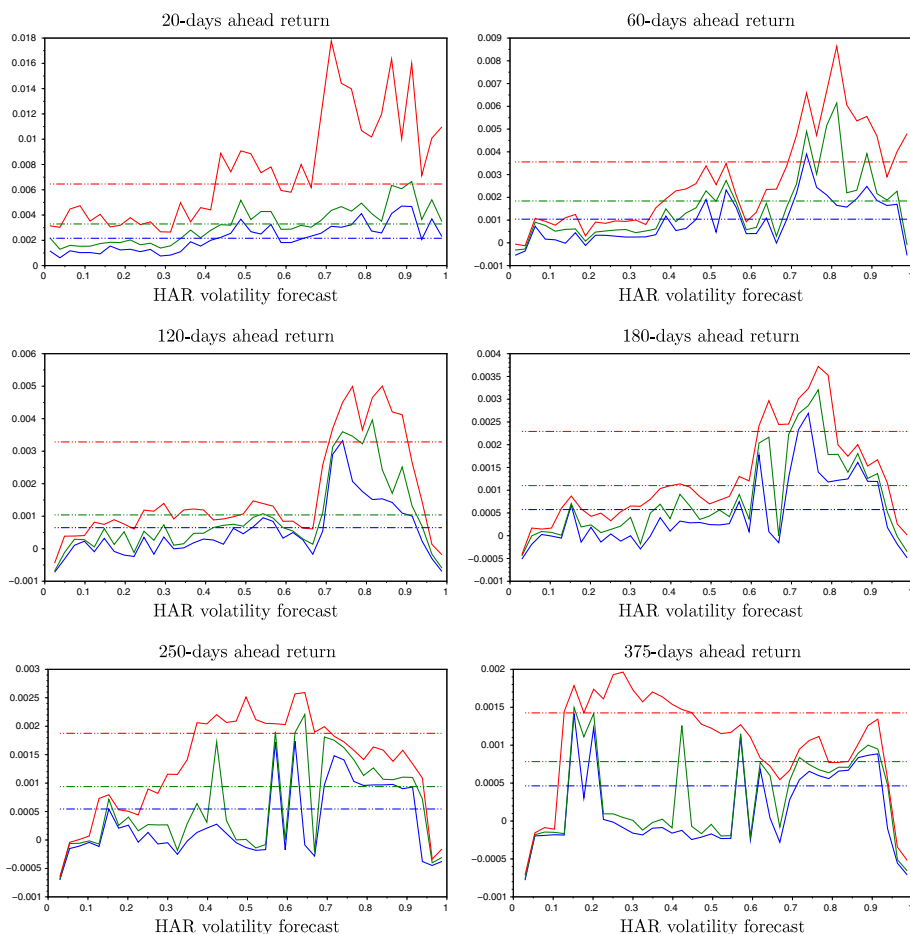


Fig. 12 Out-of-sample. These sub-graphs represent the Value at Risk for the forecast return, conditional with respect to the HAR variance forecast. Refer to the caption of Fig. 9 for the details

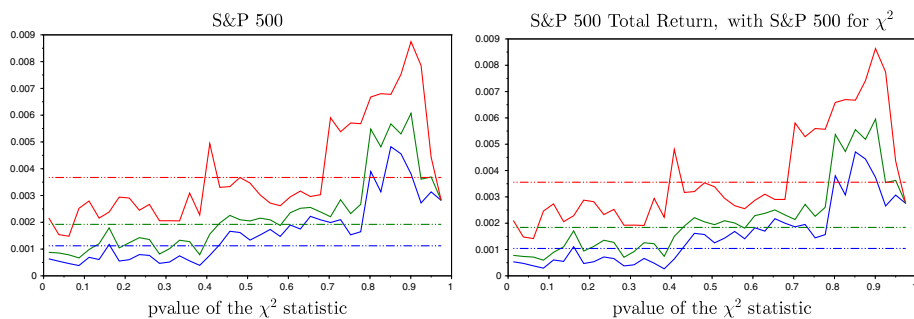


Fig. 13 Out-of-sample. Value at Risk for the forecast return, conditional with respect to the p value of the χ^2 statistic. Forecasting horizon of about 3 months. S&P 500 index without dividends (left); S&P 500 index without dividends for the computation of the χ statistics, S&P 500 Total Returns index for the computation of the VaRs (right). The continuous lines represent the conditional VaR at levels 10% (blue), 5% (green), 1% (red); the dashed lines show the corresponding unconditional values. (Color figure online)

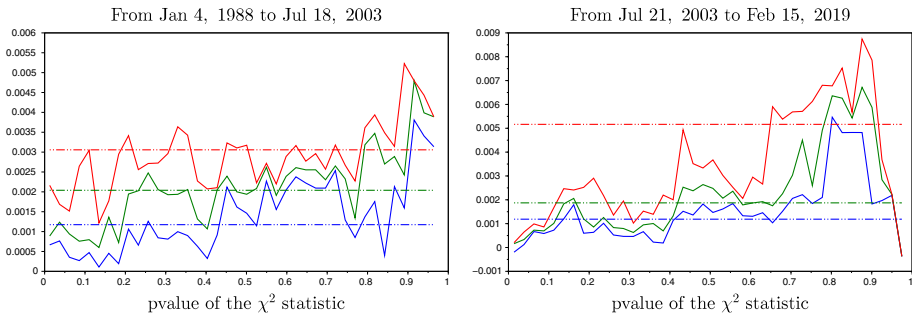


Fig. 14 Out-of-sample. Here, the Value at Risk for the forecast return is conditional with respect to the p value of the χ^2 statistic. Forecasting horizon of about 3 months. From Jan 4, 1988 to Jul 18, 2003 (left), from Jul 21, 2003 to Feb 15, 2019 (right)

period, but they cover together a longer period. Figure 14 shows that the forecasting power is robust across time periods and the effect is sharper on the recent years (the abrupt drop in the rightmost side of the right plot can be due to the low number of observations in the last intervals—28 and 14 observations only).

4.1.4 Discretization of the p_t values

Another robustness check regards the number of observation in each interval on which the possible values of p_t is divided: in our case we consider 40 intervals containing a number of elements ranging from 50 to 452, and generally decreasing from left to right (see Fig. 5). To discuss this point we perform two different checks: first, we consider Gaussian random data with the same number of observation as p_t , for each one of the 40 intervals; second, given t , we divide the range of p_t into 40 intervals of different length, but with the same number of observation in each one. This last case corresponds to work, in place of p_t , with its quantiles.

From the left panel of Fig. 15 we can see that, with independent Gaussian data, the different number of observation per interval, per se, has not any noticeable effect. In fact, the three lines are overall flat, with a noise, and without any visible pattern. In addition, the right panel of Fig. 15 clearly shows that, using p_t , the clustering of its range in intervals with the same number of observations can even made clearer the phenomenon we study.

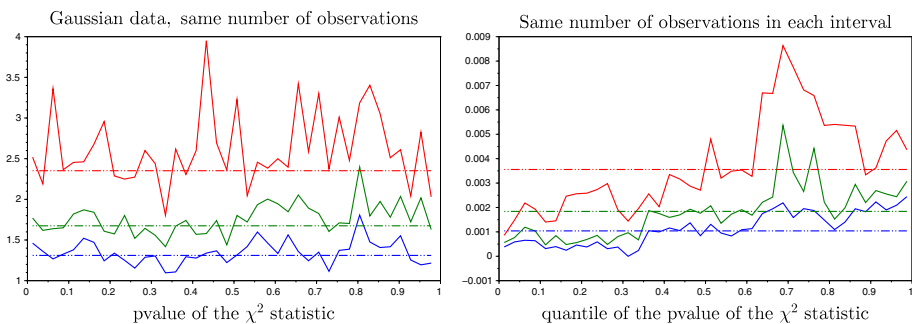


Fig. 15 Out-of-sample. Same conditioning as the previous figure. Random Gaussian data, with the same number of observation as p_t in each of the sub-intervals (left), real data, with uniform number of observation in each interval. The continuous lines represent the conditional VaR at levels 10% (blue), 5% (green), 1% (red); the dashed lines show the corresponding unconditional values. (Color figure online)

Following all these checks, we can conclude that our results are consistent with existing risk predictors, and rather robust, with respect to financially reasonable changes that might be introduced in the considered dataset.

5 Conclusions

This paper presents a novel view of BL as suitable decision support tool for investors and institutions, to detect and forecast the level of the risk of a series of financial returns. At this aim, an exploration of a large set of daily returns of the S&P 500 has been carried out.

A in- and out-of-sample analysis have been implemented, to better illustrate also the attitude of BL of leading to good forecasting experiments of daily returns.

We have found evidence of the direct relation between the compliance with BL and the riskiness of the series. More specifically, some facts emerge: first, means and standard deviations of the returns exhibit opposite behaviors when conditioned to the validity of the BL, so that the mean decreases as the standard deviation increases. In this respect, the case of financial distress with high standard deviation and low expected value occurs for high values of the p -values of the χ^2 statistics. This empirical finding is confirmed either in the in-sample as well as in the out-of-sample analysis, being much more evident for short forecasting horizons in this latter case; second, VaRs have large values in the case of high p -values in the out-of-sample case and when the horizon is short-middle. This outcome is partially confirmed in the in-sample case, where VaRs are above their unconditioned values for large p -values when they are taken at 5% and 10% levels—being more questionable at 1% level.

Two final targets have been achieved through this paper: on one side, we have shed light on an unexplored feature of BL when applied to financial data; on the other side, we have advanced a further decision support instrument in the context of financial risk assessment.

Interestingly, the results of this paper lead to relevant insights on how financial risk can be effectively predicted. Policymakers can take daily financial returns and test over them the validity of the BL. On the basis of the compliance with BL, this study offers a reliable estimation of the future risk level of the financial instruments associated with the investigated returns. Notice that the grounding methodological device is simple to be used, and financial data are usually easily accessible and available for long periods. This is a remarkable positive aspect of the proposed statistical instrument. Furthermore, results are shown to be robust with respect to financially reasonable changes that might be introduced in the considered dataset.

It is important to observe that the statistical device used for BL compliance may lead to biases in the final outcome; this is due to the scale invariance of this statistical regularity, and occurs mainly in presence of large datasets (see e.g. Druica et al., 2018; Nigrini, 2017). More specifically, the quoted papers correctly argue and statistically test that a large sample can be multiplied by a suitably selected random variable or any given scalar, without changing the compliance of the data with BL through checking null hypothesis. Analogously, one is able to remove data from the sample without radically changing the p value of statistical test when controlling for the validity of BL. We also notice that the χ^2 statistics is at the center of a debate about the severity of rejecting the BL, as claimed by Kossovsky (2014). Facing the limitations of the procedures for checking BL deserves an accurate analysis in future research, in accord to the recommendations of the above-mentioned literature contributions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abrantes-Metz, R. M., Kraten, M., Metz, A. D., & Seow, G. S. (2012). Libor manipulation? *Journal of Banking & Finance*, 36(1), 136–150.
- Akyildirim, E., Goncu, A., & Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research*, 297(1), 3–36.
- Al Janabi, M. A. (2013). Optimal and coherent economic-capital structures: Evidence from long and short-sales trading positions under illiquid market perspectives. *Annals of Operations Research*, 205(1), 109–139.
- Alali, F. A., & Romero, S. (2013). Benford's Law: Analyzing a decade of financial data. *Journal of Emerging Technologies in Accounting*, 10, 1–39.
- Alexander, C. (2008). *Market risk analysis, practical financial econometrics* (Vol. 2). John Wiley & Sons.
- Ausloos, M., Castellano, R., & Cerqueti, R. (2016). Regularities and discrepancies of credit default swaps: A data science approach through Benford's Law. *Chaos, Solitons & Fractals*, 90, 8–17.
- Ausloos, M., Cerqueti, R., & Mir, T. A. (2017). Data science for assessing possible tax income manipulation: The case of Italy. *Chaos, Solitons & Fractals*, 104, 238–256.
- Avdoulas, C., Bekiros, S., & Boubaker, S. (2018). Evolutionary-based return forecasting with nonlinear STAR models: Evidence from the Eurozone peripheral stock markets. *Annals of Operations Research*, 262(2), 307–333.
- Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1), 5–59.
- Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4), 551–572.
- Bernard, P., De Freitas, N. E. M., & Maillet, B. B. (2019). A financial fraud detection indicator for investors: An IDEa. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-019-03360-6>
- Bhattacharya, S., Xu, D., & Kumar, K. (2011). An ANN-based auditor decision support system using Benford's law. *Decision Support Systems*, 50(3), 576–584.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327.
- Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, 52(1–2), 5–59.
- Bollerslev, T., Hood, B., Huss, J., & Pedersen, L. H. (2018). Risk everywhere: Modeling and managing volatility. *The Review of Financial Studies*, 31(7), 2729–2773.
- Borges, M. R. (2010). Efficient market hypothesis in European stock markets. *European Journal of Finance*, 16, 711–726.
- Boros, E., Crama, Y., Hammer, P. L., Ibaraki, T., Kogan, A., & Makino. (2011). Logical analysis of data: Classification with justification. *Annals of Operations Research*, 188(1), 33.
- Carrera, C. (2015). Tracking exchange rate management in Latin America. *Review of Financial Economics*, 25(C), 35–41.
- Castellano, R., Cerqueti, R., & Rotundo, G. (2020). Exploring the financial risk of a temperature index: A fractional integrated approach. *Annals of Operations Research*, 284(1), 225–242.
- Cerqueti, R., & Ausloos, M. (2015). Evidence of economic regularities and disparities of Italian regions from aggregated tax income size data. *Physica A*, 421, 187–207.
- Cerqueti, R., & Maggi, M. (2021). Data validity and statistical conformity with Benford's Law. *Chaos, Solitons & Fractals*, 144, 110740.
- Clippe, P., & Ausloos, M. (2012). Benford's law and Theil transform of financial data. *Physica A*, 391(24), 6556–6567.
- Cooper, W. W., Kingyens, A. T., & Paradi, J. C. (2014). Two-stage financial risk tolerance assessment using data envelopment analysis. *European Journal of Operational Research*, 233(1), 273–280.

- Corazza, M., Ellero, A., & Zorzi, A. (2010). Checking financial markets via Benford's law: The S&P 500 case. In M. Corazza & C. Pizzi (Eds.), *Mathematical and statistical methods for actuarial sciences and finance* (pp. 93–102). Springer-Verlag.
- Corsi, F., Audrino, F., & Renó, R. (2012). HAR modeling for realized volatility forecasting. In L. Bauwens, C. M. Hafner, & S. Laurent (Eds.), *Handbook of volatility models and their applications* (Vol. 3, pp. 363–382). John Wiley & Sons.
- D'Ecclesia, R. L., & Clementi, D. (2021). Volatility in the stock market: ANN versus parametric models. *Annals of Operations Research*, 299(1), 1101–1127.
- De Ceuster, M. J. K., Dhaene, G., & Schatteman, T. (1998). On the hypothesis of psychological barriers in stock markets and Benford's Law. *Journal of Empirical Finance*, 5, 263–279.
- Deleanu, I. S. (2017). Do countries consistently engage in misinforming the international community about their efforts to combat money laundering? Evidence using Benford's Law. *PLoS ONE*, 12(1), e0169632.
- Desai, V. S., & Bharati, R. (1998). A comparison of linear regression and neural network methods for predicting excess returns on large stocks. *Annals of Operations Research*, 78, 127–163.
- Dhesi, G., Shakeel, B., & Ausloos, M. (2021). Modelling and forecasting the kurtosis and returns distributions of financial markets: Irrational fractional Brownian motion model approach. *Annals of Operations Research*, 299(1), 1397–1410.
- Ding, Z., & Granger, C. W. (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, 73(1), 185–215.
- Druica, E., Oancea, B., & Valsan, C. (2018). Benford's law and the limits of digit analysis. *International Journal of Accounting Information Systems*, 31, 75–82.
- du Jardin, P., & Severin, E. (2012). Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time. *European Journal of Operational Research*, 221(2), 378–396.
- Duan, L., Xu, L., Liu, Y., & Lee, J. (2009). Cluster-based outlier detection. *Annals of Operations Research*, 168(1), 151–168.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987–1007.
- Feldman, D., & Xu, X. (2018). Equilibrium-based volatility models of the market portfolio rate of return (peacock tails or stotting gazelles). *Annals of Operations Research*, 262(2), 493–518.
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669.
- Gavrishchaka, V. V., & Banerjee, S. (2006). Support vector machine as an efficient framework for stock market volatility forecasting. *Computational Management Science*, 3(2), 147–160.
- Groth, S. S., & Muntermann, J. (2011). An intraday market risk management approach based on textual analysis. *Decision Support Systems*, 50(4), 680–691.
- Gu, W., & Peng, Y. (2019). Forecasting the market return direction based on a time-varying probability density model. *Technological Forecasting and Social Change*, 148, 119726.
- Guerard, J. B., Xu, G., & Markowitz, H. (2021). A further analysis of robust regression modeling and data mining corrections testing in global stocks. *Annals of Operations Research*, 303(1), 175–195.
- Günneel, S., & Tödter, K. H. (2009). Does Benford's Law hold in economic research and forecasting? *Empirica*, 36(3), 273–292.
- Hales, D. N., Chakravorty, S. S., & Sridharan, V. (2009). Testing Benford's Law for improving supply chain decision-making: A field experiment. *International Journal of Production Economics*, 122(2), 606–618.
- Hales, D. N., Sridharan, V., Radhakrishnan, A., Chakravorty, S. S., & Siha, S. M. (2008). Testing the accuracy of employee-reported data: An inexpensive alternative approach to traditional methods. *European Journal of Operational Research*, 189(3), 583–593.
- Hand, D. J. (2010). Fraud detection in telecommunications and banking: Discussion of Becker, Volinsky, and Wilks (2010) and Sudjianto et al. (2010). *Technometrics*, 52(1), 34–38.
- Hill, T. P. (1995). A Statistical derivation of the significant-digit law. *Statistical Science*, 10(4), 354–363.
- Hill, T. P. (1995). Base-invariance implies Benford's Law. *Proceedings of the American Mathematical Society*, 123, 887–895.
- Hill, T. P. (1995). The significant-digit phenomenon. *The American Mathematical Monthly*, 102(4), 322–327.
- Hill, T. P. (1998). The First Digit Phenomenon: A century-old observation about an unexpected pattern in many numerical tables applies to the stock market, census statistics and accounting data. *American Scientist*, 86(4), 358–363.
- Holz, C. A. (2014). The quality of China's GDP statistics. *China Economic Review*, 30, 309–338.
- Huang, S. M., Yen, D. C., Yang, L. W., & Hua, J. S. (2008). An investigation of Zipf's Law for fraud detection (DSS# 06–10–1826R (2)). *Decision Support Systems*, 46(1), 70–83.
- Huang, Y., & Kou, G. (2014). A kernel entropy manifold learning approach for financial data analysis. *Decision Support Systems*, 64, 31–42.

- Jana, R. K., Ghosh, I., & Das, D. (2021). A differential evolution-based regression framework for forecasting Bitcoin price. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-021-04000-8>
- Jiang, C., Wang, Z., Wang, R., & Ding, Y. (2018). Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending. *Annals of Operations Research*, 266(1), 511–529.
- Juergens, J. L., & Lindsey, L. (2009). Getting out early: An analysis of market making activity at the recommending analyst's firm. *The Journal of Finance*, 64, 2327–2359.
- Karavardar, A. (2014). Benford's law and an analysis in Istanbul stock exchange (BIST). *International Journal of Business and Management*, 9, 160–172.
- Kossofsky, A. E. (2014). *Benford's Law: Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications* (Vol. 3). World Scientific.
- Kyriakou, I., Mousavi, P., Nielsen, J. P., & Scholz, M. (2021). Forecasting benchmarks of long-term stock returns via machine learning. *Annals of Operations Research*, 297(1), 221–240.
- Kürüm, E., Weber, G. W., & Iyigun, C. (2018). Early warning on stock market bubbles via methods of optimization, clustering and inverse problems. *Annals of Operations Research*, 260(1–2), 293–320.
- Ley, E. (1996). On the peculiar distribution of the U.S. stock indices digits. *The American Statistician*, 50, 311–314.
- Ley, E., & Varian, H. R. (1994). Are there psychological barriers in the Dow-Jones index? *Applied Financial Economics*, 4(3), 217–224.
- Liu, Y. (2019). Novel volatility forecasting using deep learning-long short term memory recurrent neural networks. *Expert Systems with Applications*, 132, 99–109.
- Meng, X., & Taylor, J. W. (2020). Estimating value-at-risk and expected Shortfall using the intraday low and range data. *European Journal of Operational Research*, 280(1), 191–202.
- Michalski, T., & Stoltz, G. (2013). Do countries falsify economic data strategically? Some evidence that they might. *Review of Economics and Statistics*, 95(2), 591–616.
- Mir, T. A. (2016). The leading digit distribution of the worldwide illicit financial flows. *Quality & Quantity*, 50(1), 271–281.
- Mir, T. A., & Ausloos, M. (2018). Benford's law: A sleeping beauty sleeping in the dirty pages of logarithmic tables. *Journal of the Association for Information Science and Technology*, 69(3), 349–358.
- Mir, T. A., Ausloos, M., & Cerqueti, R. (2014). Benford's Law predicted digit distribution of aggregated income taxes: The surprising conformity of Italian cities and regions. *The European Physical Journal B*, 87(11), 12–8.
- Neely, C. J. (2011). A survey of announcement effects on foreign exchange volatility and jumps. *Federal Reserve Bank of St. Louis Review*, 93(5), 361–407.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2), 347–370.
- Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, 4(1), 39–40.
- Ngai, E., Hu, Y., Wong, Y., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3), 559–569.
- Nigrini, M. J. (1996). A taxpayer compliance application of Benford's Law. *The Journal of the American Taxation Association*, 18(1), 72–91.
- Nigrini, M. J. (1999). Fraud detection: I've got your number. *Journal of Accountancy*, 187, 79–83.
- Nigrini, M. J. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud detection*. John Wiley & Sons.
- Nigrini, M. J. (2015). Persistent patterns in stock returns, stock volumes, and accounting data in the U.S. Capital markets. *Journal of Accounting, Auditing & Finance*, 30(4), 541–557.
- Nigrini, M. J. (2017). Audit sampling using Benford's law: A review of the literature with some new perspectives. *Journal of Emerging Technologies in Accounting*, 14(2), 29–46.
- Noakes, M. A., & Rajaratnam, K. (2016). Testing market efficiency on the Johannesburg Stock Exchange using the overlapping serial test. *Annals of Operations Research*, 243(1), 273–300.
- Nye, J., & Moul, C. (2007). The political economy of numbers: On the application of Benford's Law to international macroeconomic statistics. *The BE Journal of Macroeconomics*, 7(1), 1–14.
- Oztek, A., Kizilaslan, R., Freund, S., & Iseri, A. (2016). A data analytic approach to forecasting daily stock returns in an emerging market. *European Journal of Operational Research*, 253(3), 697–710.
- Patton, A. J., Ramadorai, T., & Streatfield, M. (2015). Change you can believe in? Hedge fund data revisions. *Journal of Finance*, 70(3), 963–999.
- Pietronero, L., Tosatti, E., Tosatti, V., & Vespignani, A. (2001). Explaining the uneven distribution of numbers in nature: The laws of Benford and Zipf. *Physica A*, 293(1–2), 297–304.

- Rauch, B., Götttsche, M., Brähler, G., & Engel, S. (2011). Fact and fiction in EU-governmental economic data. *German Economic Review*, 12(3), 243–255.
- Rauch, B., Götttsche, M., Brähler, G., & Kronfeld, T. (2014). Deficit versus social statistics: Empirical evidence for the effectiveness of Benford's law. *Applied Economics Letters*, 21(3), 147–151.
- Realdon, M. (2008). Credit default swap rates and stock prices. *Applied Financial Economics Letters*, 4(4), 241–248.
- Riccioni, J., & Cerqueti, R. (2018). Regular paths in financial markets: Investigating the Benford's law. *Chaos, Solitons & Fractals*, 107, 186–194.
- Santos, D. G., & Ziegelmann, F. A. (2014). Volatility forecasting via MIDAS, HAR and their combination: An empirical comparative study for IBOVESPA. *Journal of Forecasting*, 33(4), 284–299.
- Satchell, S., & Knight, J. (2011). *Forecasting volatility in the financial markets*. Elsevier.
- Shi, J., Ausloos, M., & Zhu, T. (2018). Benford's law first significant digit and distribution distances for testing the reliability of financial reports in developing countries. *Physica A*, 492, 878–888.
- Sudjianto, A., Nair, S., Yuan, M., Zhang, A., Kern, D., & Cela-Diaz, F. (2010). Statistical methods for fighting financial crimes. *Technometrics*, 52, 5–19.
- Tödter, K. (2009). Benford's Law as an indicator of fraud in economics. *German Economic Review*, 10(3), 339–351.
- Venkatesh, K., Ravi, V., Prinzie, A., & Van den Poel, D. (2014). Cash demand forecasting in ATMs by clustering and neural networks. *European Journal of Operational Research*, 232(2), 383–392.
- Vortelinos, D. I. (2017). Forecasting realized volatility: HAR against Principal Components Combining, neural networks and GARCH. *Research in International Business and Finance*, 39, 824–839.
- Yap, B. W., & Sim, C. H. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12), 2141–2155.
- Zumbach, G. (2004). Volatility processes and volatility forecast with long memory. *Quantitative Finance*, 4(1), 70–86.