ISSN: 2755-0214

Journal of Economics & Management Research

Review Article

Open Access

Assessing the Similarity Between the q-Gaussian Distribution and Surprises in Earnings Forecasting

Yuri Becaleti dos Santos

Master of Business Administration PhD Student at University of São Paulo, Brazil

***Corresponding author**
Yuri Becaleti dos Santos, Master of Business Administration PhD Student at University of São Paulo, Brazil.

Received: July 06, 2025; **Accepted:** July 10, 2025; **Published:** August 05, 2025

Introduction

The proposal of a generalization of statistical mechanics by has been widely applied across various scientific fields, including financial studies. Investigated the distributions of inter-nucleotide intervals in the DNA of various organisms and concluded that these intervals can be described by a q-exponential distribution proposed by [1]. In the same study, the authors suggested the application of these findings to the financial market. One possible application of the generalization of statistical mechanics to the financial sector may lie in the accuracy of earnings forecasts made by analysts.

Earnings forecasts can be analyzed as a time series. This time series is extraordinarily complex and contains a nonlinear memory that may help explain its events [2]. Recent studies dedicated to the accuracy of earnings forecasts have employed parametric methodologies and artificial intelligence [3,4]. Furthermore, proposed the strong form of market efficiency, in which prices can be described as a function of available information [5]. Thus, it is expected that, over time, the amount of information available about a company increases, which may be reflected in greater accuracy in earnings forecasting.

Demonstrate that the financial market can be analyzed as a Complex System (CS). A system is said to be complex when the properties of its individual components are not equal to the properties of the system as a whole [6,7]. Thus, the information available in the market is not capable, per se, of explaining the behavior of earnings, requiring the consideration of possible interactions among such information. The behavior of Complex Systems is a topic investigated in areas of physics such as thermodynamics. Investigated the dynamics of a laser passing through boiling water bubbles and concluded that the intervals of this light beam do not follow an exponential and symmetric probability density distribution like the Gaussian normal, but rather resemble a q-Gaussian distribution as proposed by [8].

Proposed a generalization of the laws of thermodynamics that can be applied to Complex Systems and used to describe the Probability Density Function (PDF) of their events. In this way, considering the complexity of the capital market, it is expected that the accuracy of analysts' earnings forecasts follows a non-exponential distribution, as observed in other complex systems, such as those studied by [9]. In this regard, analyzed the normalized absolute return of the 100 largest American companies and identified that, for different time intervals, the Probability Density Function of these returns resembles the q-Gaussian (or q-normal) PDF proposed by [10].

In addition to absolute returns, another investigation that can be applied to the financial sector is the surprise associated with earnings forecasts [11]. Investigated the association between investor attention and stock market returns in the context of extreme earnings surprises. The earnings surprise can be measured by the difference between the consensus of analysts' earnings forecasts and the earnings actually reported. By identifying the behavior of earnings surprises, it may be possible to observe the complexity of the capital market system. Therefore, the following research question is proposed: Can the Probability Density Function of earnings surprise be described by a q-exponential distribution?

To answer this question, data on earnings surprises were collected from four major companies: BlackRock, Microsoft, Petrobras, and Vale. The data were organized into time series. Time series analysis techniques were applied to assess the effectiveness of potential predictive models for earnings surprise. The Probability Density Function of earnings surprise was estimated using a hypergeometric distribution. The estimated Probability Density Function for each of these companies was then compared to the q-Gaussian Probability Density Function proposed by.

The results of this study may benefit the research literature on earnings forecasting, since the earnings surprise is associated with the accuracy of analysts' forecasts—that is, the greater the accuracy of the analyst's forecast, the smaller the earnings surprise. Thus, the earnings surprise can be analyzed as the forecast error of the analyst, regardless of the model applied.

In this sense, identifying that the earnings surprise follows a q-normal distribution may benefit the earnings forecasting literature, as normality tests of residuals are commonly used as robustness checks. Furthermore, the results may be useful for investors and creditors who rely on analysts' earnings forecasts for decision-making and may also impact finance research involving ex ante cost of equity estimation models, such as, for example [12].

Theoretical Framework

Suggest that traditional financial theories (the orthodox thinking that dates back to Bachelier's 1900 thesis and underpins models such as Black & Scholes, CAPM, and their derivatives) fail to capture the singularities present in time series representative of this system. Furthermore, these theories often produce models whose results prove inadequate to describe the reality of this Complex System, which is influenced by multiple forces and heterogeneous agents [13].

The term Complex Systems can be used to refer to systems with a large number of interactive components [14]. In this sense, models that account for the complexity of agents and market structures within the economic system, as well as the nonlinear connections among events that permeate this environment, would be capable of capturing the influence of supply and demand cycles on prices, which imply situations of dependency and constant price reversals [15].

Still regarding Complex Systems, Macau (n.d.) explains that, unlike Simple Systems (SS), Complex Systems may exhibit different behavioral rules as the scale of a given parameter changes. For example, when observing a glass of water over a short period of time, it is possible to measure the pressure and temperature acting on the fluid. However, when the observation period is extended, the water in the container may evaporate and dissipate into the environment. In this case, the rules used to describe the behavior of the water during a short, stable interval in the glass are different from the rules applied when the observation period is lengthened and one intends to explain the behavior of the water vapor dispersed in the environment (Macau, n.d.)

These concepts of rule changes as a function of scale variation are the subject of study in thermodynamics. Proposed a generalization of the physical laws applied to thermodynamics, thereby enabling the extension of these theories to Complex Systems and, consequently, to other fields of knowledge such as finance, biology, geology, and astronomy. Specifically with regard to finance, one possible application is earnings.

In this regard, a company's returns can be predicted through the use of publicly available information, such as financial and macroeconomic data series and the company's business cycle [17]. The efforts applied to such forecasting may result in estimates that are as accurate as the quality of the estimation method used, and these results can have a significant impact on the financial world by influencing the efficient allocation of resources and the estimation of the ex-ante cost of equity [18].

Examined the robustness of earnings forecasts made by analysts and demonstrated that, during the 1960s, earnings predictability was low, and that it increased during the 1970s, reinforcing the hypothesis that the financial market can be studied as a Complex System. For example, assume that an analyst seeking to forecast the abnormal earnings (ρ) of a company in period t employs model M and decides to include k vectors in this model. The selection or exclusion of a vector can be calculated by multiplying k by a binary variable that takes the value one when the selection criterion is satisfied and zero otherwise. Thus, considering that the analyst has access to information available up to t–1, it is expected that a model such as the one below will be developed:

Equation 1: General Abnormal Earnings Forecasting Model Proposed by Pesaran and Timmermann (1995)

$$M_I: \rho_{t+1} = \beta_i' X_{t,i} + \varepsilon_{t+1,i}$$

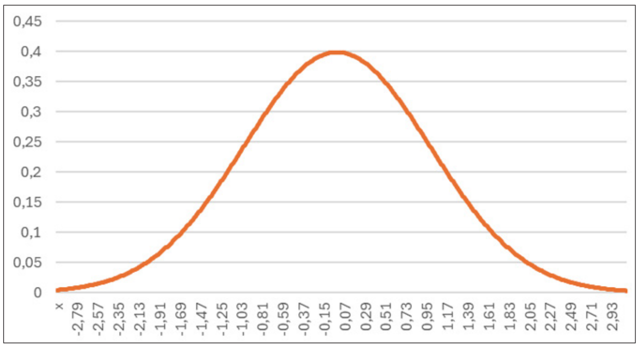
Where: t = 1, 2, ..., t–1; $X_{(t,i)}$ is a $(k_i+1)\times 1$ vector selected by the investor; β is an estimated coefficient, usually obtained through the Ordinary Least Squares (OLS) method; and ε is the model's error term.

Thus, if we consider the financial market as a Simple System, it can be expected that, over time, more information and more parameters will be included in the model, causing the error term to adopt a random behavior with equiprobable values based on their distance from the mean indicating the normality of the residuals. The normality of the residuals means that they can be described by a Gaussian normal Probability Density Function (PDF). The Gaussian PDF has a mean equal to zero and a standard deviation equal to one, and can be described by the following equation:

Equation 2: Gaussian Normal Probability Density Function

$$FDP(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

When applying the Gaussian normal PDF to random numbers and plotting the results, the following graph is observed:



Source: Prepared by the Author
Figure 1: Gaussian Normal PDF Applied to Random Numbers

The normal PDF is symmetric around its mean and exhibits flattening proportional to its standard deviation. In this sense, the greater the standard deviation, the flatter the curve, which means that values farther from the mean may occur more frequently, although the sum of all events tends toward the mean. Alternatively, there is the q-exponential Probability Density Function proposed by.

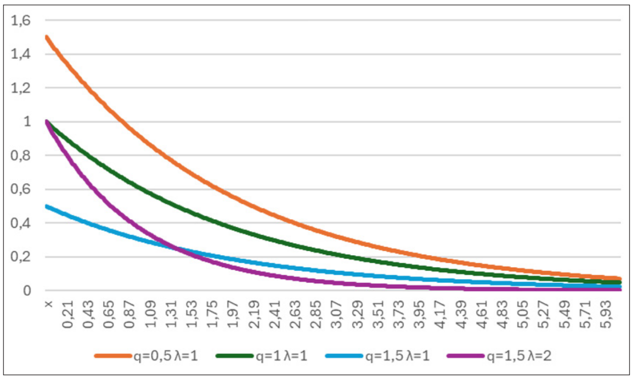
The q-exponential is a generalization of the Boltzmann-Gibbs entropy and has two parameters: q and λ, with $1 \leq q \leq 2$. When q = 1, the Boltzmann-Gibbs entropy is recovered. This Probability Density Function can be described by the following function:

Equation 3: q-Exponential Probability Density Function

$$FDP(x) = (2-q)\lambda e_q^{(-\lambda x)}$$

Once again, when applied to random numbers, the plotted points of the q-exponential distribution may form the following figure:

Citation: Yuri Becaleti dos Santos (2025) Assessing the Similarity Between the q-Gaussian Distribution and Surprises in Earnings Forecasting. Journal of Economics & Management Research. SRC/JESMR-403. DOI: doi.org/10.47363/JESMR/2025(6)304



Source: Prepared by the Author

Figure 2: q-Normal PDF Applied to Random Numbers

In the previous figure, the points of the q-exponential function of were plotted using random numbers for different parameter sets. The q parameter is associated with the slope of the curve, while the other parameter (λ) is associated with its curvature. Unlike the normal PDF, the q-normal distribution is not symmetric around its mean and is not affected by changes in the sample mean or standard deviation. When applied to the context of the financial market, the asymmetry of the PDF may help explain the fact that positive outcomes (to the right of the zero mean) may have different probabilities than negative outcomes (to the left of the zero mean), depending on the company's stage.

According to the Corporate Life Cycle Theory (LCT), it is possible to classify companies into five stages, as proposed by Introduction, Growth, Maturity, Turbulence, and Decline [19,20]. Applying this framework to earnings forecasting, demonstrated that analysts' projections are affected by companies in the Introduction and Decline stages. The fact that the Life Cycle Stages influence analysts' forecasts may support the hypothesis of the complexity of the capital market system. The objective of this study is to identify whether the Probability Density Function of earnings forecast surprise can be described by a q-normal Probability Density Function. To this end, the following research hypothesis is proposed:

H1: The Probability Density Function of earnings forecast surprise can be described by a q-normal Probability Density Function.

To evaluate H1 of this study, data were collected from companies whose earnings forecasts form relatively long time series. The selected companies were: BlackRock Inc. (BLRK), Microsoft Inc. (MSFT), Petrobras S.A. (PETR), and Vale S.A. (VALE). The following section presents the statistics and tests applied in this study.

Data and Methodology

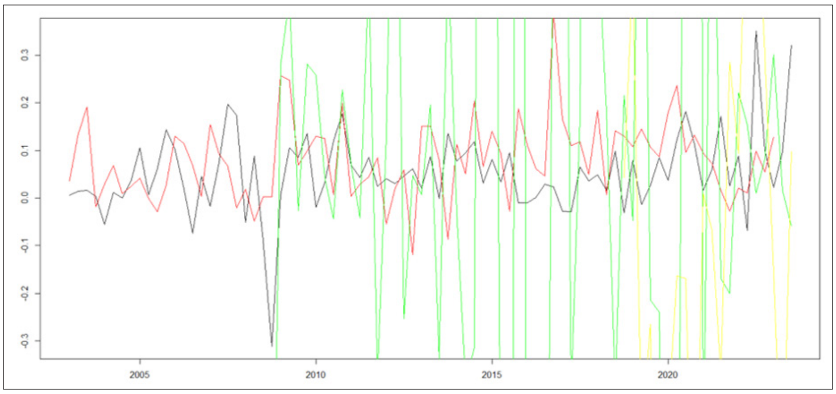
Using the Refinitiv database from Thomson Reuters®, data on the quarterly earnings surprise in percentage terms were collected for the earliest available period for the companies studied. Vale and Petrobras presented missing information in some periods. The longest continuous period (without missing values) for each company was organized into an uninterrupted time series. The data were analyzed using R® software. The following table summarizes the data collected:

Table 1: Summary-Observation Framework

	BLRK	MSFT	PETR	VALE
First Observation	03/2003	03/2003	12/2008	12/2018
Last Observation	09/2023	09/2023	09/2023	09/2023
Observation Count	83	81	60	20
Mean	0,5255	0,0802	0,3188	-0,1060
Standard Deviation	0,0849	0,0835	1,9264	0,4987

Source: Prepared by the Author

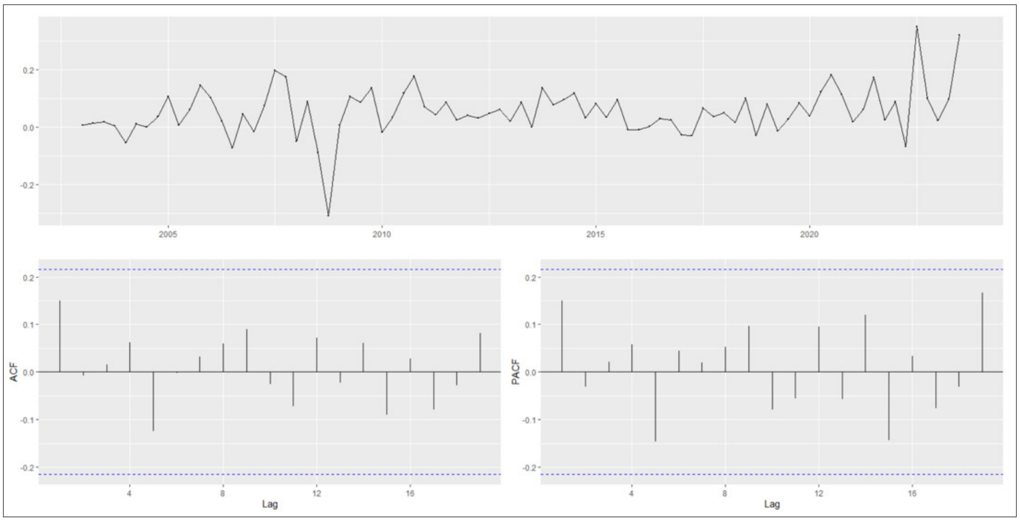
After being organized into time series, the observation data were plotted in the following graph:



Source: Prepared by the Author

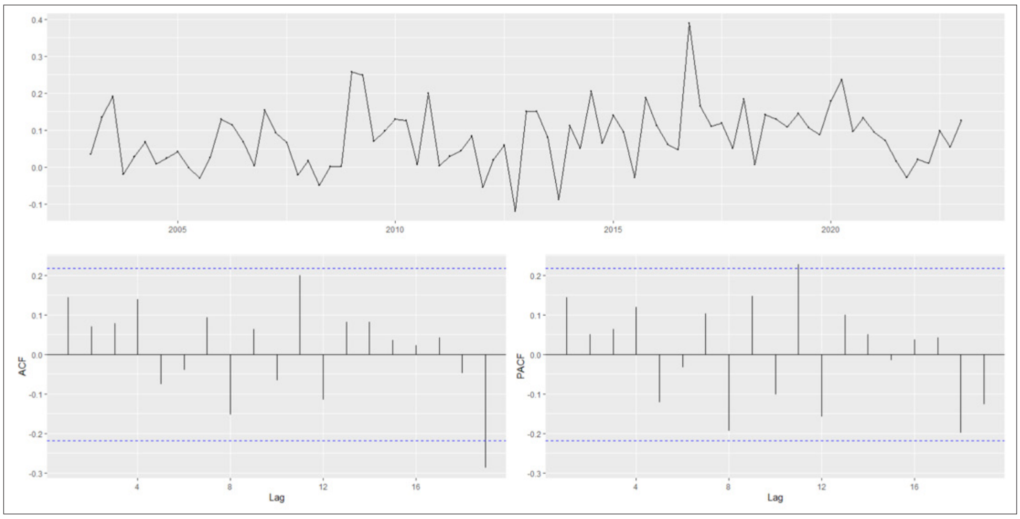
Figure 3: Earnings Surprise in Percentage for the Companies Studied from 2003 to 2023

The black line represents the earnings surprise of BLRK. MSFT, PETR, and VALE are represented by the red, green, and yellow lines, respectively. The following figures provide a more detailed view of each of these series and their Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF). Histograms of these series were also generated. Given their similarity, these images were analyzed jointly:



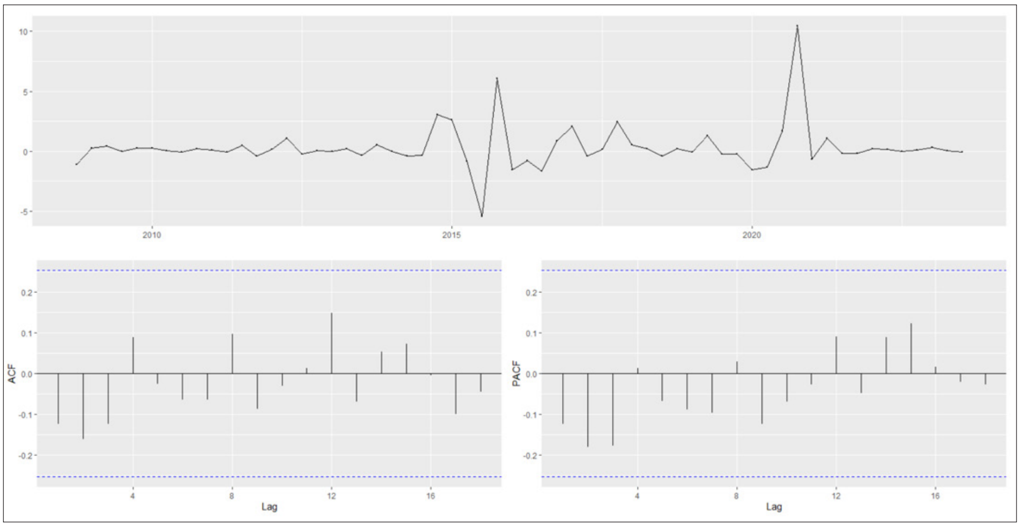
Source: Prepared by the Author

Figure 4: Plot, ACF, and PACF of BLRK



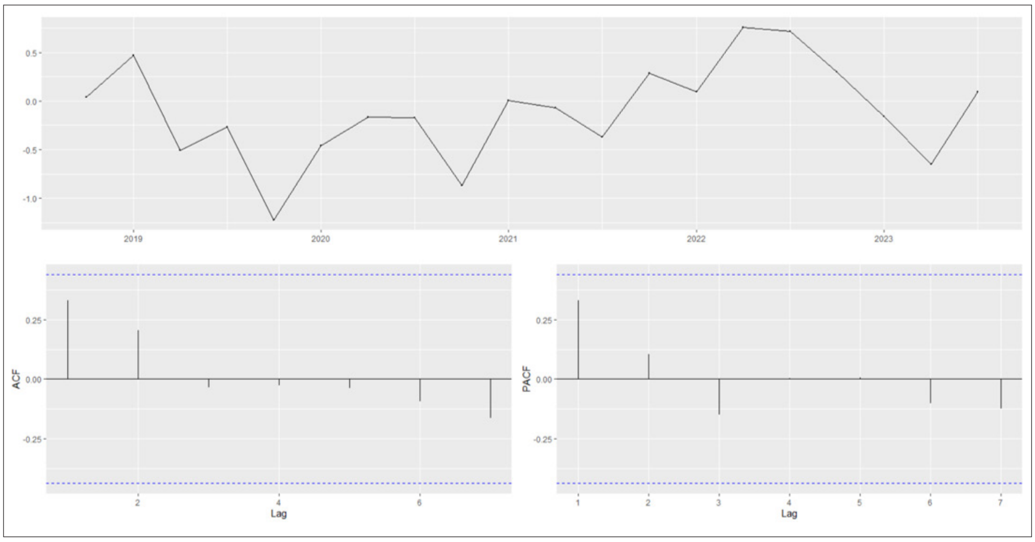
Source: Prepared by the Author

Figure 5: Plot, ACF, and PACF of MSFT



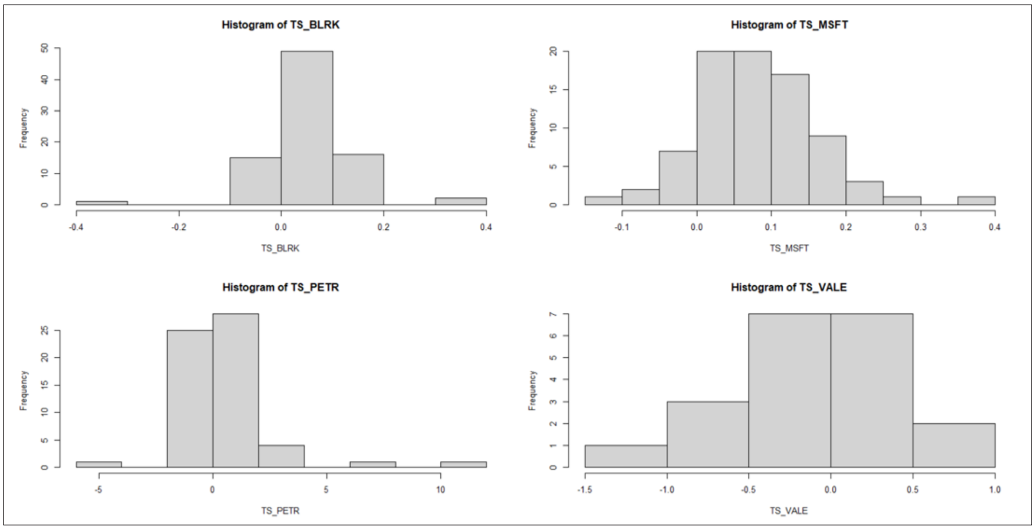
Source: Prepared by the Author

Figura 6: Plotagem, ACF e PACF da PETR



Source: Prepared by the Author

Figure 7: Plot, ACF, and PACF of VALE



Source: Prepared by the Author

Figure 8: Histogram of BLRK, MSFT, PETR, and VALE Data

With the exception of Microsoft Inc. (MSFT), the series did not show evidence of significant autocorrelation. There is evidence of autocorrelation in the MSFT series at a lag of 19 periods, which corresponds to a time span of 4 years and 3 quarters. The partial autocorrelation analysis does not indicate the presence of a fixed parameter in the series. Except for BLRK, which may exhibit an apparently normal distribution, the data appear to follow a non-exponential, long-tailed distribution. Stationarity and normality tests were then applied to each series. The results are presented in the following table:

Table 2: Stationarity and Normality Tests

	BLRK	MSFT	PETR	VALE
Stationarity Test	Dickey-Fuller	Dickey-Fuller	Dickey-Fuller	Dickey-Fuller
p-value	3,638e-10	1,722e-06	1,565e-10	0,0355
Conclusion	Estacionária	Estacionária	Estacionária	Estacionária
Normality Test	Shapiro-Wilk	Shapiro-Wilk	Shapiro-Wilk	Shapiro-Wilk
p-value	1,83e-05	0,1386	3,043e-10	0,9606
Conclusion	Provém de uma distribuição normal	Não provém de uma distribuição normal	Provém de uma distribuição normal	Não provém de uma distribuição normal

Source: Prepared by the Author

The applied tests indicate that the analyzed series contain a unit root and, in the cases of BLRK and PETR, originate from a normal distribution. Subsequently, parameter estimation techniques for predictive models were applied. To this end, the original time series was divided into two parts: testing and forecasting. The forecasting part contains three observations for each company. The testing part consists of the remainder of the time series after excluding the three most recent observations, which were used for forecasting. The following models were applied: NAIVE, MEAN, DRIFT, SNAIVE, SES, additive HOLT-WINTERS, ARIMA, and NEURAL network models. The Mean Absolute Percentage Errors (MAPEs) were calculated for the models applied in both the testing and forecasting phases. The results of the MAPEs for the forecasting tests are presented in the following table:

Table 3: MAPE of the Applied Tests

	BLRK	MSFT	PETR	VALE
NAIVE	141,86124	86,75871	240,4303	217,1269
MEAN	84,60861	33,80898	881,9619	106,5555
DRIFT	144,10798	87,51016	361,6756	238,9553
SNAIVE	158,47128	105,83464	335,6801	339,6527
SES	85,22874	33,85739	881,4708	287,0296
HW ADT	91,88791	102,64597	1.642,6681	818,1072
ARIMA	84,60861	26,91065	100,0000	217,1269
NEURAL	103,65707	52,63699	778,8994	176,8298

Source: Prepared by the Author

The lowest MAPE for BLRK, MSFT, and VALE was observed with the MEAN model, yielding 84.60%, 33.80%, and 106.55%, respectively. For PETR, the model with the lowest MAPE was ARIMA. None of the applied models were able to perfectly predict the three consecutive observations at the end of the testing period. A zero error would indicate a perfect earnings forecast. In this context, the probability of observing a surprise equal to zero (i.e., a perfect earnings forecast by the analyst) was estimated using the Hypergeometric Distribution, as follows:

Equation 4: Hypergeometric Distribution

$$P(x=0) \; = \frac{\binom{N-K}{n}}{\binom{N}{n}}$$

Where: N is the population count, n is the sample count, and K is the number of successes in the population

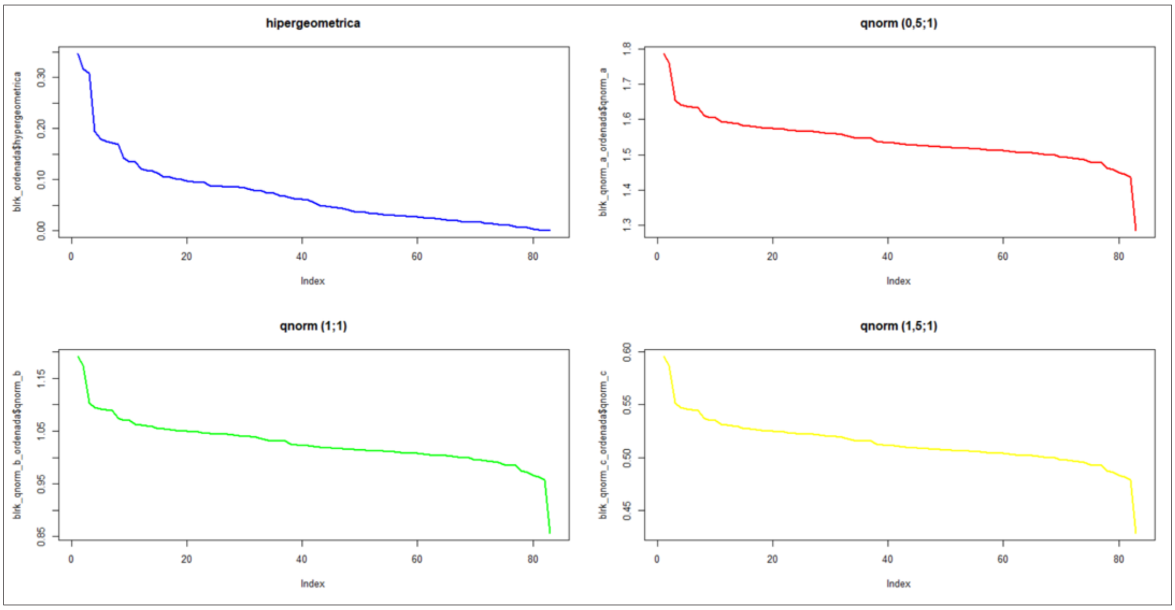
Assuming the possibility of any number being randomly drawn with the removal of a single observation, we have K and n equal to 1. Thus, by substituting into the previous equation, we obtain:

Equation 5: Hypergeometric Distribution with n and K Equal to 1

$$P(x=0) = \frac{\binom{N-K}{n}}{\binom{N}{n}} = \frac{\binom{N-1}{1}}{\binom{N}{1}} = \frac{\frac{(N-1)!}{1!\,(N-1-1)!}}{\frac{N!}{1!\,(N-1)!}} = \frac{\frac{(N-1)!}{(N-2)!}}{\frac{N!}{(N-1)!}}$$

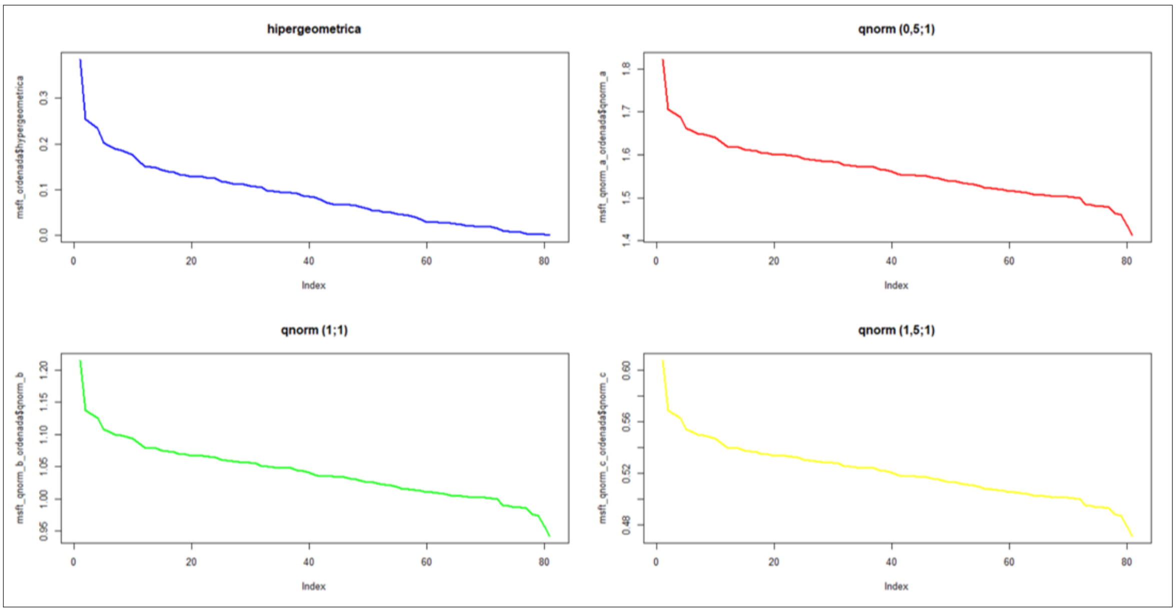
$$= \left(\frac{(N-1)!}{(N-2)!} \times \frac{(N-1)!}{N!}\right) = \frac{[(N-1)!]^2}{(N-2)!\;N!}$$

By substituting x in the previous equation with the observed values of the earnings surprises, it was possible to estimate the Probability Density Function of each observation individually. The results of these probabilities were arranged in ascending order, and over the function's data points, the Probability Density Function of a q-normal distribution was plotted using the following parameter combinations (q; λ): (0.5; 1), (1; 1), and (1.5; 1). The results are presented in the following figures:



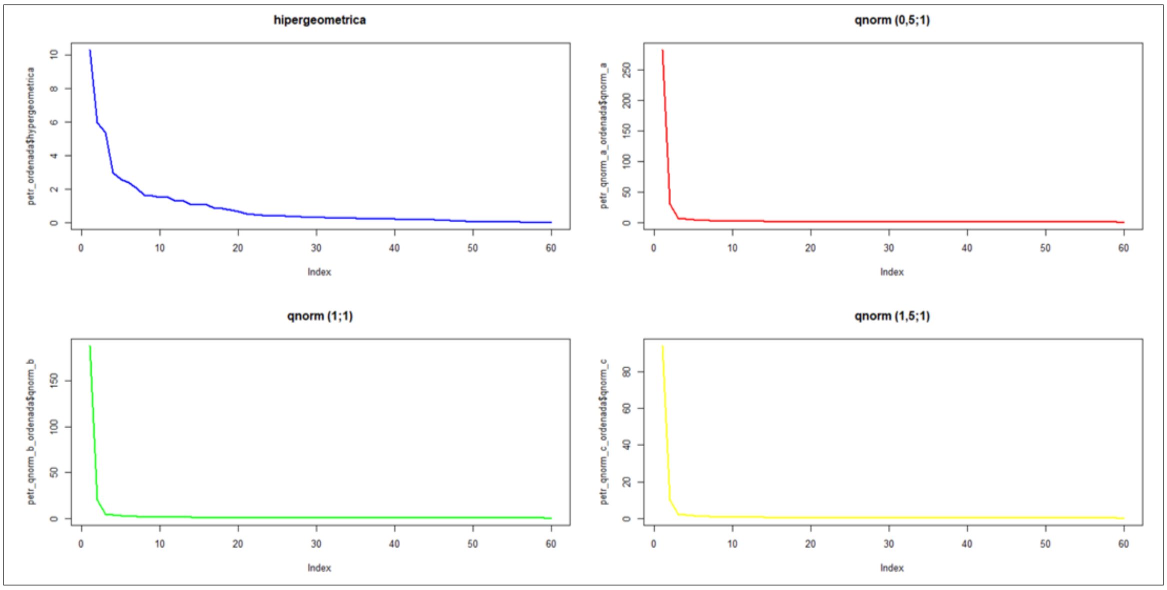
Source: Prepared by the Author

Figure 8: Comparison Between the Hypergeometric Distribution of BLRK and q-Normal Distributions with Different Parameters



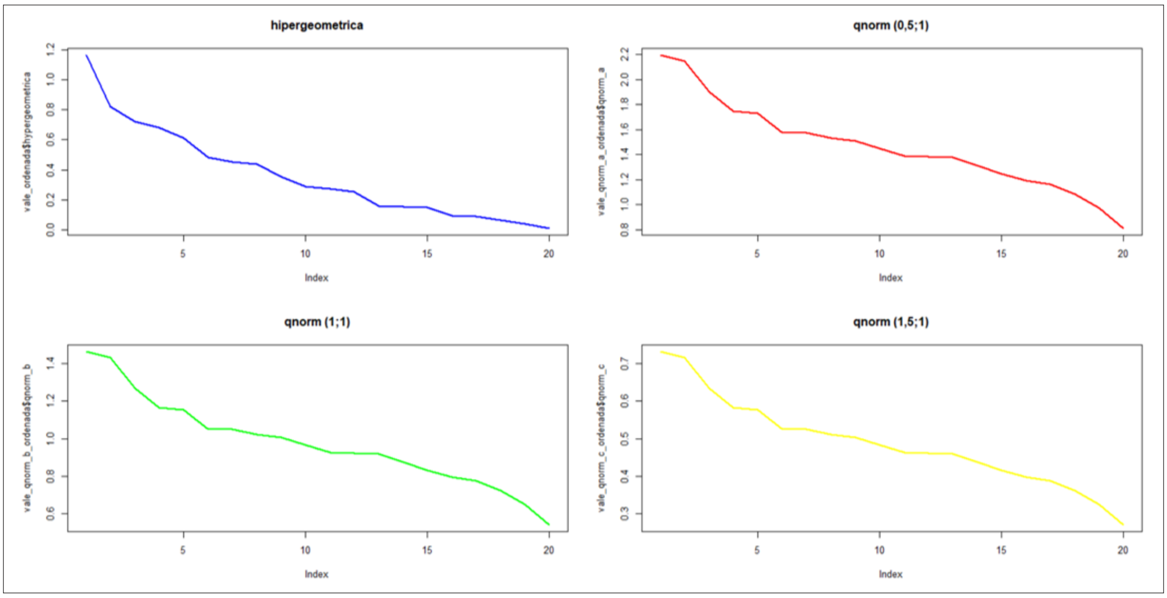
Source: Prepared by the Author

Figure 9: Comparison Between the Hypergeometric Distribution of MSFT and q-Normal Distributions with Different Parameters



Source: Prepared by the Author

Figure 10: Comparison Between the Hypergeometric Distribution of PETR and q-Normal Distributions with Different Parameters



Source: Prepared by the Author

Figure 11: Comparison Between the Hypergeometric Distribution of VALE and q-Normal Distributions with Different Parameters

The blue lines always represent the Probability Density Functions applied to the absolute values of the earnings surprises for all the companies analyzed. All of them, without exception, exhibit behavior similar to that observed in Figure 2, which may indicate that the earnings surprise follows a q-normal probability distribution. The red, green, and yellow lines represent the plotted probabilities of earnings surprises based on a q-normal function, and they show similar behavior even with variations in the parameters q e λ.

Conclusion

This study evaluated the possible similarity between the Probability Density Function of earnings surprise estimated by the hypergeometric function and the Probability Density Function of a q-normal function. The results suggest that there may be a similarity between the distributions. These findings provide evidence of the complexity of the financial market system, which-unlike non-complex systems-requires analysis focused on the possible interactions among the parameters that compose such a system. Future research may focus on developing statistical tests aimed at identifying the significance of the hypothesis that the observed data come from a q-normal distribution [21-24].

References

1. Bogachev MI, kayumov AR, Armin Bunde (2014) Universal Internucleotide Statistics in Full Genomes: A Footprint of the DNA Structure and Packaging? PLOS One 9: 112530-112534.
2. Ludescher J, Tsallis C, Bunde A (2011) Universal behaviour of interoccurrence times between losses in financial markets: An analytical description. EPL 95: 6780-68002.
3. Calster VT, Baesens B, Lemahieu W (2017) ProfARIMA: A profit-driven order identification algorithm for ARIMA models in sales forecasting. Applied Soft Computing 60: 775-785.
4. McGuinness PB (2016) Voluntary profit forecast disclosures, IPO pricing revisions and after-market earnings drift.

International Review of Financial Analysis 46: 70-83.
5. Luo J, Zhu G, Xiang H (2022) Artificial Intelligent based day-ahead stock market profit forecasting. Computers and Electrical Engineering 99: 107830-107837.
6. Fama EF, French KR (1992) The cross-section of expected stock returns. J. of Finance 47: 427-465.
7. Wu T, Gao X, An F, Xu X, Kurths J (2024) Forecasting the dynamics of correlations in complex systems. Chaos, Solitons and Fractals 178: 114330-114332.
8. Huang A, Tian L, Li Y, Xiong B, Yu J, et al. (2024) Regional complex system simulation optimization through linking governance and environment performance: A case study of water environmental carrying capacity based on the SDES model, Environmental Impact Assessment Review 104: 170350-170356.
9. Ribeiro HV, Mendes RS, Lenzi EK, Belancon MP, Malacarne LC (2011) On the dynamics of bubbles in boiling water. Chaos, Solitons & Fractals 44: 178-183.
10. Zhang H, Liu X, Wang P, Wang Q, Lu L, et al. (2024) Hydrogen-rich carbon recycling complex system establishment and comprehensive evaluation. Applied Energy 355: 122340-122347.
11. López RG, Fernández de Marcos A (2018) Evidence for criticality in financial data. arxiv https://arxiv.org/abs/1702.06191.
12. Reyes T, Batista JA, Chacon A, Martinez D, Kausel EE (2023) Attention-driven reaction to extreme earnings surprises Quarterly Review of Economics and Finance 92: 230-248.
13. Easton PD (2004) Pe Ratios, Peg Ratios, And Estimating the Implied Expected Rate of Return on Equity Capital. SSRN 79: 73-95.
14. Mandelbrot B, Hudson R (2004) The (Mis)Behavior of Markets: A Fractal View of Risk, Ruin, and Reward. Basic Books Cambridge https://www.scirp.org/reference/referencespapers?referenceid=1917498.
15. Zhang Q, Zhao B, He W, Zhu H, Zhou G (2024) A behavior prediction method for complex system based on belief rule base with structural adaptive Applied Soft Computing Journal 151: 111110-111118.
16. Santos dos AG (2013) Fractal Structure in Time Series: An Investigation into the Random Walk Hypothesis in the Brazilian Agricultural Commodities Spot Market. Master's dissertation https://repositorio.fgv.br/server/api/core/bitstreams/2287ecf0-6116-4c67-b75c-7e414c937306/content.
17. Pesaran MH, Timmermann A (1995) Predictability of Stock Returns: Robustness and Economic Significance. The Journal of Finance 50: 1201-1228.
18. Fama Eugene F, French Kenneth R (1997) Industry costs of equity. Journal of Financial 43: 153-193.
19. Mueller DC (1972) A life cycle theory of the firm. Journal of Industrial Economics 20: 199-219.
20. Dickinson V (2011) Cash flow patterns as a proxy for firm life cycle. The Accounting Review 86: 1969-1994.
21. Cambridge: v. 47, n. 2, p. 427-465, jun. 1992
22. Elbert E. N. Macau. Complex Systems. I don't know where, if or when he published this.
23. Economics. Amsterdam: v. 43, n. 2, p. 153-193, feb. 1997
24. Tsallis C (1996) Generalized entropy-based criterion for consistent testing. Physical Review E 58: 1440-1442.