Survey Castaing Distribution on Iranian Stock Market: An Econophysic Approach

Hashem Zare[*], Zeinab Rezaei Sakha[**], Mohammad Zare[***],

Abstract: *The purpose of the present study is to analyze the behaviors of the Iranian stock market using a non-Gaussian distribution function developed by a physicist called Castaing. Data were gathered on a daily basis from December 2008 to August 2016. In order to estimate the Castaing function, the Bayesian method and the Marcov chain Mont Carlo simulations were adopted. The results indicate a non-Gaussian distribution function with fat tails. Based on the findings, the probability of unexpected crises in the Iranian stock market could not be ignored.*

Keywords: *Econophysic, Iranian Stock Market, Castaing Distribution, Fat Tail*

1. Introduction

During the past few years, in order to investigate complicated social phenomena, various complex systems have been used as theoretical frameworks. It is deeply rooted in advances in physics, math, statistics, economy, finance, and other scientific fields. Despite the diversity and complexity of these systems, scholars believe that a universal and comprehensive set of rules governs all natural and human complex systems. For instance, movement in mechanics and molecule and cell mechanisms in biology are considered as universal laws. Nowadays, in order to investigate the complex system of stock market, it is vital to use modern models. Having a dynamic nature due to systematic and nonsystematic risks, numerous factors in the domestic and foreign scenes, global economic conditions, and a host of other factors interacting with each other simultaneously in a multidimensional context and in a random manner, stock markets are among the most complicated systems in economy.

---

[*] Department of Economic, Faculty of Economic and Management, Shiraz Branch, Islamic Azad University, Corresponding Author, hashem.zare@gmail.com

[**] Young Researchers and Elite Club, Shiraz Branch, Islamic Azad University, Shiraz, Iran University, Zrezaei_sakha@yahoo.com

[***] Young Researchers and Elite Club, Shiraz Branch, Islamic Azad University, Shiraz, Iran University, zaremo@yahoo.com

In order to analyze a complicated system, it is necessary to understand how each element works and is related to others. This makes analyzing a complicated system seem impossible. Then, if we are to have a systematic view toward stock market, we have to consider such qualities as the constituting elements of the market, market interactions, organization of activities and goals of such a market. The purpose of the study is to investigate how statistical physics is used in economic contexts to analyze the behavior of stock markets.

2. Literature review

A large body of research has been conducted on stock market, each investigating a particular aspect of the market. Among those studies, investigating the stock market in terms of random models is a turning point in all analyses conducted so far. After the discovery of the Brownian motion by Robert Brown and then proving this motion by Norbert Wiener, the first steps were taken to use this finding in other fields of science. In this regard, some scholars like Shreve & Karatzas (1991), [1], Duffie (1992), [2] and Karatzas & Kou (1996), [3] conducted systematic and complicated studies on the performance of random portfolios. The introduction of the random portfolio theory by Fernholz (1998), [4] was a fundamental step taken in this regard. Then the random portfolio theory was extended in the works of Fernholz, Karatzas and Kardaras (2005), [5], Banner, Fernholz and Karatzas (2005), [6] Fernholz and Karatzas (2006), [7] and Karatzas and Kardaras (2007), [8]. In these studies the effect of the Brownian motion on portfolio's behavior is shown as follows:

$$\frac{dZ_\pi(t)}{Z_\pi(t)} = \sum_{i=1}^{n} \pi_i(t) \frac{dX_i(t)}{X_i(t)}.$$

Where, $\frac{dZ_\pi(t)}{Z_\pi(t)}$ is the percentage of changes in the value of investment in a portfolio, and $\frac{dX_i(t)}{X_i(t)}$ is the percentage of changes in the share prices. The price of each share has a function as follows:

$$X_i(t) = X_0^i \cdot \exp\left(\int_0^t \gamma_i(s)ds + \int_0^t \sum_{v=1}^{n} \xi_{iv}(s) dW_v(s) \right) \qquad t \in [0 \quad \infty)$$

Where X_i (t) is the share price at the time of t and W (s) indicates the Brownian motion. A random variable $W(t)$ constant in the time t

$(W = \{W(t), \circ \le t < \infty\})$ is called the Brownian motion or the Norbert Wiener process, if and only if

1. $W(\circ) = \circ$
2. $W(t)$ has independent changes, so that for each change $W(t) - W(s)$, $s < t$ is independent from $\{W_r\}_{r<s}$.
3. For $\circ < s \le t, W(t) - W(s)$ has the normal distribution $N(\circ, t-s)$
4. W is constant in T with the probability of 1.

It should be noted that the interest of physicists in economics is not accidental. For instance, Bernoulli, who had a significant role in the development of the theory of statistics and probabilities, wrote numerous articles on measuring and quantifying risk. Econophysics scholars have also found that the power law can account for income distribution in most countries. This was first discovered by Pareto (1987), [9]. It is remarkable that the graph of the distribution function of the power law has longer sequences than normal ones, which is due to the large differences among various groups. For instance, in income distribution, there are large differences among incomes of upper classes and medium to lower classes. This will cause the distribution function to have a longer kurtosis compared to normal ones.

Since the market phenomena are the results of interactions among various factors, it might be possible to find similarities between statistical mechanics, which studies interactions among particles, and market economy. Therefore, physicists have considered the possibility of explaining market interactions in terms of methods commonly adopted in physics. From this perspective market is considered as a complicated system which the researcher tries to explain by finding experimental laws. Capital market is among those markets in which physicists are interested. Therefore, this study tries to account for distributive features of price index using the Castaing, Gagne and Hopfinger distribution function which can be estimated without the Gaussian presupposition. Furthermore, the parameters resulting from the estimation of the function can result in features regarding the market.

In this regard, various studies have been conducted including: Lihn (2010), [10], Stanley et al (2006), [11], Pisarenko and Sornette (2006), [12], Borland (2005), [13], Farmer et al (2005), [14], Durlauf (2005), [15], Newman (2005), [16], Bouchaud et al (2004), [17], Stanley (2003), [18], Plerou et al (2002), [19], Levy and Solomon (2000), [20], Mantegna and Stanley (2000), [21], Bak et al (1997), [22]. In the present study, we try to identify the distribution function of the price index in the Iranian stock market, using the Castaing, Gagne and Hopfinger distribution function, [23], which was selected for some reasons described later.

3. Model

In this section we introduce the Castaing distribution function and describe the tools used in estimating this function.

3.1. Non-Gaussian distribution function

In order to introduce the Castaing distribution function, we need to introduce the topic as well as estimating methods. In a theory developed by Kolmogorov, [24], it is stated that fluctuations in energy movements start like cascades and in large scales, and then continue with smaller fluctuations. These fluctuations are characterized by the probability distribution function $P(u = \delta v_r)$ for longitudinal changes on the r interval. These changes are defined this way:

$$U = \hat{e}_r . [\vec{v}(x+r) - \vec{v}(x)]$$

Where $\hat{e}_r$ can be derived separately along the path r. the changes in this set with r intervals can be shown as:

$$u = \delta v_r(x) = [\vec{v}(x+r) - \vec{v}(x)]$$

The purpose of detrending is to extract the cyclical component from the growth component. Detrending can be achieved using different methods including the Hodrick-Prescott method which is used in the present study. In this method, a time series like U is broken into two cyclical (c_t) and growth (g_t) components as $U = c_t + g_t$, and by minimizing the following equation, it detrends the time series:

$$\min_{\{g_t\}_t^T} \left\{ \sum_{t=1}^{T} c_t^2 + \lambda \sum_{t=1}^{T} [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})]^2 \right\}$$

Kolmogogove and Obukhov (1962), [25, 26] presumed that there is a normal logarithm distribution for the standard deviation variable of σ_r and all their experimental estimates approved this presumption. Furthermore, Landau (1944), [27], proposed that the fluctuations of σ along the path r for the vector x plays a crucial role in fluctuations. These fluctuations in the interval r could be represented as follows:

$$\sigma_r(\vec{x}, t) = \int_{\vec{x}}^{\vec{x}+\vec{r}} \sigma(\vec{x}', t) d\vec{x}'$$

In order to find the non-Gaussian distribution $P(\delta v_r)$, Castaing et al presumed that subsystems have Gaussian probability distribution functions δv_r. Of course, this presumption was proved by experimental functions. The Castaing method is based on the Bayesian theory. The Bayesian logic is a deduction method based on probabilities. The underlying idea is the fact that for each quantity there is a probability distribution for which it is possible to make optimal decisions by observing a new datum and by rationalizing about its probability distribution.

Since the beginning of 1990s, the Bayesian statistical methods and the Markov Chain Monte Carlo (MCMC) were increasingly used in various scientific researches while the Bayesian approach was more of a philosophical and conceptual one up to then and using that approach for functional purposes was not impossible but inapplicable. This trend changed by the introduction of the MCMC which allowed researchers to use simulation. In this section, we try to briefly introduce the Bayesian method; however, thorough and precise deriving of all relations requires a series of long algebraic operations which could not be summarized because it will be more complicated. Therefore, the readers are advised to refer to Gill (2014), [28], Gelman et al (2004), [29], and Jackman (2000), [30], for more detailed information.

The purpose of statistical inference is to learn about parameters which characterize the data production process based on observed data. In common and traditional approaches to statistical inference, it is supposed that the parameters are constant and have indefinite values. The probability function is represented as $f(y|\theta, x)$ or $(y|\theta)$, where y is a random time series variable of the random time series $y = \{y_1, y_2, ..., y_n\}$. θ is the vector of indefinite parameters $\theta = \{\theta_1, \theta_2, ..., \theta_k\}$ and x is the vector of independent parameters $x = \{x_1, x_2, ..., x_n\}$.

In order to estimate parameters, the maximum likelihood estimation method is used. This method uses the joint density function as a function of constant and indefinite parameters and is represented as follows:

$$L(\beta, \sigma^2 | y) \propto f(y | \beta, \sigma^2) = \prod_{i=1}^{n} \phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)$$

In order to solve the likelihood function $L(.)$ *with Respect* to the parameters, the maximum likelihood function must be maximized. In fact, based on the given time series, parameters are estimated. In order to inference about the parameters, given the estimation of parameters from a unique sample, it would be possible to estimate standard error, test the hypothesis, make confidence intervals and so on, based on the form of the sampling distribution.

While conducting the Bayesian inference, the fundamental suppositions all greatly change so that the vector of indefinite parameters θ is considered as random variables, while the data of the time series y are considered as constant and definite variables. Unobservable parameters are considered probabilistically, while the observed data are considered deterministically. In fact, the distribution of parameter θ is given after observing the series y. This posterior distribution can be written under the Bayesian theory as $f(\theta \mid y)$:

$$f(\theta \mid y) = \frac{f(y \mid \theta) f(\theta)}{f(y)}$$

Posterior distribution is a conditional distribution of parameters which is produced after observing data, as opposed to prior distribution which is produced after observing data. Therefore, posterior distribution is a probabilistic hypothetical statement about the probable value of parameters after observing data. The Bayesian theory directly follows the primary hypotheses of the probability theory and is used to relate the conditional distributions between two variables. $f(y \mid \theta)$ is the likelihood function produced by the probability model. The likelihood function has a crucial role in the Bayesian approach for analyzing data. $\square(\theta)$ is called the prior distribution. It includes all information about the value of parameters after observing data. $\square(y)$ is the prior predictive distribution which could be represented as:

$$f(y) = \int_{\theta} f(y \mid \theta) f(\theta) d\theta$$

It should be noted that some studies use the following equation in order to address some issues:

$$f(y) = \int_{\theta} f(y \mid \theta) f(\theta) d\theta$$

Basically, the posterior distribution turns the likelihood function into a probabilistic distribution with indefinite parameters, which means that the likelihood function can turn into any probabilistic distribution for deriving favorable parameters. It is made possible by the presumption of prior information existing in analyses. In order to conduct the Bayesian inference, we need a linear regression to consider prior information about both parameters. The analyzer can decide to consider such prior information for any time period.

Based on the above- mentioned discussions, in order to gain a non-Gaussian probability function $P(u)$, Castaing use the Bayesian technique and merging Gaussian probability distribution functions.

3.2. The Castaing, Gagne and Hopfinger Distribution Function

Castaing et al started off with experimental observations and the results indicated that the distributions of fluctuation's differences for a given σ_r is Gaussian and σ_r can be Describable by a log normal distribution. In order to represent the change of $P(u)$ from non-Gaussian in small scales into Gaussian in large scales, Castaing et al produced the following equation:

$$P(u_r) = \int P(\sigma_r) P(u_r \mid \sigma_r) d\sigma_r$$

The probability distribution function $P(u_r \mid \sigma_r)$ is Gaussian based on hypotheses and is experimentally proved, which can be represented as:

$$P_\varepsilon(u_r \mid \sigma_r) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma^2} \right)$$

Kolmogrove, Obukov, and Castaing presumed σ_r with log normal distribution as:

$$Q_\lambda(\sigma) d\sigma = \frac{1}{\lambda\sqrt{2\pi}} \exp\left(-\frac{\ln^2\left(\sigma / \sigma_0 \right)}{2\lambda^2} \right) d \ln \sigma$$

Where λ is the standard deviation from $\ln \sigma_r$. Finally, the integral of the following equation as the final form of the Castaing distribution function is introduced:

$$\pi_\lambda(u) = \frac{1}{2\pi\lambda} \left[\exp\left(-\frac{u^2}{2\sigma^2} \right) \exp\left(-\frac{\ln^2\left(\frac{\sigma}{\sigma_0} \right)}{2\lambda^2} \right) \right] \frac{d\sigma}{\sigma^2}.$$

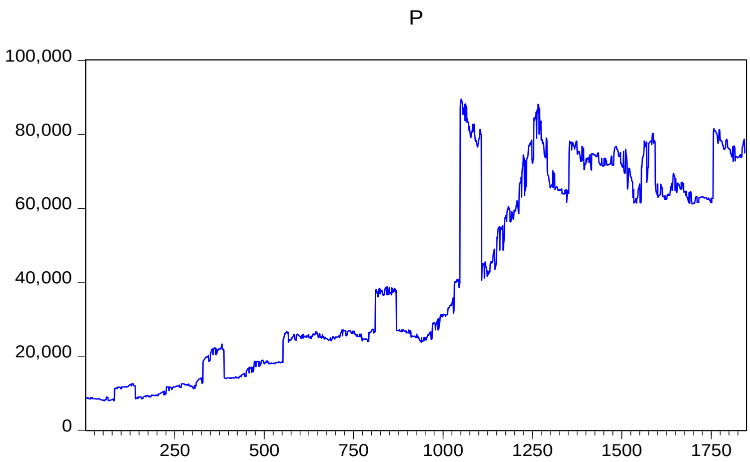
It should be noted that in addition to drawing the Castaing distribution function, another primary purpose of the present study is to measure λ . As we will show in next chapter, the value of this parameter could have an effective role in describing market conditions and qualities.

4. Results & Discussion

In this section, we provide requirements for estimating the Castaing distribution function using Hodrick Prescott, Bayessian, and MCMC methods.

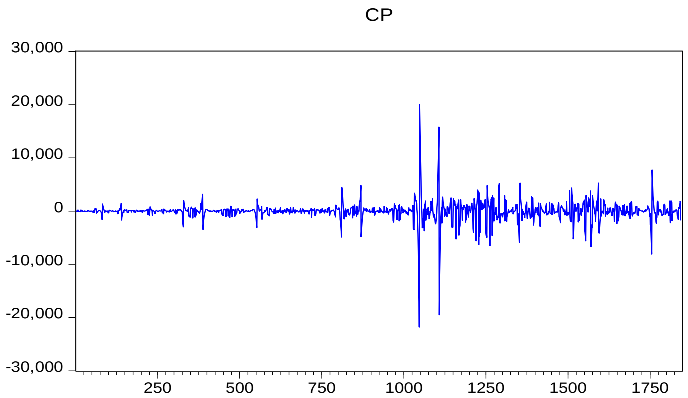
4.1. Estimating the Castaing distribution function

In this section, the data of price index in Iran is detrended. The information and statistics in this study were used from December 2008 to August 2016. The graph for the trend of changes in the price index (TEPIX) over time is represented as follows.



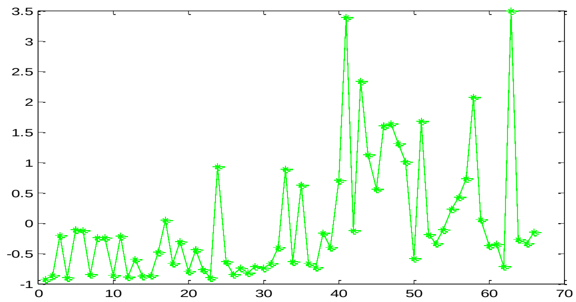
Graph 1: changes in price index.

Based on the Castaing technic, in order to estimate the distribution function, detrended data of the price index must be used. Therefore, the data of the price index are detrended using the Hodrick-Prescott method, shown in graph 2.



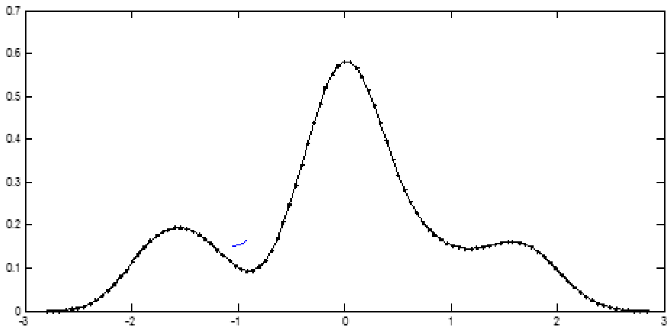
Graph 2: detrended data for the price index.

The standard deviation of the subgroups of the data is calculated in the form of a Gaussian distribution function, using the Bayesian method. In order to estimate parameters using the Bayesian method, the Markov Chain Mont Carlo (MCMC) method is used. In this method sampling must be done successively while taking into account the conditional density function for calculating parameter (mean and variance). In order to calculate those parameters, the powerful software Win-Bugs, developed by Spiegel et al, [31], was used. This software uses the Bayesian method under the simulation of MCMC to estimate parameters. In the experimental estimation using the detrended data of price index, the standard deviations of all subgroups of data are calculated using the Monte Carlo method. Graph 3 represents calculated standard deviations of the Bayesian method.



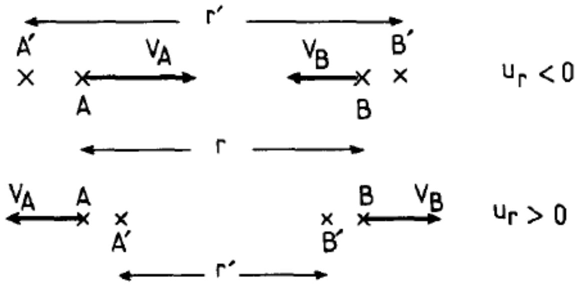
Graph 3: standard deviation calculated using the MCMC.

After calculating standard deviations of different samples using the Bayesian method and their respective distribution function, the numerical value of λ was estimated 0.66. Finally, using the MATLAB software, the Castaing distribution function was estimated, shown in Graph 4.



Graph 4: estimated Castaing function.

As shown in graph 4, the results of the Castaing distribution function are estimated in a cascade manner. A deep investigation in statistical relations reveals that the fundamental factor responsible for skewness is that in the Castaing distribution function, σ_r and u_r are correlated. Therefore, the skewness form of the distribution function for the speed of the variable is not dissimilar to the shape of energy cascade or whirlpool elongation. This feature which in physics means that energy is transferred from large scales to small scales depends on the variance correlation value of speed, respect to the scale of *r*. In addition, based on the Characteristics of this distribution function, as the speed variance increases due to changes in *r*, it is expected that the left tail is wider than the right one.

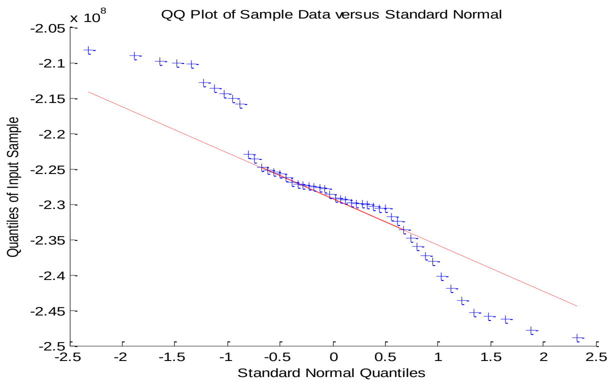


This means that as the speed of the price movement increases toward positive direction, the right-hand sequence must be longer. However, if the speed of price adjustment decreases, it is expected that the distribution form has a longer left-hand sequence. This means that if adjusted speeds either in price increase or decrease exists in the market, both ends of the function sequence must lengthen to similar extents. In other words, if emotional movements leading in intense fluctuations are quickly adjusted, it is expected that the form of the distribution function approaches normal distribution.

4.2. Characterizing the Castaing distribution function

The findings of experimental estimation indicate that the price index does not have Gaussian distribution. Since the kurtosis coefficient of normal distribution is 3, the kurtosis coefficient of the Castaing function for various samples is estimated about 6.9, therefore, the behavior of the price index in Iran is similar to those of global markets. This means that compared to normal distribution, the Castaing function has a wider tail and higher peak. In other words, compared to normal distribution, this function reaches the extreme points with a higher probability value.

The difference in the kurtosis of the normal distribution function and that of the distribution function with wide tail can be represented by the Quantile-Quantile Plot. Graph 5 indicates riskier conditions in the Iranian stock market than normal situations.



Graph 5: comparison of normal distribution with the Castaing distribution (QQ)

The question which arises is how we can provide conditions for stock markets to grow more stable. We could answer that although we are far yet to reach this goal, the results show that it is possible to reach stability by providing effective management and astute controlling of major conditions. In other words, by taking systematic and non-systematic risks into consideration and developing a mechanism which prevents market relations from getting hurt in critical situations, it could be achieved. Crucial factors leading to systematic risk include political, social and economic changes and business periods. In this regard, high market executives must provide conditions for protecting market activists from major fluctuations. Non-systematic risk is also called reducible risk. This type of risk can also be controlled and reduced by providing astute and strategic market management, improving cost structures and flexibility of institutions, reducing asymmetric information using technological instruments and so on. Therefore, although it is vital that fundamental steps be taken toward reaching stability in distribution over the country, what supports its consequences is the achievability of this goal.

5. Conclusions

In this study, we tried to characterize the Iranian stock market index using the distribution function introduced by Castaing. First, the Castaing distribution function was estimated under accidental behavior and using a unique method. The results of estimation in the studied period indicate that the Castaing distribution function in various samples was cascade with right and left fat tails. Given the kurtosis index estimated for this function, a wide sequence is proved for the castaing distribution function. In fact, a cascade wide sequence in the Castaing distribution function indicates frequent occurrences of unexpected events in the Iranian stock market. Therefore, in this market there is always a possibility of systematic or nonsystematic shocks. However, what needs to be addressed realistically is the large distance between the present conditions from ideal ones. What is obvious is that by careful planning and employing experts in different fields and utilizing interdisciplinary knowledge, it is possible to reach ideal conditions.

REFERENCES

[1] Shreve, S. E., & Karatzas, I. (1991), Brownian motion and stochastic calculus.*Graduate Texts in Mathematics*, *113*.

[2] Durlauf, S. N. (2005), Complexity and empirical economics. *The Economic Journal*, *115*(504), F225-F243.

[3] Karatzas, I., & Kou, S. G. (1996), On the pricing of contingent claims under constraints. *The Annals of Applied Probability*, 321-369.

[4] Fernholz, E. R. (1998), Portfolio generating functions, *Available at SSRN 139549*.

[5] Fernholz, R., Karatzas, I., & Kardaras, C. (2005), Diversity and relative arbitrage in equity markets. *Finance and Stochastics*, *9*(1), 1-27.

[6] Banner, A. D., Fernholz, R., & Karatzas, I. (2005), Atlas models of equity markets. *The Annals of Applied Probability*, *15*(4), 2296-2330.

[7] Fernholz, R., & Karatzas, I. (2006), The implied liquidity premium for equities.*Annals of Finance*, *2*(1), 87-99.

[8] Karatzas, I., & Kardaras, C. (2007), The numéraire portfolio in semimartingale financial models. *Finance and Stochastics*, *11*(4), 447-493.

[9] Pareto, V. (1987), *Escritos sociológicos* (No. 301 P35y).

[10] Lihn, S. H. (2010), Comment on PK Clark's Distribution of Lognormal-Normal Increments and the Lognormal Cascade Distribution. *Available at SSRN 1537686*.

[11] Stanley, H. E., Gabaix, X., Gopikrishnan, P., & Plerou, V. (2006), Economic fluctuations and statistical physics: the puzzle of large fluctuations. *Nonlinear Dynamics*, *44*(1-4), 329-340.

[12] Pisarenko, V., & Sornette, D. (2006), New statistic for financial return distributions: Power-law or exponential?. *Physica A: Statistical Mechanics and its Applications*, *366*, 387-400.
[13] Borland, L. (2005), Long-range memory and nonextensivity in financial markets.*europhysics news*, *36*(6), 228-231.
[14] Farmer, J. D., Smith, E., & Shubik, M. (2005), Economics: the next physical science?. *arXiv preprint physics/0506086*.
[15] Durlauf, S. N. (2005), Complexity and empirical economics. *The Economic Journal*, *115*(504), F225-F243.
[16] Newman, M. E. (2005), Power laws, Pareto distributions and Zipf's law.*Contemporary physics*, *46*(5), 323-351.
[17] Bouchaud, J. P., Gefen, Y., Potters, M., & Wyart, M. (2004), Fluctuations and response in financial markets: the subtle nature of 'random'price changes.*Quantitative Finance*, *4*(2), 176-190.
[18] Stanley, H. E. (2003), Statistical physics and economic fluctuations: do outliers exist?. *Physica A: Statistical Mechanics and its Applications*, *318*(1), 279-292.
[19] Plerou, V., Gopikrishnan, P., Gabaix, X., & Stanley, H. E. (2002), Quantifying stock-price response to demand fluctuations. *Physical Review E*, *66*(2), 027104.
[20] Levy, H., Levy, M., & Solomon, S. (2000), *Microscopic simulation of financial markets: from investor behavior to market phenomena*. Academic Press.
[21] Mantegna, R. N., & Stanley, H. E. (2000), An introduction to econophysics: correlation and complexity in finance. *Cambridge, UK: Cambridge University*.
[22] Bak, P., Paczuski, M., & Shubik, M. (1997), Price variations in a stock market with many agents. *Physica A: Statistical Mechanics and its Applications*,*246*(3), 430-453.
[23] Castaing, B., Gagne, Y., & Hopfinger, E. J. (1990), Velocity probability density functions of high Reynolds number turbulence. *Physica D: Nonlinear Phenomena*, *46*(2), 177-200.
[24] Kolmogorov, A. N. (1941, January), The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. In *Dokl. Akad. Nauk SSSR* (Vol. 30, No. 4, pp. 301-305).
[25] Kolmogorov, A. N. (1962), A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *Journal of Fluid Mechanics*, *13*(01), 82-85.
[26] Obukhov, A. M. (1962), Some specific features of atmospheric turbulence.*Journal of Geophysical Research*, *67*(8), 3011-3014.
[27] Landau, L. (1944), On the energy loss of fast particles by ionization. *Originally published in J. Phys*, *8*, 201.
[28] Gill, J. (2014), *Bayesian methods: A social and behavioral sciences approach* (Vol. 20). CRC press.
[29] Gelman, A., & Meng, X. L. (Eds.). (2004), *Applied Bayesian modeling and causal inference from incomplete-data perspectives*. John Wiley & Sons.
[30] Jackman, S. (2000), Estimation and inference via Bayesian simulation: An introduction to Markov chain Monte Carlo. *American Journal of Political Science*, 375-404.
[31] Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., & Cowell, R. G. (1993), Bayesian analysis in expert systems. *Statistical science*, 219-247.