UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO COPPEAD DE ADMINISTRAÇÃO

GUILHERME DA SILVA GONÇALVES

A HIGHER ORDER PORTFOLIO OPTIMIZATION MODEL INCORPORATING INFORMATION ENTROPY

Orientador:
Peter Fernandes Wanke, Ph.D.

Rio de Janeiro
2021

GUILHERME DA SILVA GONÇALVES

A HIGHER ORDER PORTFOLIO OPTIMIZATION MODEL INCORPORATING INFORMATION ENTROPY

Master's Dissertation presented to the Instituto COPPEAD de Administração, Universidade Federal do Rio de Janeiro, as part of the mandatory requirements in order to obtain the degree of Master in Business Administration (M.Sc.).

SUPERVISOR: Peter Fernandes Wanke, Ph.D.

Rio de Janeiro

2021

CIP - Catalogação na Publicação

da Silva Gonçalves, Guilherme

G635h A HIGHER ORDER PORTFOLIO OPTIMIZATION MODEL INCORPORATING INFORMATION ENTROPY / Guilherme da Silva Gonçalves. -- Rio de Janeiro, 2021.

47 f.

Orientador: Peter Fernandes Wanke.
Coorientador: Otavio Henrique dos Santos Figueiredo.

Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Instituto COPPEAD de Administração, Programa de Pós-Graduação em Administração, 2021.

1. Teoria Moderna do Portfolio. 2. Ibovespa. 3. Markowitz. 4. modelo MVSKE. 5. Entropia da Informação. I. Fernandes Wanke, Peter , orient. II. Henrique dos Santos Figueiredo, Otavio , coorient. III. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

GUILHERME DA SILVA GONÇALVES

A HIGHER ORDER PORTFOLIO OPTIMIZATION MODEL INCORPORATING INFORMATION ENTROPY

Master's Dissertation presented to the Instituto COPPEAD de Administração, Universidade Federal do Rio de Janeiro, as part of the mandatory requirements in order to obtain the degree of Master in Business Administration (M.Sc.).

Approved on February 25th, 2021

Peter F. Wanke

Peter Fernandes Wanke, D.Sc - Advisor
COPPEAD

P/ Peter F. Wanke

Otavio Henrique dos Santos Figueiredo, D.Sc
COPPEAD

P/ Peter F. Wanke

Jéfferson Augusto Colombo, D.Sc
FGV/EAESP

Rio de Janeiro

2021

ABSTRACT

The modern portfolio theory, introduced by Harry Markowitz in 1952, is a mean-variance (MV) model to evaluate a portfolio of assets, where the objective function is to maximize the expected return (mean) for a given risk (variance), or to minimize that risk for a given return. Given the Brazilian context of high volatility where the study case takes place, this paper aims to provide a framework for dealing with skewed and leptokurtic data, conditions that seems more present in emerging markets, deviating from normality assumptions of the classic Markowitz model. The objective of this paper, then, is to expand the model for higher order moments (evolving into a MVSK model with skewness and kurtosis analysis) and compare this apparently more robust model to the classic quadratic objective function of Markowitz. Along with the MVSK analysis, we add an information entropy variable in the model hoping to take into account asset's informational efficiency and diversity, trying to encompass the high uncertainty intrinsic to market's returns and to increment the models' validity. By this means, we analyze the practical effectiveness and the complexity of creating such multi-objective portfolio model to see if we can indeed provide more information to the investor with the new framework. In particular, the results obtained by MVSKE from B3 case in the 2011-2019 period shows fragile performance out-of-sample, but in good parallel with the markowitz benchmark.

Key-words: Modern Portfolio Theory, Mean-Variance, Markowitz, MVSK model, MVSKE Portfolio, Information Entropy, Ibovespa.

LIST OF FIGURES

LIST OF TABLES

SUMMARY

1. INTRODUCTION

1.1. Problem Context

The construction of a portfolio model will always be a theme of greater importance in financial modeling. Given the fact that most models used to this day are still the very ones that opened this field of study in the 1950's (such as Markowitz model of portfolio selection and CAPM model of asset pricing), and as the strong-form efficient market hypothesis gets increasingly contested by behavioral economists, new approaches to the model seems necessary. With the development of quantitative finance, more models arise trying to better explain market's behavior. And in this thesis, we propose a statistical model based mostly around assets' prices and its central moments, with a method that encompasses not only the classic risk-return duo, but also higher moments around the mean, which best explain asymmetry and tail risk. And along with this, entropy is added to the variables, seeking to efficiently price the information efficiency of a portfolio of assets.

The data used for the training set of our model was from the last decade on *B3 S.A. Exchange* (Brazil Stock Exchange and Over-the-Counter Market; formerly BM&FBovespa), consisting of daily prices fluctuation from 2011 until 2018, extracted by Yahoo Finance. After the model's training set and backtesting, we seek to validate our predictive model testing it on the 2018-2019 market, and then comparing with the Markowitz model as the benchmark, leading to conflicting but interesting results, better seen in the results section.

1.2. Objectives and Study Limitations

The main goal of this work is to gather the vast literature on portfolio modeling of higher order moments, and from that create a slightly modified model, one including the characteristics more observed in the Brazilian context (CAVALVANTE and ASSAF, 2004), such as higher volatility and positive skewed distribution of returns. Since evidence suggests that stock exchanges in emerging markets are more volatile (ADCOCK and SHUTES, 2005) and less efficient (in EMH sense), showing a more asymmetric distribution of returns, investors inserted in such environments may need better tools to operate in this scenario.

Some of the study limitations are regarding utility theory. The standard concave utility function will be employed representing a risk-averse approach (ARROW, 1971), the same assumption used by Markowitz in his classic model and by many others after him. However, this model can have flaws on modeling investor's preferences, as recent studies in behavioral economics' prospect theory suggests (KAHNEMAN and TVERSKY, 1979). For example, the loss aversion characteristic of some investors cannot be defined by expected utility theory (not with a trivial function at least), since this would represent an asymmetrical and convoluted utility function, which is out of the scope of the model. Other examples of inconsistencies with expected utility theory and the risk aversion function are the reflection effect (different preferences whether the outcomes are losses or gains) and the certainty effect (to prefer a certain return at the expense of losing some expected value), both established by Kahneman and Tversky (1979). Consequently, their empirical findings contest the prior concepts of investor rationality, but our model won't directly adress such aspects, focusing on the statistical aspects of price fluctuation and to which extent its asymmetry, that are out of classic models, can affect explained performance.

Other limitations concerns computational and mathematical complexity about polynomial multi-objective functions (of 5^{th} degree), what makes exponentially difficult to simulate for a big enough quantity of assets in the portfolio, depending on the tools used in the model. This kind of complexity goes out of the scope of the paper to be a management framework for investors to assemble a diversified and low-risk portfolio of assets in emerging countries.

Another important limitation going in that same direction is due to the already large number of variables of the model. Which results in important aspects of assets that are left out, such as liquidity and transaction costs, as a trade-off situation. Since the main objective of study is to compare higher order moments against the classic model, the focus will be on skewness and kurtosis risk and information efficiency in Brazilian markets. Transaction costs as restraints in the portfolio model would be a good addition that unfortunately will be out of the scope presented, but stays as recommendations for future models.

Finally, this study's purpose is not to create a portfolio with sophisticated constraints or elaborate algorithms but simply to expand the knowledge on certain gaps about higher order moments and the relation between assets' risk and its informational efficiency.

1.3. Hypothesis and research question

The research's aspiration is to investigate if the proposed model will add incremental value to the investor comparing to the classic models that relies on less information about assets' characteristics. In creating this model, we will also be able to make assumptions about validity of the efficient-market hypothesis (EMH) contrasted with adaptive-market one (AMH) for the Brazilian paradigm.

2. LITERATURE REVIEW

The bibliography review starts with a historical perspective on portfolio modern theory, going from Markowitz classic model to the most recent studies in the field. Then, the literature on information entropy is presented, with arguments for measuring market efficiency. In the end, fuzzy set theory is discussed (even though not present in our model in the first moment) as a valuable technique for better representing the return values of assets and used by some of the seminal works in the field. All this pieces together will be vital parts in the creation of our model, to be discussed in the methodology chapter.

2.1. Historical Perspective

One of the foundational works in the field of financial modeling is the Markowitz's mean-variance model, the 1952 seminal work which opened that field of study in modern financial economics. The Harry Markowitz MV model is the basis for a variety of regression models as mentioned by Fang et al. (2008), including the most famous one, the Capital Asset Pricing Model (CAPM), elaborated by William Sharpe, from whom Markowitz shared the 1990's Nobel in Economics for their pioneering work in financial economics theory. In fact, CAPM is uniquely a model for asset pricing in a context where investors have mean-variance efficient portfolios (KRAUS and LITZENBERGER, 1976). Both models work side by side.

The HM model is a quadratic function model with the following assumptions: an investor must be risk averse (with the risk here represented by the variance of the expected returns). The investor is rational (therefore, portfolios that lies below the frontier will never be chosen, since they will always be dominated in risk-return by the ones on the frontier). And assets' return are normally distributed. This last one is a vital assumption, in the sense that *mean* and *variance* are sufficient to describe return behavior (with no skewness or kurtosis asymmetry involved, which is hardly the case as suggested by Yang and Hung (2010). Leland (1999) showed that when normality and mean-variance preferences are violated, the market portfolio is mean-variance inefficient and the CAPM mismeasures the investment performance.

On the other hand, what is not an assumption in neither case but represents ideas compatible with CAPM and Markowitz assumptions lies in the efficient market hypothesis proposed by Fama (1965). The EMH categorized three hypothetical stages of efficiency: weak-form, semi-strong form and strong. In the weak form, prices reflect all information in historical data of assets' return. In the semi-strong form, prices reflect all public information available, including past returns or recent forecasts of future returns, like financial statements, information about the economy or any other relevant public information. In the strong form, prices would reflect not just public info but also private information from insiders (we can note that in the strong form, inside information wouldn't have value since no way to search or process that information would yield abnormal returns, therefore, the acquisition cost of such information would be zero). Thus, it's noticed that the strong-form is not reasonable to address markets present state.

A better definition of EMH would be of prices reflecting all information in a way that the marginal benefit of using this information wouldn't exceed the marginal cost of acquiring it. However, according to Shleifer (2000), economic theory does not indicate efficient financial markets, on the contrary: significant and systematic deviations in efficiency are expected, and such deviations can go for a long period.

Going back to the mean-variance model, the main concept in Markowitz model is the Efficient Frontier, a segment in the risk-return spectrum where the optimal portfolio should be. Along this frontier, it is up to the investor to choose the portfolio that best suits their preferences for each combination of assets' portfolio that makes the boundary. The Efficient Frontier is Pareto-Efficient in the sense that you will not find a better combination of risk and return altogether. Therefore, if an investor seeks higher returns, he has to increase his exposure to volatility, increasing the variance of such returns. It establishes how investors can make decisions that maximize returns for a given variance or how they minimize that variance for a given return, depending on where the investor's preferences lies.

With risk in the MV portfolio defined as the variance of the sum of all the assets in the portfolio, and observing statistical properties for central moments, we have the following equations for the risk and return of a portfolio consisting of two assets:

$$E(R_p) = \bar{R}_p = WR_x + (1 - W)R_y \tag{2.1.1}$$

$$\sigma_p^2 = W_x^2\sigma_x^2 + W_y^2\sigma_y^2 + 2W_xW_yCov_{x,y} \tag{2.1.2}$$

$$\sigma_p^2 = W_x^2\sigma_x^2 + W_y^2\sigma_y^2 + 2W_xW_yCorr_{x,y}\sigma_x\sigma_y \tag{2.1.3}$$

Where:

$\bar{R}_p$ is the portfolio's expected return;

σ_p^2 is the portfolio's variance;

W is the weight of an asset ;

$Cov_{x,y}$ or $(\sigma_{x,y})$ is the covariance between assets x, y.

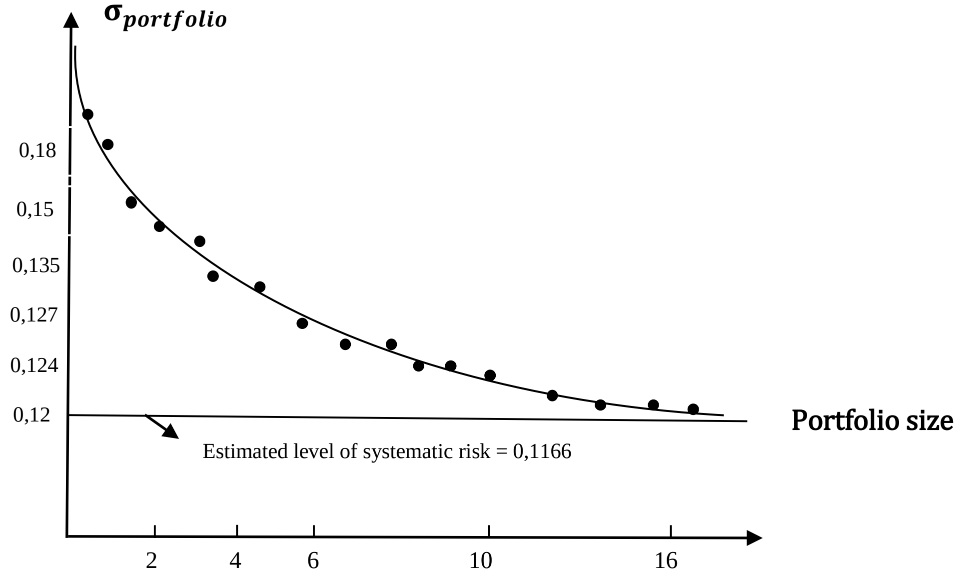
$Corr_{x,y}$ or $(\rho_{x,y})$ is the correlation between assets x, y.

Seeing the equations above, we notice how often the concepts of covariance and correlation appears. One of the key insights behind Markowitz model is the importance of diversification (diversifiable risk). If the covariance between assets are negative (meaning a negative relationship between the returns of such assets), we notice that the overall variance of the portfolio of two assets reduces. In fact, as long as $\rho_{x,y} < 1$, the standard deviation of a portfolio will be lower than the weighted average of the standard deviations of individual assets (and this concept can be expanded to a portfolio of any size).

It can be noted that the contribution of an individual asset's variance to the overall portfolio's variance tends to zero when n is big enough, with the covariance converging to the mean (contributing to the total risk of the portfolio). This implies in the diversification of individual risk reaching a limit, known as systemic risk, which is non-diversifiable. This notion can be verified from another angle: if the number of assets in a portfolio grows in the direction of the number of assets in the market, it is expected that the total variance of the portfolio will converge to the variance of the market itself, the market risk (or systematic risk).

Empirical tests by Evan and Archer (1968) shows that a number as lower as 16 assets already makes for great diversification in the portfolio. In the reproduced graph below we can see the real tendency of diversification shrinking idiosyncratic risk (denoted by the points on the graph, each one representing actual portfolios of different sizes with their individual standard deviations) converging to systematic risk as the portfolio size increases.

Figure 1 – Convergence of idiosyncratic risk to systematic risk



Source: adapted from EVANS and ARCHER, 1968

Generalizing the mean and variance equations for n assets:

$$\bar{R}_p = \sum_{j=1}^{n} R_j W_j \tag{2.1.4}$$

$$\sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} W_i W_j Corr_{i,j} \sigma_i \sigma_j \tag{2.1.5}$$

Formulating it, for example, as a general constraint minimization problem (to assist us in later steps with other operations research techniques) we have the following equation:

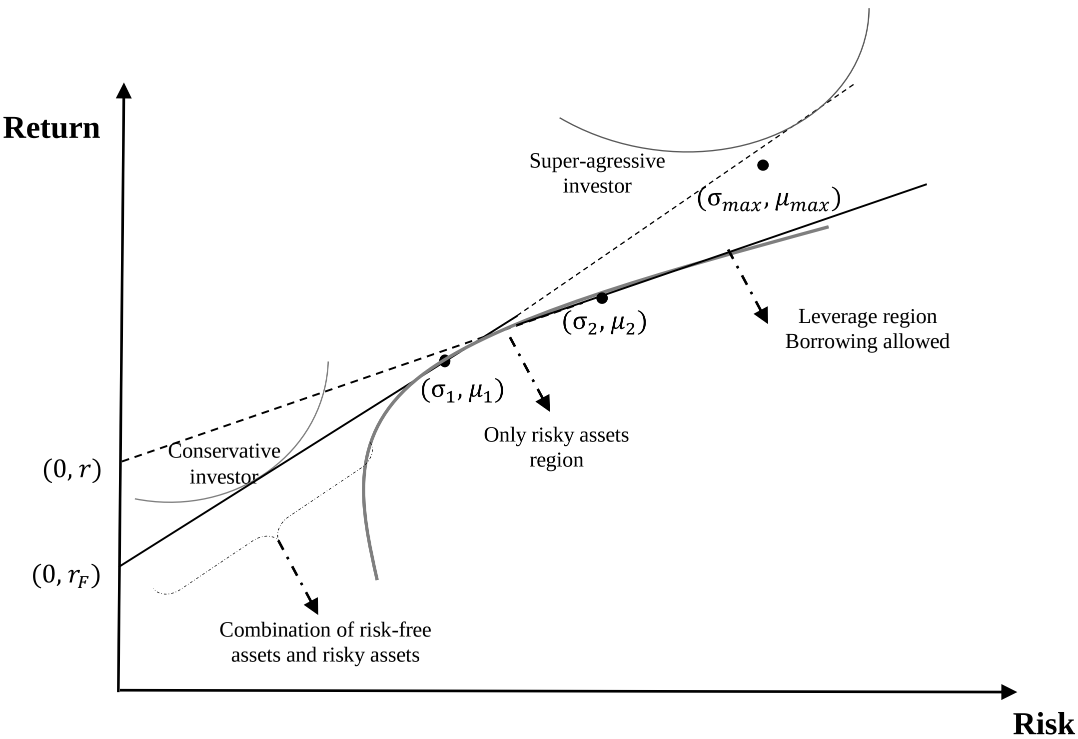
$$\text{Min} \quad \sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} W_i W_j Cov_{i,j} \tag{2.1.6}$$

$$\text{Subject to} \quad \bar{R}_p \geq \sum_{j=1}^{n} R_j W_j$$

$$\sum W_i = 1, \; W_i \geq 0$$

As we can see by the restrictions in the optimization function not allowing negative weights, this is a MV model without short-sales, where W_i would assume negative values representing the fact of an asset being sold short. This positive constraint helps to avoid highly leveraged portfolios with large short positions for example. A sensitivity analysis comparing both options to see if there is distortions in short and long positions is a reasonable approach. In the figure below, we can see all of these options and how they play out in the construction of the portfolio.

Figure 2 – Modified Capital Allocation Line with Efficient Frontier



Source: adapted from DOMINGUES, 2003

A slightly modified curve representing a junction of Capital Allocation Lines (CAL) can be observed in the mean-variance graph above by the junction of lines formed by the points $(0, r_F)$, (σ_1, μ_1), (σ_2, μ_2) and $(\sigma_{max}, \mu_{max})$. The possibility of borrowing expands the investor's opportunity set from the efficient frontier's red parabola to the right side of the Capital Allocation Line. Points between (σ_1, μ_1) and (σ_2, μ_2) represents a portfolio of risky assets only (and because of that, lies in the efficient frontier). The left side of CAL describes a indifference curve for conservative investors (high risk aversion), since we have a combination of risk-free assets (r_F) with risky assets shaping the portfolio (each indifference curve represents a utility function, measuring the investor's preference over a choice set). The right side of CAL above the efficient frontier parabola represents levered portfolios of super aggressive investors (low risk aversion), where short-sales occur. The dashed lines represents fixed Sharpe ratios tangent to (σ_1, μ_1) and (σ_2, μ_2), indicating relationships of reward to variability to be compared with the efficient frontier.

Apart from being largely employed even to this day (in pure or enhanced versions), that are also criticisms concerning the Markowitz model. Michaud (1989) criticism of Markowitz articulates that unconstrained MV optimization can yield results that are inferior to those of simple equal-weighting schemes (MICHAUD, 1989). However, one can argue that equal-weighting allocation can be done as some sort of rebalancing in the MV portfolio itself, with a model for portfolio selection of those stocks previously equally weighted (even if the model is replicating some index fund). In such circumstances, MV optimization is said to be superior to equally weighting arrangements in terms of integration of portfolio objectives with client constraints.

In this regard, Demiguel (2009) went beyond and showed that the estimation window needed for the sample-based mean-variance strategy to outperform the 1/N equally weighted benchmark is around 3000 months for a portfolio with 25 assets and about 6000 months for a portfolio with 50 assets. That goes against even the most skeptical supporters of MPT, that usually advocates on a convergence of non normal investment returns to normality within a period of roughly 30 years due to the Central Limit Theorem.

2.2. MVS Model

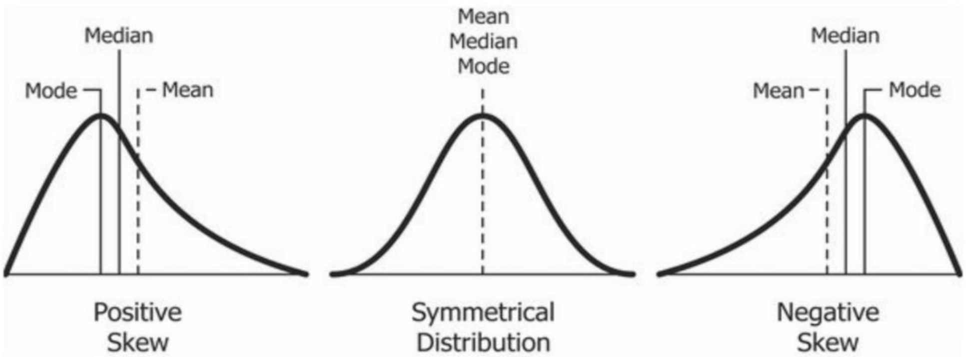
The MVS model is an expected evolution of the MV model that incorporates skewness into the portfolio construction, creating a cubic utility function for portfolio selection alongside with a multidimensional efficient set. Evidence suggests that prior inconsistencies in the MV model can be attributed (along with other factors) to the omission of systematic skewness (KRAUS and LITZENBERGER, 1976). Skewed data also undermines the results of some statistical tools such as the analysis of variance (ANOVA) for example, since this estimation model assumes normality and homoscedasticity (homogeneity of variance across the distribution for its random variables), being better employed for balanced data with independence of observations. In this case, nonparametric tests are recommended (such as the Kruskal-Wallis test), since these tests do not assume anything about the parameters of the underlying distribution.

The utility function with risk averse features (non-increasing absolute risk aversion) and decreasing marginal utility of wealth applied in MV models are still accurate in MVS, as stated by Kraus and Litzenberger (1976), since aversion to increasing standard deviations and preference for positive skewness are common characteristics of investors modeled by such utility functions. Even Markowitz (1959) showed that nonquadratic utility functions can be locally approximated with a quadratic function.

Skewness is the degree to which returns are asymmetric around the mean. If the portfolio has positive skewness, this implies in several small negative returns and scarce but big positive ones (what explains investors' preference for positive skewness). On the other hand, negative skewness means several small positive returns with rare but large losses in between (REGENSTEIN, 2018).

A common misconception about skewness is its direct relation between mean and median of a distribution, as stated by Von Hippel (2005). That is correct for the concept of nonparametric skew ($\frac{\mu-\vartheta}{\sigma}$), a naiver notion that simply calculates the mean minus median over the standard deviation, not requiring prior knowledge about the distribution's shape. Tabor (2010) declares this concept of nonparametric skew as a weak statistical tool for detecting shifts from normality (for example, a distribution with negative skew can have its mean greater than the median).

Figure 3 – Older notion of nonparametric skew for unimodal examples



Source: DOANE, 2011

The skewness (S_p) formula is derived below:

$$S_p = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} E(R_i R_j R_k) W_i W_j W_k}{(\sum_{i=1}^{n}\sum_{j=1}^{n} W_i W_j Cov_{i,j})^{3/2}} \quad (2.2.1)$$

The third order multi-objective function of the portfolio selection will be:

Min $\sigma_p^2(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} W_i\, W_j\, Cov_{i,j}$

$$\text{Max} \qquad \bar{R}\,(\mathrm{x}) = \sum_{j=1}^{n} R_j\, W_j$$

$$\text{Max} \qquad S_p(x) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} E(R_i R_j R_k) W_i W_j W_k}{(\sum_{i=1}^{n}\sum_{j=1}^{n} W_i W_j Cov_{i,j})^{3/2}}$$

$$\text{Subject to} \qquad \bar{R}_p \geq \sum_{j=1}^{n} R_j\, W_j$$

$$\sum W_i = 1, \;\; W_i \geq 0$$

Some criticism attested to the MVS model (and that can be extended for the MVSK) is about the stability of the correlation matrix. In normal distributions, the correlation matrix covers all information about the statistical dependence between assets. However, this cannot be generalized for higher order moments, as stated by Canela and Collazo (2007), and not even for MV models when not resampled often enough to capture the differences in variance over time.

Maringer and Parpas (2007) shows that even for a mean-variance-skewness space with a universe of feasible portfolios, the majority can be inefficient with the usual assumptions of risk aversion. That suggests MVS can be a less powerful model than one would consider given the increment in input data and complexity in comparison with output results. Maringer and Parpas focus on a single period model which generates an efficient surface of portfolios with every point on the surface corresponding to some investor's preference. Their global optimization approach uses to algorithms: Differential Evolution (an evolutionary algorithm for continuous optimization) and Stochastic Differential Equation (a method that penalizes deviation from the feasible set). A critique of this model is that is very difficult to achieve convergence and diversity with genetic algorithms for multiobjective optimization problems (YUE; WANG, 2016).

2.3. MVSK Model

Here, we have the kurtosis added to the MVS model, a measure that describes the shape of a distribution's tails in relation to its overall shape. Kurtosis is sometimes confused with a measure of peakedness of a distribution. A normal distribution has a kurtosis of 3, what means that some of its distribution's mass is indeed in the tails. Consequently, excess kurtosis is another concept often used, that expresses kurtosis minus 3, comparing it to the Normal distribution.

A distribution with positive excess kurtosis is called leptokurtic and negative excess named platykurtic. According to Costa et al. (2005), financial markets' variables show strong leptokurtic characteristics, what causes kurtosis risk when the model assumes normality (commonly referred as fat tail risk since leptokurtic distributions has fatter tails).

High kurtosis of the return distribution implies that the investor will experience occasional extreme returns (either positive or negative, since more mass will be distributed in the left or right tails). That higher kutosis represents more extreme returns comparing to the usual three standard deviations from the mean (which accounts for 99,73% of return outcomes according to the bell-shaped distribution). So, high kurtosis scenarios means less mass on the shoulders of the distribution and more on the tails, what the investor will try to minimize to reduce volatility.

Adding kurtosis (K_p) to the already presented MVS model, we have the following MVSK objective function:

$$\text{Min} \quad \sigma_p^2(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} W_i \, W_j \, Cov_{i,j}$$

$$\text{Max} \quad \bar{R}\,(\text{x}) = \sum_{j=1}^{n} R_j \, W_j$$

$$\text{Max} \quad S_p(x) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} E(R_i R_j R_k) W_i W_j W_k}{(\sum_{i=1}^{n} \sum_{j=1}^{n} W_i W_j Cov_{i,j})^{3/2}}$$

$$\text{Min} \quad K_p(x) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} E(R_i R_j R_k R_l) W_i W_j W_k W_l}{(\sum_{i=1}^{n} \sum_{j=1}^{n} W_i W_j Cov_{i,j})^{2}}$$

$$\text{Subject to} \quad \bar{R}_p \geq \sum_{j=1}^{n} R_j \, W_j$$

$$\sum W_i = 1, \;\; W_i \geq 0$$

Araciolglu et al, (2011) constructed the MVSK portfolio allied with the polynomial goal programming (PGP) model for maximizing expected return and skewness and minimizing risk and kurtosis simultaneously of 30 stocks on Istanbul Stock Exchange. PGP is often used for multiple competing min-max portfolio objectives with successful empirical results. It is a nonlinear improvement of goal programming for higher orders, being a branch of multiobjective optimization that deals with such conflicting objective functions. Thus, the advantages of PGP consists in the existence of an optimal solution and the possibility of incorporating investor desires to the model, like higher central moments, solving the trade-offs on conflicting objectives of maximizing return and skewness simultaneously on minimizing variance and kurtosis.

Yue and Wang (2016), in its turn, presented a MVSK model with a new proposed algorithm designed to take into consideration population decomposition, crossover operations, selection strategy and update strategy seeking to achieve the two goals of convergence and diversity in the multiobjective optimization, what is difficult to achieve with goal programming techniques. Due to computational complexity, it was chosen only 12 assets from Shanghai Stock Exchange. Their MVSK results outperformed the classic model in the tests provided; resulting in well-diversified Pareto optimal solutions, what illustrates the practicality and effectiveness of the proposed model.

2.4. Information Entropy

Information theory, originally proposed in Claude Shannon's 1948 article *"A mathematical theory of communication"* studies the quantification, storage and communication of information. In this model, entropy is defined as the amount of uncertainty contained in a random variable or process.

Shannon defined the entropy H of a discrete random variable as:

$$H(x) = \sum_{i=1}^{n} P_i log_b \left(\frac{1}{p_i}\right) \quad (2.4.1)$$

Where P_i is the probability of the random variable and b is the corresponding units of entropy (or bits). For an application in portfolio theory, P_i is

A way of thinking is that if the entropy of an information source reduces, we can ask fewer questions to guess the outcome, with Shannon's bit acting as a measure of surprise that this information brings to the system (as in the amount of information that can be transferred, with the surprise value being the mean informational amount).

When all events in a system had the same probability of occurrence, the entropy in that system is maximum, and therefore, the level of available information in that system will be minimum. So, high levels of entropy require vast amounts of information as well, in the way that information can reduce entropy. When predictability is introduced, the entropy goes down.

This concept of uncertainty is related to how much choice about an outcome there is in the selection of an event, providing measurement for this information in terms of probabilities, which can be applied for financial market information (PHILIPPATOS; WILSON, 1974).

For non-normal distributions, entropy as a measure of uncertainty works well because it is more dynamic and general than variance and does not depend on normality assumptions, allowing its financial data to be modeled in any distribution (ZHOU, 2013). In another study called *"A Portfolio optimization model based on information entropy and fuzzy time series"* from 2015, Zhou investigates entropy risk models applying fuzzy forecasting, discussing its effectiveness in the Chinese financial market and how the new model outperformed the traditional ones. Our MVSKE model for the brazilian case will neither rely on fuzzy techniques, nor on FCM clustering algorithm but this difference would be interesting to compare in future works on brazilian markets in a fuzzy environment, to see if performance would change.

FAMA (1965), and PHILIPPATOS & NAWROCKY (1973) already used the concepts of information entropy to test the hypothesis of market efficiency, analyzing the reaction of prices to the flow of information. BUCKLEY (1985) discusses the principles of minimum and maximum entropy, where maximum entropy is expected in a system where you know only a few bits of information but no further knowledge of it. If it were on a state of lower entropy it would contain more information than previously specified WEHRL (1978).

The Principle of Maximum Entropy (MEP) was successfully implemented in option pricing as well. BUCHEN & KELLY (1996) created a significant method using MEP to estimate the maximum entropy distribution of assets, simulating their option prices at different strike values.

In succession, CASSETTARI (2003) proposed a model for portfolio allocation based on the maximum entropy principle, where entropy acts as a measure of financial risk, comparing it to the classic Markowitz MV model, but instead of adding information entropy as another objective function, the author makes a mean-entropy model, pulling variance out of the model, showing interesting results for going out of mean-variance tautology critics.

Mutual information, a different form of conditional entropy model, is also used in new portfolio management approaches. Nassim Taleb (2020) mentions that metrics linked to entropy such as mutual information are vastly more potent than correlation, one of the reasons being the uncovering of nonlinearities.

An advantage of the information entropy approach is that it does not make assumptions about normality of population's parameters in the distribution. Moreover, in the case of continuous distributions, the entropy is sensitive to any oscillation in the variable, thus, being adequate to financial modeling (DIONÍSIO and MENEZES, 2003).

The case in favor of using Shannon Entropy for quantifying informational efficiency levels in financial markets is that symbolic analysis is useful in detecting the dynamic of highly noise time series as asset returns are, and the application of entropy recovers the information in the series detecting the formation of patterns (RISSO, 2009). Therefore, assuming that in an efficient market hypothesis (EMH) it is not possible to predict future prices using past prices, the probability of having positive (negative) returns the next day is ½, denoting maximum uncertainty about future predictions, what relates to Shannon maximum entropy. So, if IBOVESPA for example shows certain level of inefficiency, it will be measured by an H significantly less than 1.

Other aspect concerning models that uses entropy as an objective function to determine portfolio weights, is that such weights become automatically non-negative due to the mathematical formulation of entropy models (USTA and KANTAR, 2011). This means no short-selling, which is preferable most of the time, as stated by Zheng et al (2011).

2.5. Fuzzy Set Theory

One common critique of Markowitz's mean-variance model is of underperforming out of sample (DEMIGUEL et al., 2009). If the samples do not truly mirror the population parameters of assets returns and volatility, one conceptual tool that can help reduce such real-world uncertainty is the fuzzy theory framework.

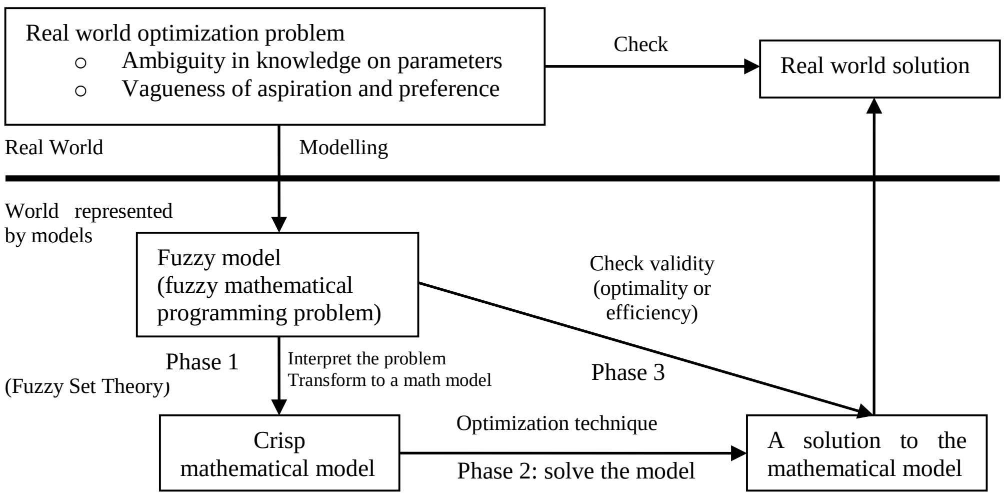
Portfolio optimization under fuzziness also makes sense as management restraint tool (for example, as expecting a return over x% or a risk lower than y%), where the investor choice prevails for a desired risk-return preference (GUPTA et al., 2014). This is difficult to obtain with crisp numbers or even interval numbers. To capture this vagueness and create flexibility to meet the investor's preferences, the fuzzy set theory can be applied as a financial tool for decision under uncertainty.

Lukasiewicz created the notion for what would be the concept of fuzzy membership in the future, with values varying from 0 (not a member of a specific set) to 1 (a member) and everything between given some possibility of membership. Zadeh (1970) later introduced the concept of fuzzy mathematical programming for decision making, with flexibility of target values of the objective function. Far ahead, Bojadziev (2007) exemplified it by asserting that future events cannot be considered binary, as true or false statements. This logic isn't enough to describe the future.

According to this definition, other category of fuzzy programming, called possibilistic programming, treats ambiguous coefficients as fuzzy (with ambiguity in the sense of how much imprecision there is around the center of such coefficient), analyzing possibility distributions on such coefficient values (INUIGUCHI, 2000).

The general approach to fuzzy programming comes in steps. First, a real-world problem is fuzzified. At this point, that fuzzy model created is transformed to a mathematical model and solved through optimization methods. Then, the optimal solution, after checked for validity, is deffuzified into a real-life solution.

Figure 4 – Fuzzy programming approach



Source: INUIGUCHI, 2000

For the portfolio selection problem in a possibilistic programming approach, we assume asset's return as a possibilistic variable instead of a random variable and then proceed to the fuzzification of Markowitz equations to define a membership function, seeking to reduce the uncertainty in the forecast of future returns.

2.6. Other models

In the previous section was shown fuzzy programming techniques but in a linear optimization approach. For example, we could transform a Markowitz quadratic function into a linear programming problem with techniques like the spread minimization model. Inuiguchi showed that a possibilistic linear programming problem with a quadratic membership function is equivalent to a stochastic programming problem with a multivariate normal distribution. But some problems arise when the model is adjusted for a non normal 5^{th} degree objective function. Fuzzy neural networks are useful in this context when input/output information is available, as stated by Zhang and Tao (2018), with genetic algorithm aimed to optimize the structure and parameters of the network. Fuzzy non linear programming (FNLP) don't appear often in academic models due to computational power involved but it's a good topic for future works on the subject.

Another models worth mentioning for future studies concerns portfolio dynamic rebalancing with resampled efficient frontiers, as in the works of Michaud (2007). As well as fuzzy-entropy models, as defined by DE LUCA & TERMINI (1974) to describe non-probabilistic entropy models with the use of fuzzy set theory.

3. METHODOLOGY

3.1. Dataset and modelling structure

The dataset contains B3's assets return information in the last decade (2011-2019). Longer time windows were not chosen due to the fact that financial data for brazilians markets are mostly scarce (YahooFinance, the bigger place for financial public data and the most used on R projects, don't have any data before 2007 for most brazilian assets for example). And 2020 was excluded due to pandemic effects that could greatly skew the results, making it harder to do proper corrections to the dataset without the correct econometric tools and statistical adjustments. So, the data will consist on assets' daily and monthly prices from the 2011-2019 period.

We will focus on a portfolio allocation model instead of allocation and selection, to make the computational efforts manageable since the difference between a fixed portfolio of 20 assets compared to that of a 75 asset portfolio (as IBOVESPA index for instance) makes it exponencially more complex when dealing to a 5-factor model. As shown in the literature, 20 assets already mitigates most of idiosyncratic risk of individual assets. For example, the 7 most negotiated assets of Bovespa embodies 34,8% of all volume traded yearly, and a portfolio with 15 of the most negotiated assets already represents more than half of the stock exchange's total size, at around 59,7% of total IBOV volume, as of 2020.

To choose these assets putting the portfolio together, the historical performance will be taken into account through backtesting simulations to generate candidate potfolios. The portfolio produced from this process will be compared in terms of risk and reward to the one consisting of the fifteen most traded papers. Then, after this selection, the model parameters will be aplied (as the MVSKE multi objective structure) coming to a desired portfolio allocation.

With this method, the newly shaped MVSKE model will try to answer some hypothesis about the brazilian financial market such as: does the high asymmetry of IBOV assets leads to higher returns? Higher moment models that capture such movements generate better portfolios? Is that statistical aspect significant enough or other variables could best explain IBOV idiosyncrasies? How does this model compares to the Markowitz classic model? And to the simple equally weighted allocation strategy?

The performances of the MVSKE model will be assessed in terms of the following portfolio performance measures: the Sharpe ratio (using the standard and adjusted for skewness ratios), Sortino ratio, Rachev ratio and mean absolute deviation ratio. Then, performance hyphotesis testing will be computed to evaluate statistical significance for the difference in ratios between the models analyzed.

3.2. MVSKE Structure

The model's structure will have five variables: mean, variance, skewness, kurtosis and entropy. The multi objective function will be modeled with the help of R software (RStudio 1.3.1093). The main packages used are: Caramel (for non linear multiobjective optimization purposes), GPareto (for Pareto Front Estimation and Optimization), Quantmod alongside with PerformanceAnalytics (for portfolio analysis), and RTransferEntropy alongside NonLinearTseries (for entropy analysis). Thinking in terms of reproducible results, the algorithm used is "L-BFGS-B", with set.seed 2 for random number generator and 20.000 repetitions for portfolio simulations.

The methodology for the training set will consist in 5/6 of the total data window, to be tested on the last 1/6 of the sample. This out-of-sample analysis will be important for the validation of the model's performance.

The objective function of the model is:

Max $f_1(x) = \bar{R} = \sum_{j=1}^{n} R_j \, W_j$

Min $f_2(x) = \sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} W_i \, W_j \, Cov_{i,j}$

Max $f_3(x) = S_p = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} E(R_i R_j R_k) W_i W_j W_k}{(\sum_{i=1}^{n} \sum_{j=1}^{n} W_i W_j Cov_{i,j})^{3/2}}$

Min $f_4(x) = K_p = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} E(R_i R_j R_k R_l) W_i W_j W_k W_l}{(\sum_{i=1}^{n} \sum_{j=1}^{n} W_i W_j Cov_{i,j})^{2}}$

Min $f_5(x) = E_P = -\sum_{i=1}^{n} W_i \, (log_n W_i)$

Subject to $\bar{R}_p \geq \sum_{j=1}^{n} R_j \, W_j$

$$\sum W_i = 1, \; W_i \geq 0$$

4. ANALYSIS AND DISCUSSION OF RESULTS

In this topic we evaluate the results of each model, comparing its similarities and dissimilarities. The data is basically historical prices from Yahoo Finance collected through RStudio, and from there we apply the statistical procedures for each model.

After the effciency of each model is evaluated through financial ratios, we compare those numbers to that of passive portfolios allocation's strategy, as done by Andrino & Leal (2018), to see if such new active models generate some excess return.

The framework for our portfolio selection model used mostly entropy to rank assets, since we're constructing an model whose distinction is the optimization of information entropy and transfer entropy between assets. But before we pick the entropy of the top components of each index studied, we filtered assets by some factors like value and momentum (as done in Fama&French models), and amount of data as well. The reason being that if we choose assets by entropy alone, would have too small of a sample, and other assets with good performance would stay out of the basket since entropy is not the only variable important to selection. Other filters to reach our model's selected assets were if they are growth stocks, defensive stocks (low beta low risk) along with other aspects like momentum and liquidity as pointed by Asness and Moskowitz (2015).

That pre-filters presented themselves necessary since just entropy ranking would turn out in portfolios with one or two assets weighting for the whole portfolio along other risk-return inefficiencies. This is due to known characteristics like small growth assets, low sharpe and other non-desirable aspects that would clearly affect the portfolio performance. In this selection framework, 25 pre-filtered assets were chosen:

Table 1: Pre-selected assets

ABEV3	BRKM5	CVCB3	IVVB11	RADL3
AZUL4	USDBRL	ECOO11	LAME4	SMAL11
BBDC4	BRML3	EMBR3	MGLU3	USIM5
BOVA11	BTOW3	ENBR3	MRFG3	VVAR3
ITSA4	CMIG4	GOLL4	MULT3	WEGE3

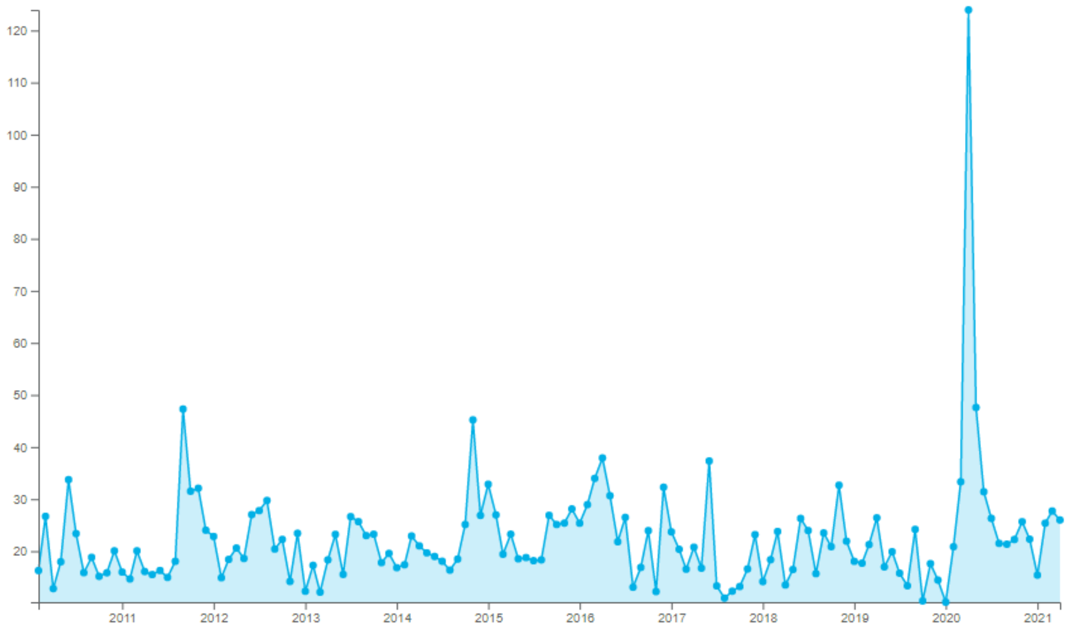
Source: from the author

From the 25 assets pre-selected, we don't have enough data provided by YahooFinance as needed for our model's test data with: AZUL4, CVCB3, ECOO11, IVVB11, VVAR3. That leaves our model with 20 selected assets between currency (USDBRL), ETFs (BOVA11 and ECOO11) and mostly, stocks.

Another limitation lies in the selection of the assets and its classes. Even though we have a diversified set of classes with currency, ETFs and stocks; some classes like fixed income were difficult to pass the filter, since the most popular Brazilian "bonds", fixed income ETFs like IMAB11, FIXA11, IB5M11 (among others) doesn't provide us with enough data to a proper analysis.

As stated in the methodology section, our model will have its training set going from 2011 to first half of 2018, consisting of 5/6 of total dataset; to be tested on the last 1/6 of the sample (second half of 2018 to 2019). This out-of-sample analysis will leave 2020 off the table due to black swan events caused by Covid crashing the markets, which took volatility to extremes, as seen in the standard deviation graph below.

Figure 5: Ibovespa Monthly Volatility

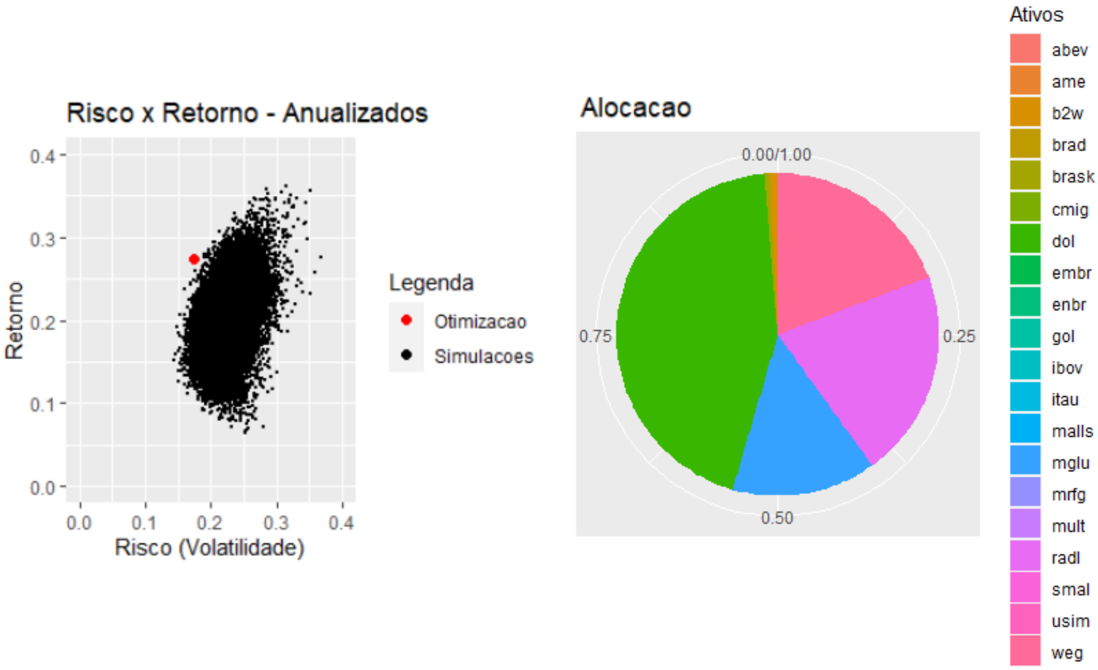


Source: BM&FBovespa

4.1. MV Model Analysis

From the 20 selected assets, there were six assets with positive weights in the optimally weighted MV in-sample analysis (considering in-sample analysis with whole dataset), as can be seen in the figure below.

Figure 6: opt MV in-sample analysis



Source: from the author

Composition of MV "In-Sample" Portfolio: 46.5% USDBRL ("dol" in the figure 6), 20.9% RADL3, 18.1% WEGE3, 14.2% MGLU3, 0.1% BRKM5 and 0.2% BTOW3 (B2W).

Constructing the portfolio and optimizing with the classic mean-variance approach, we have the "efficient frontier" above. What strikes first is that it doesn't look like an classical efficient frontier since the vast majority of stocks combinations below are underneath efficiency threshold. However, if we construct a line with only the portfolios along the top left border, we will take notice of the efficient frontier curve's concave aspect.

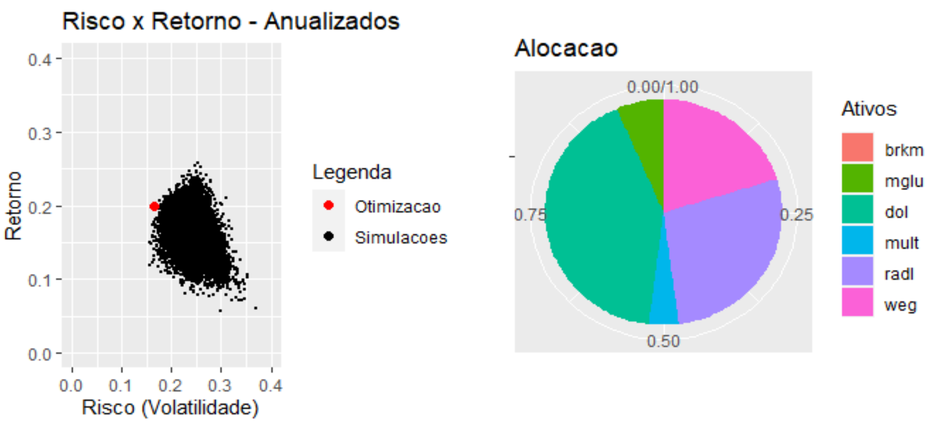
Analyzing this MV Portfolio composition, those were the six assets with the best performance (in terms of MV analysis) among the 20 selected through the decade of data in the 2011-2019 period. However, we are looking solely to the past here. To construct a robust model we have to go out-of-sample and test the variables performance without results already on the table.

To give more robustness to the model, a training dataset from 2011 to the first half of 2018 were created as stated before, to test out-of-sample performance in second half of 2018 to 2019 (totaling a 90-to-18month training-to-test ratio; 5/6 to 1/6). It was done that way instead of the classic 2/3 training 1/3 test for some reasons. The first being the need for a big training dataset since in this case 72 months to predict 36 would be insufficient for financial data and its asymmetrical properties (what demands lots of data for validity), and we do not have access to older data for most assets. The second reason being it eliminates the need for yearly rebalancing of the portfolio since with the 2/3 approach rebalancing would be vital to keep the portfolio on track with momentum of the whole market (for example in 2011-2016, Ibovespa remained on a flattening bear market but 2017 started its bull market for the index). 5/6 training, 1/6 test loses some level of robustness in comparison to 2/3-1/3 but adds simplicity in an already complex and dynamic model.

Besides, Markowitz MV model is already known for weak performance out-of-performance (DEMIGUEL, 2009). The goal is to see how MVSKE performs in comparison to the benchmark, so that 5/6 approach in comparison to MV models would serve its purpose.

Below we have the model results for out-of-sample MV analysis (for reproducible results, the algorithms and packages used are listed in the methodology section):

Figure 7: optimized MV portfolio out-of-sample



Composition of MV "Out-of-Sample" Portfolio: 41.9% USDBRL (dol), 3.2% MULT3, 27.5% RADL3, 21.2% WEGE3, 6.1% MGLU3, 0.1% BRKM5 (B2W).

With the 5/6 training set 1/6 test set method, we arrived at the performance above out-of-sample. It's easily perceived that MGLU3 is just starting getting traction (since most of the stock results came just after 2018, out of training set) and with MULT3 losing traction in comparison with the whole dataset. Another aspect concearns the overall sharpe for possible portfolios, which are lower since in comparison to "in-sample" results (as expected). The bigger part of the bull market from 2018-2019 is off in the training set, which resulted in the lower overall Sharpe Ratio as well. However, to see how significant are those differences we arrive at the tables below.

Even though the model was a good fit (since the portfolio is very similar in and out-of-sample), the weights differ in a significant degree, which can change a lot the results for each portfolio.

Table 2. MV data based on annual portfolio returns (in and out-of-sample)

Sample	Portfolio	Mean	Std Dev.	Skewness	Kurtosis	Entropy
IN	1/N	0.271	0.224	-0.105	-0.203	0.613
IN	OPT	0.246	0.164	-0.139	-0.333	0.529
OUT	1/N	0.196	0.181	0.041	-0.483	0.682
OUT	OPT	0.204	0.152	-0.229	-0.554	0.597

Source: from the author

Since this is a Mean-Variance analysis, has to be noted that the values for skewness, kurtosis and entropy did not receive any optimization. The OPT portfolio of Table 2 refers to mean-variance optimization only. In the next section we'll discuss full MVSKE optimization.

Below we have the comparison in MV performance for in and out-of-sample portfolios:

Table 3. MV ratio performance based on annual portfolio returns (in and out-of-sample)

Samp	Models	SR	MADR	SSR
IN	$MV_{(1/N)}$	0.7651	1.0372	1.2814
IN	$MV_{(OPT)}$	0.8927	1.1872	1.4235
OUT	$MV_{(1/N)}$	0.5326	0.7025	0.8672
OUT	$MV_{(OPT)}$	0.6868	0.8948	1.0274

Note: The SR, SSR and MADR denote the Sharpe ratio; Sortino-Satchell Ratio and Mean Absolute Deviation Ratio, respectively. Mean CDI (2011-2019) used as Brazilian risk free rate in the period.

We can note how naïve portfolios are underperforming in relation to optimized ones. The Sortino Ratio, that investigates downside deviation, seems to be in line with mean deviation ratio and Sharpe, what shows that data sample do not have large enough drawdowns and SR here can be a good performance evaluator.

4.2. MVSKE Model Analysis

Now we have to compare all those results in terms of risk, return and ratios to the next model, MVSKE, to see if that really exists excess return by modeling tail risk more specifically.

Going into MVSKE models, we have below the data for rolling skewness and rolling kurtosis for a 2-year moving window of some of the our portfolios to investigate before looking at the whole picture.

Two of our 5-objective-function in this model are: maximize skewness and minimize kurtosis. We can see in the graphs below how this can be a conflicting objective. And along with those two variables and the classic mean-variance ones, there is now entropy to the mix.

Figure 8: Skewness and Kurtosis for MVSKE portfolio

Source: from the author

Besides entropy, which is a new addition explored more and more by post-modern portfolio theory, this fact of MVSK variables being intrinsic related to each other (as central moments of higher powers) is a fact of criticism in the related models.

One of main critics of using such models solely based on price movement without other accounting factors based on fundamental analysis is Richard Roll. Roll's critique (1977) to MV models unfortunately can be partially extended to MVS and MVSK models for the validity of empirical tests of CAPM equation, in terms of mean-variance tautology, such that testing the CAPM equation is equivalent to testing mean-variance efficiency of the portfolio, requiring no model assumptions. And when we add two more constraints focusing on higher order moments of the same equation, we don't get much more explanatory power then we had before. We are in fact creating more factors utterly related to the first one, as skewness and kurtosis tells an expanded story, but a very similar comparing to variance, statistically speaking; going into data dredging territory.

In that same sense, the 5-factor model proposed by Eugene&Fama (2014) seeks a way out of this tautology by pursuing additional explanatory power through independent accounting factors, such as profitability and investment, besides the 3 old factors of market risk, outperformance of small versus big firms and P/B ratios. Asness (2014) goes further and proposes momentum as a sixth-factor. But even those models have their own critiques, although being more accepted than mean-variance based ones.

To try a path out on higher moments around the mean, we can add entropy to the model seeking more explanatory power. That's the reason behind moving from MVSK to MVSKE models.

The analysis of information entropy in MVSKE expands the model's likelihood to detect financial information flows, since nonlinear relationships can also be measured. It's true that small stocks have lower impact on the index while large caps dictate most of the flow. On the other hand, it is unclear how the market environment as a whole (as measured by the index) might provide information to individual stocks. Thus, to measure the extent to which information flows between the index and the stocks, we can use transfer entropy, with the help of RTransferEntropy package shaped by Behrendt, Simon, et al. (2019).

Transfer Entropy derives from Shannon's entropy $E_P = -\sum_{i=1}^{n} W_i\,(log_n W_i)$, as a measure for uncertainty, but expanding to measure the amount of directed transfer of information between two processes.

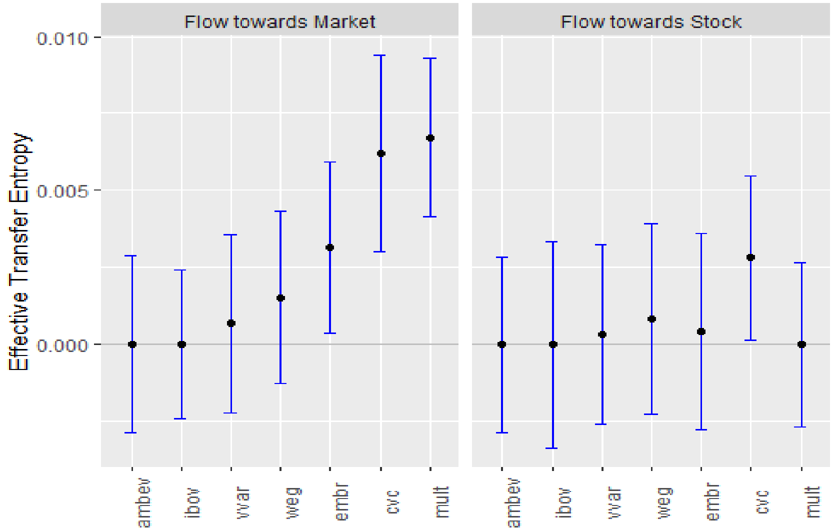
Transfer entropy can be written as:

$T_{X\rightarrow Y} = H(Y_t|Y_{t-1:t-L})$ - $H(Y_t|Y_{t-1:t-L}, X_{t-1:t-L})$, Where:

$H(X)$ is Shannon's entropy of the asset X;

X *and* Y are transfer processes where conditional mutual information takes place ;

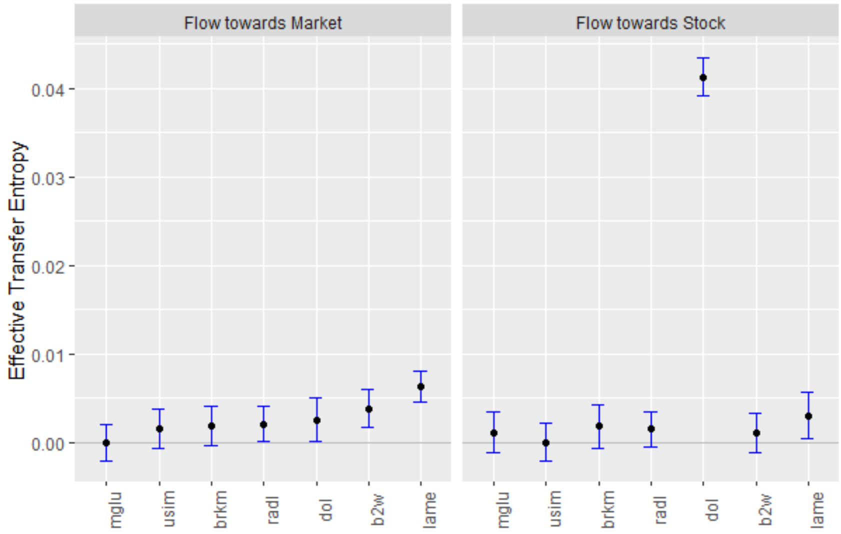
Figure 9: Transfer entropy for brazilian portfolio



Source: from the author

The graphic above reports transfer entropy with 95% confidence intervals for selected stocks. We can see that information flow from stocks to the market is higher for most stocks than in the other direction, what makes sense, even though there is indeed bi-directional information flow. Also, as can be seen from the confidence bounds, the information flow towards the stock is actually not statistically significant in all cases for our brazilian portfolio in the example, except for CVC that is slightly above the 5% significance level. When analyzing flow towards market, we have EMBR3, CVC and MULT3 with significant levels of information entropy.

Figure 10: Transfer entropy for another brazilian portfolio



Source: from the author

In this other portfolio, we can see clearly how USD/BRL currency's information flow from market towards stock is different from the other class of assets. What makes sense due to the negative correlation dolar usually has in comparison to Ibovespa (what could be seen in a greater magnitude in the pandemic due to carry trade currency effects) serving as hedge in a balanced portfolio.

A further study would consist of testing such assets with Rényi entropy instead of Shannon's. The former is particularly useful if the tails are assumed to be more informative than the centre of the distribution, what surely happens in finance with such volatile markets. Overall,this analysis show us how entropy and its different variables can find new relation between stocks that higher-order moments can't, vastly increasing the explanatory power of the whole model.

Finally, for the MVSKE 5-factor optimization, we have the following results on the positive optimized weights between the 20 assets selected:

Table 4: MVSKE Portfolio Opt Weights (In and Out-of-Sample)

	BBDC4	BOVA11	BTOW3	MGLU3	MULT3	RADL3	WEGE3	USDBRL
IN	0	0.136	0	0.252	0.038	0.202	0.153	0.219
OUT	0.057	0.112	0.053	0.172	0.095	0.177	0.149	0.185

Source: from the author

Table 5. MVSKE data based on annual portfolio returns (in and out-of-sample)

Sample	Portfolio	Mean	Std Dev.	Skewness	Kurtosis	Entropy
IN	1/N	0.241	0.215	-0.047	0.180	0.577
IN	OPT	0.286	0.227	0.001	0.434	0.509
OUT	1/N	0.226	0.234	-0.189	0.236	0.602
OUT	OPT	0.233	0.187	-0.119	0.376	0.524

Source: from the author

As we can see, the results are very widely spread out, with some portfolios improving their skewness and kurtosis risk along with better sharpe performance and others not so much (we're reviewing those numbers along with performance ratios on the table at the end of section). One aspect the model succeed in minimizing across all portfolios was to minimize entropy risk. But as can be seen below, those results do not changed the overall scenario that much, with very similar results among MV and MVSKE performances.

For the portfolio's entropy calculations, approx entropy was used (from pracma package on R). The option for using approx instead of sample entropy is that being bounded in the [0,1] range, it's easier to interpret as the amount of regularity and the unpredictability of fluctuations in a time series. With a low entropy value indicating that the time series flows towards being more deterministic; and a high value indicating more randomness.

Table 6. MVSKE ratio performance based on annual portfolio returns (in and out-of-sample)

Sam	Models	SR	ASR	SSR	MADR	FTR	GRR
IN	$MVSKE_{(1/N)}$	0.6577	0.6877	1.2004	0.8457	2.4571	1.8856
IN	$MVSKE_{(OPT)}$	0.8211	0.8805	1.4895	1.0348	2.9233	1.9233
OUT	$MVSKE_{(1/N)}$	0.5402	0.5491	1.0167	0.7436	2.0568	1.6589
OUT	$MVSKE_{(OPT)}$	0.7134	0.7430	1.0546	0.9102	2.6060	1.9005

Note: The SR, ASR, and SSR, MADR, FTR and GRR denote, respectively: the Sharpe ratio; Adjusted Sharpe Ratio, Sortino-Satchell Ratio, Mean Absolute Deviation Ratio, Farinelli-Tibiletti Ratio and Generalized Rachev Ratio.

Along with the Sharpe Ratio, Adj SR, Sortino Ratio and Median Deviation Ratio, were added two more ratios in MVSKE analysis to focus on quantiles distribution of non-parametric models: The Farinelli-Tibiletti Ratio (FTR) and Generalized Rachev Ratio (GRR). The FTR above was in relation to the first lower and upper partial moment for a moderate investor profile. The FTR results do not show big variations in moments across the different portfolios' performances. The GRR (Generalized Rachev Ratio) measure the quantiles of portfolio returns and in this specific analysis, we choose the 5% quantile levels as the standard. The results also do not scream abnormalities in the portfolio or big non-normal risks in the quantiles analyzed, what can be confirmed observing the the low variation in skewness and kurtosis among the portfolios analyzed (what can be further attested by seeing the minimal variation between SR and ASR ratios).

And finally, we can now compare the ratios between the previous models relating to portfolio performance:

Table 7. Out-of-sample performance between MVSKE and the benchmark models

Models	SR	ASR	MADR	SSR
$MV_{(1/N)}$	0.5326	0.5565	0.7025	0.8672
$MV_{(OPT)}$	0.6868	0.7168	0.8948	1.0274
$MVSKE_{(1/N)}$	0.5402	0.5491	0.7436	1.0167
$MVSKE_{(OPT)}$	0.7134	0.7430	0.9102	1.0546

Source: from the author

Since the MV model (which relies on Normal assumptions), is being compared with non-parametric models such as MVSKE, the ratios we can actually compare are the ones above (including ASR-Adjusted Sharpe Ratio which can be compared even though it calculates skewness and kurtosis, because the ratio is adjusted to a Gaussian setting that is where Sharpe Ratio functions). The Generalized Rachev Ratio and the Farinelli-Tibiletti Ratio, on the other hand, base its assumptions on non-Gaussian distribution settings.

We can note how naïve portfolios are still underperforming in relation to optimized ones in MVSKE, just like in the MV setting. The MVSKE optimized portfolio, even though presented a good performance over other portfolios presented, it did not have significant difference between the benchmark MV optimized portfolio to attest without hesitation of it being a superior model. However, it is in no way inferior to the benchmark as well.

Another limitation in the comparison is between MV and MVSKE portfolios, since the basket of assets in each one, even though very similar, differs, being harder to isolate the variables and evaluate. So, the optimized MVSKE portfolio does not necessarily outperforms the optimized MV portfolio, as the ratios would imply, but indicates that it is at least as useful as the benchmark model.

On the flip side, the comparison between 1/N naïve strategies and their optimized counterpart is a good one. It works since they have the same basket of assets' composition and because they clearly shows statistical significance in the superior performance of optimization against equally weighted portfolios. Other authors like Leal et. al (2018) could argue in favor of naïve strategies stating that transaction costs and more frequent rebalance would change this scenario but it's out of the scope of the models' variables.

Even with the still open hypothesis of MVSKE not beating the benchmark, this first attempt at studying such model in the Brazilian context helped create awareness for non-normal methodologies in portfolio management, and how to deal with such implied risks that are not being properly assessed otherwise. It generated the same kind of performance to an extent, but with the advantage of being a more complete tool for management of tail risks that are not being controlled on the benchmark models.

5. CONCLUSION

The necessity by investors to have realiable investment tools to technically assess uncertainty – that rises constantly in everyday markets – makes models like MV or MVSKE of vital importance for financial decision making, seeking to maximize value without taking non-analyzed risks.

In this study, an extension of Markowitz model with higher order moments and information entropy was proposed. Although our MVSKE model presented viable and efficient solutions, there were not persistent significant increase in excess return by quantifying skewness, kurtosis and entropy risk. On the matter of information flow through assets, efficient market portfolio suggests no persistent systematic mispricing of stocks that could generate abnormal returns to our portfolios, and that was confirmed by the ratios results. Even though no substantial empirical evidence to support higher-moments Markowitz was found, the model works well as a management tool for financial decision, being indeed a powerful instrument for quantitative analysis of asset selection and allocation.

The statistical power of MVSKE, although not significantly superior to the benchmark, uphold the same level of performance. Both MVSKE and the benchmark shows weaker performances out-of-sample. Since the MVSKE is almost completely derived from the same central moment variables, just expanding powers on the benchmark model, it could be expected that entropy variable wouldn't be capable of changing the results to a large extent.

For future studies, more work on refining the models constraints are needed, such as co-skewness and co-kurtosis deeper analysis over sub-sample periods, along with further tests to be stressed in order to investigate on the influence of asset class constrains and choice of personalized parameters in final results. Also, using a fuzzy framework could enhance the model's performance seeking to reduce the uncertainty in the forecast of future returns.

Extra research adressing transaction costs into the framework and giving more weight to entropy related variables like cross-entropy and mutual information as more robust variables in detriment to variance related ones see if performance changes, is highly appreciated. Another crucial recommendation is to gather more data to increase the reliability of the methods of research, since highly skewed variables demands high amounts of data.

This thesis is the first attempt to examine MVSKE types of higher order moment's portfolios in the brazilian markets and analyze its results comparing to the classic Markowitz model of Modern Portfolio Theory. The findings, although not conclusive, instigates more research on the field of quantitative finance and post-modern portfolio theory applied to brazilian cases in the years to come.

6. REFERENCES

ADCOCK, C.J.; SHUTES, K. **An analysis of skewness and skewness persistence in three emerging markets**. Emerging Markets Review, v. 6, n. 4, p. 396–418, 2005.

ANDRINO, Daniel D.; LEAL, Ricardo PC. **Equally weighted strategic allocation and balanced funds in Brazil.** Latin American Business Review, v. 19, n. 1, p. 23-53, 2018.

ARACIOĞLU, Burcu; DEMİRCAN, Fatma; SOYUER, Haluk. **Mean–Variance–Skewness–Kurtosis Approach to Portfolio Optimization: An Application in İstanbul Stock Exchange**. p. 9, 2011.

ASNESS, Clifford et al. **Investing with style.** Journal of Investment Management, v. 13, n. 1, p. 27-63, 2015.

ASNESS, C. **Our model goes to six and saves value from redundancy along the way.** Cliff's Perspective, 2014.

ASNESS, Clifford S.; MOSKOWITZ, Tobias J.; PEDERSEN, Lasse Heje. **Value and momentum everywhere.** The Journal of Finance, v. 68, n. 3, p. 929-985, 2013.

ARROW, Kenneth Joseph. **Essays in the theory of risk-bearing**. [s.l.]: Markham Pub. Co., 1971.

BELLMAN, R. E.; ZADEH, L. A. **Decision-Making in a Fuzzy Environment.** Management Science, v. 17, n. 4, p. B-141, 1970.

BEHRENDT, Simon et al. RTransferEntropy—**Quantifying information flow between different time series using effective transfer entropy**. SoftwareX, v. 10, p. 100265, 2019.

BHATTACHARYYA, Rupak; KAR, Samarjit; MAJUMDER, Dwijesh Dutta. **Fuzzy mean–variance–skewness portfolio selection models by interval analysis**. Computers & Mathematics with Applications, v. 61, n. 1, p. 126–137, 2011.

BUCKLEY, James J. **Entropy Principles in Decision Making Under Risk**. Risk Analysis, v. 5, n. 4, p. 303–313, 1985.

CANELA, Miguel Ángel; COLLAZO, Eduardo Pedreira. **Portfolio selection with skewness in emerging market industries**. Emerging Markets Review, v. 8, n. 3, p. 230–250, 2007.

CAVALCANTE, Jorge; ASSAF, Ata. **Long range dependence in the returns and volatility of the Brazilian stock market**. European review of Economics and Finance, v. 3, n. 5, p. 22, 2004.

CASSETARI, Ailton. **The Maximum Entropy Principle and the Modern Portfolio Theory.** Brazilian Review of Finance, v. 1, n. 2, p. 271–300, 2003.

CASTRO, F. Henrique et al. **Fifty-year history of the Ibovespa**. Brazilian Review of Finance, v. 17, n. 3, p. 47-65, 2019.

COSTA, Michele; CAVALIERE, Giuseppe; IEZZI, Stefano. The Role of the Normal Distribution in Financial Markets. *In*: BOCK, H.-H.; GAUL, W.; VICHI, M.; *et al* (Orgs.). **New Developments in Classification and Data Analysis**. Berlin/Heidelberg: Springer-Verlag, 2005, p. 343–350. (Studies in Classification, Data Analysis, and Knowledge Organization). Disponível em: <http://link.springer.com/10.1007/3-540-27373-5_41>. Acesso em: 17 jul. 2020.

DE LUCA, Aldo; TERMINI, Settimo. Entropy of L-fuzzy sets. **Information and Control**, v. 24, n. 1, p. 55–73, 1974.

DEMIGUEL, Victor; GARLAPPI, Lorenzo; UPPAL, Raman. **Optimal Versus Naive Diversification: How Inefficient is the 1/ *N* Portfolio Strategy?** Review of Financial Studies, v. 22, n. 5, p. 1915–1953, 2009.

DIONÍSIO, Andreia; MENESES, R.; MENDES, Diana. A **Entropia como Medida de Informação na Modelação Económica. Temas em Métodos Quantitativos**, edited by Elizabeth Reis and Manuela Magalhães Hill, Edições Sílabo, 2003.

DOANE, David P.; SEWARD, Lori Welte. **Applied statistics in business and economics**. 3rd ed. New York, NY: McGraw-Hill/Irwin, 2011. (The McGraw-Hill/Irwin series, operations and decision sciences).

DOMINGUES, Elder Geraldo. **Análise De Risco Para Otimizar Carteiras De Ativos Físicos Em Geração De Energia Elétrica**. p. 222, 2003.

EVANS, John L.; ARCHER, Stephen H. **Diversification and the Reduction of Dispersion: An Empirical Analysis.** The Journal of Finance, v. 23, n. 5, p. 761, 1968.

FAMA, Eugene F. The Behavior of Stock-Market Prices. The Journal of Business, v. 38, n. 1, p. 34–105, 1965.

GUPTA, Pankaj; MEHLAWAT, Mukesh Kumar; INUIGUCHI, Masahiro; *et al*. **Fuzzy Portfolio Optimization**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. (Studies in Fuzziness and Soft Computing). Disponível em: <http://link.springer.com/10.1007/978-3-642-54652-5>. Acesso em: 2 jul. 2020.

INUIGUCHI, Masahiro; RAMÍK, Jaroslav. **Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem.** Fuzzy Sets and Systems, v. 111, n. 1, p. 3–28, 2000.

KAHNEMAN, Daniel; TVERSKY, Amos. **Prospect Theory: An Analysis of Decision under Risk.** Econometrica, v. 47, n. 2, p. 263–291, 1979.

KRAUS, Alan; LITZENBERGER, Robert H. **Skewness Preference and the Valuation of Risk Assets.** The Journal of Finance, v. 31, n. 4, p. 1085–1100, 1976.

MARINGER, Dietmar; PARPAS, Panos. **Global optimization of higher order moments in portfolio selection**. Journal of Global Optimization, v. 43, n. 2–3, p. 219–230, 2009.

MARKOWITZ, Harry. **Portfolio Selection**. The Journal of Finance, v. 7, n. 1, p. 77–91, 1952.

MICHAUD, Richard O. The **Markowitz Optimization Enigma: Is 'Optimized' Optimal?** Financial Analysts Journal, v. 45, n. 1, p. 31–42, 1989.

MICHAUD, Richard O.; MICHAUD, Robert. **Estimation Error and Portfolio Optimization: A Resampling Solution.** SSRN Electronic Journal, 2007. Disponível em: <http://www.ssrn.com/abstract=2658657>. Acesso em: 14 jul. 2020.

PERMUTER, Haim H.; KIM, Young-Han; WEISSMAN, Tsachy. **Interpretations of directed information in portfolio theory, data compression, and hypothesis testing.** IEEE Transactions on Information Theory, v. 57, n. 6, p. 3248-3259, 2011.

PHILIPPATOS, George C.; NAWROCKI, David N. **The Information Inaccuracy of Stock Market Forecasts: Some New Evidence of Dependence on the New York Stock Exchange.** The Journal of Financial and Quantitative Analysis, v. 8, n. 3, p. 445–458, 1973.

PHILIPPATOS, George C.; WILSON, Charles J. **Entropy, market risk and the selection of efficient portfolios: reply**. Applied Economics, v. 6, n. 1, p. 77–81, 1974.

RAMLI, Suhailywati; JAAMAN, Saiful Hafizah. **Markowitz portfolio optimization model employing fuzzy measure**. In: AIP Conference Proceedings. AIP Publishing LLC, 2017. p. 040004.

REGENSTEIN JR, Jonathan K. **Reproducible Finance with R Code Flows and Shiny Apps for Portfolio Analysis**. CRC Press, 2018.

RISSO, Wiston Adrián. **The informational efficiency: the emerging markets versus the developed markets**. Applied Economics Letters, v. 16, n. 5, p. 485–487, 2009.

ROLL, Richard. **A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory.** Journal of financial economics, v. 4, n. 2, p. 129-176, 1977.

SHI, NingZhong; LAI, Min; ZHENG, ShuRong; *et al*. **Optimal algorithms and intuitive explanations for Markowitz's portfolio selection model and Sharpe's ratio with no short-selling**. Science in China Series A: Mathematics, v. 51, n. 11, p. 2033, 2008.

SHLEIFER, Andrei. **Inefficient markets: An introduction to behavioural finance**. OUP Oxford, 2000.

TABOR, Josh. **Investigating the investigative task: Testing for skewness: An investigation of different test statistics and their power to detect skewness.** Journal of Statistics Education, v. 18, n. 2, 2010.

TALEB, Nassim Nicholas. **Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications.** arXiv preprint arXiv:2001.10488, 2020.

USTA, Ilhan; KANTAR, Yeliz Mert. **Mean-Variance-Skewness-Entropy Measures: A Multi-Objective Approach for Portfolio Selection**. Entropy, v. 13, n. 1, p. 117–133, 2011.

VON HIPPEL, Paul T. **Mean, median, and skew: Correcting a textbook rule**. Journal of Statistics Education, v. 13, n. 2, 2005.

WEHRL, Alfred. **General properties of entropy**. Reviews of Modern Physics, v. 50, n. 2, p. 221–260, 1978.

YANG, Chin W.; HUNG, Ken. **A generalized Markowitz portfolio selection model with higher moments.** International Review of Accounting, Banking and Finance, v. 2, n. 1, p. 1-7, 2010.

YUE, Wei; WANG, Yuping. **A new fuzzy multi-objective higher order moment portfolio selection model for diversified portfolios.** Physica A: Statistical Mechanics and its Applications, v. 465, p. 124–140, 2017.

ZHANG, Ridong; TAO, Jili. **A Nonlinear Fuzzy Neural Network Modeling Approach Using an Improved Genetic Algorithm**. IEEE Transactions on Industrial Electronics, v. 65, n. 7, p. 5882–5892, 2018.

ZHOU, Rongxi; CAI, Ru; TONG, Guanqun. **Applications of Entropy in Finance: A Review.** Entropy, v. 15, n. 12, p. 4909–4931, 2013.

ZHOU, Rongxi et al. **A portfolio optimization model based on information entropy and fuzzy time series.** Fuzzy Optimization and Decision Making, v. 14, n. 4, p. 381-397, 2015.