University in Novi Sad
Faculty of Sciences
Department of Mathematics and Informatics

Some applications of information theory to investments

Master thesis by

Željana Knežević

Mentor:
Dr Mladen Kovačević

2024, Novi Sad

I would like to express my heartfelt gratitude to my mentor Dr Mladen Kovačević, whose guidance and support helped me during my journey of writing the thesis. Additionally, I extend my thanks to the members of the board for their valuable insights and feedback throughout the process.

To my college friends, you have been the vibrant colors brightening my academic journey, turning challenges into cherished memories, making the years not just bearable but truly unforgettable.

To my entire family, your collective encouragement has been the foundation upon which I've built my academic pursuits. A special thank you to my mom, thank you for being the force that propelled me forward, you gave me the strength and persistence that brought me to this moment. This work is as much yours as it is mine.

Table of Contents

"The first rule of an investment is don't lose (money). And the second rule of an investment is don't forget the first rule. And that's all the rules there are."

Warren Buffett

Chapter 1

Introduction

In 1956, a pair of Bell Labs scientists brought to light the scientific recipe for achieving wealth. One of them was Claude Shannon, often regarded as *the Einstein of the digital era*, MIT professor and inventor of information theory. The second scientist was John L. Kelly Jr., best known as the creator of The Kelly criterion. Together, they applied the science of information theory to the problem of the wealth maximization in the long run. Afterwards, Shannon and scientist Edward O. Thorp, besides co-creating the first wearable computer, they applied the "Kelly formula" in Las Vegas, and it proved to be effective. They did not just formulate theoretical concepts, but also tested them in a real life situation, earning money along the way. Inspired by their experiences in casino, it had led them to believe that there are some other spheres where potential financial gains can be realized in a similar manner, and the stock market was a logical and natural progression. The mathematics used in casino is surprisingly similar to the mathematics used on Wall Street. With a good risk and goal assessment, if used in a right way, it can give nice and successful results. Also, it is not exclusive to gambling and Stock Markets, it has a wide range of applications in various fields.

This study will explore fascinating connections between seemingly unrelated topics such as gambling, information theory, stock investing, portfolio theory and applied mathematics.

The thesis is organized into 7 chapters, where the first chapter is the introduction and the last is the conclusion. Chapter 2 introduces us to the basics concepts and properties of information theory and betting. The third chapter will provide us with application of utility function regarding wealth maximization and defining and applying Kelly criterion in order to determine the optimal portion of capital for investment. A couple of different cases will be shown. In the fourth chapter we face more serious tasks, considering the application of information theory concepts to horse racing betting, which are more sophisticated and require more conditions that must be taken into account. We deal with crucial concepts of information theory such as relative and conditional entropy and mutual information. Chapter 5 brings us to dynamic field of portfolio theory that offers a comprehensive framework for optimizing investment strategies. We will explore basic and key concepts such as modern portfolio theory which will yield us to demonstrate models that use entropy as a good substitute to traditional models. Practical application and achievements on real data will be addressed. In the final chapter we draw attention to generalization of Shannon's entropy, named Rényi Entropy. We delve into innovative ways in which Rényi entropy can be integrated into portfolio analysis, offering a new perspective on risk assessment and diversification. Through empirical examples and practical insights, we illustrate benefits of incorporating Rényi entropy into this field.

Generally, the thesis highlights the potential for improving traditional investment and gambling strategies through the lens of information theory.

Chapter 2

Overview of the basic concepts

2.1 Overview of basic concepts from information theory

This section relies mostly on [7].

First, let us explain the term information theory itself, also known as *the mathematical theory of communication* (Shannon, 1948). It is easy to conclude that without communication there is no information either. The essential problem of communication is to ensure that a message from one point can be either exactly or approximately reproduced at another point. Practically, when one person sends a message, person who receives it needs to understand it as sender intended.

Information theory does not have a single definition, but it can be explained as mathematical representation of the concepts, parameters and conditions affecting the transmission of messages through communication systems. The techniques used in information theory are described using probability theory, therefore, prior knowledge of that field is assumed.

We are interested in quantifying how much information a message is carrying. The answer will give a central notion of information theory – Entropy.

- **Entropy** - a concept that has countless definitions. It was first mentioned in the field of thermodynamics in the 19^{th} century. Many define it as a measure of disorder in a system.

However, in terms of information theory, we will start with a quite intuitive definition. Entropy attempts to measure the amount of information contained and conveyed within a probability distribution. Events with low probability are more surprising, hence, carry more information than those that are frequent. We can think of the amount of information as a surprisal of some event, for example E. Therefore, $I(E) = \log\frac{1}{P(E)} = -\log P(E)$. If probability that event will happen is 1, i.e. it will surely happen, $I(1) = 0$, and conversely $I(0) = +\infty$. Intuitively we see that an unlikely event carries much more information. According to that, we can define entropy as expected or average amount of information (surprising). A formal definition follows.

<u>Definition 1</u>: The entropy $H(X)$ can be calculated for a discrete random variable X as follows:

$$\boldsymbol{H(X) = -\sum_{x \in X} p(x) \log p(x)} \qquad 2.1$$

Note: Logarithm function uses base 2 and the units are bits. Henceforth it will be assumed that the logarithm is base 2 unless otherwise stated.

Also, entropy is measure of uncertainty. The lowest entropy is obtained when event is certain, with probability 1. Conversely, the largest entropy will be if all events are equally likely. In order to represent it graphically, we will introduce a special case of a discrete probability distribution, and also the simplest case of a random variable - the Bernoulli distribution. Bernoulli random variable can take only two possible outcomes, 1 for success and 0 for failure, where probability of success is p and probability of failure is $1 - p = q$. Because of its nature, it forms the basis for understanding entropy. Formal definition follows.

Definition 2: The entropy of a Bernoulli random variable is defined as follows:

$$H(p) = -p \log p - q \log q \qquad 2.2$$

The entropy reaches maximum for $p = q = \frac{1}{2}$. It can be clearly seen in Figure 1.

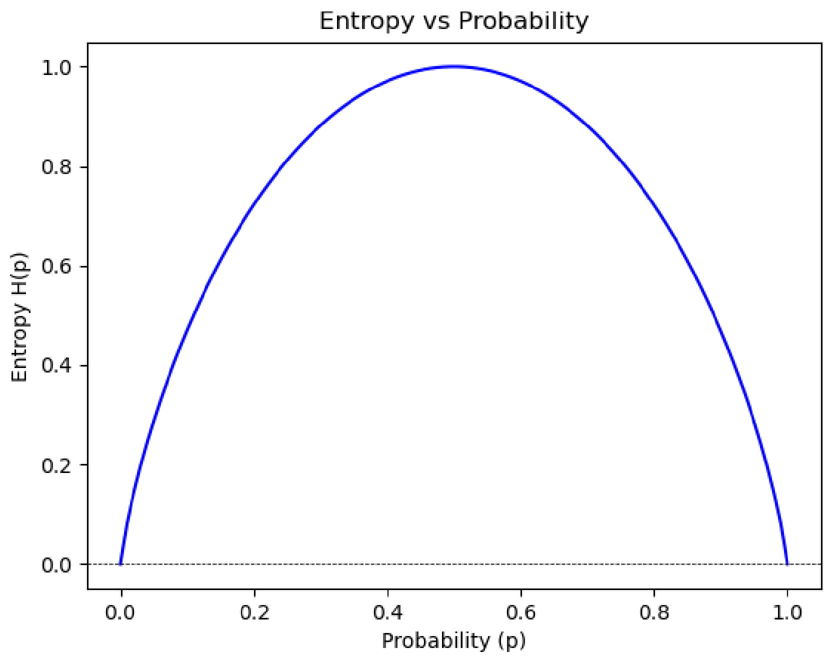

Figure 1 - Entropy of a Bernoulli random variable

In case of a continuous distribution, the so-called differential entropy is defined as follows.

Definition 3 (Continuous Entropy): If X is a continuous random variable and $p(x)$ is its probability density function (*pdf)*, entropy is calculated as:

$$H(X) = \int_X p(x) \log \frac{1}{p(x)} dx = -\int_X p(x) \log p(x)\, dx \qquad 2.3$$

- **Joint and conditional entropy**

The joint entropy of a pair of random variables (X, Y) is defined as:

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x,y) = -E \log p(X,Y) \qquad 2.4$$

Intuitively, conditional entropy quantifies uncertainty about X when Y is known, i.e. we use knowledge about Y to update our knowledge of X.

Mathematically, it is presented as:

$$H(X|Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(y)} = -E \log p(X|Y) \qquad 2.5$$

where X and Y are random variables, $p(x, y)$ joint probability, $p(x)$, $p(y)$ marginal probabilities.

- **Informational diagram**

The connection between joint, conditional and marginal entropy is presented by chain rule.

Chain rule is defined as follows:

$$H(X,Y) = H(X) + H(Y|X) \qquad 2.6$$

Graphically, it is graphically presented by so-called informational diagram as we can see in Figure 2.

Another concept is introduced here, and that is *mutual information* $I(X;Y)$. We can calculate it in different ways, depending on the given data. One way is for example:

$$H(X) - H(X|Y) = I(X;Y) \qquad 2.7$$

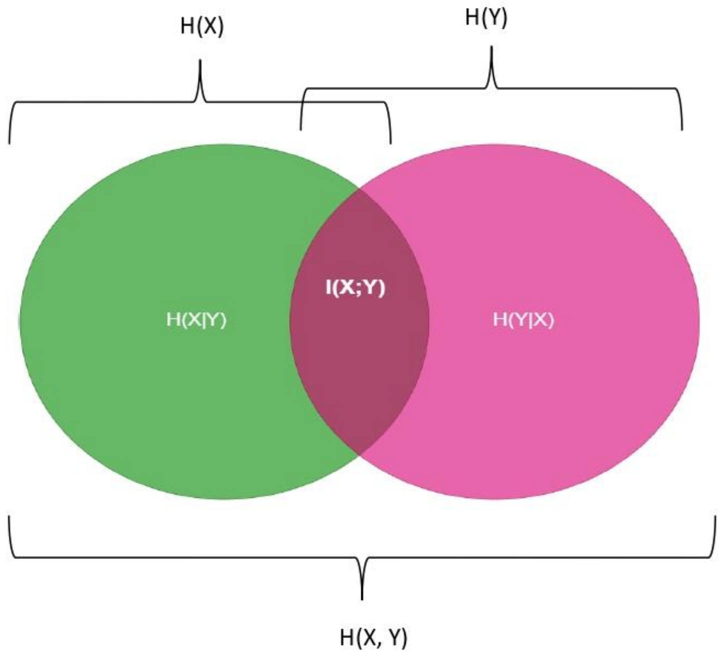

Figure 2 - Venn diagram of the relationships between information measures

- **Relative entropy**

<u>Definition 4</u>: The relative entropy is defined as follows:

$$D(p||q) = \sum_{x \in \mathrm{X}} p(x) \log \frac{p(x)}{q(x)} \quad 2.8$$

The relative entropy answers the question how different two probability distributions are, i.e. measure of distance between two probabilities.

The following properties of relative entropy hold:

$$D(p||q) \geq 0 \quad 2.9$$

$$D(p||q) = 0 \text{ iff } p = q \quad 2.10$$

- **Law of large numbers (Asymptotic equipartition property)**

In terms of information theory, the law of large numbers states that observed frequencies of events converge to their true probability as sample size grows. Additionally, patterns in the data can be better used, leading to more efficient compression.

Formal statement: Let $X_1, X_2, \dots, X_n$ be a sequence of independent and identically distributed random variables with mean $E[X_i] = \mu$. The sample mean of the first n observations is defined as:

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad 2.11$$

Then, the weak law of large numbers states:

<u>Theorem 1</u>: For any $\epsilon > 0$: $\lim_{n\to\infty} P(|\overline{X_n} - \mu| \geq \epsilon) = 0$.

Many proofs of theorems are based on the use of this concept.

The law of large numbers has an impact on expected values and our understanding of the long-term behavior of random events, which in our case are related to gambling. In short, it suggests that over a large number of trials, observed outcomes are likely to approach theoretical probabilities and expected values.

Now, we transition to a slightly broader and more general description of entropy.

- **Stochastic process and entropy rate**

A stochastic process is an indexed sequence of random variables. In other words, it quantifies randomness in time in the context of evaluation or change. Family of such random variables for a stochastic process and can be denoted by $\{X_t\}$, where t represents time.

Stochastic process is a *Markov process* if:

$$P(X_{t+1} = x_{n+1}|X_1 = x_1, X_2 = x_2, \dots, X_t = x_n) = P(X_{t+1} = x_{n+1}|X_t = x_n) \qquad 2.12$$

Formula 2.3 tells us that prediction of the future requires only knowledge of the present, i.e., nothing that happened in the past is relevant for the probability what will happen in the future. Only what matters is present state.

One of the classes of stochastic processes is *stationary process*. It is a process where the distributions are invariant under time translations, so we can translate all of them with arbitrary shift a, nothing will happen with distribution function. If stochastic process is denoted by X_t:

$$P(X_{t_1}, \dots, X_{t_n}) = P(X_{t_1+a}, X_{t_2+a}, \dots, X_{t_n+a}) \quad 2.13$$

Now we are wondering what is the entropy of that random process. If we have n random variables, we want to know how does the entropy of sequence grow with n. Answer will give us the definition of entropy rate.

<u>Definition 5</u>: The entropy rate $H(\chi)$ of a stochastic process $\{X_i\}$ is defined as follows:

$$H(\chi) = \lim_{n\to\infty} \frac{H(X_1, \dots, X_n)}{n} \quad 2.14$$

when the limits exist.

Now we will define related quantity for entropy rate:

$$H'(\chi) = \lim_{n\to\infty} H(X_n | X_1, \dots, X_{n-1}) \quad 2.13$$

From this definition, we can observe the amount of randomness that X_n gives, given the past observations. Dependence between random variables reduces total amount of uncertainty.

For a stationary stochastic process holds:

$$H(\chi) = H'(\chi) \quad 2.14$$

- **Channel capacity**

Shannon developed the idea of reliable information transmission, introducing the *channel capacity* concept. It is the upper limit, below which communication rates could be approached with low probability of error.

Detailed information can be found in Shannon's work [27] and we will deal with those notions in further work on practical examples.

2.2 Basic concepts in betting

There is variety of relationships of information theory to other fields, especially in economics (investments) and it has wide spectrum of application.

Gambling is mostly math-driven, although the majority of people enjoys gambling but dislikes math. Whether it is sports betting, horse race betting, poker, blackjack, roulette or stock market, there is always a mathematical background which plays significant role in understanding and optimizing a problem.

In the following, the basic notions that will be used in paper will be explained.

First, let us define what exactly a bet is. Two people place bets on whether or not a specific event will occur under specific circumstances. The wager may be 'even money', although generally speaking, 'odds' may be stated.

- **Even money bet**

'Even money' denotes a type of bet where, if you win, a payout will be equivalent to the original bet, i.e. the invested amount will be doubled. The reason why even-money bets are appealing is due to their simple nature and 50-50 chance of winning. An example of this type of bet is roulette and choosing to bet on red or black. We can think of these bets as *fair* due to their 1:1 payout ratio, where initial and winning amount is same.

- **Odds**

Bets are determined by their odds. They are used to express the probability of an outcome occurring. In other words, it is ratio of probabilities that an outcome will or will not happen. For instance, blackjack has the best odds for players, if they know how to play.

- **Fair odds**

Mathematically speaking, a bet is considered to be fair if all outcomes are equally likely, or when you play the same game multiple times, on the long run, your total gain is zero. It can be simply proved to be true if we observe tossing a coin. Probability of getting head is the same as getting tail. If we continue to toss for a large number of times, we expect to see roughly the same number of tails and heads.

Suppose that coin used in the game is not perfectly balanced and that chances are not 50:50 but, for example, 60:40 for heads, how much would you bet now?

What is, from scientific perspective, the optimal amount of money you should bet with this level of advantage? The solution to this problem was proposed by John Larry Kelly, an American physicist.

- **Edge**

Everyone places a bet because they believe they are in a better position than the others. For example, they are more knowledgeable, have access to secret technology, or possess private information. Therefore, that kind of a gambler who has additional information and bet on something that has a higher probability than the prescribed odds, has an edge.

- **Bookmakers**

The odds are determined by bookmakers, by adjusting them in their favor. Typically, bookies serve as market makers and benefit from the event regardless of the outcome due to the various positions that you can bet on the same event.

Chapter 3

Information-theoretic gambling strategies and the Kelly criterion

For gamblers, the main goal is to make a profit by betting. In mathematical terms, to find bets with **positive expectations**, i.e. that at the end of the game, the outcome is positive. Also an important task for gamblers is to dispose of money in the right way, and to decide *how much to bet*.

This has been of interest since the eighteenth century when the St. Petersburg Paradox was introduced by Daniel Bernoulli and led to a reevaluation of the financial decisions that were formerly thought to be wise. Fresh perspectives on money, business, and the daily choices we make have been opened up by this paradox. More about St. Petersburg paradox in: [2]. Since then, different approaches to the valuation of money have been encountered, but we will focus on application of utility function. In the case of wealth, utility function increases as the wealth increases.

We would think of relationship between money and utility as linear, but actually it is not. Utility can be seen as happiness and if you have zero dollars and someone gives you 100$, you will be happier than if you have million dollars and someone gives you 100$. The more money you have the less that extra little bit of money is going to count. As we can see in Figure 3, it slowly flattens out. The initial spike and leveling off is typically represented by a log function. Another example of utility function is exponential function ($U(x) = x^b, 0 \leq b \leq \infty$).

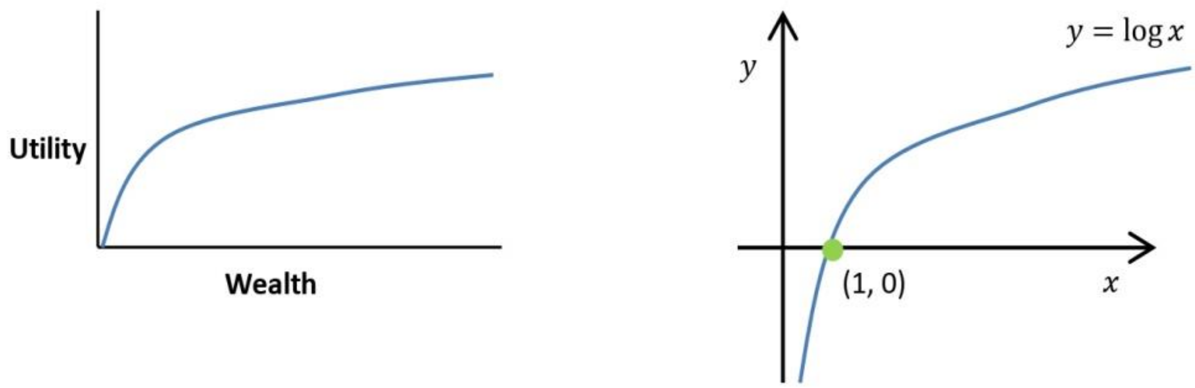

Figure 3 - Utility and logarithmic functions

This was an introduction to the Kelly's criterion, since it is frequently explained through the use of logarithmic function. It is suggested that assuming long-term goals, the logarithmic utility function is the only reasonable choice for a utility function. What we will do is actually maximization of logarithmic utility function, or maximization of the expected value of the logarithm of the random variable. In our case, random variable will be capital at the disposal of the gambler.

3.1 Kelly criterion

Let us look at some realistic and some a little less realistic situations that examine the fundamental characteristics of a communication system. The chosen scenario involves a situation in which a gambler leverages information about the received symbols from a communication channel to place advantageous wagers on the transmitted symbols.

Imagine a communication channel used to share the results of an unpredictable event before it becomes widely known. That gives the gambler an advantage as he can still gamble at the original odds. As we said earlier, the key factor that determines the amount of money a gambler could profit from is the size of his bet. Given this advantage of knowing the game results, the question arises: **How much should the gambler bet to maximize his profit, or what is the maximum he can bet without going bankrupt?**

The scenario essentially discusses an event in which an individual can profitably place bets if they have advance knowledge of game results that is communicated through a trustworthy channel. Finding the right amount to gamble for optimum profit is the difficult part.

The solution to this problem was suggested by John Larry Kelly [12]. He discovered that in order to make the most of the information that gives you an advantage, you need to bet a portion of your money according to a certain formula. There are more approaches depending on the level of advantage the gambler has.

First, we will consider the case when the gambler has a private wire.

3.1.1 The gambler who possesses private information

3.1.1.1 Noiseless binary channel

Let us start by investigating a case where we have *noiseless binary channel*. An example where this kind of channel might be used is to transmit baseball game results between teams that are evenly matched. In such a scenario, the gambler would wager the maximum amount possible, as they are confident in doubling their money due to the reliable, noiseless nature of the channel. Therefore, the gambler's money would multiply rapidly, reaching 2^n times the initial amount after n bets, showcasing exponential growth of his capital, i.e.

$$C_n = 2^n C_0, \tag{3.1}$$

where:

C_0 – Initial capital

C_n - The capital after n trials (bets)

For consistency, this convention will be applied throughout the whole work.

The situation described above is far from realistic, and is only possible in extreme situations that are mostly illegal, but it is easiest to present as a starting point.

3.1.1.2 Noisy binary channel

The more common situation is the gambler who has a *noisy channel*. In this scenario, the outcome is conveyed accurately through the channel with a probability p, and inaccurately with a probability q. And again, the gambler retains the option to wager his entire capital. But, due to the nature of the noisy channel, if he continues to wager with entire capital each time without a fixed limit, the probability that he loses everything tends to 1.

Optimal strategy implies that for such conditions it is wiser to bet only a **fraction** f of the initial capital. The bets are made with even odds, starting with an initial wealth C_o. The wealth after n trials, utilizing a betting fraction f of the initial capital (expressed in %), is expressed as follows:

$$C_n = C_0(1+f)^A(1-f)^B \quad 3.2$$

where A is number of wins (Achievements), and B number of losses (Breakdowns) in n bets.

We are interested in understanding how the gambler's capital grows over time. If the capital is growing exponentially, it means it is increasing at a certain percentage rate with each bet.

<u>Definition 6</u>: The exponential rate of growth of a gambler's capital, denoted by

$$Gr = \lim_{n\to\infty} \frac{1}{n} \log\left(\frac{C_n}{C_0}\right) \quad 3.3$$

represents the average compounded growth per bet as the number of bets approaches infinity.

Formula Breakdown:

- Practically, this is ratio which represents how much the gambler's capital has changed over all the bets, i.e. the final amount compared to the initial amount
- Limit is used because we want to see what happens to the average growth rate as we consider more and more bets. We want to see how the growth behaves in the long run
- The formula uses the logarithm to express the average growth rate in a more interpretable manner, capturing the long-term trend in the capital growth over an infinite number of trials

The exponential rate of growth can be restated according to 3.2 and 3.3

$$Gr_n(f) = \frac{1}{n}\log\left(\frac{C_n}{C_0}\right) = \frac{1}{n}\log\left[(1+f)^A(1-f)^B\right] = \frac{A}{n}\log(1+f) + \frac{B}{n}\log(1-f) \quad 3.4$$

To optimize the gambler's capital after N bets, it is imperative to maximize growth rate, but instead of maximizing Gr, we will actually maximize the expected value of Gr, denoted by gr

$$gr(f) = E\big(Gr_n(f)\big) = p\log(1+f) + q\log(1-f), \quad 3.5$$

using the fact $E(A) = np$ is expected value of success in n trials with success probability p, and $E(B) = nq$ is expected value of failure in n trials with failure probability q.

In probability theory, an expected value tells us what outcomes to expect in the long run, using the probability of each outcome.

We now turn to the maximization process and firstly, calculate the first derivative of gr which will indicate a potential maximal value. For the sake of simplicity, the natural logarithm will be used.

$$gr'(f) = \frac{p}{1+f} - \frac{q}{1-f} = \frac{p(1-f) - q(1+f)}{(1+f)(1-f)} = \frac{p - pf - q - qf}{(1+f)(1-f)} = \frac{p - q - f(p+q)}{(1+f)(1-f)} = \frac{p - q - f}{(1+f)(1-f)} \quad 3.6$$

where $gr'(f) = 0$ when $f = f^* = p - q$, and $p \geq q > 0$. Then, the second derivative

$$gr''(f) = \left(\frac{p}{1+f} - \frac{q}{1-f}\right)' = -\left(\frac{p}{(1+f)^2}\right) - \left(\frac{q}{(1-f)^2}\right) < 0 \quad 3.7$$

Concavity of function $gr(f)$ implies that there is a unique maximum at f^*, where

$$gr(f^*) = p\log(1 + p - q) + q\log(1 - p + q) = \cdots = p\log p + q\log q + \log 2 > 0 \quad 3.8$$

$$gr(f^*) = p\log p + q\log q + 1 = 1 - H(p), \quad 3.9$$

where $H(p)$ is defined by formula 2.2.

There are many advantages of maximizing gr, but the one that will be formally defined in a simple way is one which indicates the efficacy of utilizing the Kelly method of maximization expected value of capital by choosing f^*on each trial.

Proposition 1: For a binomial random variables, where probability of success is p, and probability of failure is q, expected value of the final capital is maximized by fixed fraction f^*, where $f^* = p - q$.

We will present this with a practical example in order to plastically explain the story so far. We can look at this scenario through the example of tossing a coin with biased distribution.

Example 1: Imagine a coin toss game with a biased coin, where the payoff is even money, player A enters the game with an initial capital C_0 that we assume to be infinitely divisible, and the probability that player A wins is $p = 0.6$. In order to find the optimal fraction on which to invest so that the desired outcome in the end would be as large as possible and grow as fast as possible, applying Proposition 1, $f^* = 0.6 - 0.4 = 0.2$. That means player will bet with 20% of current wealth on each play.

Under those conditions, as we can see in Figure 3, the growth rate in $f^*(= 0.2)$ will be 0.03. That means after n bets, his wealth will tend to $2^{0.03 \times n}$ times his initial capital.

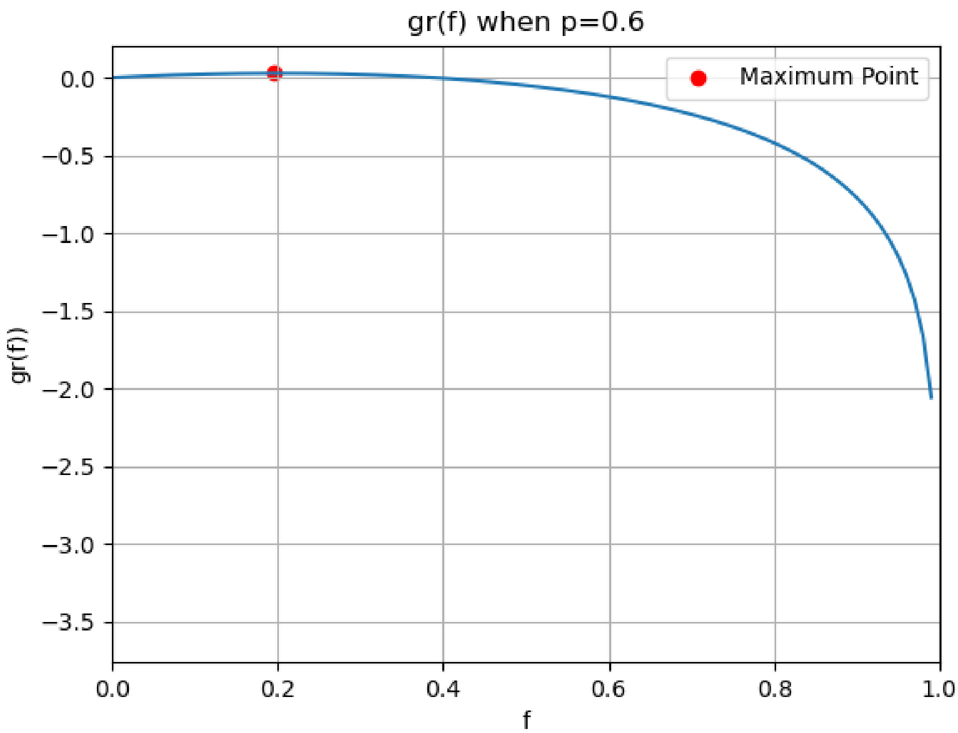

Figure 4 - Logarithm of the wealth growth rate depending on fraction with fixed winning probability

If we want to know how long will it take for the investment to double, we will set the growth rate to be equal to log2, i.e. we will solving equation $0.02\text{n} = \log 2$, which will give us estimation of time (n), $n = \frac{\log 2}{0.02} \approx 35$.

Moreover, as it can be seen in Figure 3, expected growth rate cannot be negative for all $f \geq 0$ and $f \leq f^*$, but for $f > f^*$, it becomes negative. It occurs when the probability of winning is overestimated.

Note: Everything was done with the assumption that capital is infinitely divisible, which is not applicable in the real world.

3.1.1.3 Uneven payoff games

The concept of uneven payoff games means the odds may not accurately reflect the true probabilities. Therefore, the narrative could be extended and it can be assumed the odds in favor of winning is $a: 1$. The game is favorable if $pa - q > 0$. So formula

$$gr(f) = E\big(Gr_n(f)\big) = p\log(1+f) + q\log(1-f) \tag{3.10}$$

is changed to

$$gr(f,a) = E\big(Gr_n(f,a)\big) = p\log(1+af) + q\log(1-f) \tag{3.11}$$

With a simple calculation of maximization process, optimal fraction $f^*$is calculated

$$f^* = \frac{ap-q}{a} \tag{3.12}$$

Until now we focused on analyzing binary cases. Now, we will broaden the scope to more complicated situation, particularly those involving non binary outcomes, such as horse races. It will enable us to investigate the details that lead to a more comprehensive understanding of the dynamics of gambling and information theory.

Chapter 4

Wealth formula for horse race

Now that we have familiarized ourselves with betting strategies and goals using simple examples, we will move on to a slightly more sophisticated game of chance, with rich history and cultural tradition, and which requires certain skills. It is game of chance regarding its essential uncertainty. There are numerous unpredictable factors, some factors are related to horses, but some also concern weather conditions, and other unpredictable events that can affect the result and performance of game. It is considered attractive due to its availableness for serious, but also for casual gamblers. Experienced bettors like to analyze different factors (such as recent performance, historical information, etc.) and to use it to make better decisions which attract a higher realized profit. There are also statistical models and algorithms that can help to predict more accurately. Good connoisseurs of mathematics and statistics, who are also enthusiastic gamblers, have an advantage using effective bankroll management strategies to minimize losses and maximize profits over the long term. Unlike some sports, another attractive fact is frequent races, even more in one day, which gives the possibility of consecutive betting and faster winnings.

4.1 Choosing strategy

Each strategy has its advantages and disadvantages, but we have to decide on something in order to interpret the results. Betting on one horse might lead to a higher potential profit in case that horse wins, but it is also considered risky for many reasons. Some of the reasons are very high uncertainty, great potential for losing wealth entirely, low odds for favorites etc.

On the other hand, there is another approach, widely known as ‘covering the field’. Although this also has its good and bad sides, it distributes risk across multiple outcomes, and despite potential reducing of profit, there is more consistent returns which can be reinvested again. Therefore, moving forward, ‘Dutching’ technique will be considered as principal approach in this thesis, i.e. we are covering all possible outcomes of an event. Furthermore, we will consider flat racing, i.e. horses run on a level track, and our chosen betting option is to bet on the horse to finish first.

4.2 Multi-horse betting strategy and wealth growth in repeated races

We will consider a series of horse races where gambler spreads all of his wealth over k horses. This means that he has to divide his capital and that fraction of the capital that he invests in the horse i will be denoted the same as in the previous section by f_i, with constraint that the sum of all f_i is equal to 1. That ensures that the entire capital is invested. Also, we will assume ***odds*** d_i, where the payout is related with that specific outcome. In simple example, if you bet 2$ on horse i with odds d_i, your payoff is $(2 \times d_i)$$ if horse i wins and you receive 0$ if that victory does not occur. When we apply it to our problem, we can define function $V(X) = f_X d_X$ which represents attribute that will increase gambler's capital if horse X wins, with probability p.

Let us discuss a single race. We will denote gambler's gain after first race with C_1, and $\boldsymbol{C_0}$ initial capital, where X is a winning horse in the first race:

$$C_1 = C_0 V(X) \tag{4.1}$$

If we extend this to multiple races (n races), considering that the gambler reinvests his money, his wealth after n races, including only wins is:

$$C_n = C_0 \prod_{i=1}^{n} V(X_i) \tag{4.2}$$

where X_i is the winning horse in the ith race. Product $\prod_{i=1}^{n} V(X_i)$ is accumulated gains over all n races.

For better clarity and results, using the logarithm function provides more intuitive insights. Consequently, we will define the logarithmic return $r_i = \log V(X_i)$. It is useful firstly because of basic properties of logarithms, which transform product to sum, and it is easier to work with additive terms.

As discussed in the previous chapter, expected value plays an important role, and it is used to estimate the average growth of the gambler's wealth over multiple races. Therefore,

$$E(r) = E(\log V(X_i)) \tag{4.3}$$

is expected value of the logarithmic return.

Then,

$$\log C_n = \log\left(C_0 \prod_{i=1}^{n} V(X_i)\right) = \log C_0 + \sum_{i=1}^{n} \log V(X_i) \qquad 4.4$$

If we take the average of the logarithm of wealth, $\frac{1}{n}\log C_n$, we can use *law of large numbers* which describes the convergence in probability of the sample mean to the population mean as the sample size increases. In the other words, sample mean converges in probability to the expected value. Since $\log C_0$ is a constant term, representing the logarithm of initial wealth, neglecting it does not significantly alter the accuracy of conclusions drawn from the further analysis about the convergence behavior. Therefore, we continue on omitting $\log C_0$, and taking the average of only the second part of the expression from 4.4:

$$\frac{1}{n}\sum_{i=1}^{n} \log V(X_i) \rightarrow E(\log V(X_i)) \qquad 4.5$$

where

$$E(\log V(X_i)) = \sum_{i=1}^{k} p_i \log V(X_i) = \sum_{i=1}^{k} p_i \log(f_i d_i) \qquad 4.6$$

We will denote this expression as $g(f)$.

Therefore,

$$\frac{1}{n}\log C_n \approx g(f) \qquad 4.7$$

$$\Updownarrow$$

$$\log C_n \approx ng(f) \qquad 4.8$$

$$\Updownarrow$$

$$C_n \approx 2^{ng(f)} \qquad 4.9$$

This formula suggests that the gambler’s wealth experiences exponential growth and the rate of this growth is determined by expected value of the logarithmic return. This also indicates that as larger the expected value ($g(f)$), the faster the wealth tends to grow over repeated races.

Although, the central limit theorem remind us that there will be short-term fluctuations that can impact gambler's wealth negatively.

That indicates, in order to optimize wealth, we have to maximize $g(f)$. So, we are maximizing

$$\max_f g(f) = \max_{\sum f_i = 1} \sum_{i=1}^{k} p_i \log(f_i d_i) \tag{4.10}$$

It will be done by solving the optimization problem with constraints using the Lagrange multipliers method with the objective function

$$z = \sum_i p_i \log(f_i\, d_i) \tag{4.11}$$

and constraint

$$\sum_i f_i = 1. \tag{4.12}$$

To simplify, we change the base of the logarithm to e and we set up the Lagrangian function:

$$L = \sum_i p_i\, ln(f_i\, d_i) - \lambda(\sum_i f_i - 1) \tag{4.13}$$

where $\boldsymbol{\lambda}$ is Lagrangian multiplier. Next, we need to find partial derivative of L with respect to f_i and to set it equal to zero:

$$\frac{dL}{df_i} = \frac{p_i}{f_i} + \lambda = 0, \qquad i = 1, .. k \tag{4.14}$$

which implies

$$f_i = -\frac{p_i}{\lambda}. \tag{4.15}$$

Back to the formula 4.13 and replacing f_i , we obtain the following:

$$\sum_i -\frac{p_i}{\lambda} = 1 \quad \rightarrow \quad \lambda = -1, \qquad p_i = f_i \tag{4.16}$$

Therefore,

$$\boldsymbol{f} = \boldsymbol{f}^* = \boldsymbol{p} \tag{4.17}$$

is critical point of function L, and p is probability of winning. Taking second derivative can help us determine whether that critical point is local maximum, so

$$\frac{d^2L}{df^2} = -\frac{p}{f^2} < 0 \tag{4.18}$$

Concavity of L implies that f^* is a maximum, where

$$g(f^*) = \sum_{i=1}^{k} p_i \log(p_i d_i) = \sum_{i=1}^{k} p_i \log p_i + \sum_{i=1}^{k} p_i \log d_i = \sum_{i=1}^{k} p_i \log d_i - H(p) \tag{4.19}$$

where $H(p)$ is entropy.

We can notice that the same strategy holds regardless the odds if we bet entire capital.

We have considered so far the situation when the outcomes of individual races are independent of each other, the probability distribution of outcomes for each race is the same, which indicated to i.i.d. horse races. It certainly makes the analysis more reducible. However, in reality, races may not always meet these assumptions. Also, we did the analysis above under the assumptions that gambler invests entire capital which he can reinvest again.

However, there are cases where the gambler should consider whether he wants to invest the whole wealth. He should make wise choices and must be aware of market changes and betting trends because bookmakers constantly modify their odds. Bookmakers adjust the odds to ensure they make a profit. They build margins into the odds, and the sum of the probabilities is greater than 1 (100%). That ‘overbalance’ is known as ‘vig’. The odds offered by bookmakers may deviate from fair odds. When the odds are fair, the sum of the probabilities is one.

4.2.1 The impact of knowing the probability distribution

We will define by $b_i = \frac{1}{d_i}$ bookmakers probability distribution; this is a way of converting odds into a probability-like measure. For example, if the odds for horse i are $d_i = 3$, then $b_i = \frac{1}{3}$, i.e. horse has a $\frac{1}{3}$ probability of winning.

Whether the $\sum b_i$ is greater or smaller than 1 determines whether the game is fair and whether the strategy is optimal. When sum is smaller than 1, the conditions are more favorable for the bettors because the bookmaker's odds do not cover all possible scenarios. However, the best presenting situation when it comes to the real world, is case when sum is greater than 1, which indicates that bookmaker makes a profit regardless of the actual outcome, and this is the least favorable and most challenging situation for the bettors, but at the same time, an integral part of the betting ecosystem, which cannot be avoided. Achieving perfectly fair odds in real-word is almost unattainable.

In order to make the whole concept more comprehensible, and to remove the complexities, we are going to break away from reality a bit. We will consider the case when $\sum b_i = 1$ which indicates that the game is fair, i.e. no additional margin applied by bookmaker. If we play proportional gambling ($f = p$) then:

$$g(f) = \sum_i p_i \log(f_i d_i) = \sum_i p_i \log\left(\frac{f_i}{b_i}\right) = D(f||b) = D(p||b) > 0 \qquad 4.20$$

Here we come to the term from information theory we introduced at the beginning, which is *relative entropy*. It is clear that gambler who knows the true probability cannot lose. And if and only if $p = b$ (casino's odds are correct) relative entropy will be 0. But in reality, achieving this situation is challenging due to the many factors. $D(p||b) = 0$ iff $p = b$ means that true distribution is equal to bookie's estimate of distribution. That implies the odds offered by the casino are fair. Under such conditions, in the long run, the gambler will neither win nor lose, he will be at zero.

A more realistic situation is when the true probability is unknown for both gambler and bookmaker, so proportional gambling cannot be applied.

$$g(f) = \sum_i p_i \log(f_i d_i) = \sum_i p_i \log\left(\frac{f_i}{b_i}\right) = \sum_i p_i \log\left(\frac{f_i}{p_i}\frac{p_i}{b_i}\right) = \sum_i p_i \log\left(\frac{f_i}{p_i}\right) + \sum_i p_i \log\left(\frac{p_i}{b_i}\right) = \sum_i p_i \log\left(\frac{p_i}{b_i}\right) - \sum_i p_i \log\left(\frac{p_i}{f_i}\right) = D(p||b) - D(p||f) \qquad 4.21$$

Then the gambler will be in the plus if his estimation is better than the bookmaker's.

4.2.1.1 Uniform distribution

If the odds are "t-for-1" for each horse in a race, it implies that the potential profit is consistent across all horses. Each horse has the same $t{:}1$ odds format. For instance, if the first horse has 3-for-1 odds, the second horse, the third, and so on, would also have 3-for-1 odds. Now, again with assumption of perfectly fair market, we set equal chance of winning for each horse. That implies the odds are fair, and distribution is uniform. If we say that odds for each horse is *t*,

$$g(f^*) = \sum_i p_i \log(p_i d_i) = \sum_i p_i \log(p_i t) = \sum_i p_i \log p_i + \sum_i p_i \log t = \log t - H(p) \qquad 4.22$$

As we said above, relative entropy measures how one probability distribution diverges from another. Odds are expressed as t, so implied probability is $\frac{1}{t}$, and true probability is p. Difference between the implied probability based on odds and the true probability is

$$D\left(p||\frac{1}{t}\right) \qquad 4.23$$

From formula 4.23 we got the important claim.

Theorem 2 ***(****Entropy-wealth conservation)*: If odds are uniformly distributed and fair

$$g(f^*) + H(p) = \log t \qquad 4.24$$

We can see the connection with data compression which reflects trade-off between uncertainty and information efficiency. The higher the entropy, the slower wealth grow, and vice versa. Entropy has an impact on wealth; the more unpredictable the outcomes, the less wealthy the gambler. Hence, it is advisable to minimize entropy as much as possible. The question is how to do it, and the answer lies in *conditioning*.

4.2.2 Strategic betting with exclusive knowledge

The conditional probability captures the updated likelihood of a horse winning based on the additional information. As we know entropy decreases by decreasing uncertainty, but also if we have any additional information (*side information*). Various types of information can reduce uncertainty and increase wealth, for example historical data, weather condition etc.

The following notation will be used:

$x \in K$ - symbols representing horses

$s \in S$ - side information

$p(x)$ - win probability

$p(s)$ - probability of side information s occurring

$d(x)$ - odds

$p(x,s)$ - joint probability of x and s

$p(x|s)$ - conditional probability of horse x winning given side information s

$p(s|x)$ conditional probability of side information s given that horse x wins

$f(x|s)$ fraction of gambler's wealth he decides to bet on x after receiving s

$d(x) = \frac{1}{p(x)}$ - meaning the reciprocal of the winning probability gives the odds

We still consider fair odds, and that gambler bets on full capital

$$\sum_{x} f(x|s) = 1 \qquad 4.25$$

We want to maximize g with respect to $f(x|s)$. Using the formula 4.11 we can construct conditional grow rate by:

$$g(X|S) = \sum_{x,s} p(x,s) \log f(x|s) \cdot d(x) \qquad 4.26$$

$$\max g(X|S) = \max_{f(x|s)} \sum_{x,s} p(x,s) \log f(x|s) d(x) \qquad 4.27$$

We are interested in what difference in the growth rate this additional information contributes and we will denote that desired difference by:

$$\Delta(g) = \max_{f(x|s)} g(X|S) - \max g(X) \qquad 4.28$$

Considering proportional gambling, i.e. $f^*(x|s) = p(x|s)$:

$$\Delta(g) = \max_{f(x|s)} \sum_{x,s} p(x,s) \log f(x|s) d(x) - \left(\sum_x p(x) \log d(x) - H(X) \right)$$
$$= \sum_{x,s} p(x,s) \log p(x|s) d(x) - \left(\sum_x p(x) \log d(x) - H(X) \right)$$
$$= -H(X|S)$$
$$+ \sum_{x,s} p(x,s) d(x) - \left(\sum_x p(x) \log d(x) - H(X) \right) = H(X) - H(X|S) \qquad 4.29$$

Using the formula 2.7 we can conclude

$$\Delta(g) = I(X;S) \qquad 4.30$$

So we can see that the difference between maximum exponential growth rate for proportional betting in the presence of side information and maximum growth rate without side information is equal to the mutual information between horse race X and side information S. It can be concluded if the odds are fair, uncertainty is the highest, and profit can be obtained only if gambler has some side information. Understandably, if side information is independent, it does not affect rate of growth.

We said in the opening part of this chapter that experienced gamblers analyze previous races in order to gamble smarter, but this makes sense if those races are truly interdependent and the results of the current race can be correlated with the historical performance of the participating horses.

The next thing we will see is how the dependent races affect the growth rate of wealth, and therefore the knowledge of stochastic processes and entropy rate that we defined in Chapter 2, Section 2.1 will help us a lot.

4.2.3 Dependent bets on horse racing

Making a strategy for future bets is greatly influenced by the past, if these events are somehow connected. It can be betting on sports matches or investing in the stock market. In sports matches as well as in horse races, a team or a horse that often wins, the probability that it will win again is high. In the same manner, stocks that are showing signs of instability are undesirable for investors, because there is a greater chance that they will be less resistant to market uncertainty in the future as well. Managing the risk of uncertainty is the topic of the next chapter. So, it is smart to use knowledge of previous results in our betting strategy for horse race. We assume the horse races form a stochastic process where the sequence $\{X_i\}$ contains race results and the odds are uniform, $r:1$. We are defining bet on the last race as $f(X_n|X_{n-1},\dots,X_1)$. Therefore the growth rate is calculated:

$$\max_f g(f) = \max_f \sum p(X_1,\dots,X_n)\log(\mathrm{f}(\mathrm{X}_n|\mathrm{X}_1,\dots,\mathrm{X}_{n-1})\mathrm{r}) = \log r - H(\mathrm{X}_n|\mathrm{X}_1,\dots,\mathrm{X}_{n-1}), \quad 4.31$$

when $f^* = p$. For a stationary stochastic process, using formula 2.16

$$\max_f g(f) = \log r - H(X_n|X_1,\dots,X_{n-1}) = \log r - H(\chi) \quad 4.32$$

where $H(\chi) = \lim_{n\to\infty} \frac{H(X_1,\dots,X_n)}{n}$. That implies

$$\max_f g(f) + H(\chi) = \log r \quad 4.33$$

We are coming again to conservation statement, but for entropy rate. That indicates for a lower entropy rate, growth rate is larger. In other words, low entropy rate indicates that the outcomes are more deterministic. This is logical since with reduction of uncertainty outcomes become more predictable and certain. That means that initial conditions are well-defined. This predictability can contribute to achieving higher maximum rate.

So far we have seen that the best strategy in terms of long-term betting and reinvestment is to bet proportionally to the probability distribution. Although our intuition may tell us otherwise, we should not deviate from the original strategy. There may be better options if the wagers are independent or if you intend to keep your earnings. For instance, single sport wagers and lottery tickets. You need to be careful and familiarize yourself with the idea because the circumstances in which Kelly should be used rely on the particular qualities.

One can continue in this narrative and generalize the previous strategy to the stock market and portfolio theory. Log-optimal portfolios, connection with side information, stationary markets, universal portfolios etc. can be found in [7].

"It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so. "

Mark Twain

Chapter 5

Measure of entropy in portfolio theory

In this chapter, we dig further into the progress and shift from gambling's unpredictable results to investment management's strategic approach. As we have seen in previous chapters, randomness is inherent in gambling activities. On the other hand, portfolio theory provides a more structured framework for optimizing risks and returns on the stock market. Thus, this chapter will logically be suitable for a more methodical investigation of risk management and financial decision-making.

Although both the casino and the stock exchange originated and developed in the 17th century, in a similar period, their purpose is certainly different. In contrast to casinos which have traditionally been primarily for games of chance, enjoyment and entertainment, the stock market is used to purchase and sell financial bonds that allow businesses to generate funds and investors to trade in ownership. The key difference is that stock market is legitimate form of investment, contributing to economic growth, which does not apply to gambling.

Regardless of the fact that they have different goals, gambling and stock market share common features such as risk and uncertainty. Another common thing is investing money that can potentially lead to profit or loss. As expected, speculation is involved in both cases.

Although we have seen that there is an optimization solution for classic gambling as well, with certain conditions that may conflict with reality, but despite all, luck is still the key factor, and we will see that in stock market there are many more different factors which affect the outcome.

Why is the stock market appealing is a frequent question and we will answer it with some basic facts.

- Building wealth over time. When an investor has shares in a company, along with the company's growth, the investor's capital also grows.
- When investor buys stocks, he owns a share of a company.
- Anyone can easily access the stock market through online platforms.

In addition to the fact that everyone can follow the stock market, they must be aware that every investment carries with it risks, and what is the basic and main goal of everyone who invests is to minimize the risk and maximize the profit. Therefore, before making an investment decision, everyone who finds investing intriguing and wants to give it a try should conduct study or consult an expert.

Conducting a study is exactly what we will do, based on the principles of information theory and related concepts. But before we scratch the surface, let us familiarize ourselves with the basic factors and concepts needed to build and manage a portfolio. There are many definitions of a portfolio, but in the simplest way it can be defined as a collection of financial assets. Those assets may belong to an individual, but also to investment institution or firms. Financial assets can include stocks, bonds, funds, even real estate. Recent years have seen a significant amount of research on the advantages of incorporating unconventional assets into conventional stock and bond portfolios. *Bitcoin* as an alternative investment option has attracted enormous attention from the media and investment communities. More about this topic is in [26]. Stocks or shares in a company are the most typical part of a portfolio.

In creating a portfolio, it is important to identify the objective, risk tolerance, choose the asset allocation, but perhaps most important is to *diversify* the investments, in order to spread risk. (It is less hazardous to own a portfolio of assets from diverse classes than it is to maintain a portfolio of assets that are similar). Having assets in your portfolio that are not exactly positively connected is essential to diversification.

5.1 Modern portfolio theory

The desire to reduce risk for a given level of expected return or maximize expected return for a given level of risk was framed mathematically in 1952 by Harry Markowitz [19]. To put it differently, it is a theory that addresses how to build a diversified investment portfolio while striking a balance between risk and return. It is called modern portfolio theory (MPT).

Key concepts of MPT:

- Return – financial gain or loss made on an investment over some period of time expressed in percentages

$$stock\ return = \left(\frac{ending\ stock\ price - beginning\ stock\ price + dividends}{begging\ stock\ price}\right) \times 100$$

- Risk – probability that an investment will lose money or deviate from expected outcome

- Risk-Return tradeoff - Investments with a high degree of risk are more likely to experience higher returns (*higher risk - higher reward*)
- Diversifiable Risk – risk that can be reduced through diversification
- Correlation – measure how two assets move in relation to each other
- The risk-free rate - return on an investment with zero risk of financial loss
- The efficient frontier – the best combination of investments to maximize returns for minimal risk

5.1.1 The efficient frontier – graphical representation

Risk can be represented as standard deviation. To present efficient frontiers, we need to understand the relationship between risk and *standard deviation*. Standard deviation is a common statistical measure that quantifies the amount of variation or dispersion of a set of values. In the context of finance and portfolios, standard deviation is used to quantify the risk of a certain investment. The standard deviation indicates how much the actual return is expected to deviate from the average expected return. Generally, a higher the standard deviation is associated with higher risk, and vice versa.

Expected return is actually the *mean* return.

One of the most significant graphs in the field of finance is the curve represented in Figure 5.

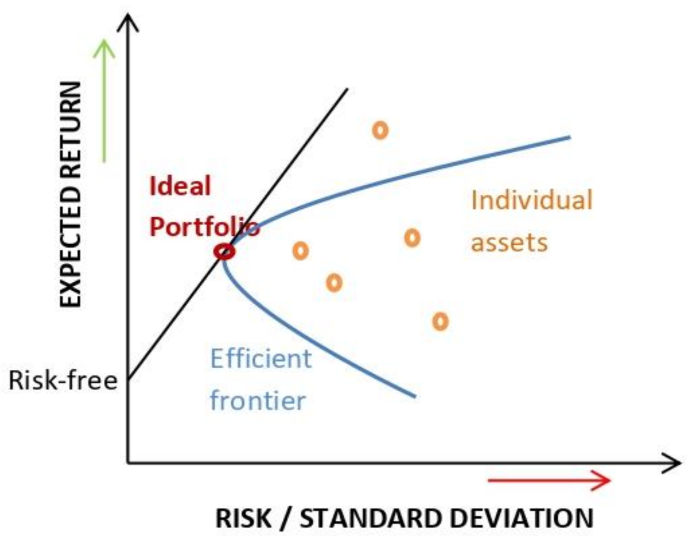

Figure 5 - ***Efficient frontier curve***

If the market is efficient, this curve exists. An explanation of the curve follows:

- A portfolio that lies on the efficient frontier is superior to the one below. For a given risk, we are aiming for the best possible return (risk and return are efficiently traded). If we look at portfolios that are on the efficient frontier, we cannot say what the best portfolio allocation is, because it depends on the level of risk tolerance of the investor. If we have two portfolios that have the highest return for a given risk, we cannot say that one is better than the other. Investors with a higher risk tolerance will prefer one with a higher risk, while more conservative investors will look for less risky investments.
- A portfolio that lies above the curve means that the portfolio is considered underestimated (for the same level of risk, the expected return is higher). If one invests in the portfolio of that kind, its value will rise and adjust to fair value.
- A portfolio that lies below the curve is considered overestimated because the risk is greater for a lower return that the investor can get. Therefore, if one invests in those portfolios, losses are expected.

Portfolios that are outside the curve lead to opportunity and to another concept in finance, which is *arbitrage*. This happens due to temporary market inefficiencies. The investor concludes whether the price is fair or not fair, and accordingly decides whether to buy or sell something. The idea that lies behind it is ‘buy low, sell high’. That's why undervalued and overvalued portfolios mean to us, that overvalued stocks are sold and undervalued stocks are bought. It is important to know that such arbitrage opportunities do not last long, due to active tracking and monitoring.

In addition to all this, the fact is that risk cannot be completely avoided, due to the constantly changing market. Therefore, besides knowing the terminology and mastering the subject matter, it is crucial to stay informed about economic trends, company performance and global events that may affect the financial market.

5.1.2 Mean-variance approach

Now that we have familiarized ourselves with the basic concepts and principles used in finance, or more specifically in portfolio theory, we have made a nice introduction for the approach which marks the beginning of modern portfolio theory, and it is exactly the mean-variance approach. This theory continues to be the standard, or baseline methodology used by scientists, researchers, and all professionals in the field of finance.

The traditional answer to the basic question that interests every investor is how to find the optimal way to invest in a certain set of assets. The technical name for strategies that solve this problem is Portfolio selection model, same as the title of the first paper by Markowitz on the subject [19].

Before Markowitz's work, portfolio selection was primarily based on the return generated by an investment. Risk was not given much attention and Markowitz considered it mistake. Therefore, he raised risk to the same level of importance as return. And so the idea of portfolio risk emerged. Variance has become an accepted measure of risk. Markowitz was the first to show how the variance can be reduced by the influence of diversification.

Mathematically, portfolio can be presented by vector of weights $v = (v_1, v_2, \dots, v_n)$, where v is containing of investor's allocation of wealth where

$$\sum_{i=1}^{n} v_i = v' 1_n = 1. \tag{5.1}$$

As we said earlier, the sample variance represents a measure of risk, and a sample mean a measure of return. Weights refer to the proportion (in %) of investment allocated to each asset within a portfolio, i.e. weights indicated how much of your total investment is placed in each particular asset. Therefore, MV (mean-variance) problem seeks to choose those weights where the variance is the lowest of portfolio returns, while aiming for a certain target.

We define vector of the returns of individual assets in the portfolio as $R = (R_1, \dots, R_n)$, covariance matrix $var(R) = C$ and expected (average) return $m = E(R)$.

Now we can formulate the objective function of MV problem as follows:

$$\min_{v} v' C v \tag{5.2}$$

subject to the constraints:

$E(v'R) = v'$

$m = \tau_0$, where τ_0 is target return

$v' 1_n = 1$, investment is fully allocated

We can also go the other way by maximizing expected portfolio return ($v'm$) instead minimizing the portfolio variance.

Understanding assumption about investor behavior and financial markets is very important for using this model in a proper way, because model is built upon certain expectation. Deviations from these assumptions in a real world can impact the reliability. A couple of example assumptions will be listed below:

1. Investors can make probabilistic assessments of the future returns of assets using for example historical data, models etc., and those estimated probability distributions is a key input in mean-variance analysis. The accuracy of these assumptions depends on the availability and quality of the data, as well as the investor's ability to use the given data in the best way and predict the future.
2. Investors measure their satisfaction represented by a utility function and they aim to maximize it, making decision based on their financial state over a specific period of time, they are not looking a long-term perspective.
3. Investors focus only on the two first moments of a probability distribution: expected value and variance.

This type of analysis provides a useful framework and fundament od modern portfolio theory, but makes certain assumptions that may not hold in the real world, such as a normal distribution, and observing only the first two moments. The problem is that financial returns are not typically normal. They mostly have negative symmetry, skewness, kurtosis, etc.

The success of a portfolio depends in part on the correct assessment of risk. However, it happens that the risk is correctly assessed on the basis of historical data, but the problem is still not solved because no attention is paid to the uncertainty of data, because MV is often concentrated only on a few assets. Therefore, the optimal portfolio may be less diversified. Some even suggest that the standard deviation cannot perfectly illustrate the risk.

5.2 Entropy Approach

Having established a foundation with methodology in the previous section, we are now entering the pivotal point of this chapter. That is how information theory relates to this topic.

Although there are still various new applications of the mean-variance method and it is still actively used for various researches, we will devote our attention to research from a newer and different angle based on information theory which will try to overcome the problems that existed in the traditional approach. 20 years had passed since Markowitz method, and in 1972, entropy was introduced as a measure of risk to replace variance. A model was considered to maximize expected return and minimize entropy. In the last section, we saw variance as a measure of risk, which we will replace with entropy in this section. Although used as measures of risk, these two terms are fundamentally different when it comes to interpretation. Special case is when random variable X is Gaussian, entropy and variance become connected:

$$H(X) = \frac{1}{2}\log(2\pi eVar(X)) \quad (5.3)$$

Another type of comparison is given in the paper [22], which compares the efficient frontier of mean-variance and mean-entropy. This is another reason why entropy should be considered, and mean-entropy model will be given below. The first thing that is exemplified in the picture is the difference in shape. The main difference is that entropy balances risk and return better.

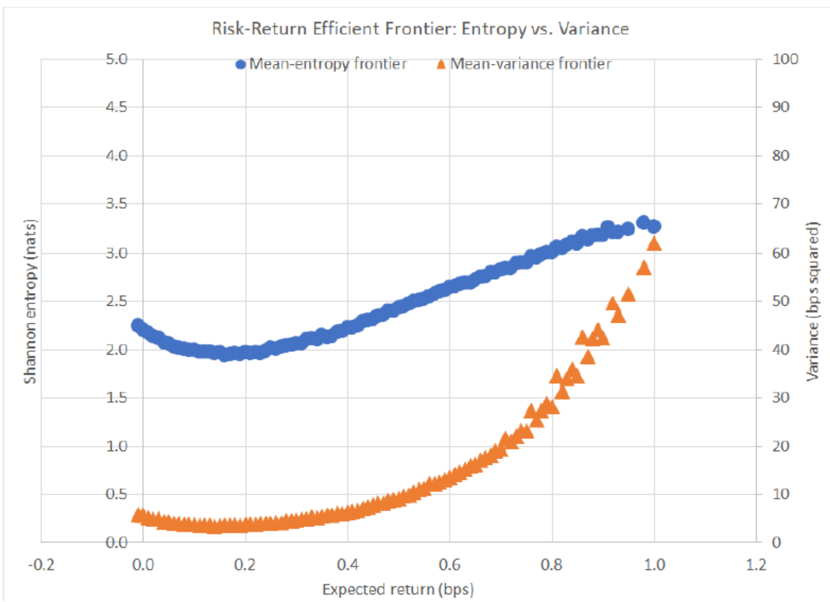

Figure 6 – Difference between entropy and variance. Soruce: [22]

In order to overcome the problems and limitations we had in the traditional model, we will employ a variety of different models and see if they show better performance when we are not limited to the normal distribution.

5.2.1 Single objective model - ME (Mean-Entropy) model

When the transition from MV to ME happened, an index-based framework for construction a portfolio was described. The portfolio's entropy is computed based on the relationship between individual asset returns and a market index return.

Market index return R_M – an indicator of how well a set of assets performed in a specific market measured in percentage over a given period of time

Asset's return $(R_1, R_2, \dots, R_n)$ – depends on R_M

We can describe this dependence by conditional entropy:

$$H(R_i|R_M) \tag{5.4}$$

In the context of portfolio, conditional entropy measures the uncertainty of individual returns taking into consideration the impact of the market.

Another notation is *portfolio weights* (v_i) – proportion of the portfolio allocated to each asset.

e.g. v_1 – proportion of the portfolio allocated to asset 1.

The objective function combines these weights and conditional entropies in the following way:

$$\min_{v} \sum_{i=1}^{n} v_i^2 H(R_i|R_M) \tag{5.5}$$

with respect to the constraints:

$E(v'R) = \tau_0$ – expected return of the portfolio is equal to a specific target return

$v'1_n = 1$ – sum of all portfolio weights is equal 1 (no cash left unallocated)

What is interesting and innovative about this approach is observing not only the individual asset but also its connection with the wider market, allowing for a more sophisticated and potentially more realistic representation.

5.2.2 Single objective versus multi-objective portfolios

Although ME is more interesting and brings something new by observing conditional entropy, it is still a single objective problem – which has the task of optimizing only one goal or objective. Its benefit is that it is simple and easy to understand, but it probably won't be able to adequately represent the complexity of the real-world. Most of the real-life problems, including portfolio optimization, have multiple conflicting objectives. The basic and at the same time the most important example is profit maximization with minimal risk. Due to this problem, and problems similar to it, multi-objective optimization is used, which considers several objectives at the same time. Decision-making in the real world is often reduced to an ideal balancing act between multiple goals. Therefore, unless there is a cause to assume that higher moments have no impact on the investor's decision, higher moments must not be disregarded.

5.2.3 Portfolios with multiple objectives

5.2.3.1 Mean-variance-entropy (MVE) portfolio

As we said earlier, diversification is key goal of asset allocation since it lowers the unsystematic risk. Furthermore, when diversification rises, variance falls. By combining the variance of assets with the entropy of weights, it is created diversified portfolio know as mean-variance-entropy (MVE) portfolio.

In this portfolio, entropy is not used as a risk indicator, instead, it is incorporated to produce weights that are nearly uniformly distributed. This model can balance the risk and diversity of the portfolio with the right selection of the momentum factor.

Model is defined as follows:

$$\min_{v} v'Cv + \mu \sum_{i=1}^{n} v_i \log v_i \tag{5.6}$$

Adjusting the parameter μ allows to control the importance if entropy in the overall objective.

If we take $\mu = 0$, we get MV model.

5.2.3.2 Mean-variance-skewness-entropy (MVSE) portfolio

Skewness describes the asymmetry in a probability distribution, as it is presented in Figure 7.

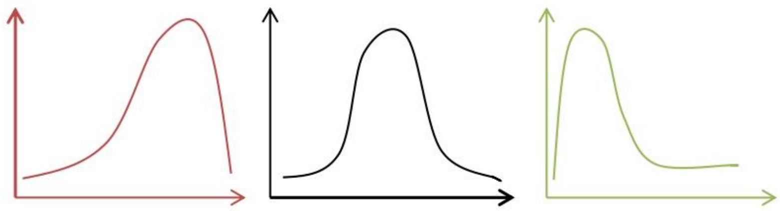

Figure 7 - Types of skewness

Positively skewed – higher probability of large gains than large losses

Negatively skewed – higher probability of large losses

It is stated that investors should incorporate the skewness of the return distribution as a part of decision-making process. The reason is investors' awareness that there is a chance for asymmetric returns and they have the ability to recognize the impact of extreme events. By making investors better provided, it can potentially lead them to higher returns.

Skewness can be added to mean-variance model, and if, on the top of that, we add entropy, we get a well-diversified portfolio – MVSE. This is a model where the aim is to maximize the skewness and entropy of weights, while minimizing the variance, presented as follows:

$$\begin{cases} \min v'Cv \\ \max v'S(v \times v) \\ \max -v' \ln v \end{cases} \quad (5.7)$$

Here we have introduced various measures that improve portfolio performance. Transaction costs (turnover) have also been shown to be lowest when using this model.

5.2.4 Diversification regarding the terms of entropy

As we were introduced at the beginning of the chapter, Markowitz defined the concept of portfolio diversification. In his model, when number of asset increases, the variability of expected return decreases, as the specific risk within the portfolio. We can see that in Figure 8.

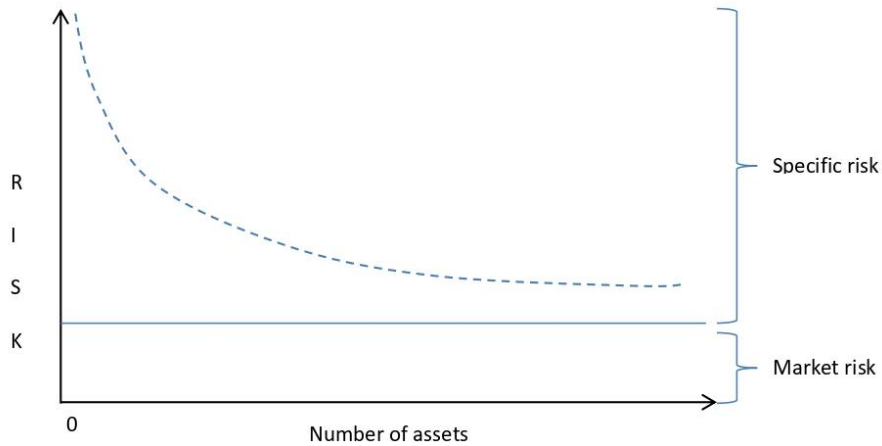

Figure 8 - Market and specific risk

As we can see, we divided total risk into specific and market risk. Specific risk can be diversified, because it is a part of risk which is specific to the firm, while market risk is connected to the whole economy. The risk that comes from specific sources is more uncertain, has limited possibilities of risk protection and often unpredictable factors specific to each company. Nevertheless, investors must consider both types of risks to make right decisions and manage overall portfolio risk.

The reason we started the topic of diversification again lies in the interesting fact. The division of the total risk described above can be represented with *entropy*. More precisely, by the chain rule of entropy.

We will observe:

- A – an asset
- M – market index
- $H(A)$ –Risk associated with the asset X
- $I(A;M)$ – mutual information between A and M which measures shared information, i.e. shared risk, so we can see it as market risk.

- $H(A|M)$ – specific risk associated with A given the market index M

So, we can decompose $H(A)$ into mutual information $I(A, M)$ and conditional entropy $H(A|M)$, i.e.

$$H(A) = H(A|M) + I(A; M) \tag{5.8}$$

Now, we inspect how entropy reacts to diversification and how sensitive is in respect to this.

The concept of convexity is closely related to diversification in the context of portfolio management and risk analysis because it influences how the risk of a portfolio changes as more assets are added. The convexity of the standard deviation function implies that the effect on total risk becomes smaller as you diversify the portfolio by adding more assets. Convexity contributes to the sensitivity of portfolio risk as it decreases with further diversification. As we aware, the entropy is a concave function, but negative entropy is a convex.

One important property of entropy is subadditivity:

$$H(X,Y) \leq H(X) + H(Y) \tag{5.9}$$

It can be seen there is a reduction in uncertainty (risk) when both variables are consider jointly. This property supports the idea that considering multiple variables together can lead to risk reduction or increased sensitivity to diversification.

Some scientists claim that using entropy as a measure of diversity has some advantages. Entropy is not limited to numeric data and can be applied to various types of information.

5.2.5 Empirical verification

It was empirically confirmed that when variance is used as a measure of risk, diversification can be a risk-reducing factor. In the experiment, an equally weighted portfolio was used.

An equally weighted portfolio is type of investment portfolio where the investor allocates an equal amount of money or an equal percentage of the total portfolio value to each individual asset. Giving each asset the same significance is the aim of an equally weighted portfolio.

Now, using a similar experiment, it can be confirmed how entropy responds to diversity. In experiment, the analysis is based on 15 different stocks from the New York Stock Exchange. The information used is the monthly closing prices of these stocks (last traded price at the end of the day). The data covers period from 2004 to 2013. For each of these stocks, there are 107 monthly observations. Then we started adding assets to the portfolio ranked by the amount of riskiness, starting with the riskiest. The conditional entropy is used to quantify the specific risk. The term ‘normal entropy’ was introduced as the entropy calculated based on normal distribution. The idea is to use normal entropy for comparing portfolios instead of using standard deviations, since it is equivalent to standard deviation, but offers comparison across different datasets.

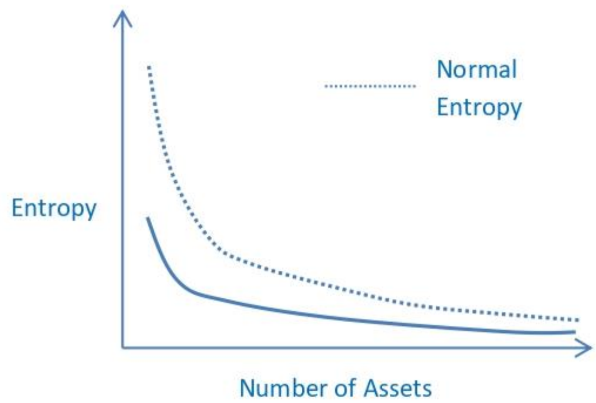

Figure 9 - Comparison of the empirical and normal entropy

A sketch of the results is shown in the image above. We can see that both empirical and normal entropy (standard deviation) decreases as the numbers of assets increases. Consequently, entropy can also be used to quantify diversification because it is highly responsive to the effects of diversification.

Moreover, we can see that normal entropy is greater than empirical. It may indicate that risk is overvalued or predictability is undervalued when assets are normally distributed. This research leads us to the conclusion that variance is less informative than entropy. Furthermore, distribution-free property is advantageous, because doesn’t require knowledge of the exact distribution.

Additionally, this experiment offers empirical support for the entropy subadditivity rule.

$$H[\varphi X + (1-\varphi)Y] \leq H[\varphi X] + H[(1-\varphi)Y] \tag{5.10}$$

where $\varphi = \frac{1}{N}$.

5.2.6 Best performing multi objective portfolio model – MEE model

The first and second moments are typically unsatisfactory and incomplete to explain portfolios in the situation of non-normal return distribution, since it is empirically demonstrated that returns on practically all asset classes and portfolios are not normally distributed. The logical consequence is that performance of a single objective is poorer than of a multi-objective model. Moreover, it has been discussed that if in such models entropy is much more useful to use than variance. In this chapter, we provide an entropy-based multi-objective portfolio model that guarantees a well-diversified portfolio and appropriate utilization of historical risk.

5.2.6.1 Entropy based multi-objective portfolio model

For a well-diversified portfolio model, at least two objectives ought to be accomplished: diversification and risk minimization. That can be formulated as:

$$\begin{cases} \min \sum_{i=1}^{n} {v_i}^2 A_i \\ \max\left(-\sum_{i=1}^{n} v_i \ln v_i\right) \end{cases} \quad 5.11$$

with respect to the constraints

$$\sum_{i=1}^{n} v_i R_i = \tau_o \quad 5.12$$

$$\sum_{i=1}^{n} v_i = 1 \quad 5.13$$

where

A_i- risk associated with the ith asset

R_i – return from the ith asset

τ_o- expected return

v_i – portfolio weight for the ith asset

This can be reconstructed, combining these two objectives into the single one objective as:

$$\min \sum_{i=1}^{n} v_i^2 A_i + \xi \sum_{i=1}^{n} v_i \ln v_i \qquad 5.14$$

where ξ determine the trade-off between risk and diversity.

- For ξ = 0, and risk A_i is changed to corresponding variance estimate, this model becomes single objective, well-known MV model.
- For ξ = 0, and A_i is changed with corresponding entropy estimate, model becomes ME model
- For ξ > 0 and $A_i$is changed to corresponding variance estimate, we got MVE model
- For ξ > 0 and A_i is changed to corresponding entropy estimate, ***MEE*** (mean-entropy-entropy) model is obtained.

In the MEE model, entropy is used as a measure of both risk and diversity.

5.2.7 Measure of performance

The most straightforward way of measuring portfolio performance is by using the Sharpe ratio, which is calculated as ration between expected value of the return and standard deviation of the return. Given how it is calculated, we conclude that it is only applicable for the normal distribution. The use of the Sharpe ratio in situations where return distributions are non-normal might result with false conclusions and unresolved contradictions.

In order to properly assess the performance of the models we mentioned, alternative measures were used. Some of them are: ASR (Adjusted for Skewness Ration), MADR (The Mean Absolute Devitation Ration), etc. More about them and their formulas can be seen in [28].

5.2.8 Application of the model to stock market data

In this section, we will represent the results of comparing the models we discussed above (ME, MEE, MV, MVE) which were applied on four stock markets (Shanghai stock exchange (SEE), Korea exchange (KRX) and New York stock exchange (NYSE)). The observed time period is from 2009 to 2012. Dataset is consisted of 15 stock prices and the market index.

Summary statistics of stock returns of four stock exchanges was calculated, as well as performance of different portfolio models for each stock exchange. Both in-sample and out-sample periods were considered.

In the paper [28] you can see detailed tables with results for each stock exchange, here will be represented one as an example, and later we will be discussed the results and the best performances.

Stock Name	Mean	Variance	Entropy	Skewness	Kurtosis	Shapiro-Wilk normality test	p-value
Hankuk Steel Wire Co Ltd	1.4061	385.6402	1.6333	0.5499	6.2177	0.9184**	0.0000
TaihanFiberoptics Co Ltd	0.1265	324.4033	1.6923	1.0514	6.5448	0.9298**	0.0000
Kocom Co Ltd	0.5341	223.9532	1.7166	0.9036	6.4239	0.9240**	0.0000
ATLASBX COMPANY LIMITED	3.5648	301.7039	1.7425	0.7121	5.3548	0.9513**	0.0010
KB Autosys Co Ltd	1.3184	301.3374	1.7094	1.3196	6.8080	0.9200**	0.0000
Austem Co Ltd	1.6799	189.7367	2.1202	0.3181	2.7821	0.9827	0.2142
AZTECH WB COMPANY LIMITED	1.9220	443.5969	1.6025	0.7829	6.3898	0.9342**	0.0001
Korea Real Estate Investment & Trust Co Ltd	0.9285	186.0678	1.9208	-0.1930	3.7473	0.9839	0.2639
Aurora World Corp	3.2377	262.7860	1.9382	0.5813	3.4991	0.9618**	0.0054
Atec Co Ltd	1.5973	297.7796	1.4837	2.3486	12.3691	0.8003**	0.0000
KODACO Co Ltd	0.1387	310.6263	1.8783	0.5002	4.0770	0.9785	0.1014
Komelon Corp	0.1290	83.1496	1.6696	0.4031	6.5324	0.9456**	0.0004
AnamInformationTechnology Corp	1.0806	480.9879	1.7874	0.2714	4.2859	0.9606**	0.0044
Kona I Co Ltd	2.0014	396.8853	1.9038	0.5366	4.3331	0.9740*	0.0448
AfreecaTV Co Ltd	2.5056	223.4037	1.8194	-0.0066	3.9551	0.9785	0.1009
KOSPI composite Index	0.7559	38.7713		-0.8696	5.6053	0.9566**	0.0023

Table 1 - Summary NYSE

			In-Sample					Out-of-Sample	
	ME	MEE	MV	MVE		ME	MEE	MV	MVE
Return	0.5417	0.5468	0.5016	0.4982		0.5384	0.5415	0.4333	0.4333
Variance	32.5262	32.0207	21.2564	21.2683		32.6766	32.2088	22.0416	22.0416
Entropy	0.1183	0.1189	0.3762	0.3606		0.1187	0.1192	0.3277	0.3277
SR(Var)	0.0950	0.0967	0.1088	0.1080		0.0942	0.0954	0.0923	0.0923
SR(Ent)	1.5751	1.5861	0.8169	0.8286		1.5629	1.5686	0.7570	0.7570
ASR	0.0925	0.0949	0.1078	0.1070		0.0925	0.0937	0.0917	0.0917
MADR	0.1423	0.1383	0.1181	0.1206		0.1423	0.1357	0.0999	0.0999
SSR	0.1241	0.1278	0.1568	0.1554		0.1241	0.1262	0.1322	0.1322
FTR	1.2915	1.2980	1.3216	1.3200		1.2915	1.2934	1.2678	1.2678
Portfolio turnover	0.0061	0.0023	0.1078	0.1017					

Table 2 - Performance of different portfolio models (NYSE)

Note: SR(Var), SR(Ent), ASR, MADR, SSR, FTR are different ratios for measuring performance.

Stock returns often show characteristics that deviate from a normal distribution. Financial markets are known for their complexity, and various factors can contribute to non-normality, asymmetry, and excess kurtosis in stock return distributions.

We will not go through each stock exchange individually, but we will draw a conclusion based on all 4 together, because the results are similar in most cases.

For all four stock exchanges, multi-objective models perform better than singe-objective models. MEE gives the highest portfolio return for both in and out of sample periods. Also, it has the best performance measures. Moreover, MEE yields the lowest portfolio turnover. It can be good or bad, depends on expectations and preferences of investors. Lower portfolio turnover is good for investors who prefer buy-and-hold strategy. This strategy requires fewer transactions, so transaction costs are lower. However, for someone who doesn't like to miss any opportunity, low turnover is not desirable.

In the early literature, entropy is employed separately, for a diversity or as a risk measure. A more comprehensive portfolio model is provided by the multipurpose use of entropy in MEE. MEE can be used to realize the potentially significant investment gain instead of a variety of portfolio investment strategies.

Chapter 6

Application of Rényi entropy in portfolio theory

This chapter is mostly based on [14].

This chapter's objective is to provide a generalization of Shannon's entropy and examine the advantages of using it as a risk measure in portfolio optimization.

We have seen that recently the use of entropy as an investment criterion has become popular. However, its application to portfolio selection is still in its infancy and there is still much to learn about the possible benefits of entropy regarding this problem. As we have seen earlier, non-normality is a challenging problem which we tried to solve by introducing entropy as an additional parameter in the models described above.

In this chapter, however, we address this problem by introducing a new term, i.e. the generalization of Shannon entropy, named Rényi entropy. The goal is to estimate portfolio risk through the amount of randomness carried by its returns. Unlike Shannon entropy, Rényi entropy gives us an additional parameter which controls how uncertainty is measured. We will see how the parameter controls both the center and tail part of distribution. We will also see how minimizing Rényi entropy yields portfolios that outperform minimum variance portfolios.

We will explore theoretical basis of the Rényi entropy, its exponential measure as a risk measure, introduce minimum Rényi entropy portfolios, and support all this with empirical evidence. The main parts will be described, and results will be presented.

6.1 Rényi entropy – definition

Thirteen years after the introduction of Shannon's entropy, Rényi suggested generalization with help of a parameter $\rho \in R^{+}$.

Definition 7: Rényi entropy is defined as:

$$H_{\rho}(X) = \frac{1}{1-\rho}\log\left(\sum_{i=1}^{n} p_i^{\rho}\right) \tag{6.1}$$

where $\rho \in [0, \infty]$, $\rho \neq 1$. In case of continuous random variable, the definition changes slightly to:

$$H_{\rho}(X) = \frac{1}{1-\rho} \ln \int f_X(x) dx \qquad 6.2$$

For $\rho \to 1$, Shannon entropy is recovered, i.e.

$$H_1(X) = \lim_{\rho \to 1} H_{\rho}(X) = H(X) \qquad 6.3$$

6.2 Exponential Rényi entropy and its connection with deviation risk measures

Despite the fact that Rényi entropy has interesting properties, its exponential transformation is considered to have more natural properties when dealing with risk. Hence, we will transform the Rényi entropy defined above into an exponential form.

Definition 8: Exponential Rényi entropy is defined as follows:

$$H'_{\rho}(X) = \exp(H_{\rho}(X) = \left(\int (f_X(x))^{\rho}\, dx\right)^{\frac{1}{1-\rho}} \qquad 6.4$$

Note: for $\rho \to 1$ this expression become exponential Shannon's entropy.

Now we apply this measure to build a portfolio with a minimal risk.

Deviation risk measures belong to the class of risk measures used to quantify deviations of a real portfolio's return as opposed to what is expected. The most common deviation risk measures are Standard Deviation, VaR (Value at Risk) and extensions of both. These measures are crucial for risk management and assessment.

Regarding the properties, a mathematical and conceptual similarity was noticed between the exponential Rényi entropy and deviation risk measures.

Deviation risk measures satisfying the following four properties: positivity, positive-homogeneity, translation-invariance and subadditivity. It has been proven that $H'_{\rho}(X)$ satisfies the same properties. More about properties and proofs are given in a paper [14].

6.2.1 Exponential Rényi entropy and its interpretation with different values of the parameter ρ

In this section we will see how the adjusting parameter ρ affects the central and tail parts of the distribution which might lead to alternative definition of risk, since adjusting ρ changes the way entropy is measured. What is certain is that the parameter ρ contributes to flexibility.

For a demonstration, we will first take ρ to be either 0 or ∞ to see how extreme cases behave.

- When $\rho = 0$ risk is measured with the support of distribution. That means that the emphasis is on the whole range of possible outcomes. Event with the lowest probabilities contribute the most to the risk measure, i.e. rare events (extreme values) have a greater impact.

- When $\rho = \infty$ risk is measured with by the maximal probability, i.e. the risk measure becomes sensitive to the most likely outcomes. Concentrating only on the most likely outcomes indicates that most of the risk is attributed to the events with the highest probability. Rare events or extreme values have little influence on the risk measure when ρ is very large.

From this we can conclude as ρ approaches infinity, the influence of tail events or extreme values is neglected. Since one of the key concerns in portfolio theory is managing the risk associated with extreme events or tail events, we want to avoid large parameter values. The ideal case is to find a balance between the most likely outcomes and extreme events. Finding the right value for the parameter ρ is very important because it can balance the investor's appetite for risk with the nature of the asset's return.

Using a graphic example, we will see how by decreasing ρ, the exponential Renyi entropy becomes more sensitive to events in the tail or extreme outcomes in the distribution. The entropy is evaluated for different values of parameter in the t-Student distribution. The number of degrees of freedom in the t-distribution influences its shape, particularly its tails. This analysis helps to assess how the Exponential Rényi Entropy reacts to changes in the behavior of the tail of the distribution.

Formula used for this example is given in [37].

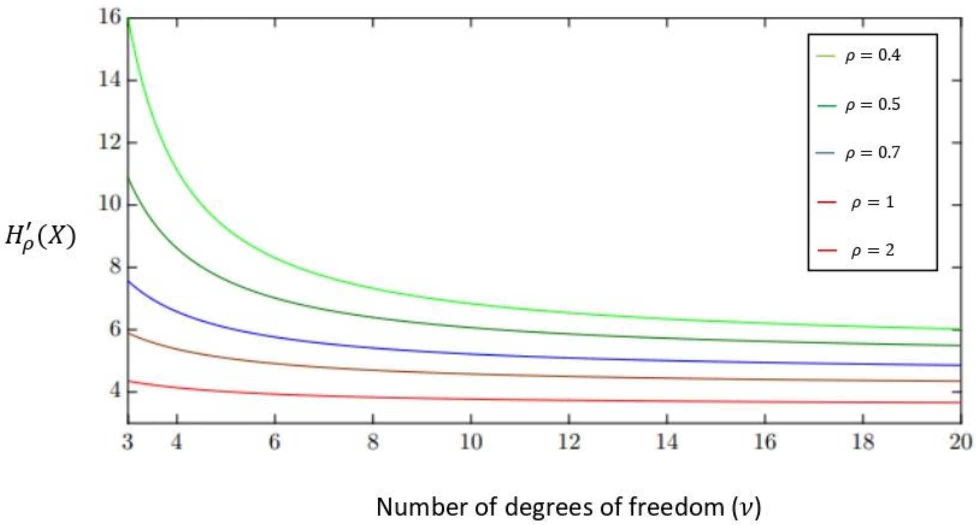

Figure 10 - Sensitivity to tail uncertainty of the exponential Rényi entropy

We can see that for $\rho = 0.4$ the sensitivity to increasing tail uncertainty is the most visible. It implies that a risk measure or investment strategy is becoming more responsive to extreme or rare events in the tails of the probability distribution. More specifically, it was concluded both through empirical results and through various arguments that favorable setting for ρ is between 0 and 1, i.e. $\rho \in [0,1]$.

6.2.2 How Rényi entropy interconnects with portfolio selection

In this section we will use exponential Rényi entropy as an objective function to support investment strategies. It is best proposed to create a minimum risk portfolio that we can call a *minimum Rényi entropy (MRE) portfolio*. In this way, the exponential Rényi entropy of the portfolio return will be minimized.

If R is the portfolio return, $v = (v_1, \dots, v_n)'$ is the vector of portfolio weights, $X = (X_1, \dots, X_n)$ random asset-return vector, we can define portfolio return as

$$R = v'X = \sum_{i=1}^{n} v_i X_i \tag{6.5}$$

Definition 9 (Minimum Rényi entropy portfolio): If we have a set of n assets, the MRE portfolio for a given ρ is defined as following:

$$v_\rho^* = \min_{v \in V} H_\rho'(R), \quad 6.6$$

where V is a set of constraint on v, and $1_n' v = 1$ holds, i.e. 100% is invested.

When someone talks about the "higher-order moments of portfolio return," they are usually referring to moments beyond the second order (variance). Skewness is considered as the third moment) measuring the asymmetry of a probability distribution, as we have shown in the previous section. Kurtosis is considered as the fourth moment measuring the "tailedness" of a distribution. High kurtosis indicates heavy tails, meaning that the distribution has more extreme values (outliers), while low kurtosis indicates lighter tails.

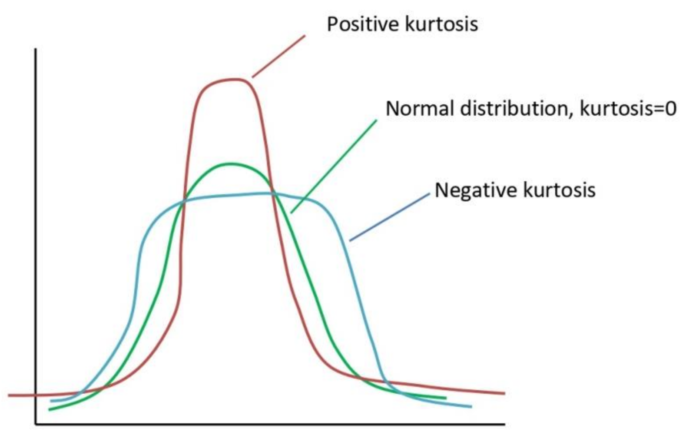

Figure 11 - Types of kurtosis

Regarding the higher-order moments, minimum Rényi entropy portfolio is more suitable than minimum-variance since it aims to incorporate information from higher-order moments into the optimization process. By addressing the risk and uncertainty brought on by the skewness and kurtosis of the portfolio returns, the minimum Rényi entropy portfolio offers a more realistic method of risk management and portfolio creation. To show this, we introduce the Gram-Charlier expansion of Rényi entropy.

The Gram-Charlier expansion is a method applied to approximate probability distributions. It's particularly useful for capturing deviations from a normal distribution, and it provides a more detailed analysis considering higher order moments. Through the next proposition we will see how exactly.

Proposition 2: Let $X \in L^4(\Omega)$ and $\check{X} = \frac{X-E(X)}{\sqrt{Var(X)}}$ standardization for a random variable X which follows a standard normal distribution. (This transformation is commonly used to compare values from different distributions or to identify outliers.)

Skewness (S) is defined as

$$S(X) = E(\check{X}^3) \tag{6.7}$$

Kurtosis (K) is defined as

$$K(X) = E\left(\check{X}^4\right) - 3 \tag{6.8}$$

The standardized form is used to eliminate the scale of the original distribution so that skewness and kurtosis are expressed in a dimensionless manner, making them comparable across different distributions

$$H_\rho^{GCE}(X) = H_\rho\left[\eta\left(0, \sqrt{V(X)}\right)\right] + b_1(\rho)K(X) + b_2(\rho)S(X)^2 + b_3(\rho)K(X)^2 \tag{6.9}$$

with coefficients

$$b_1(\rho) = \frac{1-\rho}{8\rho} \tag{6.10}$$

$$b_2(\rho) = -\frac{3\rho^2-6\rho+5}{24\rho^{3/2}} \tag{6.11}$$

$$b_3(\rho) = -\frac{3\rho^4 - 12\rho^3 + 42\rho^2 - 60\rho + 35}{384\rho^{5/2}} \tag{6.12}$$

For $\rho = 1$, the expansion of Shannon entropy is recovered:

$$H_1^{GCE}(X) = H_1\left[\eta\left(0, \sqrt{V(X)}\right)\right] - \frac{1}{12}S(X)^2 - \frac{1}{48}K(X)^2 \tag{6.13}$$

It can be noted that in this kind of setup, $b_1(\rho) = 0$, therefore Rényi entropy is in noteworthy advantage, since it can control kurtosis.

In the following picture coefficients $b_1(\rho), b_2(\rho), b_3(\rho)$ are presented.

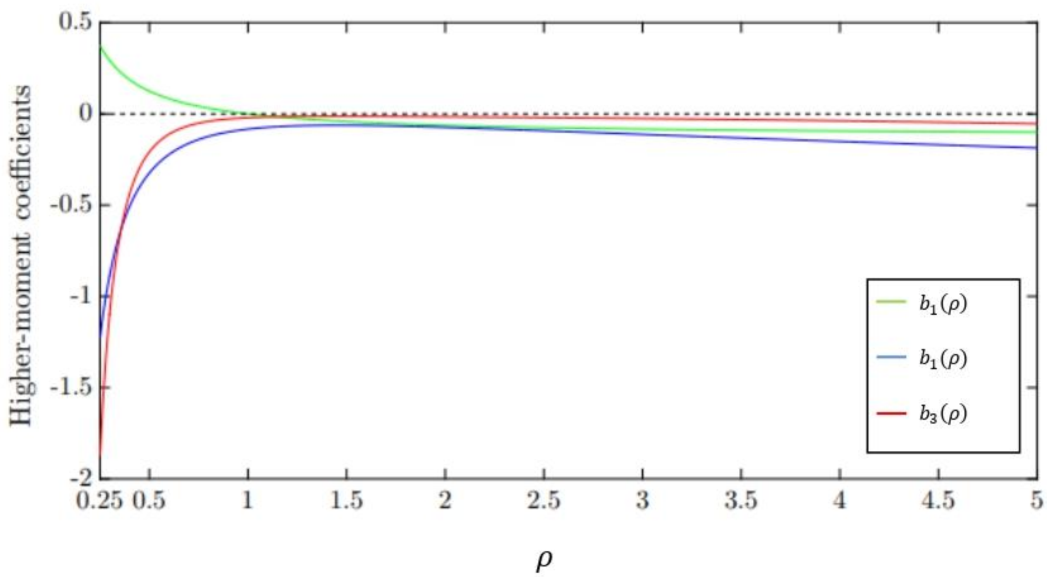

Figure 12 - Behavior of coefficients

$H_\rho\left[\eta\left(0, \sqrt{V(X)}\right)\right] = H_\rho[\eta(0,1)] + \frac{1}{2}\ln V(R)$ provides the explicit connection between MRE and higher-order portfolios

$$v_\rho^* \approx arg\min_{v \in V} \frac{1}{2}\ln V(R) + b_1(\rho)K(R) = b_2(\rho)S(R)^2 + b_3(\rho)K(R)^2 \qquad 6.14$$

What can be observed in the picture is: when f_R is close to a normal distribution, $b_1(\rho)K(R)$ will have the most contributions. It may be noted that when $\rho < 1, b_1(\rho) > 0$ MRE portfolio is similar to minimum-variance-kurtosis portfolio, which we did not mention in this paper, and more about it in [21]. However, when $\rho > 1, b_1(\rho) < 0$ and the effect is reversed.

As we know, traditional models frequently assume a normal distribution of returns. It is known that kurtosis normal distribution has a kurtosis of 3. If it is highly probable that extreme values are less likely, lower kurtosis would be preferred.

As we can see, for $\rho \in [0,1]$, the kurtosis is lower, and with assumption that investor is risk-averse, he certainly prefer lower kurtosis, therefore this setting of ρ is natural, as we mention earlier. Changing ρ changes the minimization of variance and tail uncertainty.

Next, $b_2(\rho)$ and $b_3(\rho)$ are negative for all ρ, therefore $b_2(\rho)S(R)^2$ and $b_3(\rho)K(R)^2$ can be seen as factors that moving the solution away from normal distribution skewness and kurtosis. This is logical given the nature of Shannon's entropy which is maximal for the normal distribution.

6.2.2.1 Robust estimator of exponential Rényi entropy

To avoid making bad assumptions about the data distribution, we introduce a robust estimator since it provides reliable results despite extreme values. Therefore, the objective is to estimate exponential Rényi entropy in a robust way using appropriate estimator. In this work the m-spacing estimator is used. This estimator is appropriate because it observes the differences between successive values in the data in a robust way, which implies that the influence of extreme values is reduced. The use of m-spacing ensures a better observation of the distribution characteristics. Robustness ensures resistance to outlier. Ensuring robustness is crucial for achieving a good out-of-sample performance. It is important that the construction of the portfolio be stable regardless uncertainties in the true return distribution. The detailed properties and implementation of this estimator is covered in paper [14]. Here we will report only the final results for the sake of brevity.

6.2.3 Empirical study – comparison of MV and MRE portfolios

6.2.3.1 Methodology

Now it is time to demonstrate practically the interest of the strategy proposed in this chapter in comparison with several different minimum-variance strategies. Six datasets were used, which are often benchmarks in the literature.

First, MRE portfolios are compared with each other for given different values of the parameter $\rho \in \{0.3, 0.5, 0.7, 1, 1.5, 2\}$. Second, MRE portfolios were compared with five different MV portfolios.

The portfolios are constructed by *dynamic rebalancing* which involves periodically reviewing and adjusting portfolio to respond to the certain changes. The portfolio is reviewed and adjusted every year. Historical data from the past ten years was gathered for making decisions. The total period of analysis is from 1963 to 2016, that is 43 years. We want to see how well portfolios perform over this period of time, using out-of-sample data.

Regarding robust estimator, it is concluded that with increasing m, stability of portfolio weights is improved.

In order to improve stability and performance in real situations, it is good to limit the range of possible solutions. Without limitation, optimized portfolios can have high turnover, which means that the composition of the portfolio changes a lot, which is not desirable when there are a small number of assets. Also, the constraint is introduced based on the explanation that stocks with higher standard deviations are more prone to estimation errors. Consequently, global variance-based constraint (GVBC) established by [15] was introduced. GVBC is a method that imposes more stringent constraints on stocks with higher standard deviations in order to improve the stability and performance of out-of-sample portfolios.

Lastly, for measuring portfolio performance, three criterions are used: well-known The Sharpe Ratio, Adjusted Sharpe Ration (ASR)

$$ASR = SR\left(1 + \frac{S(R)}{3!}SR - \frac{K(R)}{4!}SR^2\right) \tag{6.15}$$

and turnover. Portfolio turnover is a measure of how quickly assets are either bought or sold over a given period of time. High turnover rate will have higher fees to reflect the turnover costs. Higher turnover can introduce more uncertainty and potential risks, but it can also present opportunities for gain. Lower turnover is generally associated with lower transaction costs. [6]

6.2.3.2 Results

Since we have 6 values of paremeter ρ, we comparing performance of 6 MRE portfolios. Regarding SR, the best results give values of parameter between zero and one. Also, the parameter selected in this way enables the reduction in portfolio turnover. If we take the average of all six datasets, $\rho = 2$ provides monthly turnover of 86.7%, and $\rho = 0.3$ provides 32.7%. Moreover, the smaller the alpha and the closer to zero, the more the turnover drops. However, although the advice is to use small values of ρ, for skewness and kurtosis, it may lead to high tail risk. Smaller the ρ leads to decrease of skewness (extreme events are less likely) and increase of excess kurtosis (extreme events are more likely). For $\rho = 0.3$ it is best trade off between risk, return and turnover.

Next, we compare MRE and MV. There are several noteworthy observations that can be made. For instance, for $\rho = 0.3$ MRE shows better performance than all five MV portfolios. MRE outperforms MV for both SR and ASR. But, regarding turnover, MV portfolios are more stable than MRE. This can be expected since MRE consider higher-order moments, which are more influenced by extreme events, leading to potentially higher turnover as the portfolio reacts more strongly to unexpected or extreme market conditions.

The choice between these approaches depends on many factors. But overall, it can be thought that Rényi entropy is a more appropriate risk criterion than variance, more notably for a low values of ρ.

	Sharpe ratio										
	MRE portfolios						*MV portfolios*				
	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 2$	$\hat{\Sigma}$	$\hat{\Sigma}_{CC}$	$\hat{\Sigma}_{SF}$	$\hat{\Sigma}_I$	MP
6BTM	0.844	0.845	0.846	0.841	0.832	0.815	0.835	0.821	0.833	0.836	0.841
25BTM	0.985	0.980	0.973	0.961	0.937	0.906	0.953	0.913	0.951	0.957	0.956
6Mom	0.768	0.763	0.753	0.750	0.716	0.700	0.738	0.741	0.738	0.737	0.731
25Mom	0.936	0.948	0.957	0.960	0.954	0.928	0.902	0.920	0.904	0.903	0.916
10Ind	0.995	1.001	1.014	1.013	0.971	0.936	0.977	0.970	0.983	0.973	0.976
17Ind	0.938	0.947	0.936	0.965	0.905	0.891	0.936	0.924	0.935	0.935	0.918
Average	**0.911**	**0.914**	**0.913**	**0.915**	**0.886**	**0.863**	**0.890**	**0.882**	**0.891**	**0.890**	**0.890**
	Skewness										
	MRE portfolios						*MV portfolios*				
	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 2$	$\hat{\Sigma}$	$\hat{\Sigma}_{CC}$	$\hat{\Sigma}_{SF}$	$\hat{\Sigma}_I$	MP
6BTM	-0.591	-0.601	-0.612	-0.622	-0.643	-0.621	-0.520	-0.497	-0.520	-0.523	-0.590
25BTM	-0.533	-0.538	-0.568	-0.566	-0.609	-0.631	-0.447	-0.308	-0.430	-0.447	-0.532
6Mom	-0.451	-0.448	-0.439	-0.425	-0.397	-0.365	-0.418	-0.415	-0.416	-0.417	-0.403
25Mom	-0.536	-0.521	-0.505	-0.491	-0.436	-0.372	-0.491	-0.452	-0.478	-0.494	-0.513
10Ind	-0.208	-0.226	-0.243	-0.219	-0.212	-0.187	-0.192	-0.173	-0.182	-0.195	-0.176
17Ind	0.037	0.049	0.056	0.040	0.080	0.072	0.118	0.167	0.145	0.117	0.078
Average	**-0.380**	**-0.381**	**-0.385**	**-0.381**	**-0.369**	**-0.351**	**-0.325**	**-0.280**	**-0.314**	**-0.327**	**-0.356**
	Excess kurtosis										
	MRE portfolios						*MV portfolios*				
	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 2$	$\hat{\Sigma}$	$\hat{\Sigma}_{CC}$	$\hat{\Sigma}_{SF}$	$\hat{\Sigma}_I$	MP
6BTM	2.561	2.586	2.612	2.645	2.723	2.623	2.375	2.292	2.378	2.391	2.540
25BTM	2.557	2.617	2.701	2.537	2.445	2.641	2.243	1.907	2.182	2.253	2.571
6Mom	2.120	2.092	2.026	1.986	1.824	1.717	2.048	2.004	2.043	2.050	2.010
25Mom	2.780	2.783	2.763	2.718	2.507	2.313	2.524	2.472	2.507	2.532	2.678
10Ind	1.355	1.292	1.280	1.207	1.101	1.216	1.298	1.251	1.277	1.300	1.232
17Ind	2.614	2.452	2.423	2.107	2.669	2.411	3.102	3.337	3.224	3.083	2.701
Average	**2.331**	**2.304**	**2.301**	**2.200**	**2.212**	**2.153**	**2.265**	**2.210**	**2.269**	**2.268**	**2.289**
	Turnover										
	MRE portfolios						*MV portfolios*				
	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 2$	$\hat{\Sigma}$	$\hat{\Sigma}_{CC}$	$\hat{\Sigma}_{SF}$	$\hat{\Sigma}_I$	MP
6BTM	0.166	0.164	0.164	0.172	0.219	0.250	0.149	0.148	0.148	0.144	0.151
25BTM	0.454	0.497	0.588	0.945	1.324	1.514	0.394	0.389	0.387	0.389	0.433
6Mom	0.148	0.152	0.165	0.178	0.233	0.268	0.152	0.157	0.151	0.151	0.162
25Mom	0.383	0.391	0.484	0.599	0.845	1.165	0.350	0.348	0.342	0.348	0.363
10Ind	0.328	0.324	0.353	0.416	0.567	0.744	0.252	0.246	0.244	0.253	0.295
17Ind	0.481	0.530	0.648	0.790	0.994	1.260	0.381	0.357	0.362	0.381	0.417
Average	**0.327**	**0.343**	**0.400**	**0.517**	**0.697**	**0.867**	**0.280**	**0.274**	**0.272**	**0.277**	**0.303**

Table 3 – Results

Note: α is ρ in our case and $\hat{\Sigma}, \hat{\Sigma}_{cc}, \hat{\Sigma}_{SF}, \hat{\Sigma}_I$ and MP are different types of minimum-variance portfolios we mentioned above.

From those results, it can be concluded that we have achieved reliable out-of-sample performances. Moreover, Rényi entropy has proven to be a powerful alternative to existing risk criteria and outperform traditional MV model, but it also opened the door to other applications and new ideas. For instance, this research gives an idea for a different and potentially more attractive way of implementing entropy, which is to *maximize the entropy of portfolio returns*.

Chapter 7

Conclusion

To conclude, uncertainty is one of the main characteristics of gambling, but also in investing in the stock market. What is certain is that the 50s and 60s of the last century marked enormous achievements in finance and information theory. Shannon, Kelly and Markowitz are proven geniuses of their time, since their fundaments still remain popular and fertile ground for research.

In the first half of the thesis, we got to know the ground breaking theory in financial field and showed the basics of the Kelly criteria which highlight close connections with information theory. We tackled the theory through its application in gambling. We started from the simplest things to the more complex ones, but again not enough as wide range of application. In recent times, the theory that originated in the 60s, has flourished again in 21st century and this is shown by the fact that variations of the Kelly criteria are used by the most famous investors, among who is Warren Buffet [8]. We have seen how entropy and its modifications play a key role, but also stochastic processes and their entropy rate. Generally, we have seen how different notions from information theory are related to different betting strategies. The aim of this thesis was to show how information theory contributes to this topic, but if we want to go further and deeper we could see how experts advise never to use the full Kelly criterion, but modifications or 'fractions'. This problem has been widely elaborated in recent researches as well [33] where different experiment was conducted. Also, one of the greatest applications of Kelly Criterion is by using it for card counting in blackjack, discovered by Edward O. Thorp, where the subject is elaborated in a popular book 'Beat the Dealer', and more about game of blackjack in [9]. It can be said that it was information theory that helped Edward beat the casino. Although it sounds mathematically perfect, this theory still has its flaws and limitations. One of the main problems is the assumption of an ideal scenario that often conflicts with reality. When theory and practice are in agreement it means that the probability distributions are correct and well known. In [16] both positive and negative sides are highlighted. In short term period, gambling is mostly based on luck, making the game difficult to predict. When looking at the long-term goal, the good qualities still dominate, but it must be remembered that the background must be well known and that the criterion is carefully applied, paying attention to the risks.

Although there are various scientific works that have continued the narrative in direction extending Kelly criterion to managing portfolios of investments, we nevertheless took a different path with the desire to see in what other ways information theory can be applied.

The purpose of the second part is to analyze several models to solve some of the challenges faced in the field of portfolio selection, focusing on risk and how to better measure it. We actually focus on upgrade and improvement of a well-known Markowitz mean-variance model. Markowitz received the Nobel Prize for his work, so the problem was well set even at the very beginning of the 70s, it was only supplemented with additional parameters or objective functions, and the technology used to test the model advanced. Everything that exists today is based on what was created 70 years ago. But our aim is to go beyond and rely on information theory and try to improve traditional model and also introduce strategies that accounts for higher moments. Attention was drawn to the model Mean-Entropy-Entropy that showed the best performance and we saw that entropy is a good measure of risk as well as diversification. Entering the world of entropy, we investigated Rényi entropy as a more flexible measure of Shannon's entropy and again showed how this way of approaching risk is better than the traditional one. The traditional model is still in use because of its clarity, mathematical ease, and is based on a lot of researches. Entropy is somewhat more complex to calculate so in this work but also in any other the simplified real complexity of financial markets was used.

The application of information theory to investment, exemplified through the exploration of the Kelly Criterion, portfolio theory, and entropy-based risk measures, underscores the profound impact that harnessing principles of information theory can have in optimizing decision-making processes within the realm of finance, ultimately paving the way for more efficient and effective investment strategies in an increasingly complex and dynamic market environment.

Bibliography

[1] Acharya, J., Orlitsky, A., Suresh, A. T., and Tyagi, H. (2015). “The Complexity of Estimating Rényi Entropy”. *Society for Industrial and Applied Mathematics*, 1855–1869.

[2] Bell, R.M., and Cover, T. M. (1980). “Competitive Optimality of Logarithmic Investment”. *Mathematics of Operations Research*, 161-166.

[3] Bera, A.K., and Park, S.Y. (2008). “Optimal portfolio diversification using the maximum entropy principle”. *Econometric Reviews*, 484-512.

[4] Bernoulli, D. (1954). “Exposition of a New Theory on the Measurement of Risk”. *Econometrica*, 23-36.

[5] Breiman, L. (1961). “Optimal gambling systems for favorable games”. *Berkeley Symp. on Math. Statist. and Prob.*, 65-78.

[6] Chen, J. (2022). “Portfolio Turnover Formula, Meaning, and Taxes”. Retrieved February 22, 2024, from https://www.investopedia.com/terms/p/portfolioturnover.asp.

[7] Cover, T.M, and Thomas, J.A. (2006). *Elements of Information Theory*. Wiley-Interscience.

[8] Forbes. (2024). “Warren Buffett Profile”. Retrieved February 24, 2024, from https://www.forbes.com/profile/warren-buffett/?sh=21190b094639.

[9] Griffin, P. A. (1996). *The Theory of Blackjack*. Huntington Press.

[10] Haghani, V., and Dewey, R. (2016). “Rational Decision-Making Under Uncertainty: Observed Betting Patterns on a Biased Coin”. *SSRN*.

[11] Kang, Y.L., Tian, J.-S., Chen, C., Zhao, G.-Y. , Li, Y.-F., and Wei, Y. (2021). “Entropy based robust portfolio”. *Physica A: Statistical Mechanics and Its Applications*, 583:126260.

[12] Kelly, J. L. (1956). “A New Interpretation of Information Rate”. *The Bell System Technical Journal*, 917-926.

[13] Kempton, C. (2011). *Horse Play - Optimal Wagers and the Kelly Criterion*. University of Washington Math Department.

[14] Lassance, N., and Vrins, F. (2021). “Minimum Rényi entropy portfolios”. *Annals of Opererations Research*, 299:23–46.

[15] Levy, H., and Levy, M. (2014). “The benefits of differential variance-based constraints in portfolio optimization”. *European Journal of Operational Research*, 372-381.

[16] MacLean, L. C., Thorp, E. O., and Ziemba, W. T (2010). "Good and bad properties of the Kelly criterion". *Word Scientific*, 563-572.

[17] MacLean, L. C., Thorp, E. O., and Ziemba, W. T , Zhao, Y. (2011). "How does the Fortune's Formula-Kelly capital growth model perform?" *The Journal of Portfolio*, 96-111.

[18] MacLean, L. C., Thorp, E. O., and Ziemba, W. T. (2011). *The Kelly Capital Growth Investment Criterion: Theory and Practice*. World Scientific.

[19] Markowitz, H. (1952). "Portfolio Selection". *Wiley for the American Finance Association*, 77-91.

[20] Marsh, C. (2013). "Introduction to Continuous Entropy". *Department of Computer Science, Princeton University*.

[21] Martellini, L., and Ziemann, V. (2010). "Improved Estimates of Higher-Order Comoments and Implications for Portfolio Selection". *The Review of Financial Studies*, 1467–1502.

[22] Mercurio, P. J., Wu, Y., and Xie, H. (2020). "An entropy-based approach to portfolio optimization". *Entropy*, 22(3):332.

[23] Levitt, S. D. (2021). "Heads or Tails: The Impact of a coin toss on major life decisions and subsequent happiness". *Review of Economic Studies*, 378-405.

[24] Novais, R.G., Wanke, P., Antunes, J., and Tan Y. (2022). "Portfolio Optimization with a Mean-Entropy-Mutual Information Model". *Entropy*, 24(3):369.

[25] Ormos, M., and Zibriczky, D. (2014). "Entropy-Based Financial Asset Pricing". *Plos one*, 9(12): e115742.

[26] Platanakis, E., and Urquhart, A. (2019). "Should Investors Include Bitcoin in Their Portfolios? A Portfolio Theory Approach". *The British Accounting Review*, 52(4):100837.

[27] Shannon, C. (1948). "A Mathematical Theory of Communication". *The Bell System Technical Journal*, 379-423.

[28] Shirazi, Y. I. (2017). *Entropy in Portfolio Optimization*. Kuala Lumpur: Institute of Mathematical Sciences, University of Malaya.

[29] Song, R., and Chan, Y. (2020). "A New Adaptive Entropy Portfolio Selection Model". *Entropy*, 22(9):951.

[30] Song, Z. (2017). *Kelly Investing with iteratively updated estimates of probability of success*. Ontario, Canada: The University of Guelph.

[31] Thorp, E. O. (2006). “The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market”. *Handbook of Asset and Liability Management*, 385-428.

[32] Thorp, E. O. (2008). “Understanding the Kelly criterion”. *A Mathematician on Wall Street in Wilmott Magazine*.

[33] Uhrin, M., Sourek, G., Hubavcek, O., and Zelezny, F. (2021). “Optimal sports betting strategies in practice: an experimental review”. *Czech Technical University in Prague*.

[34] Wesselhofft, N. (2016). *The Kelly Criterion: Implementation, Simulation and Backtest*. Humboldt-Universitat zu Berlin.

[35] Ziemba, W.T. (2016). “Understanding the Kelly Capital Growth Investment Strategy”. *Investment Strategies*, 49-55.

[36] Ziemba, W.T., and MacLean, L. C. (2011). “Using the Kelly Criterion for Investing”. *Stochastic Optimization Methods in Finance and Energy*, 3-20.

[37] Zografosa, K., and Nadarajah, S. (2005). “Expressions for Rényi and Shannon entropies for multivariate distributions”. *Statistics & Probability Letters*, 71-84.

Biography

Željana Knežević was born on the 21st of October 1998 in Novi Sad, Serbia. In her home town Kikinda, she attended "Đura Jakšić" elementary school and "Dušan Vasiljev" grammar school. In 2017, she enrolled at University of Novi Sad, Faculty of Sciences, Department of Mathematics and Informatics, where she finished her Bachelor studies in Mathematics in 2021, with a GPA 8.48. In the same year, she continued with her Master studies in Applied Mathematics: Data Science at the same faculty and passed all exams in 2023 with a GPA of 8.69.

UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:
RBR

Identifikacioni broj:
IBR

Tip dokumentacije: Monografska dokumentacija
TD

Tip zapisa: Tekstualni štampani materijal
TZ

Vrsta rada: Master rad
VR

Autor: Željana Knežević
AU

Mentor: dr Mladen Kovačević
MN

Naslov rada: Neke primene teorije informacija na investicije
NR

Jezik publikacije: Engleski
JP

Jezik izvoda: Engleski
JI

Zemlja publikovanja: Republika Srbija
ZP

Uže geografsko područje: Vojvodina
UGP

Godina: 2024.
GO

Izdavač: Autorski reprint
IZ

Mesto i adresa: Novi Sad, Prirodno-matematički fakultet, Departman za matematiku i informatiku, Trg Dositeja Obradovića 4
MA

Fizički opis rada: 7 poglavlja, 61 strana, 37 lit. citata, 12 slika, 3 tabele
FO

Naučna oblast: Matematika
NOS

Naučna disciplina: Primenjena matematika
ND

Ključne reči: Proporcionalno klađenje, Keli kriterijum, konjske trke, entropija, teorija portfolija, MEE model, diverzifikacija, upravljanje rizikom, Renjijeva entropija
KR

Univerzalna decimalna klasifikacija:
UDK

Čuva se: Biblioteka Departmana za matematiku i informatiku, Prirodno-matematičkog fakulteta, u Novom Sadu
ČU

Važna napomena:
VN

Izvod: Ovaj rad istražuje primenu teorije informacija u kontekstu investicija. U prvom delu pružamo osnovni pregled teorije informacija i definišemo Kelijev kriterijum kao alat za određivanje optimalne veličine ulaganja radi maksimizacije dugoročnog rasta bogatstva. Analiziramo trke konja kako bismo potvrdili primenljivost ovog koncepta i otkrili zanimljive uvide. Drugi deo obuhvata teoriju portfolija, osnovne pojmove i osvrt na tradicionalni model, koji je podstakao potrebu za alternativnom merom rizika - entropijom, umesto standardne varijanse. Pokazujemo kako uključivanje teorije informacija može poboljšati ove modele, kako u upravljanju rizikom, tako i u optimizaciji diverzifikacije. Ova analiza ističe ključnu ulogu pojmova teorije informacija u optimizaciji investicionih strategija, naročito zbog nedovoljnog sagledavanja složenosti finansijskih rizika u tradicionalnom pristupu. Proučavamo različite modele i identifikujemo MEE model kao najperformantniji. U poslednjem delu, uvodimo Renjijevu entropiju kao generalizaciju Šenonove entropije, koja uključuje dodatni parametar za kontrolu repnih i centralnih delova distribucije. Ponovno, kroz praktični primer, demonstriramo superiornost ovog pristupa u odnosu na tradicionalne modele.
IZ

Datum prihvatanja teme od strane NN veća: 9.2.2024.
DP

Datum odbrane:
DO

Članovi komisije:
KO
Predsednik: dr Dušan Jakovetić, vanredni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu
Mentor: dr Mladen Kovačević, vanredni professor, Fakultet tehničkih nauka, Univerzitet u Novom Sadu
Član: dr Dejan Vukobratović, redovni profesor, Fakultet tehničkih nauka, Univerzitet u Novom Sadu

UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCES
KEY WORDS DOCUMENTATION

Accession number:
ANO

Identification number:
INO

Document type: Monograph type
DT

Type of record: Printed text
TR

Contents code: Master's thesis
CC

Author: Željana Knežević
AU

Mentor: Dr Mladen Kovačević
MN

Title: Some application of information theory to investments
TL

Language of text: English
LT

Language of abstract: English
LA

Country of publication: Republic of Serbia
CP

Locality of publication: Vojvodina
LP

Publication year: 2024.
PY

Publisher: Author's reprint
PU

Publ. place: Novi Sad, Department of Mathematics and Informatics, Faculty of Science and Mathematics, University of Novi Sad, Trg Dositeja Obradovića 4
PP

Physical description: 7 chapters, 61 pages, 37 references, 12 figures, 3 tables
PD

Scientific field: Mathematics
SF

Scientific discipline: Applied mathematics
SD

Key words: Proportional gambling, Kelly criterion, horse racing, entropy, portfolio theory, MEE model, diversification, risk management, Rényi entropy
KW

Universal decimal classification:
UDC

Holding data: The Library of the Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad
HD

Note:
N

Abstract: This paper explores the application of information theory in the context of investments. In the first part, we provide a basic overview of information theory and define the Kelly criterion as a tool for determining the optimal investment size to maximize long-term wealth growth. We analyze horse racing to confirm the applicability of this concept and uncover interesting insights. The second part includes portfolio theory, basic concepts and a review of the traditional model, which prompted the need for an alternative measure of risk - entropy, instead of the standard variance. We show how incorporating information theory can improve these models, both in risk management and diversification optimization. This analysis highlights the key role of information theory concepts in the optimization of investment strategies, especially due to insufficient understanding of the complexity of financial risks in the traditional approach. We study different models and identify the MEE model as the best performer. In the last part, we introduce the Rényi entropy as a generalization of the Shannon entropy, which includes an additional parameter to control the tails and central parts of the distribution. Again, through a practical example, we demonstrate the superiority of this approach compared to traditional models.
AB

Accepted by the Scientific Board on: 9.2.2024.
ASB

Defended on:
DE

Thesis defend board:
DB
Chairperson: Dr Dušan Jakovetić, associate professor, Faculty of Sciences, University of Novi Sad
Mentor: Dr Mladen Kovačević, associate professor, Faculty of Technical Sciences, University of Novi Sad
Member: dr Dejan Vukobratović, full professor, Faculty of Technical Sciences, University of Novi Sad