FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

OPTIMAL PAIN INDEX PORTFOLIOS

Word count: 15,128

Lars Van den Berghe
Student number: 01605502

Supervisor: Prof. dr. Kris Boudt

Master's Dissertation submitted to obtain the degree of:

Master in Business Engineering: Finance

Academic year: 2020-2021

UNIVERSITEIT
GENT

This page intentionally left blank.

Abstract

The pain index is a risk measure defined as the sum of the drawdowns divided by the length of the time period under consideration. In this paper, we study its properties as well a method of constructing portfolios which minimizes it. This method is based on linear programming. We argue that the pain index is attractive for two main reasons: it focuses on downward risk, which matters to most investors, and can be optimized easily. We demonstrate the optimization using a small sample and make analogies to mean-variance optimization.

Keywords— pain index, pain ratio, iVaR, linear programming, portfolio optimization

This page intentionally left blank.

Foreword

This paper is a bit different than those that most of my colleagues have been working on. This is not an analysis of some data, but rather an exploratory study. I have tried to keep it concise - you will notice that I do not refer to it as a thesis but rather as a paper. I attempted to create something that sparks the interest of the reader and is easily digestible. I do not cover any advanced calculus which is present in most research that works on Brownian motions, martingale sequences or any other tongue-twisting theories - I decided not go down that rabbit hole as my head might have gotten stuck in it. Mathematicians often like writing things when results make sense and are elegant - not because they are useful per se. Some researchers are able to balance pragmatism and diligence perfectly. You will see that I love quoting them - they are able to express themselves way better than I am.

You will also notice that I do not use 'I' after this foreword: 'we' refers to myself, prof. dr. Kris Boudt and Nabil Bouamara. I am very thankful for the time they invested in me, their guidance and the knowledge they shared with me. Without them, this would not have been possible at all. Of course, all errors are my own. I also want to thank the professors of the Business Engineering program in general, as they all played an important role in the completion of this work.

Almost five years have passed since I attended my first lecture at the University of Ghent and now my life as a student is almost finished. Writing this work was rather challenging to me: as a jack of all trades but master of none I am interested in everything that the world has to offer, but inherently had to dedicate to one subject for this paper. Therefore, I am very grateful the people who inspired, guided and supported me along the way. I want to thank:

- *My parents, for supporting me on all dimensions: mentally, physically and financially, as well as for letting me pursue my passion.*
- *My sister and my brother-in-law, for making me forget about the things that stress me out.*
- *My companions at university, for bringing out the best in me.*
- *My friends, for all of the fun times and their everlasting support.*
- *Most importantly, my beautiful girlfriend Elisabeth. Her support has been invaluable. I wish everyone the luck to find someone as lovely to share their life with. Also her family has always been an important support.*

Lastly, this paper was written during the corona pandemic. I was fortunate enough not to have been directly affected, nor was my research obstructed in any significant way (this is the only line I copied from someone else without a reference - I promise).

Enjoy the read,

Lars

Si hoc legere scis nimium eruditionis habes.
- Anonymous

This page intentionally left blank.

Table of contents

List of used abbreviations

CAL Capital Allocation Line

CDD Conditional Drawdown

CVaR Conditional Value-at-Risk

ES Expected Shortfall

EW Equally Weighted

FRC Fractional Risk Contribution

GMV Global Minimal Variance

IS In-Sample

iVaR Investsuite Value-at-Risk

LP Linear Program(ming)

MAD Mean-Absolute Deviation

MPI Minimal Pain Index

MPT Modern Portfolio Theory

MRC Marginal Risk Contribution

OOS Out-Of-Sample

PI Pain Index

PR Pain Ratio

RC Risk Contribution

VaR Value-at-Risk

List of figures

List of tables

1 Introduction

Investors need to distribute their wealth over the set of available assets, and want to do so optimally. This is called the problem of *portfolio optimization.* What they consider optimal depends on their personal objectives and the measures used. We consider a framework in which the investor wants to minimize risk and attain at least a certain level b of expected return. By modelling the preferences of an investor in such a way, we implicitly assume (s)he is a rational investor, as proposed by Markowitz (1952). In general, the problem can thus be written as:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & f(\mathbf{w}), \\ \text{subject to} \quad & g(\mathbf{w}) \geq b, \end{aligned} \tag{1.1}$$

where f is an expression of the risk in function of the vector of weights $\mathbf{w}$, and g denotes the function of expected return.

We study a method the use of the pain index as a measure of risk. The pain index was introduced in 2006 by Zephyr Associates and defined as follows:

$$\textit{pain index} = \sum_{t=1}^{T} \frac{d_t}{T}, \tag{1.2}$$

where d_t is the drawdown at time t and T is the length of the time period under consideration. It did not receive much attention in research, but has recently been picked up and popularized by Investsuite, a Belgian fintech start-up, under the name *iVaR*, short for *Investsuite Value-at-Risk*.

In this article, we contribute to the current body of research in two ways. Firstly, we cover properties of the pain index which may not be evident at first sight and are not covered by previous research. Secondly, we study in detail the use of a linear program to minimize the pain index. We base our analysis on Markowitz' original framework and try to take an analogue approach.

This paper proceeds as follows. We start by describing current literature in section 2. In section 3, we study the pain index in more detail. We examine it both mathematically as well as visually, try to decompose the risk in multivariate cases, and show why it does (not) fit in different risk classification frameworks. We study how the measure can be minimized by using a simple linear program in section 4, which has interesting and useful properties. Finally, section 5 concludes.

2 Literature overview

2.1 Modern Portfolio Theory

Before diving in the details of optimization, we need to introduce the reader to our framework that was covered briefly in the introduction. We start from the very basics by building our framework from the ground up.

2.1.1 Model development

Throughout this paper, we consider a universe of N assets (or securities) to invest in. Assets can theoretically be almost anything, but it is common practice to mainly consider financial assets under the form of stocks or shares. An investor has a certain amount (s)he wants to invest and will allocate a proportion of his or her available capital to each of the N assets. This proportion is what we call the *weight* assigned to each asset and we refer to this as w_i, where i is an index referring to one of those assets: $i = \{1, \ldots, N\}$. Naturally, these weights should sum up to one as we assume the investor completely uses his or her capital: $\sum_{i=1}^{N} w_i = 1$. In most research, this weight is allowed to vary between 0 and 1: $0 \leq w_i \leq 1$, but as of 2021 even the most financially illiterate know that there is a thing called *shorting*. This allows weights to be smaller than 0, and thus also larger than 1. Assuming whether shorting is possible does not significantly influence most analyses, thus we keep it simple, make abstraction of its existence and assume a traditional long-only investor.

Every asset has a certain price at each point in time. The points in time under consideration are defined by the past time horizon we choose to examine. We refer to a certain point in time as t and our time period under consideration is of length T: $t = \{1, \ldots, T\}$. We consider each point in time to be a certain day, but theoretically these could also be minutes or even months. We try to make predictions of the future by looking at the past, thus this period should have ended somewhere in the past, preferably as recent as possible, i.e. yesterday. We denote the price of asset i at time t as $p_{t,i}$. As a consequence, this can range from $p_{1,1}$, denoting the price of our 'first' asset on the first moment in time, to $p_{T,N}$, which is the price of our 'last' asset N at time T.

Let us now assume our investor has an amount of 100 (euro, dollar...) to invest in four available stocks. The investor has heard the paradigm to *'never put all your eggs in one basket'* and decides to split his/her capital equally among the four assets, i.e. 25 in each. If the four assets are by chance all worth 5 at the time of investment, the investor can buy 5 shares of each, but this will probably not be the case. To solve this issue, we will assume that assets are *infinitely divisible*, which means one can buy any fraction of a share. For example, if one of the assets were worth 50, we are able to buy 0.5 of it.

An investor invests money with the goal of making more money. To quantify this reward, we define the *return* of asset i from time $t-1$ to time t as

$$r_{t,i} = \frac{p_{t,i} - p_{t-1,i}}{p_{t-1,i}}. \tag{2.1}$$

We now have the building blocks to introduce the problem of *portfolio optimization*, as mentioned in the introduction and formulated by Markowitz (1952): we want to invest our money in an optimal way. As the economic paradigm goes, we assume that investors like return and dislike risk, which is the ever-present trade-off in investing. We know what return is, but we are still faced with the question of is exactly represented by the concept of risk.

According to the dictionary, risk is *the possibility of something bad happening*, yet this abstract definition can take many forms in a financial world. One can make an analogy to eating cookies and being unhealthy. We assume that everyone likes eating cookies (return): the more cookies, the better. However, we do not like being unhealthy (risk), but what does that in fact entail? The abstract concept can be quantified by blood pressure, fat percentage, or simply your weight. Using either of those has its advantages and disadvantages, just like different measures of risk measures do.

We cannot with full certainty predict how the prices of the securities will develop in the future, we can only make educated guesses based on the past. Therefore we should normally state that we maximize *expected* return and minimize *expected* risk, but we will often omit the adjective, as we assume that the past performance is indicative for the future. It is common practice to denote the expected return of asset i as μ_i. This 'past' is the time period we introduced before and the investor invests all of his capital at time $T+1$, defined by a vector of weights $\mathbf{w} = \{w_1, \dots, w_N\}$. Notice how we denote vectors in bold. This problem thus considers a single period of investment. Simply said, an investor allocates his or her capital among different securities by assigning a weight to each, based on the past performance. Every security generates a random return during the single period, which leads to a change in the total portfolio value.

The duality in the trade-off should be clear now. We can either minimize a certain measure of risk f while attaining at least a level b of return g (as shown in the introduction):

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & f(\mathbf{w}) \\ \text{subject to} \quad & g(\mathbf{w}) \geq b, \end{aligned} \tag{2.2}$$

or maximize the return while attaining no more than a certain level a of risk. This second possibility in fact equal to the former from a theoretical perspective on optimization:

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & g(\mathbf{w}) \\ \text{subject to} \quad & f(\mathbf{w}) \leq a. \end{aligned} \tag{2.3}$$

An alphabetical overview of the notation which will be used throughout this paper is provided in table 1. Vectors are always indicated in bold and so are matrices, but additionally, those will be in capital letters. Note that when no time index t is added to the price or return variable, it refers the past average (e.g. r_i = average return of asset i) and that the generic f for risk will often be denoted by ρ, as is common in research. The subscript ptf is used when referring to the portfolio instead of to an individual asset i, and lastly, it is also assumed that there is a risk-free asset available, returning a (constant) risk-free rate.

Notation	**Meaning**
i	Index referring to one of the assets, $i = \{1, \ldots, N\}$
N	Number of assets
$p_{t,i}$	Price of asset i at time t
$r_{t,i}$	Return of asset i from $t-1$ to t
r_f	Risk-free return
t	Index referring to point in time, $t = \{1, \ldots, T\}$
T	Length of time period = moment of most recent observation
w_i	Weight of asset i

Table 1: Overview notation

2.1.2 Mean-variance optimization

Markowitz (1952) not only introduced the framework of portfolio optimization, but even more so suggested using the standard deviation of the portfolio as a measure of risk, denoted σ_{ptf}, or equivalently, the variance: σ^2_{ptf}. Doing so leads to very elegant and mathematically attractive results. In order to express the portfolio variance, we need to introduce the covariance between the returns of asset i and j (where j also represent an index in the set $\{1, \ldots, N\}$) as σ_{ij}. This is equal to the the multiplication of the standard deviations of the two assets and their Pearson correlation coefficient ρ_{ij}: $\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$. The covariance between an asset and itself is equal to its variance, as the correlation coefficient is then equal to one: $\sigma_{ii} = \sigma_i^2$. All of the aforementioned values can be summarized in a matrix called the *covariance matrix*. The variance of a complete portfolio is then the objective function of the following optimization:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{ij}w_iw_j, \\ \text{subject to} \quad & \sum_{i=1}^{N} w_i\mu_i \geq b, \\ & \sum_{i=1}^{N} w_i = 1. \end{aligned} \tag{2.4}$$

This is a quadratic programming (QP) problem, as the decision variables in $\mathbf{w}$ are multiplied by each other. The minimal return constraint is expressed as the overall return of the portfolio, which can be found by multiplying the weight w_i of each asset by its expected return, which has to be smaller than the minimal yield required by the investor b. This is thus one of the possible configurations of model 2.2. This is however not the real Markowitz model, as it simply minimizes variance, while Markowitz optimisation entails a real trade-off between risk and reward. In this paper the focus is on this minimization, as it is in line with the covered model for the pain index.

Markowitz' framework is often called *mean-variance optimization* (MV-optimization), but even more so *Modern Portfolio Theory* (MPT), which technically speaking is the combination of the former with ideas of Sharpe (1964), who developed the Capital Asset Pricing Model based on the work of Markowitz. For further analysis of applications and a complete overview of Markowitz theory, please refer to Guerard (2009), an extensive reference book which has been approved by Markowitz himself in a later publication (Markowitz, 2010).

2.1.3 Efficient frontier

In the aforementioned framework, the set of all possible portfolios can be graphically represented on a graph where the x-axis normally depicts risk under the form of standard deviation, while the y-axis shows the return. The representation in figure 1 is borrowed from Lintner (1965). All of the possible portfolios lay in the area bounded by the large curve. As investors are rational, they will prefer those portfolios that, for a given level of risk, provide the highest return. Those can be found on the upper part of the curve. We call that segment the *efficient frontier*.

Under some assumptions, the optimal portfolio on that frontier is then that with the highest Sharpe ratio (Sharpe, 1966), denoted by the letter M on the figure (here with risk σ_r and return

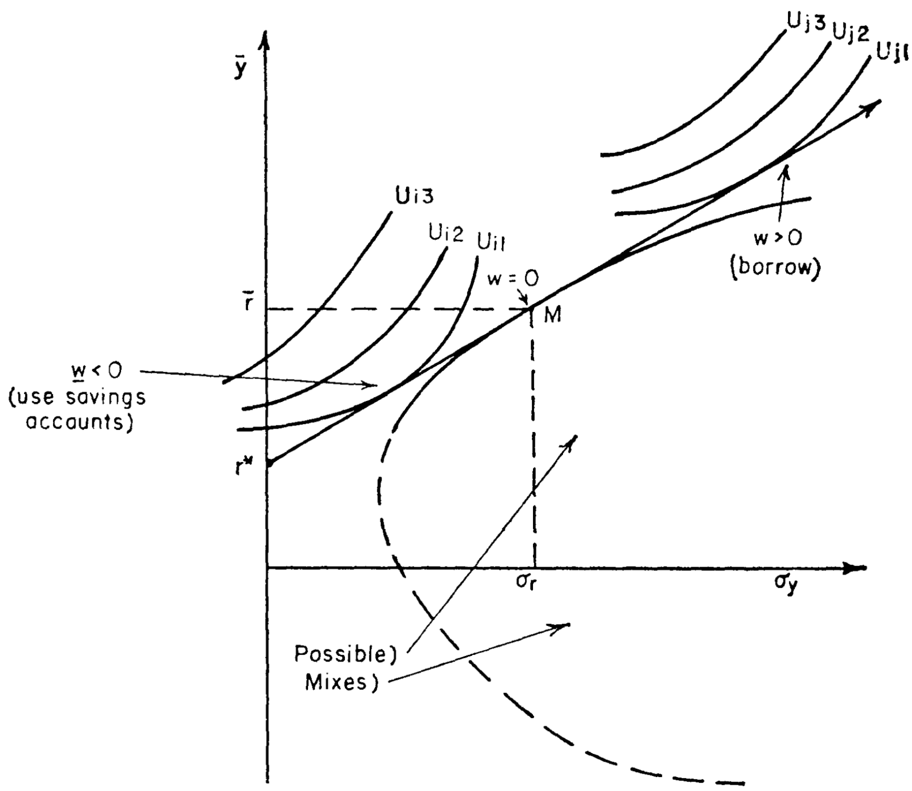


Figure 1: Efficient frontier

$\overline{r}$). The assumptions include the fact that there is a risk-free asset, returning the risk-free rate (here denoted r^M), and that investors can borrow and lend at that risk-free rate. The two-fund seperation theorem of Tobin (1958) then states that a rational investor should take a position with a certain weight assigned to the optimal portfolio and one minus that weight assigned the risk-free asset, depending on his or her personal risk tolerance. These combinations all lay on the straight line that starts at the risk-free asset, goes through the optimal risky asset and thus touches the efficient frontier. Note that the weight w on the figure - rather unintuitively - refers to the amount that is being borrowed at the risk-free rate.

Investors then have to decide which point on that straight line they prefer and that is were utility functions come into play. This concept stems from the microeconomic *indifference curve*, which shows combinations of quantities of two goods to which a consumer is indifferent, i.e. (s)he does not prefer one above the other. A utility curve works similarly, but will explicitly put a number to it. In our current case of portfolio optimization, it would thus assign a number to a certain combination of risk and reward and will also be dependant on the consumer's risk aversion.

Analytically, one could thus argue the following for a utility function U:

$$U = f(risk,\ return,\ risk\ aversion). \tag{2.5}$$

There are two sets of utility curves on figure 1: U_i and U_j. The second index shows the ranking within each set, e.g. U_{i2} has a higher utility than U_{i1}. The investor with set U_i is clearly more risk-averse than the other: the former will assign some capital to the risk-free asset, while the latter will borrow at the risk-free rate. Using this concept, the goal of portfolio optimization can be rephrased to being the maximization of one's personal utility.

2.1.4 MPT criticism

Markowitz' theory is omnipresent in research but fueled by its success, it has also received an extensive amount of criticism. People have countered and adapted the model numerous times. The most prominent drawbacks include the following:

Computational difficulty: Generating the needed inputs as well as solving the QP problem is computationally demanding, especially for large-scale portfolios (i.e. high-dimensional problem: thousands of assets over a long period of time), which is what investors are being faced with in real-life.

Variance: By using variance as a risk measure one implicitly assumes that returns are normally distributed, which is in general not the case (see Richardson and Smith (1993) and references therein, including the work of Fama (1965)). Even if it were so, it is claimed not to reflect an investor's view on risk. Firstly, it ignores higher moments and thus assumes mean and variance fully describe return distributions. Markowitz (1952, p.90) himself already argued that *'the third moment of the probability distribution of returns from the portfolio may be connected with a propensity to gamble'*, even though his MPT does not take notice of it. On top of that, even if returns were distributed normally and they could be described by their first two moments (if by some magic force all assets would have the same higher moments), there is nothing that guarantees that investors view risk as being standard deviation. After all, why would investors dislike an asset's value going up? Markowitz (1959) acknowledges this fact himself and proposes the use of the semi-variance, which only penalizes returns that fall below the mean.

Estimation error: In order to implement the optimal asset allocation, both expected returns as well as the covariance matrix need to be estimated. This estimation will normally be based on the historical means and covariance matrix, i.e. an in-sample (IS) estimation period. Several pieces of research have addressed the fact that there is no guarantee at all for the out-of-sample (OOS) performance to be as good as the IS estimation and it even seems that the 'optimal' portfolio performs rather poorly OOS (Chopra & Ziemba, 1993; Michaud, 1989).

Several authors have suggested methods to reduce these estimation errors, or at least their impact, ranging from Bayesian estimators to portfolio allocation rules that are forced to diversify estimation risk. Others argue that the risk in terms of variance is quite predictable, but return is not. Therefore they suggest making abstraction of return and only focus on the covariance matrix and minimizing portfolio risk: these *global minimum variance* portfolios seem to perform quite well OOS (Jagannathan & Ma, 2003). This argument is supported by research showing that the effects of the mean in standard MV-optimization is significantly more influential: some even find that errors in means are up to ten times as important as errors invariance (Chopra & Ziemba, 1993). This is interesting for us as well, as we will in like manner ignore returns in the general pain index optimization.

Sensitivity: Linked to the previous problem, optimal portfolios tend to be sensitive to input changes. In other words, a small change to either the means or to the variances can result in significant adjustments to the optimal portfolio. This instability becomes even more problematic when one were to consider the effect of transaction costs. Not only changes to the seemingly correct input variables are consequential, but also the errors in estimation. Related to this is the mocking quote of *GIGO: Garbage In, Garbage Out.*

Diversification: MV-optimization seems to deliver rather concentrated positions, by assigning large weights to those assets that have performed strongly in the past. This raises the question whether or not these resulting portfolios can really be considered diversified (see i.a. Green and Hollifield (1992) and references therein). On the contrary, others argue that MV-optimization results in portfolios that are in fact 'too diversified': ignoring transaction costs becomes really problematic and investors do actually prefer to invest in a smaller number of assets (Blume & Friend, 1975).

2.1.5 Risk

Risk is inherently subjective. Most people are not keen on ever facing an aggressive bull, yet every year some daredevils seem to enjoy the bull run in the streets of Pamplona (not to be confused with a financial bull run, which we assume everyone likes). As a matter of fact, even gender seems to influence the way we consider risk (Harris & Jenkins, 2006). Traditional investment theory incorporates this effect to some extent: a portfolio should consistent of a certain fraction invested in the risk-free asset, depending on one's personal level of risk aversion.

However, investors do not act in the way traditional thinking predicts: this is the crux of the matter in the discipline of *behavioural finance* (of which the foundations were laid by Shiller, for which he received a Nobel prize; a detailed overview of the subject can be found in Montier (2009)). Simply said, traditional theory talks about what investors *should* do, while behavioural finance looks at what they *actually* do. In other words, the former assumes an investor is rational at all times, while the latter realises (s)he is not: humans are *biased* in different ways. For example, when making decisions, totally irrelevant factors seem to have value as an input: this effect is known as *anchoring*. The following example of Tversky and Kahneman (1974), as mentioned by Montier (2009, p.25), demonstrates this form of intrinsic irrationality of humans:

> *They (Tversky and Kahneman) asked people to answer general knowledge questions, such as what percentage of the United Nations is made up of African nations? A wheel of fortune with the numbers 1 to 100 was spun in front of the participants before they answered. Being psychologists, Tversky and Kahneman had rigged the wheel so that it gave either 10 or 65 as the result of a spin. The subjects were then asked if the answer was higher or lower than the number on the wheel, and also asked their actual answer. The median response from the group that saw the wheel spot at 10 was 25, and the median response from the group that saw 65 was 45! Effectively, people were grabbing at irrelevant anchors when forming their opinions.*

The world of irrational, behavioural finance is interesting, but will not be our focus. We will assume that investors are rational at all times.

In order to optimize and use risk in calculations, we need our financial risk to be *quantified* and we want this quantification to be of ratio scale. This means that we need a meaningful non-arbitrary zero point, which is in our case the amount risk a risk-free asset possesses: none (although some question whether a risk-free asset is really risk-free). By doing so, we can meaningfully state that a certain asset is twice as risky as another if its level at risk is twice that of the latter.

Some critique the way the financial industry looks at risk. A perspective that has gained popularity is to think of risk as being the small chance of something really bad happening, termed

Black Swan events by Taleb (2008). According to Taleb, financial experts like to ignore their existence while in fact those events have caused important repercussions that have shaped the world to what is today. He even argues that J.P. Morgan jeopardized the entire world by exposing their RiskMetrics system, which included Value-at-Risk, as it simply makes abstraction of what is very unlikely. Researchers have attempted to incorporate Black Swan risk (e.g. Paté-Cornell (2012)), but it has proven to be challenging and will probably take decades to be part of the standard toolbox, as is argued by Taleb (2008, p.18):

> *Go ask your portfolio manager for his definition of 'risk', and odds are that he will supply you with a measure that excludes the possibility of the Black Swan—hence one that has no better predictive value for assessing the total risks than astrology.*

MPT assumes that asset returns are normally distributed, while research has shown numerous times that this is not the case. Actual return distributions are said to be more *leptokurtic* (i.e. having fatter tails) compared to normal distributions, thus the chance of extreme Black Swan events actually taking place is vastly underestimated.

2.2 Drawdown

It is often argued that the variability of the rate of return above the mean should not be treated as its counterpart (Mansini, Ogryczak, & Speranza, 2003): investors care more about underperformance than they do about 'overperformance' - if there is such a thing. Therefore research has suggested the use of metrics that focus on downside risk, like the classical Value-at-Risk (VaR), Conditional Value-at-Risk (CVaR), but also drawdowns. Drawdowns focus on downward risk like VaR and CVaR do, yet take a multi-period perspective by considering a cumulative loss compared to a previous peak, while (C)VaR is inherently one-period.

2.2.1 Simple drawdown measures

In mathematical terms, the *drawdown* of a security i with price $p_{\tau,i}$ at time τ is often defined as follows:

$$d_{\tau,i} = \max\left[\left(\max_{t\in(0,\tau)} p_{t,i}\right) - p_{\tau,i}, 0\right]. \tag{2.6}$$

The maximum inside the brackets defines what is called the *running* or *rolling maximum* (sometimes also called *maximum-to-date*), while the outer one guarantees non-negativity, as a drawdown can only be positive or otherwise zero. However, the price at τ cannot be any larger than the running maximum as it would otherwise just be the running maximum, thus the formula can be simplified to the following:

$$d_{\tau,i} = \left(\max_{t \in (0,\tau)} p_{t,i} \right) - p_{\tau,i}. \tag{2.7}$$

The first equation is used by some authors, but is not as elegant as it could be - please never trust Wikipedia. One should normalize the prices at time 0 to 100 in order for a comparison between different assets to make sense. This first definition is what will be referred to as an *absolute drawdown*, as the drawdown represents or is defined as a monetary value. An interesting analogy is found by plotting the magnitude of the drawdown at time t in function of its price at time t, as it would resemble the payoff of being long a put option with a strike price equal to the running maximum at $t-1$.

A second possibility is to define the drawdown in relative terms compared to its running maximum:

$$d_{\tau,i}^{rel} = \frac{d_{\tau,i}}{\max_{t \in (0,\tau)} p_{t,i}}. \tag{2.8}$$

For example, if a portfolio has reached a maximum value of 5 million in the past and is now worth 4 million, the current drawdown is equal to 1 million (absolute) or 20% (relative). One could also calculate what is called a *rally* or *drawup*: this positive counterpart of a drawdown is the difference between the present value and the historical minimum, but is rarely used (Goldberg & Mahmoud, 2017).

Visually, the drawdown is conceptualized on figure 2. This graph and all of the subsequent ones are own creations, generated using the **ggplot2** package of Wickham (2011) in the programming language R[1]. The area in black shows the development of the value of a certain security or portfolio. The red area then visualizes the difference between the rolling maximum and the current value and thus shows the magnitude of the drawdown at each point in time. The graph shows a security's price starting at 100, reaching a value of around 125, but it falls back to an all time low, below its starting value. Then its price starts picking up again, but even during the period of increasing prices there is still a drawdown as the price is still below its previous maximum.

By taking the drawdowns and putting those on a seperate graph, we have what is called the *underwater curve*, shown on figure 3. Notice how the underwater curve basically shifted the red area of figure 2 up (on a rescaled y-axis). The loss is shown in absolute figures (as this matches our definition of the pain index), but you will often find this graph expressed in relative terms.

[1] https://www.r-project.org/R

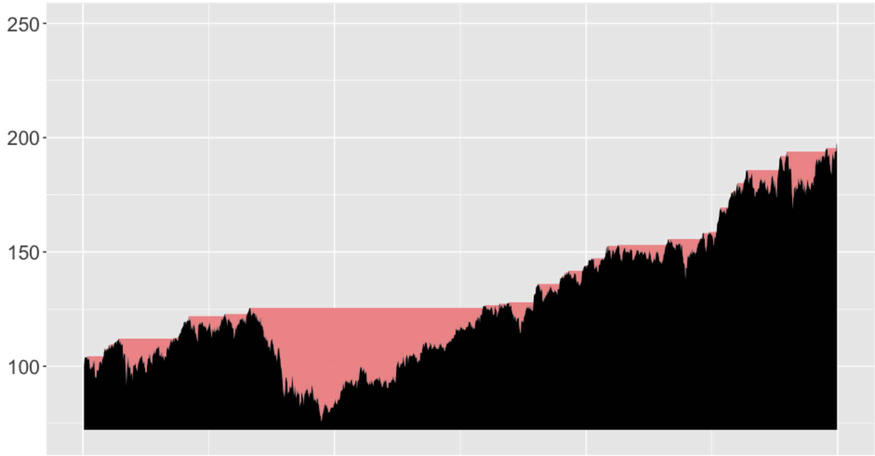


Figure 2: Sample path with drawdown visualized

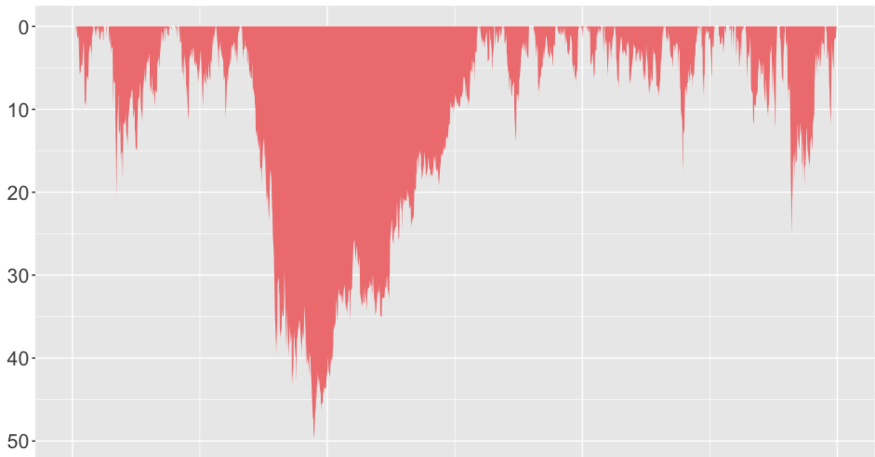


Figure 3: Example of underwater curve

All of the building blocks that are used in subsequent measures have now been introduced. These will be referred to as being *simple drawdown measures*, as they perform rather simple arithmetic operations, as opposed to the tail risk measures in subsection 2.2.3. They are easy to calculate and often used for ex-post comparisons of investment products. An overview of the different measures is provided in panel A of figure 2.

The very first drawdown risk measure to appear in research concerning portfolio management was the *maximum drawdown* (Garcia & Gould, 1987). We mention it specifically, because this conservative measure for the potential loss a portfolio could face is the one with the most extensive researched surrounding it. It has been applied numerous times, especially in studies concerning hedge funds, for which we refer to the work by Hayes (2006) or Heidorn, Kaiser, and Roder (2009).

According to Bacon (2008), drawdown measures in general are especially applicable in the hedge fund industry. This obscure asset class is a type of fund where an active, sometimes even called aggressive, style of management tries to earn alpha, i.e. beat the market. They differ most from other due to the limited set of rules they have to obey, because of their quasi non-existent regulation. Investors that allocate their capital to this type of fund are typically bound to a certain lock-up period during which they are not allowed to withdraw their money. All of these elements cause drawdowns to be a suitable measure of evaluation. For example, hedge funds are by design more volatile and thus risky from a Markowitz point-of-view. It may be an exaggeration, but one could argue that hedge funds will always have a high variance, both in the case they perform well (be it caused by upside moves) as well as in the case where they perform below expectations. An asymmetric measure is then appropriate to make a distinction between them.

The first occurrence of the pain index and pain ratio can be found in a 2006 publication by Thomas Becker and Aaron Moore of Zephyr Associates, an American investment firm. As far as our knowledge, the original report is not available anymore, but the work of Odo (2011, 2016, 2020), who is associated to the firm, is. It has also been picked up by Bacon (2008), who defines the pain index as follows:

$$pain\ index_i = \sum_{t=1}^{T} \frac{d_{t,i}}{T}, \tag{2.9}$$

where $\{1, \ldots, T\}$ are still the different points in time under consideration. The pain index is thus equal to the average drawdown over a certain horizon, assuming an investment has taken place at $i = 0$, scaled to 100. At $t = 1$, the value starts deviating from its initial magnitude and we could start observing a drawdown. Although it is not clear in any of the publications or other research, we believe it is important *not* to take into account the starting point (of 100), because if we would do so, we would introduce a bias depending on the length T. Shorter periods would naturally have a slight advantage over longer periods.

When referring to *average drawdown*, ambiguity exists in literature. Sometimes it is defined equal to the pain index, but in other pieces of research it is divided by the number of drawdowns δ (thus not taking into account periods of gain):

$$average\ drawdown_i = \sum_{t=1}^{T} \frac{d_{t,i}}{\delta}, \tag{2.10}$$

where δ is the total number of drawdowns in the entire periods (i.e. when the drawdown is equal not equal to zero). This is a crucial difference between these two measures and we believe that making the distinction is important, hence we follow the definition of Bacon (2008).

Theoretically, the drawdowns used could be defined either in an absolute way or a relative one. Even though the difference seems subtle, it does have a significant influence. For example, the LP which we will later present minimizes the former yet not the latter. Unless otherwise mentioned, we will be talking about the absolute definition. Also note that the discrete definition could easily be converted to a continuous form by replacing the summation by an integration, but this does not change the rationale of the measure.

To our knowledge only an ex-post definition has been proposed. By defining a simple ex-ante measure, the interpretation of the ex-post measure is also clarified: it can be seen as an expected drawdown:

$$pain\ index_i^{ex-ante} = \mathbb{E}(d_i). \tag{2.11}$$

For a visualisation of the pain index we again refer to figure 2. The pain index is equal to the average of the red drawdown area over time and when minimizing it (cf. infra), we will basically be minimizing this red area and thus pursue a monotonic growth. A monotonically increasing curve (i.e. always increasing or remaining constant) has a pain index of zero.

We want to mention that the pain index is currently being popularized by the firm Investsuite (2021) as part of one of their commercial applications. They refer to is the *iVaR*, short for *Investsuite Value-at-Risk*, which they employ in their *Portfolio Optimizer* product. In collaboration with the firm, Lemahieu (2020) studies the concept of diversification in an iVaR framework. A summary of some relevant aspects of his work is provided in appendix A.

A recent framework is that of the *weighted drawdown* (wDD), which is intuitively simple, yet less powerful in algebraic applications (Korn, Möller, & Schwehm, 2019):

$$wDD_i = \sum_{t=1}^{T} \omega_t d_{t,i}, \quad 0 \leq \omega_t \leq 1, \quad \sum_{t=1}^{T} \omega_t = 1. \tag{2.12}$$

The drawdown at each point in time thus receives a weight ω_t, not to be confused with w_i. By varying the weights, one can form several possible measures such as the maximum drawdown ($w_t = 1$ where d_t is maximal, $w_t = 0$ for the others) or a linearly weighted drawdown (i.e. assigning a higher weight to more recent drawdowns). This measure is to our interest because it also encompasses the pain index by setting $\omega_t = \frac{1}{T}$, but the authors do not acknowledge this.

2.2.2 Performance measures

The drawdown risk measures defined above can now be used to form *performance measures*, which are all based around the original Sharpe ratio (Sharpe, 1966), which on its turn stems from Markowitz' theory. Treynor (1965) also deserves some credit, as his paper on the Treynor ratio actually appeared one year earlier. By including both risk as well as return, these ratios are able to capture the intrinsic trade-off in one single metric. The ratios all have an excess return (those presented compare to the risk-free rate, but there are others that juxtapose a market benchmark) in the numerator, divided by a measure of risk in the denominator. An overview of drawdown-related performance measures is provided in panel B of table 2, including the two initial measures.

In a similar fashion, Odo (2011) levered the pain index in a performance measure called the *pain ratio*:

$$pain\ ratio_i = \frac{r_i - r_f}{pain\ index_i}. \tag{2.13}$$

He frames both the pain index and ratio in the context of the aftermath of the Great Financial Crisis. Contrary to before, lots of financial products started appearing which were designed not to track the market. Performance measures before were often based on performance relative to some benchmark (such as the market), but doing so did not make sense anymore. Therefore Odo refers to them as 'post-MPT metrics'.

We believe that it is important to outline the current spectrum of performance measures for two main reasons: it demonstrates how the pain ratio complements them and how this existing gap in research is filled, as well as to provide the interested reader a basis to further discover related research.

The space of available risk and performance measures is already very broad (Bacon, 2008), and more problematically, several pieces of research have shown that the difference in explanatory power between them is limited, as correlations are rather high (see Auer (2015), Auer and Schuhmacher (2013), Caporin and Lisi (2011), Eling (2008), Eling and Schuhmacher (2007), Ornelas, Silva Junior, and Fernandes (2012), and Schuhmacher and Eling (2011) to name a few).

	Source	**Formula**	**Comments**
Panel A: Simple drawdown measures			
Maximum drawdown	Garcia and Gould (1987)	$max_{t\in[0,T]} d_t$	Denoted d^{max}
Ulcer index	Martin and McCann (1989)	$\sqrt{\sum_{t=1}^{T} \frac{d_t^2}{T}}$	Punishes large losses
Average drawdown	Bacon (2008)	$\sum_{t=1}^{T} \frac{d_t}{\delta}$	δ representing the number of drawdowns
End-of-period drawdown (eopDD)	Möller (2018)	d_T	Psychologically important aspect
Weighted drawdown (wDD)	Korn, Möller, and Schwehm (2019)	$\sum_{t=1}^{T} \omega_t d_t$	Less powerful in algebraic applications
Panel B: Performance measures			
Treynor ratio	Treynor (1965)	$\frac{r_i - r_f}{\beta_i}$	
Sharpe ratio	Sharpe (1966)	$\frac{r_i - r_f}{\sigma_i}$	
Martin ratio	Martin and McCann (1989)	$\frac{r_i - r_f}{ulcer\ index_i}$	
Calmar ratio	Young (1991)	$\frac{r_i - r_f}{d^{max}}$	Sensitive to outliers
Burke ratio	Burke (1994)	$\frac{r_i - r_f}{\sqrt{\sum d_t^2}}$	Penalizes larger losses
Sterling ratio	Kestner (1996)	$\frac{r_i - r_f}{AAMD_i}$	AAMD = Average Annual Maximum Drawdown

Table 2: Overview risk measures

2.2.3 Tail risk measures

Related to the aforementioned simple measures, several more general and broader measures have been developed. These are similar to each other as they are all based on a certain probability distribution for which a cut-off value is determined based on a confidence level often denoted α. Therefore we refer to them as *tail risk measures*. These measures and their accompanying research are to our interest for two main reasons. Firstly, they include more rigorous mathematics regarding drawdowns and statistical properties which will me used to make statements about the pain index in section 3.5. Secondly, there is often related research available which provides linear programming procedures for optimization of the measures, which served as the basis for the optimization of the pain index.

The first measure mentioned here is the *Conditional Value-at-Risk* (CVaR), also referred to as the *Expected Shortfall* (ES), *mean excess loss*, *mean shortfall* or *tail VaR* (although technically speaking, the original definition of the ES was not exactly the same as the current CVaR, but in fact only was identical for most continuous distributions), introduced by Artzner, Delbaen, Eber, and Heath (1997). Although not directly related to drawdowns, its concept and intuition are applied in the subsequent measures. It is based on the original *Value-at-Risk* (VaR), which is defined as the lowest amount such that with a certain probability α, the loss of a portfolio will not exceed that amount. The CVaR is then the conditional expectation of losses above that amount.

The CVaR has gained popularity due to its advantages over the regular VaR. By definition, the CVaR is able to quantify risks beyond the VaR and estimations of the former are also more robust compared to those of the latter. CVaR is intrinsically more useful since it is considered a coherent risk measure (Acerbi & Tasche, 2002), which will be explained in more detail later, while the regular VaR is not. The interested reader is referred to Embrechts (2000) for a full overview of VaR criticism.

A succeeding measure that has become common in research is the *Conditional Drawdown-at-Risk* or CDaR (Chekhlov, Uryasev, & Zabarankin, 2004), which is conceptually quite close to CVaR. It has been referred to as the *Conditional Drawdown* or CDD in their subsequent publication (Chekhlov, Uryasev, & Zabarankin, 2005). It is defined as the mean of the worst $(1-\alpha)\times 100\%$ drawdowns, which is deducted from the empirical distribution that is formed by ranking all drawdowns over a certain period from smallest to largest. By definition this drawdown measure encompasses several others: the limiting cases are the maximum drawdown (for $\alpha = 1$) and average drawdown (for $\alpha = 0$, although we would argue that this is in fact the pain index).

It seems that only later the simpler idea of *drawdown-at-risk* or DaR was proposed by Steiner (2011). The DaR at a α%-confidence level is equal to the α%-quantile on the drawdown distribution. For example, the 95% DaR is only exceeded in 5% of all cases. The maximum drawdown is the DaR at a 100% confidence level while the median drawdown is located at 50%. It is clear that this idea encompasses both the concept of drawdowns as well as value-at-risk, hence its name: DaR is to CDaR what VaR is to CVaR. Steiner summarizes it by stating that the DaR is a quantile on the distribution of drawdowns, while the VaR is one on the distribution of returns.

The most recent tail risk measure that received a significant amount of attention is the *conditional expected drawdown* or CED (Goldberg & Mahmoud, 2017). The authors claim that the CED is in fact the first formalization of ex-ante drawdown risk. It is defined as the tail mean of the maximum drawdown distribution, again with a parameter $\alpha \in [0, 1]$. In other words, it is *'the expected maximum drawdown, given that some maximum drawdown threshold (the α-quantile of the maximum drawdown distribution) is breached'* (Goldberg & Mahmoud, 2017, p.2). It is thus again similar to CVaR and CDD, but instead of employing the distribution of returns or drawdowns, it uses the distribution of maximum drawdowns. It is obvious that one thus needs to define a certain interval to be able to construct this distribution, after which it advances on a rolling window basis.

The CED measure has two interesting properties. Firstly, it is a degree one positive homogeneous function, which allows for linear attribution to its factors (using Euler decomposition). Secondly, it meets the condition of convexity (Foellmer & Schied, 2002). This convexity is crucial because it entails the notion of diversification. All of these properties will be defined later when their application to the pain index is considered. Overall, this measure is related to drawdowns but is also associated to other generalized deviation measures (Rockafellar, Uryasev, & Zabarankin, 2006), which have properties more similar to standard deviation. Molyboga and L'Ahelec (2016) have adapted the CED to a *modified conditional expected drawdown* or MCED. By demeaning the components (i.e. subtracting the mean), the measure becomes i.a. less sensitive to sample errors.

2.3 Portfolio optimization

The framework of Markowitz (equation 2.4) is a QP problem. However, the portfolio optimization problem can also be concretized to an linear programming (LP) problem, either in the form of equation 2.2 or 2.3.

Research mostly mentions two main advantages of LP. Firstly, the classical Markowitz model, using variance as a risk measure, leads to a computationally difficult quadratic optimization problem, especially for large-scale portfolios. LP problems on the other hand can be solved using efficient solving algorithms that require little computational power. Sharpe (1971a, p.1264) already argued that *'if the essence of a portfolio analysis problem could be adequately captured in a form suitable for linear programming methods, the prospects for practical application would be greatly enhanced'*. There have been many attempts in research to linearize the problem at hand. Sharpe approximated variance with a piecewise linear function and adapted the covariance matrix, such that the inter-asset correlations are all equal to zero (this is called *diagonalization*).

LP models are also important because of their ability to loosen certain assumptions, which do not hold in real life and lead to additional constraints, yet are not considered in MPT. Examples include transaction costs, limiting the lower and upper bounds of the fractions held (*quantity constraints*), restricting the number of different assets (*cardinality constraints*)... In practice, it may be suboptimal to hold a portfolio consisting of a large number of small holdings due to transaction costs and potential minimum lot sizes, and in some cases institutional factors might simply prohibit it. These real features are thus certainly useful, yet yield mixed-integer linear programming models (MILP). These are computationally significantly more complex compared to ordinary LP models in which the decision variables can take any real value. The use of heuristic algorithms to then solve these problems is considered in research and are found to compress the computational time required (Kellerer, Mansini, & Speranza, 2000). For more details regarding the general mechanics of LP we refer to the work of Luenberger and Ye (2016). For more insights into the use of additional constraints on MV-optimization, we refer to Cesarone, Scozzari, and Tardella (2009) and references therein, especially those on the increase in computational difficulty. They refer to this adaptation as the *Limited Assets Markowitz* model.

We will first consider LP in portfolio optimization for general risk measures, and afterwards look at applications with drawdown measures. In order to keep these descriptions concise we will not elaborate on each model, nor show the explicit calculations for each measure. Doing so would not enhance the core of our research. This overview is simply provided to properly set the scene.

2.3.1 General applications

What seems to be the earliest use of an LP model in portfolio optimization is one involving *Gini's mean difference* (Yitzhaki, 1982). This risk measure quantifies dispersion and is intuitively close to a standard deviation, but it does fit inside an LP framework. However, one needs several possible future paths in order to calculate and optimize it, which restricts its usefulness. Nevertheless, the approach of using a finite number of possible future scenarios is convenient because it eliminates the need for assumptions on the distribution of returns. Overall this risk measure is not as common in the context of portfolio optimization.

The *mean-absolute deviation* as a risk measure (Sharpe, 1971b) led to the MAD model (Konno & Yamazaki, 1991), which received more attention. Their main strength was the performance compared to Markowitz's model: their resulting portfolios were quite similar, but the computational time required was minimal compared to the latter. Additionally, they did not need the normality assumption that Markowitz included. However, Simaan (1997) argued that the computational savings do not compensate for the loss of information by not considering the covariance matrix. Konno had already proposed the use of piecewise linear risk functions in portfolio optimization in one of his earlier publications (Konno, 1990).

The *minimax portfolio selection rule* proposed by Young (1998) minimizes the maximum loss or thus maximizes the minimum gain (therefore also called *maximin*). This is therefore intuitively close to minimizing the maximum drawdown, but is in fact simply a minimization of the biggest one-period loss.

The *omega ratio* (Keating & Shadwick, 2002) can also be optimized linearly (Kapsos, Zymler, Christofides, & Rustem, 2014). It was introduced as a response to the Sharpe ratio, which does not consider higher moments of the return distribution. The omega ratio on the other hand does in fact capture the distribution as a whole, by using the cumulative density distribution. The main downsides of this measure are the lack of convexity, which causes the possibility of local optima, as empirical results in earlier research had already shown (Kane, Bartholomew-Biggs, Cross, & Dewar, 2009), combined with the fact that is does not take into account the existence of serial correlation. For other examples of the usage of LP in portfolio optimization and a complete overview, we refer to publications by Mansini et al. (2003, 2014) and references therein.

2.3.2 Drawdown applications

We can now explore the union of the preceding subsections and examine research regarding optimization of drawdown measures. Note that most of these models rely on the use of auxiliary variables and other methods (e.g. greedy substitutions) to be modelled as an LP problem.

Grossman and Zhou (1993) were one of the first to study drawdown optimization. They formulate a framework for an investor who *'wants to lose no more than a fixed percentage of the maximum value'* (p.1), which is in fact a constraint on the maximum drawdown. By assuming a one-dimensional case (i.e. allocating between only one risky and the risk-free asset) and a geometric Brownian motion with constant coefficients (i.e. log-normal returns), they are able to derive an exact analytical solution. The model has later been generalized to multiple risky assets by Cvitanic and Karatzas (1997). Both of these models have gained traction, but are not LP problems.

The CVaR can be optimized using a linear program by considering a finite number of possible scenarios, as introduced by Rockafellar and Uryasev (2000). VaR also plays a role in their technique as it is calculated simultaneously. It was initially only defined for continuous distributions but later generalized to other distributions (Rockafellar & Uryasev, 2002). CVaR has also been used as a constraint and its implications were considered and compared to the usage of regular VaR constraints in optimization (Alexander & Baptista, 2004, 2006).

The first occurence of the use of drawdown measures in portfolio optimization using LP can be found in Chekhlov et al. (2004), in which they also propose the use of the CDD. They maximize the average return, subject to constraints on CDD, maximum and average drawdown respectively. They claim the problems can be reduced to linear programs by making use of auxiliary variables.

Davidsson (2012) also uses simple linear drawdown constraints which allow for LP. He compares the performance of four different methods, i.e. two QP (minimizing portfolio variance and maximizing risk-adjusted return) and two LP problems (maximizing return and maximizing a risk-adjusted return which uses maximum drawdowns). An important weakness mentioned is the fact that a percentage drawdown constraint is not linear and does not allow the usage of LP methods.

3 Properties of the pain index

In the following subsections the pain index is covered in more detail. There is currently no research available studying the characteristics to their full extent. We believe that this is important, as the true underlying mechanics of the risk measure may not be evident at first sight.

3.1 Markowitz comparison

In order to demonstrate the power of the pain index, a comparison between two paths is made, shown in figure 4. Axis labels are deliberately excluded for the sake of simplicity in this illustration. These two paths do not look similar in terms of risk, but do have the same risk from a Markowitz point-of-view: the standard deviation of their returns is equal. However, when calculating their pain indices or by simply looking at the different areas, it is clear that they are very different from a drawdown perspective. The pain index of the leftmost graph is about half of that of the path on the right.

We can still go one step further. The aforementioned paths are not fully equal yet in a Markowitz framework: their (cumulative) returns clearly differ and the leftmost path would be preferable, as it provides a higher cumulative return for the same level of risk (i.e. same standard deviation of returns). Now consider figure 5, where both paths end in the same point. On top of that, the standard deviations of their returns are identical as well. In a Markowitz framework, an investor would thus not prefer any of the two above the other. When computing their pain indices, we find that the path on the right has a pain index of almost double the one on the left and an investor would thus clearly prefer the one on the left. One can also observe that the area under the rolling maximum is smaller on the leftmost graph.

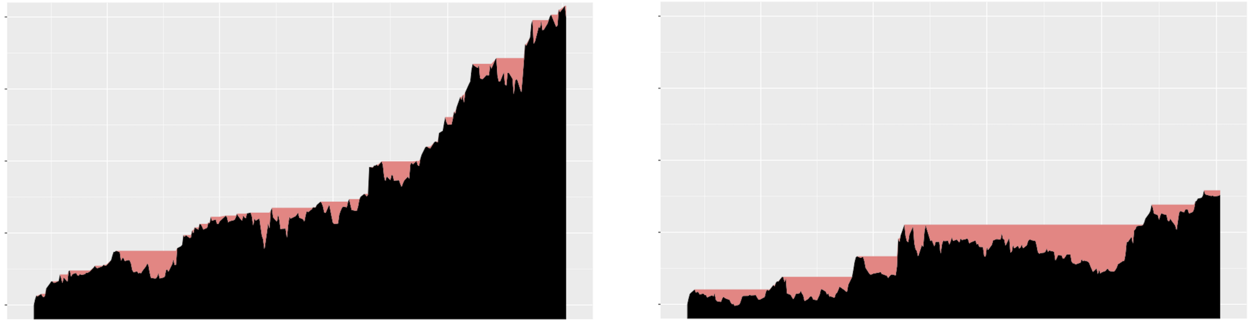


Figure 4: Two paths: same standard deviation, different pain index

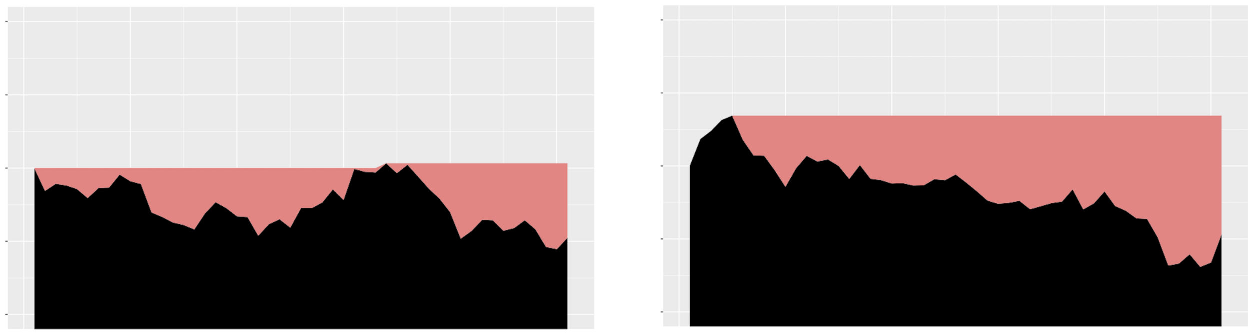


Figure 5: Two paths: identical in Markowitz framework, different pain index

What causes this behaviour is the path dependency of drawdowns (Goldberg & Mahmoud, 2017). Let us demonstrate this with a small numerical example. Consider a portfolio with a value of 100 and the following two series of returns:

$$\begin{aligned} &[+50\%,\ +0\%,\ -50\%]\,, \\ &[-50\%,\ +0\%,\ +50\%]\,. \end{aligned} \tag{3.1}$$

By assuming both start at 100, their resulting price paths are the following:

$$\begin{aligned} &[150,\ 150,\ 75]\,, \\ &[50,\ 50,\ 75]\,. \end{aligned} \tag{3.2}$$

As the constituents of the return sets correspond, they are identical in a Markowitz framework. It is however clear that their drawdowns did not simply change positions but differ significantly.

3.2 Important remarks

There are some more pitfalls to consider when using a pain index. We mention these because the effects of the seemingly simple definition will probably not be clear at first sight to the reader and help to understand what the pain index *really* measures. The first involves *normalization.* Let us consider two securities with the following evolution, with an initial investment of 100 and 10 respectively:

$$\begin{aligned} \mathbf{p}_{A^*} &= [90,\ 90,\ 90,\ 90]\,, \\ \mathbf{p}_{B^*} &= [7.5,\ 7.5,\ 7.5,\ 7.5]\,. \end{aligned} \tag{3.3}$$

When calculating the pain indices in absolute terms we find 10 and 2.5 respectively. It does not make sense however to say that security A^* is more risky from a pain index point-of-view. A remedy to this problem is to scale the securities to some common ground, like for example to 100:

$$\begin{aligned} \mathbf{p}_A &= [90,\ 90,\ 90,\ 90]\,, \\ \mathbf{p}_B &= [75,\ 75,\ 75,\ 75]\,. \end{aligned} \tag{3.4}$$

The index has become 25 for the latter example and a comparison does now make sense.

Secondly, we want to bring some clarity to the reader concerning the way the index is defined. Let us again consider two securities, both with initial investment of 100:

$$\begin{aligned} \mathbf{p}_C &= [200,\ 100,\ 100,\ 100]\,, \\ \mathbf{p}_D &= [80,\ 80,\ 80,\ 80]\,. \end{aligned} \tag{3.5}$$

The two securities are already scaled correctly and their pain indices are 75 and 20 respectively. The significant outlier in the evolution of security C causes the index to be influenced heavily. Yet this is simply the way the pain index is designed: holding security C throughout its changes will probably cause feelings of regret considering the missed selling opportunity.

A last remark is related to the distinction between an absolute and relative definition. The difference is only subtle but very important. If we again look at securities A and B, we can argue that security A is down 10 compared to its rolling maximum on average, while security B is down 25 on average. They are both scaled to 100 and we can, in this specific case, also say they are down 10% and 25% on average. This is case because neither of the two securities has hit a new rolling maximum compared to its initial value.

Securities C and D are down 75 and 20 on average and have both been scaled to 100. We can however not argue that security C is down 75% on average, due to its increased maximum over the time frame. By expressing the drawdowns on a relative base, their evolution looks like this:

$$\begin{aligned} \mathbf{d}_C &= [0\%,\ 50\%,\ 50\%,\ 50\%]\,, \\ \mathbf{d}_D &= [20\%,\ 20\%,\ 20\%,\ 20\%]\,. \end{aligned} \tag{3.6}$$

The former is thus down 37.5% on average, while the later is down 20% on average. Expressing the drawdowns and the index as a percentage leads to a more intuitive result and can be interpreted more easily, e.g. a pain index of 8% means the portfolio is down 8% compared to its maximum on average. However, the LP we will introduce minimizes the absolute pain index as relative minimization does not fit inside any LP framework. Minimizing the absolute pain index will naturally also lower the relative magnitude, but performance starts to differ significantly when large outliers are present. This effect also manifests when comparing security E and F:

$$\begin{aligned} \mathbf{p}_E &= [150,\ 130,\ 130]\,, \\ \mathbf{p}_F &= [100,\ 80,\ 80]\,. \end{aligned} \tag{3.7}$$

With both securities starting at a price of 100, they have an equal pain index of $\frac{0+20+20}{3} = 13.33$ and our LP framework will in fact consider them as equal. This demonstrates the necessity of always considering risk combined with return, possibly by comparing performance instead of simple risk measures.

3.3 Why would one use the pain index?

As mentioned before, there are already lots of different measures and ratios with subtle differences, with often high correlations and little additional explanatory power, yet we believe there is a place for the pain index and ratio. When evaluating drawdown measures in general and comparing them to other risk measures, the main advantage is the fact that they only penalize downward risk, thus under-performance. There are however some drawbacks to consider, which all apply to the pain index as well:

- Drawdown-related measures are path dependent, in contrast to measures such as volatility, skewness, and factor exposures often used in academia. Their path dependency logically leads to a significant sensitivity to serial correlation (Goldberg & Mahmoud, 2017).
- There still seems to be very little research regarding to what extent drawdown characteristics are intrinsic to a certain security. In other words, it might not be correct to bluntly assume that a certain pattern will continue in the future. We implicitly assume a pure backward-looking, myopic investor and the corresponding behavior may not be robust in different samples. There is a significant body of research which looks at the properties of drawdowns under the assumption of Brownian Motion (see i.a. the work of Meilijson (2003) or Magdon-Ismail, Atiya, Pratap, and Abu-Mostafa (2004)), but these analytical derivations are of little use in portfolio optimzation. On the contrary, there seems to be a consensus that different types of assets do have an inherent level of risk in terms of variance or standard deviation. Nevertheless, it is a stylized fact that return distributions are subject to serial correlation, which confirms the value of drawdowns.

If we look at the pain index, we believe there are some clear advantages, compared to both other drawdown measures as well general risk measures:

- The concept of the pain index is simple to grasp and easy to visualize. Ideas such as punishing larger drawdowns (like in an ulcer index) are certainly useful, but might overcomplicate things and make interpretation less intuitive. In this regard, also the value of MPT has been questioned. Horvitz and Wilcox (2007) argue that Markowitz optimization for individuals is simply to hard to explain, implement, and maintain for individual investor in the long run, while Goldstein and Taleb (2007) have shown that even finance professionals barely know what standard deviation really is.

- The pain index covers attributes of losses investors care about: the depth, duration and frequency (Odo, 2011). By doing so, it accurately reflects the goals of fund investors and managers. This is also the idea behind the name of the related ulcer index (*ulcer* - an internal or external sore), which refers to the painful stress investors feel during bear periods.

- It can be optimized using linear programming (cf. infra).

3.4 Multivariate analysis

In order to get a better understanding of the definition of the pain index, we will show some of its more technical properties in the current and subsequent subsection (3.5 Risk frameworks). We start by looking at the pain index in a multivariate case, which means there is more than one risky asset.

Firstly, one should notice that the pain index of a portfolio consisting of two securities X and Y is not equal to the sum of the individual pain indices:

$$PI_{ptf} \neq PI_X + PI_Y = \sum_{i=1}^{N} PI_i \tag{3.8}$$

This equation only holds in some very specific cases (e.g. X and Y both monotonically in- or decreasing). The total risk of a portfolio is also not equal to the weighted sum of source risks:

$$PI_{ptf} \neq \sum_{i=1}^{N} w_i \; PI_i \tag{3.9}$$

Notice how this is a more general form of equation 3.8, as there we implicitly assume that the weights are equal. Because of the method by which the pain index is computed, we can in fact not give a closed-form expression for its calculation in a multivariate case. There is thus no obvious or straightforward method to analyse the sources of risk in a portfolio and attribute to total risk to individual components of that portfolio.

What we can do is use *Euler decomposition*, also referred to as *Euler's homogeneous function theorem*, to assign the total risk to its constituents. The following argumentation is based on the work by Goldberg and Mahmoud (2017), who decompose the CED measure, and on references therein. We consider the case in which the independent variable $\mathbf{p}_i$ represents a vector of prices, but it may also be used in other contexts. In general, a function f is considered a *positive homogeneous function of degree* $k \in \mathbb{R}$ if it satisfies the following condition for a constant $C \in \mathbb{R}$:

$$f(C \cdot \mathbf{p}_i) = C^k f(\mathbf{p}_i). \tag{3.10}$$

What is most interesting is the case where $k = 1$. This means that, if f were a measure of risk, when multiplying a portfolio by a certain constant, the magnitude of risk is simply multiplied by that value. This is in fact equivalent to saying that no diversification takes place by adding identical components as the ones present. We argue that the pain index is a positive homogeneous function of degree one. It is quite clear that in its absolute definition, multiplying the value of the underlying by a certain factor would also multiply the pain index by that same factor. Note how the relative definition does not satisfy this property, as it would remain the same percentage.

As the pain index is a positive degree-one homogeneous function, it is characterized by Euler's homogeneous function theorem and the total risk of a portfolio can be attributed linearly along its factors. The theorem says that we can quantify the *marginal risk contribution* (MRC) of asset i as:

$$MRC_i^{\rho} = \frac{\partial \rho_{ptf}}{\partial w_i}. \tag{3.11}$$

It is thus the approximate change in overall portfolio risk when increasing the exposure to i by a small amount, ceteris paribus (i.e. keeping other exposures fixed). The *risk contribution* (RC) of component i is then equal to its weight multiplied by its marginal risk contribution:

$$\sum_{i=1}^{N} w_i MRC_i^{\rho} = \sum_{i=1}^{N} RC_i^{\rho} = \rho_{ptf} \tag{3.12}$$

We can also define *fractional risk contributions* (FRC) to denote the contribution a component delivers as a percentage:

$$FRC_i^{\rho} = \frac{RC_i^{\rho}}{\rho_{ptf}} = \frac{RC_i^{\rho}}{\sum_{i=1}^{N} RC_i^{\rho}} \tag{3.13}$$

However, as there is no closed form formula for the pain index, we believe we cannot derive an analytical partial derivation. One would thus need to employ numerical differentiation in order to derive the contributions (for which we refer to Burden (2011)). We simply state the possibility of risk attribution using Euler's theorem, application is beyond the scope of this paper.

3.5 Risk frameworks

In research, a broad set of definitions has been developed to classify and structure risk measures - and by extension functions in general. These might be most interesting, as Rockafellar et al. (2006) state, to formally consider the advantages of substituting the classical standard deviation for another measure. We cover the most prominent ones and demonstrate whether or not the pain index satisfies their conditions, although no rigorous mathematical proofs will be given.

The discussed frameworks are designed to fit one-period risk measures. Additionally, some also have multi-period alternatives. However, we do not believe that the multi-period definitions change the needed underlying rationale. We will show the reasoning why we believe statements do (not) hold mostly based on the research of Mahmoud (2017), who explored properties of the simple (absolute) drawdown process. We show her original proofs in appendix B. Another source that covers these properties quite extensively is the work of Chekhlov et al. (2005) on the CDD. Previously, we have already argued that this measure in fact includes what they call the average drawdown, equal to our pain index. Even though our statements will thus be similar to theirs for some properties, they do not cover the underlying reasoning but simply show some short mathematical proofs. Therefore we believe our interpretations are still valuable.

3.5.1 Coherent risk measures

The idea of coherency has been introduced in a short research note (Artzner et al., 1997), formalized in a subsequent paper (Artzner, Delbaen, Eber, & Heath, 1999) and later even generalized to a multi-period version (Artzner, Delbaen, Eber, Heath, & Ku, 2007). Ever since, it has been prominent in research, especially in the context of regulation. Its most notable application is on the CVaR (Acerbi & Tasche, 2002). We assume here that X and Y represent two future (thus expected) portfolios' P/L (profit or loss), where profit is mentioned first as it should have a positive sign in this framework. They are part of the feasible set of portfolios $\mathcal{G}$.

If a risk measure $\rho(X)$ satisfies all of the conditions, then it may be interpreted as *'the riskiness of a portfolio or the amount of capital that should be added to a portfolio X, so that the portfolio can then be deemed acceptable from a risk point of view'* (Haugh, 2010, p.1). In other words, a negative value means that a position can be considered 'safe' and the magnitude of its absolute value refers to the amount of cash that can be withdrawn while still keeping the position safe. This makes sense, as Artzner et al. (1997) defined what is called an *acceptance set* as the set of portfolios for which the measure has a value smaller than or equal to zero. Notice how this acceptance point-of-view is most intuitively applied in a regulatory setting, but could also be applicable in the case of e.g. a fund manager overlooking individual traders. It is evident that the aforementioned interpretation does not hold in case of the pain index, thus at least one of the conditions should not hold.

Definition 3.5.1 (Coherent risk measures)**.** *A risk measure has to satisfy the following four axioms in order to be considered coherent:*

1. *Subadditivity: for all* X *and* $Y \in \mathcal{G}$*,* $\rho(X+Y) \leq \rho(X) + \rho(Y)$*.*
2. *Monotonicity: for all* X *and* $Y \in \mathcal{G}$*, with* $X \leq Y$*, we have* $\rho(Y) \leq \rho(X)$*.*
3. *Positive homogeneity (of first degree): for all* $\lambda \geq 0$ *and all* $X \in \mathcal{G}$*,* $\rho(\lambda X) = \lambda \rho(X)$
4. *Translation invariance: for all* $X \in \mathcal{G}$ *and all real numbers* α*, we have* $\rho(X+\alpha \cdot r) = \rho(X) - \alpha$*.*

Subadditivity encapsulates the idea of diversification and is actually quite a natural prerequisite for a risk measure to have. If two 'merged' portfolios would have more risk than the two separate, an investor could just keep them separate. Others also refer to this as being some sort of economy of scale. The main criticism of VaR is in fact it not being subadditive, as described clearly in the book on market risk of Dowd (2007, p.34)[2]. In other words, VaR actually discourages diversification in some cases. This is the case because the underlying distribution could be discontinuous, therefore small changes in the confidence level (even by some basis points) can have significant influence on the VaR.

The pain index does satisfy the condition of subadditivity. It makes sense that the simple drawdown measure is subadditive (see attachment B): when adding an asset to a portfolio, the total drawdown could be equal to the sum of the individual ones (i.e. the case when both go down), but if an additional positive evolution compensates for other losses, the total drawdown can become lower. As subadditivity holds for the drawdown process at each point in time, we can sum over the different points in time and the following then still holds:

[2] The following example is copied from Dowd's book: *We have two identical bonds, A and B. Each defaults with probability* 4%*, and we get a loss of* 100 *if default occurs, and a loss of* 0 *if no default occurs. The* 95% *VaR of each bond is therefore* 0*, so* $VaR(A) = VaR(B) = VaR(A) + VaR(B) = 0$*. Now suppose that defaults are independent. Elementary calculations then establish that we get a loss of* 0 *with probability* $0.96^2 = 0.9216$*, a loss of* 200 *with probability* $0.04^2 = 0.0016$*, and a loss of* 100 *with probability* $1 - 0.9216 - 0.0016 = 0.0768$*. Hence* $VaR(A+B) = 100$*. Thus,* $VaR(A+B) = 100 > 0 = VaR(A) + VaR(B)$*, and the VaR violates subadditivity. Hence, the VaR is not subadditive.*

$$\sum_{t=1}^{T} d_{t,X+Y} \leq \sum_{t=1}^{T} d_{t,X} + \sum_{t=1}^{T} d_{t,Y}. \tag{3.14}$$

The pain index thus satisfies, as after division by T:

$$PI_{X+Y} \leq PI_X + PI_Y. \tag{3.15}$$

We believe that this line of reasoning (if it holds for the drawdown functional at one point in time, it also does for the sum over a period of time and thus for the pain index) could also be used for other conditions, but we will not show it explicitly.

Monotonicity simply says that if Y is in all possible cases bigger then X, then Y cannot be be riskier than X. Notice that, as this is the only requirement that uses the portfolio values X and Y as such, it is the only property where returns are taken into account. It is often considered quite confusing, as this ex-ante statement seems inconsistent with the classic adage of *high risk, high reward.* It is evident that the pain index does not fulfill this property, as it is argued by Mahmoud (2017, p.9) that *'processes that can be ordered according to their magnitudes do not necessarily imply the same or opposite ordering on the drawdown magnitudes'*.

We have already shown that the pain index is *positive homogeneous of degree one* - if and only if defined on an absolute basis. More generally, the drawdown function is as well: risk should be proportional to the size of the position. Some argue that this condition does not always hold because of the liquidity risk that relatively large position possess (Dowd, 2007).

Translation invariance says that when a risk-free amount α is added to a portfolio, the amount that should be added for the risk to be acceptable decreases by that amount α. This therefore means that the risk measure simply decreases by that amount. Within the risk functional we see $X + \alpha \cdot r$, as the future value would not only be the portfolio's, but also the added amount times a cumulative risk-free return r (which is often assumed equal to 1 and thus omitted). This condition forces the risk measure to be in monetary units. The name of this property is quite odd as a measure is not *invariant* to the change but rather *variant.* It is obvious that the pain index does not meet this requirement: adding a sure amount does in fact not change its value.

Overall, the pain index can thus not be considered a coherent measure of risk, as it does not meet the conditions of monotonicity and translation invariance.

We do want to mention the fact that there are two contradictory versions of this framework. The first being the one as just presented, the second being mentioned in e.g. Haugh (2010). In the alternative version, X does not represent a future P/L, but a future L/P: losses are defined positively. This does not change the property of subadditivity nor the positive homogeneity, yet does alter the remaining ones. For monotonicity, the sign should change: $X \leq Y$ would then imply $\rho(X) \leq \rho(Y)$ and translation invariance would be influenced in a similar way: $\rho(X - \alpha \cdot r) = \rho(X) - \alpha$. These alternative versions are the source of plenty of confusion and it seems that some authors mix up the two versions (e.g. Wilmott (2007)).

3.5.2 Generalized deviation measure

Another influential framework has been that of deviation measures (Rockafellar et al., 2006). The symbols X and Y may be interpreted in the same way as in the framework of coherency.

Definition 3.5.2 (Generalized deviation measures)**.** *By a deviation measure will be meant any functional ρ satisfying:*

1. $\rho(X + C) = \rho(X)$ *for all X and constants C,*
2. $\rho(0) = 0$, *and* $\rho(\lambda X) = \lambda\rho(X)$ *for all X and all $\lambda > 0$,*
3. $\rho(X + Y) \leq \rho(X) + \rho(Y)$ *for all X and Y,*
4. $\rho(X) > 0$ *for all nonconstant X, whereas* $\rho(X) = 0$ *for constant X.*

We recognize the properties of subadditivity (3) and positive homogeneity (2, here combined with the normalization condition which was implicitly assumed in the coherency framework). The authors emphasize the fact that the combination of condition 2 and 3 is called *sublinearity*, which implies that ρ is a *convex function.* The joint property is therefore also known as *convexity* (see subsection 3.5.3). We have already shown that the pain index satisfies both.

The first condition is the counterpart of translation invariance: notice how the two conditions are mutually exclusive. In this case, the addition of a constant (think of it as an amount of cash, i.e. risk-free) should not change the magnitude of the measure. As mentioned before, this is the case for the pain index.

However, the pain index does not comply with the first component of the fourth condition. The pain index of a nonconstant X is not necessarily strictly larger than zero, as it could be strictly increasing and have a pain index equal to zero. It can thus not be considered a generalized deviation measure.

3.5.3 Generalized path-dependent deviation measure

Finally, the pain index can be considered a generalized path-dependent deviation measure, a variation defined by Goldberg and Mahmoud (2017) with the intent to formalize drawdown-related measures.

Definition 3.5.3 (Generalized Path-Dependent Deviation Measure)**.** *A generalized path-dependent deviation measure is a path-dependent risk measure $\rho : \mathbb{R}^\infty \to \mathbb{R}$ satisfying the following axioms:*

- *Normalization: for all constant deterministic $C \in \mathbb{R}^\infty$, $\rho(C) = 0$.*
- *Positivity: for all $X \in \mathbb{R}^\infty, \rho(X) \geq 0$*
- *Shift invariance: for all $X \in R^\infty$ and all constant deterministic $C \in \mathbb{R}^\infty$, $\rho(X + C) = \rho(X)$.*
- *Convexity: for all $X, Y \in \mathbb{R}^\infty$ and $\lambda \in [0, 1]$, $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y)$.*
- *Positive degree-one homogeneity: for all $X \in \mathbb{R}^\infty$ and $\lambda > 0$, $\rho(\lambda X) = \lambda\rho(X)$.*

Most of the conditions are picked up from previous frameworks, but convexity needs some further explanation. Introduced by Foellmer and Schied (2002), it basically is a generalization and thus a stronger form of subadditivity. In fact, it is the combination of the properties of the latter and positive homogeneity of the first degree. It is a key condition particularly in optimization as it ensures that a local minimum also is a global minimum. For more details on convex analysis, we refer to Rockafellar (1970). As the drawdown function is a convex one (Goldberg & Mahmoud, 2017), the sum of drawdowns also is and thus the pain index is as well:

$$PI(\lambda X + (1 - \lambda)Y) \ \leq \ \lambda PI(X) + (1 - \lambda)PI(Y). \tag{3.16}$$

Lastly, we have seen that the homogeneity also holds and we can thus argue that the pain index is a generalized path-dependent deviation measure.

Overall, there are already tens of frameworks designed with conditions that risk measures could and in fact should adhere to. Their main purpose is to systematically study and classify measures. From a pragmatic point-of-view three of them are of utmost importance: convexity, because it is a prerequisite for mathematical optimization, sub-additivity, to correctly include diversification effects, and positive homogeneity of the first degree, in order for a measure to be attributable to its components.

4 Methodology to minimize the pain index

In this section we will show how one can minimize the pain index using linear programming. We start by showing the formulation of the LP, apply it to a real-life sample and show some of its properties later.

4.1 Minimizing the pain index

4.1.1 *Erst wägen, dann wagen*

Even though the model as mentioned in the beginning is seemingly simple, one must carefully consider the meaning of the weights. It might appeal as if these weights represent some kind of constant asset allocation, like it is the case in the framework of strategic asset allocation (SAA). Both our model as well as SAA claim have a weight that is constant through time: our weights w_i are not time-dependent (as they are not a function of t) and neither are the weights in the world of asset allocation. Nonetheless, they do represent something different.

The weights in our model represent an initial investment as a fraction of our capital, but it is easier to think of them as simply being the amount of each asset that is bought at the beginning, provided that they are normalized. Therefore, the value of the portfolio $p_{t,ptf}$ at a given moment t can be calculated as:

$$p_{t,ptf} = \sum_{i=1}^{N} p_{t,i} w_i. \tag{4.1}$$

On the other hand, if SAA states that 60% should be invested in a certain asset and 40% in another at all times, this would imply that *rebalancing* is needed in order for this ratio to remain constant. Our model on the other hand assumes an 'invest and let go'-approach. This makes the inclusion of transaction costs insignificant.

4.1.2 Optimization problem

We start by repeating some notation. Let $\{1, \ldots, N\}$ denote a set of possible assets considered for an investment and $\{1, \ldots, T\}$ be the set of points in our time horizon (thus of length T). If the subscript of a matrix has only one value, it is a square one with that amount of rows and columns. Let:

$$\begin{aligned}
&\text{weights } \mathbf{w} = [w_1, w_2 \dots w_N] && 0 \le w_i \le 1, \\
&\text{prices } \mathbf{P} = \begin{bmatrix} p_{1,1} & \cdots & p_{1,N} \\ \vdots & \ddots & \vdots \\ p_{T,1} & \cdots & p_{T,N} \end{bmatrix} && p_{t,i} \ge 0, \\
&\text{drawdowns } \mathbf{d} = [d_1, d_2 \dots d_T] && d_t \ge 0, \\
&\text{rolling maxes } \mathbf{m} = [m_1, m_2 \dots m_T] && m_t \ge 0.
\end{aligned} \tag{4.2}$$

The prices should have been divided by their price at time 0 (not included in the matrix under consideration) in order for the optimization to make sense. A portfolio is thus completely characterized by the vector $\mathbf{w} \in \mathbb{R}^N$, whose components should sum up to one.

Now the decision variables are defined using the vector $\mathbf{x} = [\mathbf{w}\ \mathbf{m}\ \mathbf{d}]^{\mathrm{T}}$ (superscript T, not in italics, denotes the transpose of a matrix), which is of dimension $(N + 2T) \times 1$. This vector is used in the goal function of the optimization, the sum of the drawdowns, which needs to be minimized. By doing so also the pain index is minimized, as it is the sum divided by the fixed number T:

$$\min_{\mathbf{w}} \sum_{t=1}^{T} \frac{d_t}{T} = \min_{\mathbf{w}} \mathbf{cx}, \tag{4.3}$$

because the $(N + 2T)$-dimensional row vector $\mathbf{c}$ is designed to pair with the drawdowns in the vector $\mathbf{x}$:

$$\mathbf{c} = [\underbrace{0 \ \cdots \ 0}_{N+T} \ \underbrace{1 \ \cdots \ 1}_{T}]. \tag{4.4}$$

The goal function is subject to several *constraints*. These are designed to coordinate well with $\mathbf{x}$. They all must be expressed as a matrix with $N+2T$ columns to be able to undergo the operation of matrix multiplication with $\mathbf{x}$ on the left-hand side, with a column vector on the right-hand side with as many rows as the former matrix.

1) Definition of the drawdowns

The first set of constraints will define the drawdowns. The drawdown at each point in time t is equal to rolling maximum minus the total portfolio value:

$$d_t = m_t - \sum_{i=1}^{N} p_{t,i} w_i. \tag{4.5}$$

The subtraction is thus equal to zero if the portfolio value is a new maximum (or if equal) and takes a positive value otherwise. In order to express this as a constraint that can be used in combination with our expression of $\mathbf{x}$, we rewrite it to $\sum_{i=1}^{N} p_{t,i}w_i - m_t + d_t = 0$. This should hold for each time period $t \in \{1, \dots, T\}$. In matrix notation this becomes:

$$\mathbf{Pw} - \mathbf{I}_T\mathbf{m} + \mathbf{I}_T\mathbf{d} = \mathbf{0}_{T\times 1}, \tag{4.6}$$

with $\mathbf{I}_T$ being an identity matrix of dimension $T \times T$ and $\mathbf{0}_{T\times 1}$ a matrix of zeros of dimension $T \times 1$. As $\mathbf{x} = [\mathbf{w}\ \mathbf{m}\ \mathbf{d}]^{\mathrm{T}}$, this can also be written as:

$$[\mathbf{P} \quad -\mathbf{I}_T \quad \mathbf{I}_T]\,\mathbf{x} = \mathbf{0}_{T\times 1}. \tag{4.7}$$

The matrix on the left-hand side (that is being multiplied by $\mathbf{x}$) is now of dimension $T \times (N + 2T)$. It is important to stress that this thus defines the relation between portfolio value, rolling maximum and drawdown for each point in time.

2) Rolling maxima

Next, the rolling maxima need to be defined. A rolling maximum can by definition not get smaller over time and thus $m_t \leq m_{t+1}$ holds for all t. We define an auxiliary matrix $\mathbf{R}$ as:

$$\mathbf{R}_{T\times T} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & 0 & -1 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{4.8}$$

If we then force $\mathbf{R}_{T\times T}\mathbf{m} \geq 0$, we get $-m_t + m_{t+1} \geq 0$, which is the specification we are trying to achieve. One should pay attention to the last row of the matrix being redundant (because the last rolling maximum does not have a succeeding value to compare against), but it is included to keep things simple and elegant. In order to fit this in our existing framework, we define the following set of constraints by adding zeros left and right of $\mathbf{R}$ to form a matrix of dimension $T \times (N + 2T)$:

$$[\mathbf{0}_{T\times N} \quad \mathbf{R}_{T\times T} \quad \mathbf{0}_T]\,\mathbf{x} \geq \mathbf{0}_{T\times 1}. \tag{4.9}$$

3) Full investment

We want the weights allocated to each of the assets to sum up to one. If we would not include this, the optimal solution would just be to assign a weight of zero to each asset. In order to assure this, we only need a matrix of dimension $1 \times (N + 2T)$:

$$[\underbrace{1 \ \cdots \ 1}_{N} \ \underbrace{0 \ \cdots \ 0}_{2T}] \ \mathbf{x} = 1. \tag{4.10}$$

The preceding constraints theoretically suffice, but assuming a traditional long-only investor, we will explicitly add some additional constraints.

4) No short sales

We do not want our model to give negative weights to any of the assets. Shorting is often considered too expensive for retail investors, while most institutional ones are not allowed to. We need a matrix of dimension $N \times (N + 2T)$ to cover all of the assets:

$$[\mathbf{I}_N \quad \mathbf{0}_{N\times 2T}] \ \mathbf{x} \geq \mathbf{0}_{N\times 1}. \tag{4.11}$$

5) Non-negativity

Most optimization packages will assume this by standard for all variables, but we show it explicitly:

$$p_{t,i}, \ d_t, \ m_t \geq 0. \tag{4.12}$$

These sets of constraints define the full linear program for minimizing the sum of the drawdowns and therefore also the pain index. Notice how the LP does not consider the return of the portfolios.

4.1.3 General LP

It is interesting for the reader to understand how our LP fits inside a standard LP framework, e.g. as defined by Luenberger and Ye (2016):

$$\begin{aligned} \text{minimize} \quad & \mathbf{cx}, \\ \text{subject to} \quad & \mathbf{Ax} \geq \mathbf{b}, \\ & \mathbf{x} \geq 0. \end{aligned} \tag{4.13}$$

We have covered the definition of the goal function before. $\mathbf{A}$ has become a large matrix with $2T + N + 1$ rows (or thus individual constraints) and $2T + N$ columns (matching the length of $\mathbf{x}$), while $\mathbf{b}$ is a column vector with $2T + N + 1$ elements. These numbers exclude the non-negativity constraints.

4.2 Example

In order to demonstrate the linear program it is applied it to a real-life example, using the Gurobi software[3] in R to solve the optimization. Part of the used code is available in attachment C for the interested reader.

4.2.1 In-sample estimation

To estimate our model, we take a sample of ten stocks traded on Euronext Brussels over the four year time period from 01/01/2009 until 31/12/2012. Data was downloaded from Yahoo Finance (daily closing prices) using the **quantmod** package of Ryan and Ulrich (2020) and visualisation are still made using **ggplot2**. The normalized paths (i.e. all starting at 100 after division by their initial value) of the securities are shown in figure 6 with their respective ticker symbols. When examining the data, we see a wide range of different paths, with some having more of a steady growth and others being more volatile.

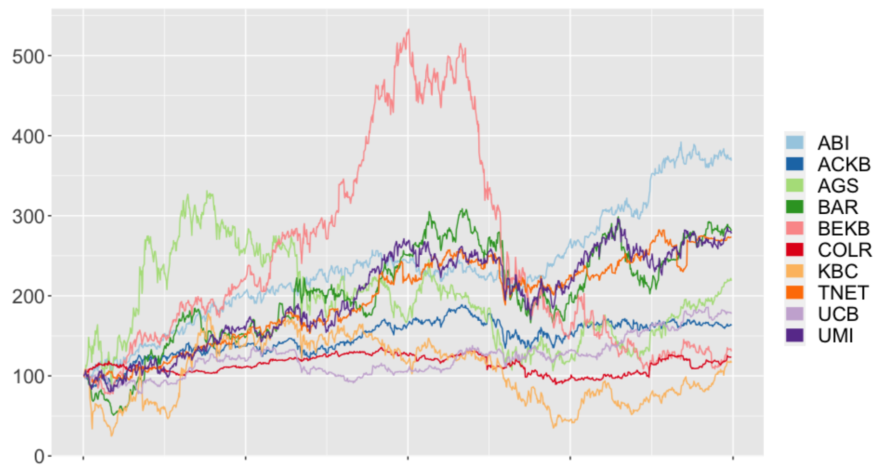
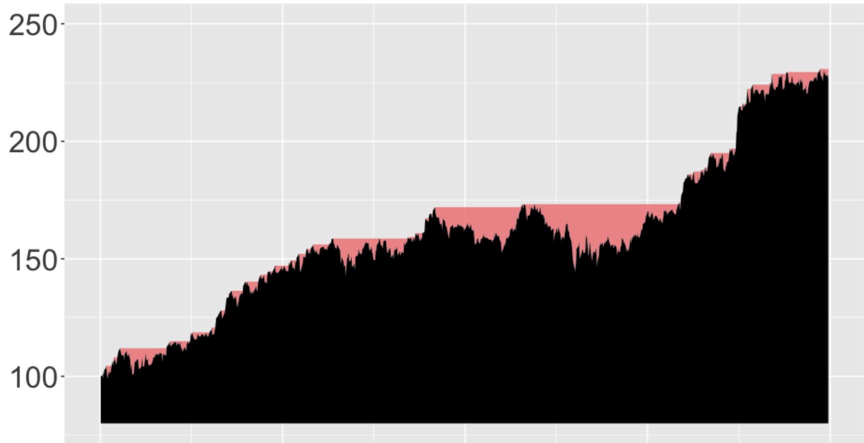


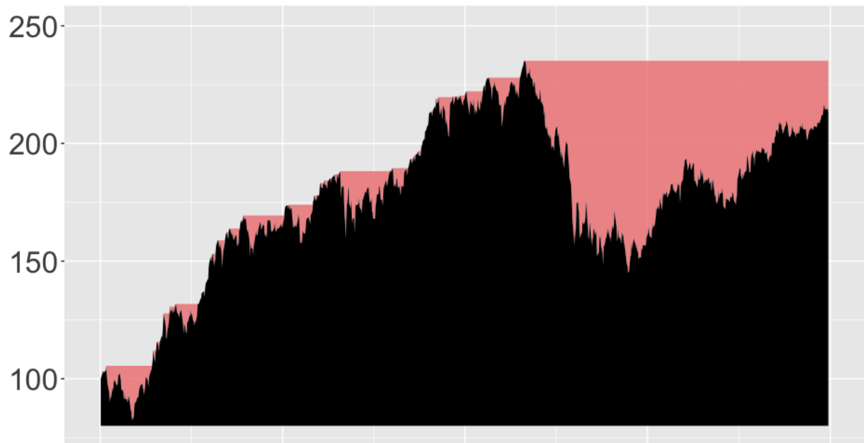
Figure 6: Normalized paths of the securities in the sample, 01/01/2009 - 31/12/2012

Next, we run our LP to minimize the pain index: the resulting weights are shown in table 3, denoted as MPI (minimal pain index). The outcome shows what can be the downside of running an unconstrained optimization: we get quite concentrated positions, with only a weight assigned to four out of ten securities. What would have been the resulting path of the optimal portfolio is shown in figure 7a. We see that even in our optimal portfolio there seems to be a dip at around the middle of our time horizon which cannot be fully 'solved'.

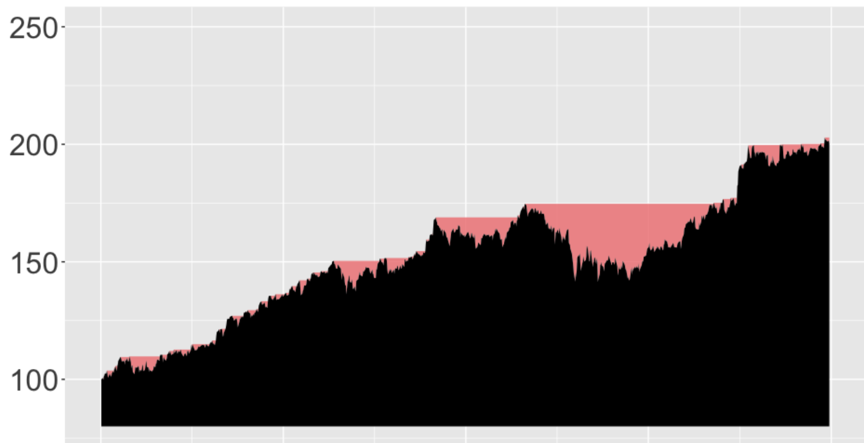
[3]https://www.gurobi.com/products/gurobi-optimizer/



(a) MPI



(b) EW



(c) GMV

Figure 7: IS performance of the different portfolios

	ABI	ACKB	AGS	BAR	BEKB	COLR	KBC	TNET	UCB	UMI
MPI	0.3144	0.0000	0.0000	0.0000	0.0000	0.3173	0.0000	0.0647	0.3036	0.0000
EW	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000
GMV	0.1427	0.0465	0.0000	0.0098	0.0000	0.4295	0.0000	0.1963	0.1752	0.0000

Table 3: Weights different portfolios

Next, a comparison is made to the performance other weighting schemes. A simple approach would be to form an *equally-weighted portfolio* (EW), also referred to as naive diversification, in which each security receives the same weight. The path generated by this portfolio (figure 7b) clearly suffers from more pain compared to our optimized portfolio and additionally, it reaches a lower end value and thus has a lower average return.

We also want to compare to the use of variance or standard deviation as a measure of risk. By for example choosing the *global minimum variance* (GMV) portfolio, we get the path on figure 7c. The result is quite close to our proposed minimal pain index portfolio (also similar in terms of weights), but this is not necessarily a problem: the disadvantages of using variance have been covered before (i.e. QP is needed to find this portfolio). Notice how neither of the three methods presented here takes into account past return, as the comparison makes most sense doing so and does not bias against any of the methods.

An overview of the IS performance of the different portfolios is presented in table 4, which includes the geometric average return, daily standard deviation and pain index for each portfolio. The numbers make sense: the MPI portfolio has the lowest pain index, but the GMV outperforms it if we would consider standard deviation to be the measure of risk.

	Average return (%)	Daily standard deviation (%)	Pain index
MPI	22.67	1.11	5.61
EW	20.79	1.62	23.98
GMV	18.96	1.00	6.67

Table 4: IS performance

4.2.2 Out-of-sample backtest

Using the results of the different optimal portfolios determined over an estimation period, we can also consider their performance in an out-of-sample (OOS) experiment. The subsequent time period of four years is displayed in figure 8 and the resulting paths for the different portfolios in figure 9. The paths seem quite similar, but what is most noticeable is the fact that the portfolio which was designed to have the lowest pain index now has the highest pain index. This is a (very) limited sample, yet it may hint that pain indices, or drawdowns in general, are not inherent to assets. At the end of the day, past performance is not indicative of future results.

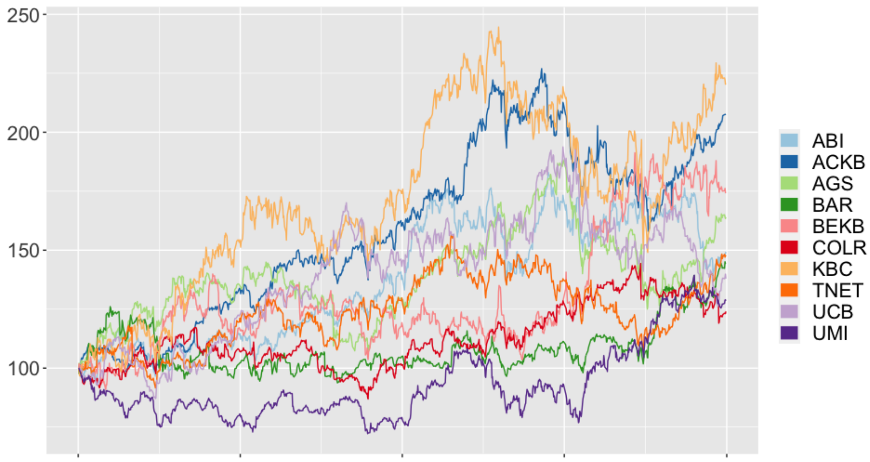


Figure 8: OOS normalized paths of the securities in the sample, 01/01/2013 - 31/12/2016

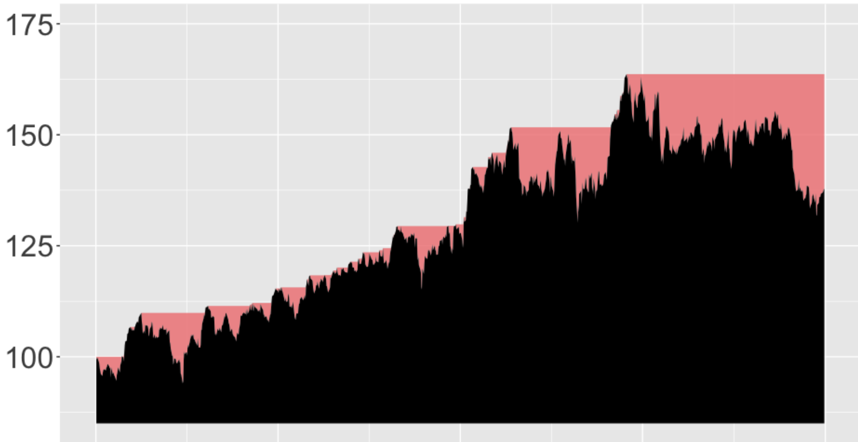
4.3 Risk-return space

4.3.1 An analogy to Markowitz

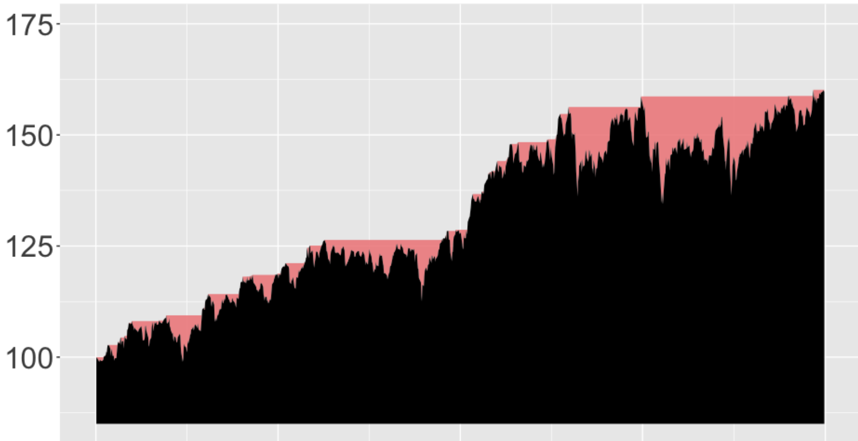
Another interesting analogy to classical portfolio theory is to plot a graph with all possible portfolios, with the x-axis representing the pain index, thus risk, and the y-axis corresponding to the return. As mentioned before, the x-axis would represent a standard deviation or variance in a Markowitz framework. We can show the feasible 'cloud' of portfolios (although an expert in nephology would argue that clouds are not to be found in (a) space) by plotting a large amount of random ones.

When 5000 portfolios consisting of random weights from a universe of securities are generated (basically performing some sort of Monte Carlo simulation), we get the graphical results as shown in figure 10. The graph includes a color scale with the pain ratio. By assuming a risk-free rate of zero, the pain ratio simplifies and becomes return divided by pain index: this relationship is also noticeable visually. The sample under examination seems to be a rather bull period, with most of the portfolios having an average return of more than 10% per year.

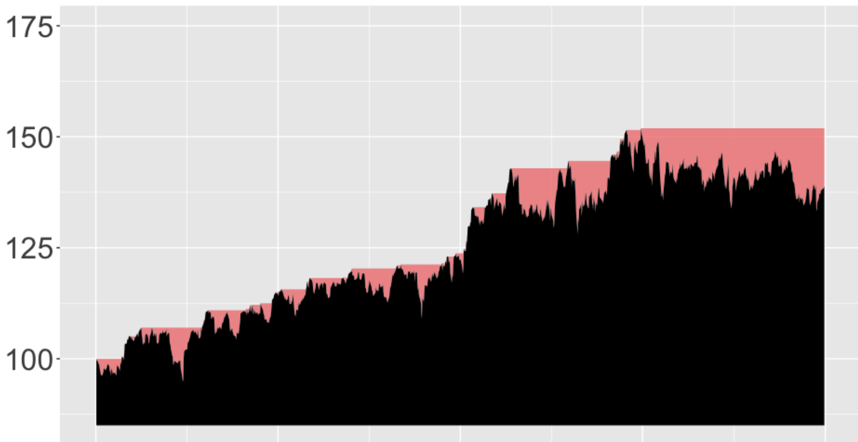
In this example, we see a clear pattern with an optimal portfolio in terms of minimum pain index on the left, similar to the bullet-like shape in MPT risk-return graphs. Deviating from this



(a) MPI



(b) EW



(c) GMV

Figure 9: OOS performance of the different portfolios

point to a point of either a higher or lower return leads to a higher pain index. The area ’in’ this curve contains all of the feasible portfolios. Other samples sometimes give more extreme (i.e. a less symmetrical shape) results, but this intrinsic trade-off in risk and return from the point of minimal pain index is persistent. Fully random portfolios deliver the expected results: the portfolios seem to be distributed with higher concentrations in the middle, yet we can observe some ’clustering’ in the area of where the efficient frontier will be drawn next.

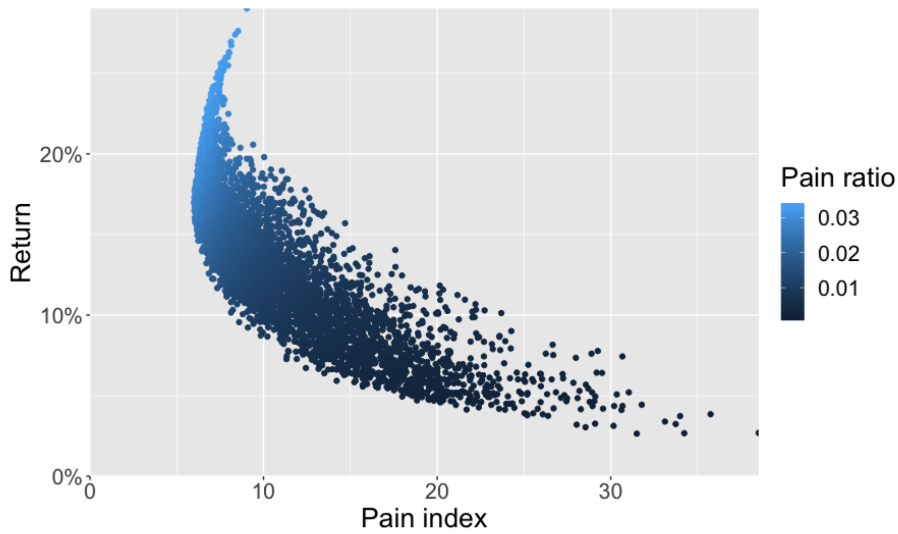


Figure 10: Risk-return space with pain ratios

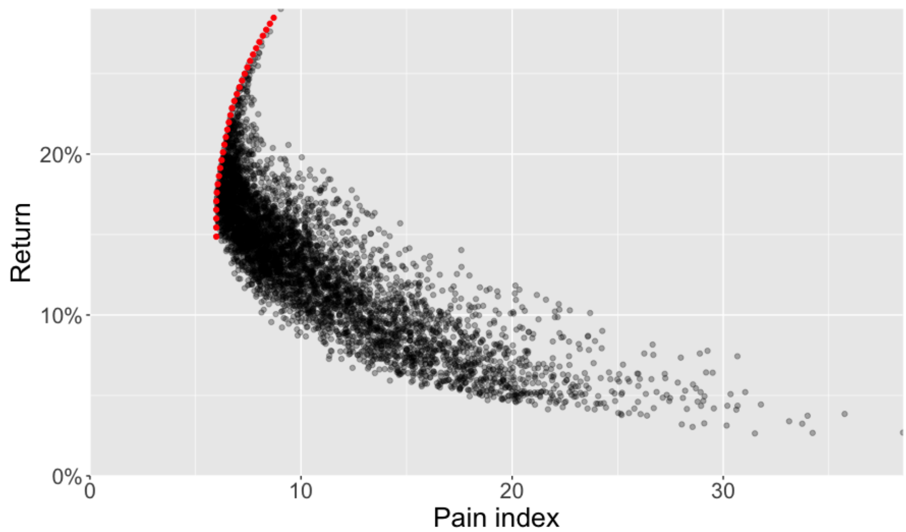


Figure 11: Risk-return space with efficient frontier

4.3.2 Efficient frontier

In the figure shown in the previous subsection, better portfolios can be found in the upper-left quadrant, as those corresponds to a high return and a low pain index concurrently. The 'best' portfolios are then located on a curve starting from our minimal pain index (MPI) portfolio, moving along our field of possible portfolios. These portfolios are called *efficient* and the resulting curve is then called the *efficient frontier*. These portfolios thus offer the highest return for a certain pain index or, vice versa, the lowest pain index for a certain return.

These portfolios can be found numerically, by adding a constraint to the original linear program which forces the return to be at least equal to a certain value: $m_T - d_T \geq goal$. This forces the end value (i.e. at time T) to be at least equal to the goal. The added constraint can thus be formulated as follows:

$$[\underbrace{0 \;\cdots\; 0}_{N+T-1} \quad 1 \quad \underbrace{0 \;\cdots\; 0}_{T-1} \quad -1]\,\mathbf{x} \quad \geq \quad goal. \tag{4.14}$$

The 1 at position $(N+T)$ is multiplied by the last rolling maximum, while the -1 is multiplied by the last drawdown. Repeating this process for different values of the return then creates the frontier. The resulting frontier can be shown in the risk-return space, see figure 11. The leftmost/lowest point on the efficient frontier thus represents the MPI portfolio.

There is also an alternative approach, as mentioned by Chang, Meade, Beasley, and Sharaiha (2000) for the framework of Markowitz. In order to determine the efficient frontier, they introduce a weighting parameter λ $(0 \leq \lambda \leq 1)$ and subsequently solve an optimization with the following objective function:

$$\min_{\mathbf{w}} \quad \lambda \left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij}\right] - (1-\lambda)\left[\sum_{i=1}^{N} w_i \mu_i\right]. \tag{4.15}$$

By doing so, the case of $\lambda = 0$ corresponds to maximizing return (and will result in a portfolio consisting of a weight of 100% assigned to the security with the highest return), while $\lambda = 1$ corresponds to minimizing risk under the form of variance of the portfolio. Values between these two thresholds include a trade-off between risk and return and thus generate the efficient frontier. An analogue approach in our framework would be to optimize the following function:

$$\min_{\mathbf{w}} \quad \lambda\left[PI_{ptf}\right] - (1-\lambda)\left[\sum_{i=1}^{N} w_i \mu_i\right]. \tag{4.16}$$

4.3.3 Two-fund separation theorem

To complete the analogy to the Markowitz framework, we can employ an approach based on the two-fund separation theorem (Tobin, 1958), as covered in subsection 2.1.3. The optimal portfolio was initially defined as having the maximal Sharpe ratio (Sharpe, 1964) and thus becomes the maximal pain ratio in our framework. The equilibrium price states of the initial two-fund theory have later led to the Capital Asset Pricing Model (Lintner, 1965). In order to visualize the possible optimal portfolios, we thus need a line connecting the risk-free asset and our optimal portfolio. This line is then called the Capital Allocation Line (CAL), as it shows how an investor should distribute his or her capital between the risk-free and optimal risky asset. The asset allocation problem has become a two-step process: start by finding the portfolio with the highest pain ratio and then decide on your personal optimal mix of that portfolio and the risk-free asset.

Figure 12 uses a risk-free rate of 0%, while figure 13 assumes a rate of 5%. This causes a change in the pain ratios (notice the change in the legend) and shifts the CAL (i.e. moving upwards as well as changing slope). The risk-free asset is assumed to have a pain index equal to zero: it should never lose value.

It may seem short-sighted to bluntly consider the portfolio with the highest pain ratio as being optimal. After all, the underlying assumption then seems to be that an investor values the trade-off of pain and return proportionally. The trade-off might not be linear and investors could require less (or more) than proportional compensation for additional risk. Consider for example the following comparison of two pain ratios:

$$\frac{5\% - 0\%}{20} \leq \frac{3\% - 0\%}{10}. \tag{4.17}$$

Although theoretically, some investors could prefer the former option, this is not a problem. Under the assumption that the investor can borrow and lend at the risk-free rate, the portfolios on the CAL outrank the whole space of feasible portfolios. This is why the portfolio with the maximal pain index (or originally the maximal Sharpe ratio (Lintner, 1965)) can be considered the optimal portfolio.

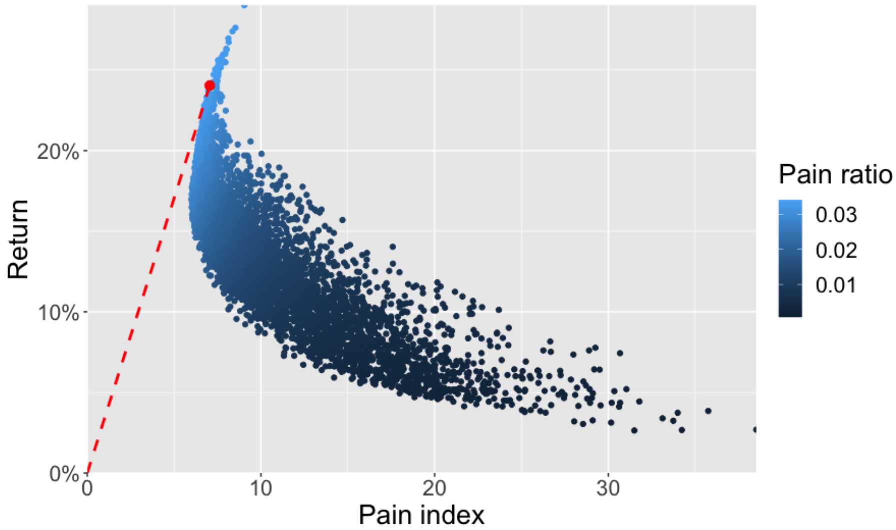


Figure 12: Optimal portfolio connected to risk-free portfolio ($r_f = 0\%$)

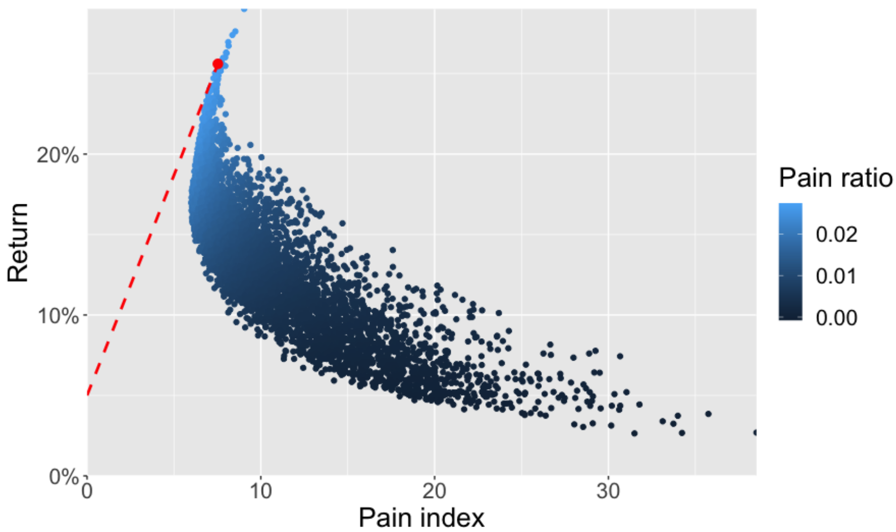


Figure 13: Optimal portfolio connected to risk-free portfolio ($r_f = 5\%$)

5 Conclusions

We believe our contribution to the literature has been twofold. Firstly, we are the first to thoroughly analyze the pain index and its properties and secondly, we have studied how the pain index can be optimized using a linear program and presented its possibilities and properties. Overall, we believe in its usefulness in portfolio management, mainly by virtue of its linear optimization, but also acknowledge its limitations.

We started by considering the current body of research, focusing on the inception of variance in the MPT and other risk measures that countered its disadvantages. There was also a section dedicated to LP, to understand its benefits over QP, but also to provide the reader a basis to discover other applications of it in portfolio optimization.

We then covered the pain index in more detail, focusing on both intuitive properties as well as its application to more formal risk frameworks. We did so with the intention to help the reader understand the effects of the definition of the pain index, as well as to situate it in the current spectrum as being a path-dependent deviation measure.

The essence of this paper was then covered in the section on the methodology to minimize in pain index. We showed its functioning in a short sample, which had small dimensions but served an illustrative purpose. Lastly, we tried to draw parallels between MPT and the pain index framework by exploring the latter's risk-return space.

Further research could look deeper into the linear programming performance in an out-of-sample experiment and compare it to other portfolio optimization methods. One could also consider how the pain index performs under certain assumptions, e.g. log-normal returns. We did not make any assumptions on the distribution of returns but instead employed a more empirical view by basing our optimal portfolio on the historical paths. Another idea would be to explore which values of the pain index and/or ratio are common or 'normal' for different types of assets during different time periods, i.e. explore whether or at least to which degree a pain index is inherent to a certain security. We also believe that the pain index can be used in the broad spectrum of optimization schemes designed for variance like *naive pain parity* or *equal pain contribution* portfolios, by leveraging the Euler decomposition.

Similarly, it may be interesting to research to what extent minimal pain index and minimal variance portfolios are alike. If it were the case, it would not be a problem but rather an advantage, considering the properties of LP and QP. We also believe that the linear program proposed may be valuable for applications outside of portfolio optimization, or could be modified to e.g. maximizing rallies instead of minimizing drawdowns. By adapting the objective function, one can also easily minimize the end-of-period drawdown, a linearly-weighted drawdown or any other weighting scheme.

We do want to end this paper with a warning note. Markowitz and the ideological offspring he provoked (including what we presented) are both built on a common assumption. The underlying idea is that the lines on a graph on past performance (or numbers in a spreadsheet - whatever you prefer) will be representative of the future, by the magic of some sort of invisible hand. In the real world, this underlying assumption rarely holds: there is absolutely nothing that guarantees this behavior. They are interesting to study, but should mainly serve an ex-post analysis function. MPT has apparently gained popularity because of the fact that quantitative portfolio managers liked it, and therefore it became part of the standard investing toolbox. This institutional factor now forces investment professionals to adhere to it because not doing so would not be appreciated by their managers. An investor should never solely rely on these frameworks, but should in the long run probably also consider the quality of the underlying asset in more of a value investing approach.

As an final reminder, we would like the reader to remember the following words by Odo (2011, p. 18), on the idea of a *one-size-fits-all* risk measure:

> *Just as no rational investor would place all of their money in a single investment, no analyst should put all of his or her faith in a single risk measure or ratio.*

References

Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, *26*(7), 1487–1503.

Alexander, G., & Baptista, A. (2004). A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model. *Management Science*, *50*(9), 1261–1273.

Alexander, G., & Baptista, A. (2006). Portfolio selection with a drawdown constraint. *Journal of Banking & Finance*, *30*(11), 3171–3189.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1997). Thinking coherently. *RISK*, *10*(11), 68–71.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203–228.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., & Ku, H. (2007). Coherent multiperiod risk adjusted values and Bellman's principle. *Annals of Operations Research*, *152*(1), 5–22.

Auer, B. (2015). Does the choice of performance measure influence the evaluation of commodity investments? *International Review of Financial Analysis*, *38*, 142–150.

Auer, B., & Schuhmacher, F. (2013). Robust evidence on the similarity of Sharpe ratio and drawdown-based hedge fund performance rankings. *Journal of International Financial Markets, Institutions and Money*, *24*(1), 153–165.

Bacon, C. R. (2008). *Practical portfolio performance measurement and attribution* (2nd ed.). New York: John Wiley & Sons.

Blume, M. E., & Friend, I. (1975). The asset structure of individual portfolios and some implications for utility functions. *The Journal of Finance*, *30*(2), 585–603.

Burden, R. L. (2011). *Numerical analysis* (9th ed.). Boston: Brooks/Cole.

Burke, G. (1994). A sharper Sharpe ratio. *Futures*, *23*(3), 56.

Caporin, M., & Lisi, F. (2011). Comparing and selecting performance measures using rank correlations. *Economics E-Journal*, *5*(10), 1–34.

Cesarone, F., Scozzari, A., & Tardella, F. (2009). Efficient algorithms for mean-variance portfolio optimization with hard real-world constraints. *G. Ist. Ital. Attuari*, *72*.

Chang, T.-J., Meade, N., Beasley, J. E., & Sharaiha, Y. M. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, *27*(13), 1271–1302.

Chekhlov, A., Uryasev, S., & Zabarankin, M. (2005). Drawdown measure in portfolio optimization. *International Journal of Theoretical and Applied Finance*, *8*(1), 13–58.

Chekhlov, A., Uryasev, S., & Zabarankin, M. (2004). Portfolio optimization with drawdown constraints. In P. Pardalos, A. Migdalas, & G. Baourakis (Eds.), *Supply chain and finance* (pp. 209–228). World Scientific.

Chopra, V. K., & Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, *2*(19), 6–11.

Cvitanic, J., & Karatzas, I. (1997). On portfolio optimization under drawdown constraints. *IMA Lecture Notes in Mathematics & Applications*, *65*(1), 35–45.

Davidsson, M. (2012). Drawdown constraints and portfolio optimization. *Journal of Finance and Investment Analysis*, *1*(1), 93–105.

Dowd, K. (2007). *Measuring market risk* (2nd ed.). Chichester: John Wiley & Sons.

Eling, M. (2008). Does the measure matter in the mutual fund industry? *Financial Analysts Journal*, *64*(3), 54–66.

Eling, M., & Schuhmacher, F. (2007). Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking & Finance*, *31*(9), 2632–2647.

Embrechts, P. (2000). Extreme value theory: Potential and limitations as an integrated risk management tool. *Derivatives Use, Trading & Regulation*, *6*(1), 449–456.

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, *38*(1), 34–105.

Foellmer, H., & Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, *6*(4), 429–447.

Garcia, C., & Gould, F. (1987). An empirical study of portfolio insurance. *Financial Analysts Journal*, *43*(4), 44–54.

Goldberg, L., & Mahmoud, O. (2017). Drawdown: from practice to theory and back again. *Mathematics and Financial Economics*, *11*(3), 275–297.

Goldstein, D., & Taleb, N. (2007). We don't quite know what we are talking about when we talk about volatility. *The Journal of Portfolio Management*, *33*(4), 84–86.

Green, R. C., & Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified? *The Journal of Finance*, *47*(5), 1785–1809.

Grossman, S., & Zhou, Z. (1993). Optimal investment strategies for controlling drawdowns. *Mathematical Finance*, *3*(3), 241–276.

Guerard, J. B. (2009). *Handbook of portfolio construction: Contemporary applications of Markowitz techniques*. New York: Springer.

Harris, C. R., & Jenkins, M. (2006). Gender differences in risk assessment: Why do women take fewer risks than men? *Judgement and Decision Making*, *1*(1), 48–63.

Haugh, M. (2010). *Risk measures, risk aggregation and capital allocation.* Retrieved May 10, 2021, from http://www.columbia.edu/~mh2078/RiskMeasures.pdf

Hayes, B. (2006). Maximum drawdowns of hedge funds with serial correlation. *The Journal of Alternative Investments*, *8*(4), 26–38.

Heidorn, T., Kaiser, D., & Roder, C. (2009). The risk of funds of hedge funds: An empirical analysis of the maximum drawdown. *The Journal of Wealth Management*, *12*(2), 89–100.

Horvitz, J. E., & Wilcox, J. W. (2007). Back to Markowitz. *The Journal of Wealth Management*, *10*(1), 43–53.

Investsuite. (2021). *Portfolio optimizer.* Retrieved April 27, 2021, from https://www.investsuite.com/portfoliooptimizer

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, *58*(4), 1651–1683.

Kane, S., Bartholomew-Biggs, M. C., Cross, M., & Dewar, M. (2009). Optimizing omega. *Journal of Global Optimization*, *45*(1), 153–167.

Kapsos, M., Zymler, S., Christofides, N., & Rustem, B. (2014). Optimizing the omega ratio using linear programming. *Journal of Operational Risk*, *17*(4), 49–57.

Keating, C., & Shadwick, W. F. (2002). A universal performance measure. *Journal of performance measurement*, *6*(3), 59–84.

Kellerer, H., Mansini, R., & Speranza, M. (2000). Selecting portfolios with fixed costs and minimum transaction lots. *Annals of Operations Research*, *99*(1), 287–304.

Kestner, L. N. (1996). Getting a handle on true performance. *Futures*, *25*(1), 44–47.

Konno, H. (1990). Piecewise linear risk function and portfolio optimization. *Journal of the Operations Research Society of Japan*, *33*(2), 139–156.

Konno, H., & Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management science*, *37*(5), 519–531.

Korn, O., Möller, P. M., & Schwehm, C. (2019). *Drawdown measures: Are they all the same?* (CFR Working Paper, No. 19-04). Retrieved from SSRN website: https://ssrn.com/abstract=3473694

Lemahieu, E. (2020). *Diversification in an iVaR framework* (Master's dissertation, King's College London, London, United Kingdom).

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, *47*(1), 13–37.

Luenberger, D. G., & Ye, Y. (2016). *Linear and nonlinear programming* (4th ed.). New York: Springer.

Magdon-Ismail, M., Atiya, A. F., Pratap, A., & Abu-Mostafa, Y. S. (2004). On the maximum drawdown of a Brownian motion. *Journal of Applied Probability*, *41*(1), 147–161.

Mahmoud, O. (2017). The temporal dimension of risk. *Journal of Risk*, *19*(3), 57–83.

Mansini, R., Ogryczak, W., & Speranza, M. G. (2003). On LP solvable models for portfolio selection. *Informatica*, *14*(1), 37–62.

Mansini, R., Ogryczak, W., & Speranza, M. G. (2014). Twenty years of linear programming based portfolio optimization. *European Journal of Operational Research*, *234*(2), 518–535.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91.

Markowitz, H. (1959). *Portfolio selection: Efficient diversification of investments*. New York: Wiley.

Markowitz, H. (2010). Portfolio theory: As I still see it. *Annual Review of Financial Economics*, *2*(1), 1–23.

Martin, P. G., & McCann, B. B. (1989). *The investor's guide to fidelity funds*. New York: Wiley.

Meilijson, I. (2003). The time to a given drawdown in Brownian motion. In J. Azéma, M. Émery, M. Ledoux, & M. Yor (Eds.), *Séminaire de Probabilités XXXVII* (pp. 94–108). Springer.

Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, *45*(1), 31–42.

Möller, P. M. (2018). Drawdown measures and return moments. *International Journal of Theoretical and Applied Finance*, *21*(7), 1–42.

Molyboga, M., & L'Ahelec, C. (2016). Portfolio management with drawdown-based measures. *The Journal of Alternative Investments*, *19*(3), 75–89.

Montier, J. (2009). *Behavioural investing: A practitioner's guide to applying behavioural finance*. New York: Wiley.

Odo, M. (2011). *Pain index and pain ratio*. Retrieved April 29, 2021, from https://www.styleadvisor.com/content/pain-index-and-pain-ratio

Odo, M. (2016). *The pain ratio — A better risk/return measure*. Retrieved April 29, 2021, from https://www.swanglobalinvestments.com/pain-ratio-better-risk-return-measure/

Odo, M. (2020). *Pain index: A better measure of risk*. Retrieved April 29, 2021, from https://www.swanglobalinvestments.com/pain-index-better-measure-of-risk/

Ornelas, J., Silva Junior, A. F., & Fernandes, J. L. B. (2012). Yes, the choice of performance measure does matter for ranking of US mutual funds. *International Journal of Finance & Economics*, *17*(1), 61–72.

Paté-Cornell, M.-E. (2012). On "black swans" and "perfect storms": Risk analysis and management when statistics are not enough. *Risk Analysis*, *32*(11), 1823–1833.

Richardson, M., & Smith, T. (1993). A test for multivariate normality in stock returns. *Journal of Business*, *66*(2), 295–321.

Rockafellar, R. T. (1970). *Convex analysis* (1st ed.). Princeton: Princeton University Press.

Rockafellar, R. T., Uryasev, S., & Zabarankin, M. (2006). Generalized deviations in risk analysis. *Finance and Stochastics*, *10*(1), 51–74.

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, *2*(3), 21–42.

Rockafellar, R. T., & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, *26*(7), 1443–1471.

Ryan, J. A., & Ulrich, J. M. (2020). *Package 'quantmod'*. Retrieved May 3, 2021, from http://cran.hafro.is/web/packages/quantmod/quantmod.pdf

Schuhmacher, F., & Eling, M. (2011). Sufficient conditions for expected utility to imply drawdown-based performance rankings. *Journal of Banking & Finance*, *35*(9), 2311–2318.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425–442.

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, *39*(1), 119–138.

Sharpe, W. F. (1971a). A linear programming approximation for the general portfolio analysis problem. *Journal of Financial and Quantitative Analysis*, *6*(5), 1263–1275.

Sharpe, W. F. (1971b). Mean-absolute-deviation characteristic lines for securities and portfolios. *Management Science*, *18*(2), 1–13.

Simaan, Y. (1997). Estimation risk in portfolio selection: The mean variance model versus the mean absolute deviation model. *Management Science*, *43*(10), 1437–1446.

Steiner, A. (2011). *Drawdown-at-risk Monte Carlo optimization* (Research note). Retrieved from SSRN website: https://ssrn.com/abstract=1782664

Taleb, N. N. (2008). *The black swan: The impact of the highly improbable* (1st ed.). London: Random House.

Tobin, J. (1958). Liquidity preference as behavior towards risk. *The Review of Economic Studies*, *25*(2), 65–86.

Treynor, J. (1965). How to rate management of investment funds. *43*(1), 63–75.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(2), 180–185.

Wilmott, P. (2007). *Paul Wilmott introduces quantitative finance* (2nd ed.). Chichester: John Wiley & Sons.

Yitzhaki, S. (1982). Stochastic dominance, mean variance, and Gini's mean difference. *The American Economic Review*, *72*(1), 178–185.

Young, M. R. (1998). A minimax portfolio selection rule with linear programming solution. *Management Science*, *44*(5), 673–683.

Young, T. W. (1991). Calmar ratio: A smoother tool. *Futures*, *20*(1), 40.

A Diversification in an iVaR framework

Lemahieu (2020) researched diversification properties of the iVaR risk measure (Investsuite, 2021), which is equal to the pain index. In this attachment, we show some of the concepts he introduced.

First, the coiVaR is defined, a measure of pairwise association (like the Pearson correlation coefficient is), as the ratio of the risk of an equally weighted portfolio of asset X and Y over the sum of the risk of holding X and Y separately:

$$coiVaR(X,Y) = \frac{iVaR(X+Y)}{iVaR(X) + iVaR(Y)}. \tag{A.1}$$

As the iVaR is subadditive and non-negative, it has to lay between the following thresholds:

$$0 \leq coiVaR(X,Y) \leq 1. \tag{A.2}$$

If two assets have a coiVaR close to 0 it thus means their combined path is close to monotonic growth (low iVaR in the numerator), while their individuals paths are not (high iVaR's in the denominator).

By calculating all possible combinations in a certain space of N assets, one can then form a coiVaR matrix:

$$\mathbf{C} = \begin{pmatrix} 1 & coiVaR_{1,2} & \cdots & coiVaR_{1,N} \\ coiVaR_{2,1} & 1 & & \\ \cdots & & \cdots & \\ coiVaR_{N,1} & \cdots & & 1 \end{pmatrix}. \tag{A.3}$$

The iVaR diversification ratio is another possible measure of diversification:

$$D_{iVaR}(\mathbf{w}) = \frac{\sum_i^N w_i \; iVaR_i}{iVaR_p(\mathbf{w})}. \tag{A.4}$$

Next, the problem of portfolio optimization is considered, starting with an inverse iVaR portfolio (IIP):

$$\mathbf{w} = (\frac{1/iVaR_1}{\sum_i^N (1/iVaR_i)}, \cdots, \frac{1/iVaR_N}{\sum_i^N (1/iVaR_i)}). \tag{A.5}$$

The author acknowledges the fact that this approach is very naive as it ignores the mechanics by which portfolio iVaR is formed, yet could be useful to serve as e.g. a transparent benchmark. On the contrary, maximizing the iVaR diversification ratio results in the most-diversified iVaR portfolio (MDIP):

$$\max_{\mathbf{w}}(D_{iVaR}(w)), \tag{A.6}$$

which does account for real diversification benefits in portfolios. To penalize the concentrations of assets in terms of iVaR, one can minimize penalized iVaR portfolios (PiVaR). By defining a correction factor $C(w)$ as:

$$C(\mathbf{w}) = \sum_{i}^{N}\sum_{j}^{N} w_i w_j coiVaR(i,j), \tag{A.7}$$

the PiVaR can be minimized:

$$\min_{\mathbf{w}}(iVaR(w) + \lambda C(w)). \tag{A.8}$$

Lastly, the MVMIP minimizes iVaR while maximizing variance:

$$\underset{\mathbf{w}}{max}\left(\frac{\sum_{i}^{N} w_i \sigma_i}{iVaR_p(w)}\right). \tag{A.9}$$

B Drawdown properties

In this attachment we show drawdown properties proven by Mahmoud (2017). We use these concepts to deduct properties of the pain index, as the needed transformations of summing and dividing rarely changes the attributes. By looking at the drawdown *process* (which refers to a vector of drawdowns over an extended period of time, instead of the single-period drawdown measure), we are able to make propositions on the pain index. Mahmoud uses a bit of a variation in the notation (retained in this attachment) by defining the drawdown as follows:

Definition B.0.1 (Drawdown process)**.** *For a horizon $T \in (0,\infty)$, the drawdown process $D^{(X)} = \{D_t^{(X)}\}_{t\in[0,T]}$ corresponding to a stochastic process $X \in \mathcal{R}^\infty$ is defined by*

$$D_t^{(X)} = \overline{X_t} - X_t,$$

where

$$\overline{X_t} = \underset{u\in[0,t]}{sup}\, X_u$$

is the running maximum of X up to time t.

Using this notation, we can now define the attributes next. Proofs for properties 1 through 3 were omitted in the final version of the paper but are retrieved from an earlier working paper.

Definition B.0.2 (Properties of drawdowns)**.** *Given the stochastic process $X \in \mathcal{R}^\infty$, let $D^{(X)}$ be* *the corresponding drawdown process for a fixed time horizon T. Then:*

1. *For all constant deterministic processes $C \in \mathcal{R}^\infty$, $D^{(C)} = 0$.*

2. *For constant deterministic $C \in \mathcal{R}^\infty$, $D^{(X+C)} = D^{(X)}$.*

3. *For $\lambda > 0$, $D^{(\lambda X)} = \lambda D^{(X)}$.*

4. *For $Y \in \mathcal{R}^\infty$ and $\lambda \in [0, 1]$, $D^{(\lambda X+(1-\lambda)Y)} \leq \lambda D^{(X)} + (1-\lambda)D^{(Y)}$.*

Proof.

1. Since $M^{(C)} = C$, we immediately get $D^{(C)} = M^{(C)} - C = 0$.

2. Since for $t \in [0, t]$, $M_t^{(X+C)} = sup_{u\in[0,t]}(X + C)_u = sup_{u\in[0,t]}(X)_u + C = M_t^{(X)} + C$, we have $D^{(X+C)} = M^{(X+C)} - X - C = M^{(X)} + C - X - C = M^{(X)} - X = D^{(X)}$.

3. For $\lambda > 0$, we have for $t \in [0, T]$, $M_t^{(\lambda X)} = sup_{u\in[0,t]}(\lambda X)_u = \lambda sup_{u\in[0,t]}(X)_u = \lambda M_t^{(X)}$, and therefore $D^{(\lambda X)} = \lambda M^{(X)} - \lambda X = \lambda D^{(X)}$.

4. For $\lambda \in [0, 1]$, we clearly have $\overline{\lambda X + (1-\lambda)Y} \leq \lambda\overline{X} + (1-\lambda)\overline{Y}$ by properties of the supremum, and therefore $D^{(\lambda X+(1-\lambda)Y)} = \overline{\lambda X + (1-\lambda)Y} - \lambda X - (1+\lambda)Y \leq \lambda\overline{X} + (1-\lambda)\overline{Y} - \lambda X - (1+\lambda)Y = \lambda D^{(X)} + (1-\lambda)D^{(Y)}$.

□

C R code: optimization

In this attachment we show the R code that is used when optimizing our sample of assets. As mentioned in the text, we use the Gurobi optimizer software. The code should be self-explanatory. In case it is not or one would be interested in any of the other code (used to e.g. generate the figures or to test the OOS performance), feel free to get in touch.

```
# Load the needed packages
library(gurobi)
library(quantmod)
library(ggplot2)
# Following packages are used for the GMV portfolio
library(ROI)
library(ROI.plugin.quadprog)
library(ROI.plugin.glpk)
library(PortfolioAnalytics)


# Function that returns vector of running maxima
```

```r
calculate_running_max <- function(prices) {
  running_max <- rep(0,length(prices))
  prices <- coredata(prices)
  drawdowns <- rep(0,length(prices))
  running_max[1] = prices[1]
  for (j in 2:length(prices)){
    if (prices[j] >= running_max[j-1]){
      running_max[j] = prices[j]
      drawdowns[j] = 0
    }
    else if (prices[j] < running_max[j-1]){
      drawdowns[j] = running_max[j-1]-prices[j]
      running_max[j] = running_max[j-1]
    }
  }
  return(running_max)
}


# Function that returns pain index. Assumes that starting point is given, but should not be taken into account.
calculate_PI <- function(prices) {
  running_max <- 0
  prices <- coredata(prices)
  drawdowns <- rep(0,length(prices))
  for (j in 1:length(prices)){
    if (prices[j] >= running_max){
      running_max = prices[j]
      drawdowns[j] = 0
    }
    else if (prices[j] < running_max)
      drawdowns[j] = running_max-prices[j]
  }
  return(sum(drawdowns[2:length(drawdowns)])/(length(drawdowns)-1))
  # Not taking starting point into account. Could start at 1 instead of 2, but should always be divided by (length - 1)
}


##############################################################

# Load data
```

```
getSymbols('ACKB.BR',src='yahoo')
getSymbols('ABI.BR',src='yahoo')
getSymbols('AGS.BR',src='yahoo')
getSymbols('BAR.BR',src='yahoo')
getSymbols('BEKB.BR',src='yahoo')
getSymbols('COLR.BR',src='yahoo')
getSymbols('KBC.BR',src='yahoo')
getSymbols('TNET.BR',src='yahoo')
getSymbols('UCB.BR',src='yahoo')
getSymbols('UMI.BR',src='yahoo')

# Remove NAs
ACKB.BR$ACKB.BR.Close <- na.locf(ACKB.BR$ACKB.BR.Close)
ABI.BR$ABI.BR.Close <- na.locf(ABI.BR$ABI.BR.Close)
AGS.BR$AGS.BR.Close <- na.locf(AGS.BR$AGS.BR.Close)
BAR.BR$BAR.BR.Close <- na.locf(BAR.BR$BAR.BR.Close)
BEKB.BR$BEKB.BR.Close <- na.locf(BEKB.BR$BEKB.BR.Close)
COLR.BR$COLR.BR.Close <- na.locf(COLR.BR$COLR.BR.Close)
KBC.BR$KBC.BR.Close <- na.locf(KBC.BR$KBC.BR.Close)
TNET.BR$TNET.BR.Close <- na.locf(TNET.BR$TNET.BR.Close)
UCB.BR$UCB.BR.Close <- na.locf(UCB.BR$UCB.BR.Close)
UMI.BR$UMI.BR.Close <- na.locf(UMI.BR$UMI.BR.Close)


# Define matrices
p_temp <- merge(ABI.BR$ABI.BR.Close,ACKB.BR$ACKB.BR.Close,AGS.BR$AGS.BR.Close,BAR.BR
    $BAR.BR.Close,BEKB.BR$BEKB.BR.Close,
                COLR.BR$COLR.BR.Close,KBC.BR$KBC.BR.Close,TNET.BR$TNET.BR.Close,UCB.
                    BR$UCB.BR.Close,UMI.BR$UMI.BR.Close)
# Using only a subset of the full dataset
p <- p_temp[index(p_temp) >"2009-01-01" & index(p_temp) < "2013-01-01"]
colnames(p) <- c("ABI","ACKB","AGS","BAR","BEKB","COLR","KBC","TNET","UCB","UMI")

# Scale the prices to all start at "scale"
scale <- 100
p <-sweep(scale*p, 2, p[1], "/")

# Plotting the sample period
autoplot(p, facet = NULL) +
  theme(text = element_text(size=20), legend.title = element_blank(), axis.text.x =
      element_blank()) + ylab("") + xlab("") +
  scale_color_manual(values=c('#a6cee3','#1f78b4','#b2df8a','#33a02c','#fb9a99','#
```

```
    e31a1c','#fdbf6f','#ff7f00','#cab2d6','#6a3d9a'))+
  guides(colour = guide_legend(override.aes = list(size=5)))

# Defining the variables
# Note that these are represented by capital letters in text,
# but 'T' already represents the boolean 'TRUE', thus we stick to 't' and 'n'
t <- length(p$ACKB)
n <- ncol(p)

##############################################################
### CONSTRAINTS
### We define the matrices on the left-hand side first, denoted a1 to a5
##############################################################

# Defining the drawdowns: a1
a1 <- cbind(as.data.frame(p),
            -diag(t),
            diag(t))
a1 <- matrix(unlist(a1), ncol = ncol(a1), byrow = FALSE)
dim(a1)     # Should be t x (n + 2t)

# Rolling maxes: a2
a2 <- cbind( matrix(0L, nrow = t, ncol = n),
             rbind(cbind(0,diag(t-1))-cbind(diag(t-1),0),0), # Auxiliary matrix
                 formed with some rbind and cbind creativity
             matrix(0L, nrow = t, ncol = t))
dim(a2)    # Should be t x (n + 2t)

# Full investment constraint: a3
a3 <- cbind(matrix(1L, nrow = 1, ncol = n),
            matrix(0L, nrow = 1, ncol = 2*t))
dim(a3)     # Should be 1 x (n + 2t)

# No short sales: a4
a4 <- cbind(diag(n),
            matrix(0L, nrow=n, ncol=2*t))
dim(a4)     # Should be n x (n + 2t)


##############################################################
### GUROBI MODEL
##############################################################
```

```
model <- list()
# We have to provide the left-hand side as one large matrix (see standard LP formulation)
model$A           <- rbind(a1,a2,a3,a4)
# Objective function
model$obj         <- cbind(matrix(0,1,n+t), matrix(1,1,t))
# Minimization problem
model$modelsense <- 'min'
# Right-hand sides of constraints
# Drawdown definition and rolling maxes need 0, full investment: 1, no shorting: 0
model$rhs         <- cbind(matrix(0,1,2*t),
                           1,
                           matrix(0,1,n))
# Relational operators
model$sense       <- cbind(t(rep('=',t)),
                           t(rep('>=',t)),
                           '=',
                           t(rep('>=',n)))
# Run LP
result <- gurobi(model)

########################################################
### Results
########################################################

# Weights
result_weights <- data.frame(name = names(p),weight = result$x[1:n])
ggplot(result_weights, aes(x="Weights", y = weight, fill=name)) + geom_col() +
  theme(text = element_text(size=20), legend.title = element_blank(), axis.text.x = element_blank()) + ylab("") + xlab("")
result_weights

# Graph
value <-  xts(p[1:t] %*% result$x[1:n],time(p))
result_max <- data.frame(date=index(p), rolmax=result$x[(n+1):(n+t)])
ggplot(result_max, aes(x=date,y=rolmax)) +
  geom_ribbon(aes(ymin=value, ymax=rolmax),color="NA", fill="lightcoral",alpha=0.8) +
  geom_ribbon(aes(ymin=80,ymax=value),color="NA", fill="black") + ylab("Value") + xlab(("")) + guides(x = "none")+
  theme(text = element_text(size=30), legend.title = element_blank(), axis.text.x =
```

```
    element_blank()) + ylab("") + xlab("") +
  scale_y_continuous(limits=c(80,250))

calculate_PI(value)

# Standard deviation. First element is NA because of calculation of returns
sd(as.numeric(CalculateReturns(value)[2:length(value)]))
# Average return
(as.numeric(value[length(value)])/100)^(1/4)-1

# Simple check: the sum of the drawdowns...
calculate_PI(value)*(t-1)
# ... is equal to the objective function of our optimization
result$objval

##################################################################
### Equally weighted portfolio
##################################################################
eq_w <- rep(0.1,10)
ptf_eq_w <- xts(p %*% eq_w,time(p))
result_max <- data.frame(date=index(p), rolmax= calculate_running_max(ptf_eq_w))
value <- ptf_eq_w # To be able to simply copy the code
ggplot(result_max, aes(x=date,y=rolmax)) +
  geom_ribbon(aes(ymin=value, ymax=rolmax),color="NA", fill="lightcoral",alpha=0.8) +
  geom_ribbon(aes(ymin=80,ymax=value),color="NA", fill="black") + ylab("Value") + xlab(("")) + guides(x = "none")+
  theme(text = element_text(size=30), legend.title = element_blank(), axis.text.x = element_blank()) + ylab("") + xlab("") +
  scale_y_continuous(limits=c(80,250))


calculate_PI(value)
sd(as.numeric(CalculateReturns(value)[2:length(value)]))
(as.numeric(value[length(value)])/100)^(1/4)-1

##################################################################
### Global min variance (GMV) portfolio
##################################################################
returns_temp <- CalculateReturns(p)
returns <- na.omit(returns_temp)
colnames(returns) <- c("ABI","ACKB","AGS","BAR","BEKB","COLR","KBC","TNET","UCB","
```

```r
    UMI")

# Get a character vector of the fund names
funds <- colnames(returns)
# Create portfolio object
portf_minvar <- portfolio.spec(assets=funds)
# Add full investment + no short constraint to the portfolio object
portf_minvar <- add.constraint(portfolio=portf_minvar, type="full_investment")
portf_minvar <- add.constraint(portfolio=portf_minvar, type="box",
                               min=rep(0,10),
                               max=rep(1,10))
# Add objective to minimize variance
portf_minvar <- add.objective(portfolio=portf_minvar, type="risk", name="var")
opt_gmv <- optimize.portfolio(R=returns, portfolio=portf_minvar,optimize_method="ROI
    ", trace=TRUE)
ptf_min_var <- xts(rowSums(p %*% unname(opt_gmv$weights)),time(p))
result_max <- data.frame(date=index(p), rolmax= calculate_running_max(ptf_min_var))
value <- ptf_min_var
ggplot(result_max, aes(x=date,y=rolmax)) +
  geom_ribbon(aes(ymin=value, ymax=rolmax),color="NA", fill="lightcoral",alpha=0.8)
    +
  geom_ribbon(aes(ymin=80,ymax=value),color="NA", fill="black") + ylab("Value") +
    xlab(("")) + guides(x = "none")+
  theme(text = element_text(size=30), legend.title = element_blank(), axis.text.x =
    element_blank()) + ylab("") + xlab("") +
  scale_y_continuous(limits=c(80,250))
round(opt_gmv$weights,4)

calculate_PI(value)
sd(as.numeric(CalculateReturns(value)[2:length(value)]))
(as.numeric(value[length(value)])/100)^(1/4)-1
```

Optimization.R