Master's thesis 2016

Prospect Utility Portfolio Optimization

Selection and optmization using gradient and regularization techniques

NIKLAS LINDEKE

Lund University
School of Economics and Management
Department of Economics
Supervisor: Birger Nilsson

Abstract

Portfolio choice theory have in the last decades seen a rise in utilising more advanced utility functions for finding optimal portfolios. This is partly a consequence of the relatively simplistic nature of the quadratic utility, which is often assumed in the classical mean-variance framework. There have been some suggestions on how to find optimal portfolios in accordance to more realistic utility functions gathered from Prospect theory. However, some of these methods suffer from practical drawbacks.

This paper proposes a method consisting of a mixture between two optimization techniques, in order to find a portfolio allocation that is optimal in relation to the first four moments. In the empirical implementation, we utilise the S-shaped and Bilinear utility functions gathered from Prospect Theory. Results hold in an in-sample testing environment. Improving expected utility, and the first three moments when tested against a standard benchmark method, as well as in measurement of the Sharpe Ratio.

Keywords— Utility Maximization, Portfolio Choice, Gradient Ascent, Sparse Group LASSO

Contents

1 Introduction

As the pioneer of modern portfolio theory, Markowitz (1952) suggests that investors should not merely place focus on the maximization of expected return, because doing so ignores the principle of diversification. His answer to that problem was to view the variance of a portfolio as something to be negatively considered, and its expected return as something positively so. Subsequently, to find a vector of weights that maximizes expected returns, while retaining a certain level of volatility, constitutes the so called Mean-Variance Maxim.

There are of course some issues with this. The MV framework is, in a sense, a special case of expected utility maximization, where the agent is either assumed to have quadratic utility, or that return distributions are assumed to be normal. It is well observed that financial data typically does not follow these standard normal assumptions. Straightforward comparisons have found that the 'fatness' of the tails, height of the bell peak, and inconsistency with zero skew, imply a statistical incoherence with the normality assumptions. This explanation is, in part, a consequence of the frequency of extreme events (Mandelbrot, 1963), and the fact that busts are more frequent than booms (Fama, 1965).

Furthermore, if one were to assume that an investor sustains a quadratic utility preference profile, variance and expected return would be sufficient to handle the problem of maximizing the utility of a portfolio within the MV framework. But, as is rather clear now, a realistic investor does not comply with a simple quadratic utility preference. Though economists have tried to verify the importance of higher statistical moments, early attempts have proved rather unsuccessful - with studies finding only ambiguous results that could not motivate much improvement for any new allocation strategy (Jondeau and Rockinger, 2006). As optimization methodologies, computational abilities, and data availability have improved, new light has been shed on the subject. This paper will, henceforth, aim to extend some of these concepts, to find a new computationally efficient method for maximizing a more realistic utility of a portfolio.

Kroll et al. (1984) showed how the maximization of expected utility – looking at various types of utility functions and allowing for a fifty percent borrowing rate – will be very evenly matched by the allocation produced by the Mean Variance method, which they accredit to the quadratic approximation of the utility function. Scott and Horvath (1980) demonstrated how the moment order for portfolio preferences holds the inequality that all moments $u^k(w) > 0$ when k is odd and $u^k(w) < 0$ when k is even. Meaning that expected utility is positively dependent on the mean and skewness, while it depends negatively on variance and kurtosis. Athayde and Flores (2002) extend this idea of portfolio optimization and proposed a theory of a higher dimensionality in a utility maximization framework. They suggest a method where even moments are optimized subject to odd moments. Laying important groundwork in the derivation of the objective function in utility maximization depending on higher moments.

In practice, a more complex utility function is necessary as many utility functions yields very little information in the approximation of higher order moments. Thanks to Kahneman and Tversky (1979), the concept of utility as piecewise functions, which in addition to some risk averse or similar utility functions, depends on the condition: whether or not the change in wealth is considered a loss or a gain. Hence, a more complex, and more realistic notion of utility is wrought. These utility functions have subsequently gained popularity over the recent decades (Cremers et al., 2005).

Handling these new ideas, Hagströmer et al. (2008), Adler and Kritzman (2006), and Cremers et al. (2005) take an approach called Full-Scale Optimization (FSO). The idea of FSO is to perform a numerical search over all possible compositions of the portfolio weights, and choose the vector which maximizes the chosen utility function. Promising results were found when the utility function is assumed to be either S-shaped or Bilinear. This is because the FSO will find the optimal solution, though at the cost of using an extremely computationally exhausting method. Some statistical properties other than the expected utility in itself also improves, in comparison to the standard MV optimization (Adler and Kritzman, 2006).

Jondeau and Rockinger (2006) and Harvey et al. (2010) adopts a method more similar to the one suggested by Athayde and Flores (2002), whereby they formulate a different kind of objective function, based on the first four moments, which also - when assuming S-shaped or Bilinear utility, yields some significant improvements in statistical properties but with less computational exhaustion in comparison to the FSO approach.

Previous methods for portfolio optimization will thus forth be considered too simplistic when only expected return and variance are used as inputs. While the FSO and moment based optimization generally seems to find better allocations in terms of utility and other statistical properties (meaning that portfolio risk metrics also improves), it falls short in atoning for issues regarding larger sets of assets. The concept of optimization in those frameworks are rather computationally exhausting, meaning that if the number of assets grows, the methods grows equally impractical at the same pace. Thus a more versatile framework for portfolio selection and optimization is needed.

Thenceforth, we extend the theoretical appeal of the philosophy of utility maximization while, at the same time, confront further impractical issues that the previous methods hold. To do this we suggest a framework which extends as a combination of two optimization processes. One to induce sparsity via selection based on the two first moments – so that the original size of the number of assets used will have less effect in the second process - which will be done using the Sparse Group LASSO methodology, effectively eliminating all non-contributing assets in terms of the expected return and variance. The second step of the process utilizes a gradient-based method for finding the optimal allocation, which will maximize expected utility over the first four moments in combination.

In this paper, we present a successful formulation of a new method for optimizing a portfolio based on utility maximization. We utilise the variable selection properties of the Sparse Group LASSO technique to subsequently incrementally improve that allocation in relation to the gradient of the portfolio utility. We managed to outperform the standard Mean Variance, both in terms of the Sharpe ratio as well as in skewness, however, not in terms of kurtosis. We tested this by doing in-sample comparisons between the two methods. From this we conclude that the added gradient method to the Sparse Group LASSO Mean-Variance improves the allocation further, while not being too computationally exhausting.

The paper will be structured as follows: firstly, in section 2, a motivation of the innate relationship between utility maximization and portfolio theory, before (in section 2.2) a subsequent motivation for using the concept of prospect utility, furthering the idea of the necessity of the use of more complex and realistic utility functions in portfolio optimization. Following this, in section 3.2, is an explanation and motivation of the reasoning behind using variable selection techniques, and the provision of an augmentation of the regularization method called the Sparse Group LASSO. In order to extend this, and incrementally improve the allocation garnered from the Sparse Group LASSO, there follows a motivation of methodology in section 4.2, with explication of a method for approximating the gradient vector – with the discrete partial derivatives such that we can follow the path of the portfolio that yields the greatest effect on the expected utility, in terms of each of the four first moments. After this, in section 5, we present an empirical implementation of the proposed methodology and investigate its relevance by testing it against the MV as a benchmark method.

2 Utility Maximization

In the following section we provide a brief background on utility theory and the most relevant and famous utility functions within the subject. We will furthermore provide some reasoning to why that type of utility function can be considered archaic in a realistic and modern setting, and also how a better utility function can be considered to atone for this. Lastly we will explain how such a theory can be applied in the concept of utility maximization of a portfolio.

2.1 Risk Aversion

The concept of Risk Aversion and its relation to concave functions was first established a long time ago by Bernoulli (1782), but it was not until Pratt (1964) and Arrow (1965), independently and almost simultaneously, pointed out the importance of the first and second derivative (respectively) of a utility function, in order to see whether or not an agent exhibits risk averse tendencies. The following Arrow-Pratt framework came later to be an important foundation in financial theory.

Kimball (1990), and Eeckhoudt and Schlesinger (1995) ventured further in this subject as they did not stop at seeing an important pattern (in regards to risk appetite) in only the first two derivatives, in means to signify non-satiation at $u'(w) > 0$, and risk aversion when $u''(w) < 0$, but also to observe the third derivative as $u'''(w) > 0$, to signify prudence. This is then related to the idea that an agent will thus forth prefer a positive skew in the return distribution. Prudence according to Kimball (1990) is a concept that states that an agent will increase its savings rate when faced with riskier situations. These savings are called precautionary savings. This furthers the concept of risk aversion in the sense that if an agent is only risk averse, it only sheds a dislike towards risk, but a prudent agent will take active steps towards decreasing its exposure to risk in times of higher volatility.

The next order risk attitude is called temperance, contrived by Kimball (1992), and Gollier and Pratt (1996) as the risk attitude constituent that plays a role when faced with exogenous background risk. However not as simple as an aversion towards variance, Eeckhoudt (2012) states more precisely that: when looking at the fourth derivative as $u^{iv}(w) < 0$, the agent is effectively considered to exhibit an aversion towards increasing kurtosis, and is so forth a temperate agent.

This is very intuitive from the perspective of an investor, as aversion towards unnecessary volatility would be considered as straight forward common sense. And happiness would most likely befall the investor who were to obtain a skew towards a higher frequency of positive returns rather than negative. And last but not least, an aversion towards extreme events would as well most likely be preferred amongst any relatively rational investor.

2.2 Loss Aversion

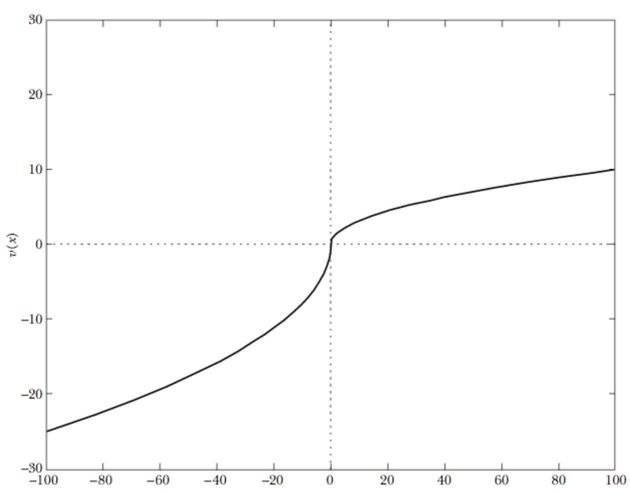
Kahneman and Tversky (1979) performed further investigations on how individuals actually act in terms of simple gambles, and in their paper they propose the concept of which they call Prospect Theory. The essence of which stems from what they call Loss Aversion.

Loss Aversion is epitomized by the idea that losses hurt the agent more than gains please. Hence, the function will be a piecewise function where there will be some kink at a specific reference point. In the case of asset returns, it is natural to place the kink around zero as changes in wealth will so forth either be positive or negative. The intuition of which can be found in equation 1,

$$u(x) = \begin{cases} f(x) & \text{if } x \geq 0 \\ \lambda g(x) & \text{if } x < 0 \end{cases} \tag{1}$$

where $u(x)$ is the utility function in which x is the change in wealth. A typical example of $f(x)$ and $g(x)$ would be when $f(x)$ is some strictly increasing continuously differentiable function and $g(x)$ would carry similar traits but have a steeper slope and some multiplier $\lambda \geq 1$. This results in a kinked curve with the kink located, in the case of financial asset returns, at the origin. In figure

Figure 1: Graphic example of the Prospect Theory value function with a kink at the origin. Figure gathered from Barberis (2013)



1 one can easier and more explicitly see how smaller losses takes a relatively greater toll than small gains increases utility. This is a much more realistic way of thinking about utility for an investor according to the experiments performed by Kahneman and Tversky (1979).

2.3 Maximizing Expected Utility

The central connection between utility theory and maximization of utility in relation to an investors exposure towards risk was made popular by Markowitz (1952). A few years later Kraus and Litzenberger (1976) furthered the concept by including the higher moments on valuation in asset pricing. In their paper they found empirical evidence that a three moment valuation technique in the Sharpe-Lintner Capital Asset pricing model framework, refutes the usefulness of quadratic utility. In this paper we will extend this idea and apply it to the loss averse S-shaped and Bilinear utility functions. We will express these preferences in a manner similar as Kraus and Litzenberger (1976), Jondeau and Rockinger (2006), and Harvey et al. (2010) where we express utility as the general Taylor series

$$u(w) = \sum_{k=0}^{\infty} \frac{u^k(w)}{k!}(w - E[w])^k \tag{2}$$

where $w = \phi r$ where r denotes asset return and ϕ denotes the weighting vector, which constitutes an investors current portfolio of assets. If taken the expectation of, and truncate at $k = 4$, we will get

$$E[u(w)] = u'(w)\mu + \frac{u''(w)}{2!}\sigma + \frac{u'''(w)}{3!}S + \frac{u^{IV}(w)}{4!}K \tag{3}$$

where S is Skewness and K is Kurtosis. As mentioned in 2.1. $u''(w) < 0$ implies risk aversion, and as Kimball (1990) and Eeckhoudt and Schlesinger (1995) extended: $u'''(w) > 0$ implies prudence. And an extension made by Kraus and Litzenberger (1976) implies that this is in direct relation to a preference towards a positive skew, whereas Eeckhoudt (2012) claims that the same conclusion is to be drawn about the fourth derivative and the kurtosis.

3 Optimal Portfolios

We will in the following section present a more formal explanation of the modern portfolio theory approach on optimizing a portfolio allocation. Then present some central issues with this method and explain how some of these issues can be solved by using regularization techniques.

For the sake of definition we define the mean, variance and covariance in this paper as the sample mean, sample variance and sample covariance. The mean is also the first k sample moment

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i^k \tag{4}$$

where $k = 1$. The following second central moment will be defined to be the variance

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \tag{5}$$

and we obviously follow by evaluating sample covariance as

$$\boldsymbol{\Sigma} = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x_i} - \boldsymbol{\mu})^T \tag{6}$$

with T signifying the transpose of the vector in the calculation that will yield a NxN covariance matrix, which we will frequently use in the standard portfolio optimization routines in the Markowitz framework.

3.1 Portfolio Optimization

In the famous essay by Markowitz, he makes the (now) obvious definition that expected return is something to covet, and risk (measured as variance or standard deviation) is something to disdain (Markowitz, 1952). From this realization, he suggests that one should construct an objective function to cover just this issue. In equation (7), all of those aspects are covered as the $\boldsymbol{\phi}$ is the $Nx1$ vector chosen such that the covariance matrix $\boldsymbol{\Sigma}$ is as small as possible, and the vector of expected return $\boldsymbol{\mu}$ is as large as possible. This yields the convex optimization problem

$$\begin{aligned}
\min_{\phi} \quad & \theta \boldsymbol{\phi}^T \boldsymbol{\Sigma} \boldsymbol{\phi} - \boldsymbol{\phi} \boldsymbol{\mu}^T \\
\text{s.t.} \quad & \boldsymbol{\phi}^T \mathbf{1} = 1, \\
& \phi_i \geq 0
\end{aligned} \tag{7}$$

where θ denotes a multiple which can be considered as a measure of the level of risk aversion, and $\mathbf{1}$ is the $Nx1$ vector of ones. In this paper we assume a non short-selling constraint. This is to make the optimization simpler as it can otherwise assume very large leveraged positions financed by short positions, which due to their nature will be very sensitive to small changes in return characteristics in individual assets, which could result in an undesirably stochastic portfolio (Ruppert, 2010).

The model assumes a risk averse investor and it is also very interpretable, parsimonious and simple to construct. It has subsequently played a central role in economic theory and asset pricing theory, and obviously been very popular as the main benchmark in portfolio theory, earning its creator a Nobel price in 1990, and John von Neumann Theory Prize in 1989.

3.2 Sparse Variable Selection

In the theoretical framework proposed by Markowitz, he assumes that the future returns are known, which consequently assumes that the input parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are also known. However, these parameters are in practice merely estimated from historical returns, meaning that small changes in the return characteristics can have a big impact on the portfolio optimization, even when the zero-short constraint is assumed. This is the so-called parameter uncertainty problem and yields estimation error, which Carrasco and Nérée (2011) claims is amplified by two facts. Firstly, the fact that the number of assets are typically high, and the more assets, the higher estimation error. Secondly, the asset returns can tend to have a relatively high degree of correlation.

To combat such a dilemma Frost and Savarino (1986) suggests a Baysian approach, however, Carrasco and Nérée (2011), Brodie et al. (2009), and Fastrich et al. (2012) suggests it can be treated using regularization techniques. Brodie et al. (2009) means that by including a penalty, which stands in proportion to the sum of the absolute values of the weights in the portfolio, and a tuning parameter to either amplify or decrease that penalty. This implies that our assets selection will be sparse, and depending on our tuning parameter have fewer non-zero weights. This means that any individual which does not contribute to the desired extent is effectively excluded from the sample, consequently leaving only a group of optimal individuals to be regarded as necessary in the optimization process, and so forth stabilizing the optimization problem at hand.

The concept of using regularization was first wrought by Tibshirani (1996) when he faced similar issues regarding accuracy and interpretability in regression models, with the only available option to perform some step wise selection or the Ridge Regression, which faced the remaining problem of having issues with interpretability. Hence, seeing how that was highly inefficient, Tibshirani proposed a method called LASSO (Least Absolute Shrinkage and Selection Operator). The idea was to force the sum of the absolute value of the coefficient to be less than some value λ. This means that some coefficients will be forced to be set to zero, yielding a much simpler model (See James et al., 2015, p 203-222).

In practice, this model will optimize a Least Square problem penalized over the l_1 norm as a convex optimization problem. The concept is mathematically illustrated as the residual sum of squares proposed in James et al. (2015) below in equation (8).

$$\arg\min_{\beta_i, c} \left(\boldsymbol{y} - c - \sum_{i=1}^{M} \beta_i \boldsymbol{X_i} \right)^2 \quad \text{s.t.} \quad \sum_{i=1}^{M} |\beta_i| < \lambda \tag{8}$$

This minimizes the magnitude of the β vector and sets any non-contributing individuals in the matrix of explanatory variables to zero.

This method has also been proven to perform relatively well in other applications, than those originally intended by Tibshirani. Implementing this into the portfolio optimization framework, one will similarly to Carrasco and Nérée (2011) and Ao et al. (2015), perform a straight forward augmentation of equation (7) with l_1 penalties in accordance to the Residual Sum of Squares (RSS) in equation (8), where the RSS is a tool for measuring the accuracy as the difference between the estimated model and the data.

In recent years issues in the variable selection process when groups of variables are known to be correlated has arose. Yuan and Lin (2007) illuminates this as an issue of the nature in the multi-factor Analysis of Variance (ANOVA). Seeing that the application of the standard LASSO has been thoroughly investigated by Fastrich et al. (2012) and Carrasco and Nérée (2011) we will extend this method further and avoid any potential issues which can occur with high correlation between certain groups in the sample. Hence we will focus mainly on the method that is called the Sparse Group LASSO (SGL). As proposed by Yuan and Lin (2007) the Grouped LASSO takes the LASSO with a grouped variable

$$\arg\min_{\beta, c} \left\| \boldsymbol{y} - c - \sum_{l=1}^{K} \beta_l^T \boldsymbol{X_l} \right\|_2^2 + \lambda \sum_{i=l}^{K} \sqrt{p_l} \|\beta_l\|_2 \tag{9}$$

where $\sqrt{p_i}$ denotes the amount in each group so that it accounts for varying sizes of groups, and $\|.\|_2$ denotes the non-squared euclidian norm. Here the $\boldsymbol{X}$ denotes the MxN matrix of features (with for example M assets over N data points) with the $Mx1$ parameter vector β. β_l and $\boldsymbol{X_l}$ denotes the data and predictor corresponding to the l^{th} group. In this formulation the LASSO acts entirely on group level, which can effectively exclude an entire group from the data set. This will however not yield any sparsity within the group. To adress this Friedman et al. (2010) proposes the SGL

$$\arg\min_{\beta,c} \left\| \boldsymbol{y} - c - \sum_{l=1}^{K} \beta_l^T \boldsymbol{X_l} \right\|_2^2 + \lambda(1-\alpha) \sum_{l=1}^{K} \sqrt{p_l} \|\beta_l\|_2 + \lambda\alpha \|\boldsymbol{\beta}\|_1 \tag{10}$$

where $\beta = (\beta_1, \beta_2, ..., \beta_K)$ is the entire vector of parameters. As equation 10 is a linear combination on α of two convex problems, it will also be a convex optimization problem.

This means that the α parameters effectively works as a weighting parameter, which if $\alpha = 1$ reduces the SGL into a regular LASSO, while if it is closer or equal to zero, will only emphasize the group factor. That would yield sparsity only between the entirety of the individual groups, and include all individuals of each group equally.

For the sake of simplicity, we will limit our selves to plainly performing the same augmentation as referred to in the previous paragraph when handling the regular LASSO. This will however be an augmentation of the convex optimization problem as proposed by Friedman et al. (2010) and Yuan and Lin (2007) and is presented in equation (11), which will effectively only change how we handle the penalizing terms. The objective function will so forth only be a Mean-Variance problem with Sparse Group LASSO penalties as

$$\begin{aligned} \arg\min_{\phi} \quad & \sum_{l=1}^{K} \left(\phi_l^T \boldsymbol{\Sigma} \phi_l - \phi_l \boldsymbol{\mu}^T \right) + \lambda(1-\alpha) \sum_{l=1}^{K} \sqrt{p_l} \|\phi_l\|_2 + \lambda\alpha \|\boldsymbol{\phi}\|_1 \\ \text{s.t.} \quad & \boldsymbol{\phi}^T \mathbf{1} = 1, \\ & \phi_i \geq 0 \end{aligned} \tag{11}$$

with K group types of assets, N amounts of assets, non-short constraints, and the general penalty amplification λ. This results in a convex portfolio optimization problem, which will yield a vector that sets any non-contributing individuals to zero on both a group level and individual level.

Hence it is very useful for an investor whom need not to own all assets at the same time but values the fact that the resulting portfolio will, in the context of the second and fourth moment, be much more stable. One can even go so far as to state industry sector preferences so that this grouping process can exclude or weight the optimization in accordance to that preference as well. We will for the time being not focus on that, as that would fall into the scope of more practical applications of a SGL portfolio optimization. Instead we will focus on furthering this concept to cover the prospect utility framework, and improve the existing allocation in regards of higher moments with the help of gradient techniques.

3.3 Optimizing Utility

Earlier in this section, we dealt with the issue of variable selection through sparsity, which would find a portfolio optimal over the first two moments, which would suffice for any standard risk-averse agent. However, the aim of this thesis is to further this framework onto more realistic and more complex utility functions. As previously mentioned we will in this thesis focus on optimizing over the Bilinear and S-shaped utility functions. These utility functions are more dependent on further moments past the expectation and variance.

This could be done by the FSO, optimizing the expected portfolio utility over an infinite amount of moments. However, the practical issues implied by this method, such as the fact that both any increased resolution or added asset increases the number of required iterations exponentially, are something that we want to overcome. According to Jondeau and Rockinger (2006), and Harvey et al. (2010) one can numerically solve a system of equation on a series of first order conditions. Simplifying this process to the first four derivatives, one will find an optima in regards to the first four moments. This means that to find this vector would mean a more accurate estimate in accordance to more complex utility functions. However, in this paper, we have chosen to look directly at the gradient vector of the expected portfolio utility. By using the MV-SGL optima as the initial starting point, which is optimized over the first two moments, we utilise a type of line-search technique based on the gradient, to find a portfolio optimal over the third and fourth moment as well.

4 Gradient Based Optimization Techniques

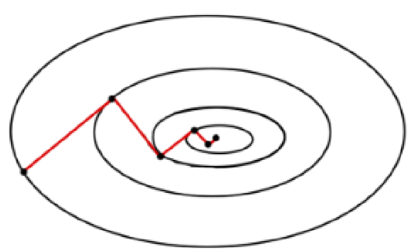
Gradient optimization techniques such as the Conjugate Gradient method, and line search methods such as Gradient Descent are members of a family of iterative optimization techniques, and is very popular within the field of numerical analysis. The general idea is to search for the optimum in a direction defined by the gradient of the objective function. The gradient can be viewed as a vector containing all the information about the change of a function in relation to each individual variable, as illustrated in expression (12).

$$\nabla f(a) = \left(\frac{\partial f}{\partial x}(a), \frac{\partial f}{\partial y}(a), \frac{\partial f}{\partial z}(a) \right) \tag{12}$$

This means that in relation to the change in the function it self, one can also observe the change in individual variables in relation to the function, which is a very useful insight and tactic when optimizing more complex non-linear functions. This means that depending on the gradient vector, one will be able to deduce which 'direction' is the fastest way to reach a local (and sometimes global) optima.

In figure 2 one can see how the iterations from the starting point finds the way towards the optima by following the direction of the largest gradient which can be simplified (for intuition) to be considered as the change in y divided by the change in x.

Figure 2: Graphic Example of a contour plot with the line signifying iterations towards the optima as the steepest descent. Figure gathered from Heggelien Refsnæs (2009)



In the following part of this section we will cover the general ideas of Gradient Based Optimization Techniques and provide the framework in which these techniques can be used to solve the final steps of the Portfolio Optimization problem in relation to the Taylor expansion of the utility function as motivated in Section 3.3.

4.1 Approximate Steepest Ascent

Gradient Ascent is an optimization technique in which you choose an initial point and then walk in the direction where the gradient is the largest. Meaning that in relation to one of the dependent variables, your function will increase the most when adjusting your path towards that individual.

The main reason for the popularity of methods such as this is the fact that they are fast and can handle larger problems more easily than comparative optimization techniques. The mathematical concept can be described as:

$$\max f(\theta) \tag{13}$$

By updating the parameters so that one follows the largest path according to the gradient $\nabla_\theta f(\theta)$ where $\theta = (\theta_i, ..., \theta_P)$ is the vector of parameters. The update function will be (in the vanilla case) defined as presented in 14

$$\theta = \theta - \eta \nabla_\theta f(\theta) \tag{14}$$

where η denotes the so called learning rate of which determines the size of each iteration or step. This is chosen such that the learning rate converges at an appropriate rate. This rate is commonly chosen by simulation. If the learning rate is too small, the time til convergence will be unnecessarily long, and if chosen too large, convergence will be hard to find as it might fluctuate around the optima and even diverge (Heggelien Refsnæs, 2009).

4.2 Application in Utility Maximization of a Portfolio

The problem at hand in the case of utility maximization by adjustment of the weighing vector calls for slight augmentation of the general idea. Given our objective function

$$\begin{aligned} &\max_{\phi} f(\theta) = E[u(\boldsymbol{\phi}^T \boldsymbol{R})] \\ &\text{s.t. } \boldsymbol{\phi}^T \mathbf{1} = 1, \ \ \phi_i \geq 0 \end{aligned} \tag{15}$$

with $\phi^T \boldsymbol{R}$, where ϕ is the $Mx1$ weight vector and $\boldsymbol{R}$ is the MxN matrix of returns, as our definition of wealth, which means that we maximize the utility of change in wealth given our allocation, and again including the same assumption of the non-short constraint as assumed in the previous examples of portfolio optimization in this paper. From 2.3 that we can approximate the $E[u(\phi^T \boldsymbol{R})]$ with the Taylor series in equation 3 in order to get all the partial derivatives of the objective function. To retrieve the gradient, i.e. the change in the objective function in regards to its partial derivatives of the parameter vector we claim the following

$$\nabla_\theta E[u(w)] = \left(\frac{\partial E[u(w)]}{\partial \mu}, \ \frac{\partial E[u(w)]}{\partial \sigma}, \ \frac{\partial E[u(w)]}{\partial S}, \ \frac{\partial E[u(w)]}{\partial K} \right) \tag{16}$$

from which we can see the how the partial derivative is directly linked to each moment. However, for practical purposes, we approximate the vector as forward differences and subsequently get

$$\nabla_\theta E[u(w)] \approx \left(\frac{\Delta E[u(w)]}{\Delta \mu}, \ \frac{\Delta E[u(w)]}{\Delta \sigma}, \ \frac{\Delta E[u(w)]}{\Delta S}, \ \frac{\Delta E[u(w)]}{\Delta K} \right) \tag{17}$$

where Δ denotes the discrete partial derivative for each moment, i.e. the incremental change of the expected utility in relation to each moment. This indicates that the gradient with respect to the mean is the change in expected utility of the portfolio divided by the change in the mean, and so on for the other moments. To create a practical framework in which to calculate this we propose an algorithm which takes its initial point at the portfolio composition that we got from the SGL method, and continues to evaluate the gradient of each statistical moment by approximating the relative change per each of said moment in relation to maximizing the objective function. We do this by looking at 18 as the that relative change we call the gradient.

$$\nabla_{\theta_i} E[u(w)] = \frac{\partial f}{\partial \theta_i} \tag{18}$$

Approximating the change as a direct relative change with the reference point as the previous point, we approximate the change in f as Δf as in equation 19 where $E[u(w)]^*$ denotes the new portfolio candidate. This means that if the new candidate has a better utility than the previous candidate, then the Δf will rise.

$$\partial f \approx \Delta f = \frac{E[u(w)] - E[u(w)]^*}{E[u(w)]^*} \tag{19}$$

Where in a very similar fashion to 19 we evaluate the relative change in a statistical moment between two candidate portfolio constitutions.

$$\partial \theta_i \approx \Delta \theta_i = \frac{\theta_i - \theta_i^*}{\theta_i^*} \tag{20}$$

Furthermore, from the assumption in 2.3 that odd moments are considered as something positive for utility, and even moments are considered as something negative. We assume that a positive increase in 20 is something bad in general. Hence, formulating this in a practical aspect we consider the numerator as $\theta_i - \theta_i^*$ when i is even and $\theta_i^* - \theta_i$ when i is odd. This means that we consider relative increases in the relative change in the even moments as something negative, and vice versa for odd moments. From this we see in equation 21 that a positive change in $\nabla u(w)$ indicates a better portfolio, and if we can follow the path of the highest $\nabla u(w)$ we will find the best portfolio in a fast pace.

In order to find the actual relative change in a statistical moment to the relative change in the utility function, we divide equation 19 by 20 to get the approximate gradient vector

$$\nabla u(w) \approx \frac{\Delta u(w)}{\Delta \theta_i} \tag{21}$$

where we will use this metric as our compass in the optimization process. We present a short summary of the algorithm we wrote to perform this optmization in algorithm 1.

Algorithm 1 Portfolio Gradient Maximization

procedure SGL-MV
  Retrieve data $\boldsymbol{X}$
  Evaluate $\boldsymbol{\Sigma}$
  Evaluate $\boldsymbol{\mu}$
  Minimize 11 **return** ϕ_{SGL}
procedure Maximize Utility
  init $\phi \leftarrow \phi_{SGL}$
  Evaluate Initial $u(\phi^T \boldsymbol{R})$
  while $\delta > \alpha$ **do**
    for i = 1 to M **do**
      Calculate equation 17
      $\phi^* = \phi_i + \eta$
      $\phi^* = \phi_j - \frac{\eta}{N-1}$
      where $i \neq j$
      for $\theta = 1$ to G **do**
        Evaluate $u(\phi^T \boldsymbol{R})$
        Re-calculate equation 17
        if $\nabla_\theta u(\phi^{T*} \boldsymbol{R}) > \nabla_\theta u(\phi^T \boldsymbol{R})$ **then**
          $\phi = \phi^*$
      Evaluate $\delta = |\nabla_\theta u(\phi^* \boldsymbol{R}) - \nabla_\theta u(\phi \boldsymbol{R})|$
      return ϕ_{GSGL}

M denotes the number of individuals in the weight vector, and G denotes the number of individuals in the approximated gradient vector. In this process δ denotes the absolute value of the change in the size of the change in the gradient, meaning that as this decreases, we are closing in on the sought after maximum. The critical value α where we determine convergence is arbitrarily chosen at such a level so that we believe that convergence is found without to much unnecessary running around the maxima. This gives us a good and intuitive method for numerically optimizing the portfolio problem over the first four moments in a relatively fast and parsimonious way. To summarize our aim with this process we conclude this section with the following:

- 22 Firstly perform the Sparse Group LASSO optimization technique. This is to find a viable starting point for the next gradient based optimization technique. Doing this will also effectively eliminate any unwanted assets relative to the first two moments.
- Moving on we will have the initial vector of weights. From this we evaluate the utility in accordance to the S-shaped and Bilinear utility function.
- As we now have the utility, and are able to evaluate the sample moments with ease we will evaluate our approximate gradient in accordance to equation 21.
- Then we utilise algorithm 1 that will follow the path where the gradient is the highest, so as fast as possible find the best portfolio at our predefined convergence.

We will then also compare this to standard methods of portfolio optimization in-sample framework where we try these methods in various trading horizons, with portfolio metrics as our metric of measurement against the MV portfolio as the benchmark.

5 Empirical Application

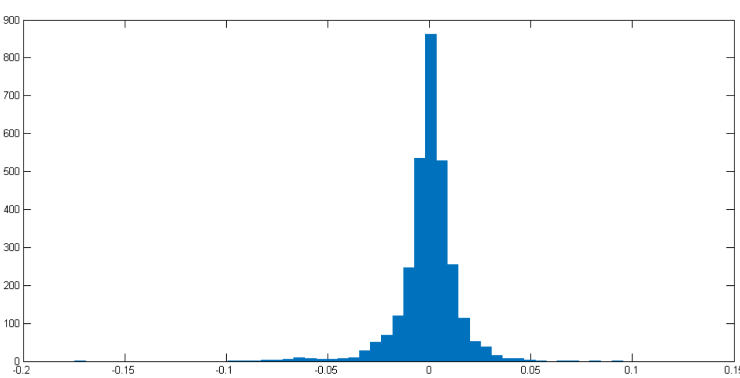
We will investigate the application of our Gradient Ascent SGL method of portfolio optimization by comparing its performance against a regular straight forward Mean Variance portfolio. In this section we will cover the data set and how we implement the techniques discussed in the previous sections, as well as testing them in an in-sample analysis environment.

5.1 Data and Statistics

As one of the main goals of our optimization technique is to make it more computationally efficient than other techniques which relies on maximizing a utility. We have so forth chosen to randomly select fifty companies from the SP 500 index over the last 3000 days from march 2016 and back. We have covered so that we have assets within ten different industry sectors and controlled so that they are all traded on the same day in the entire data set. If an asset was not traded on a specific day for some reason, we copied the price quote from the previous trading day.

From this we calculate the asset return series as the percentage change between each trading day and yielded a $50x3000$ matrix of returns. As it would be quite cumbersome to summarize the statistical properties of all the individual assets we evaluated the average return on each trading day to the following index to capture the statistical properties of the entire dataset without any optimization procedures done. The distribution is illustrated in figure 3.

Figure 3: Histogram over the average daily return of the return matrix



From this one can see a slight negative skew and far reaching tails, which is in line with the theoretical assumptions on financial return series as discussed in section 1. These statistical properties are furthermore summarized in table 1 below.

N	Mean	Standard Deviation	Skewness	Kurtosis	JB-test
50	-0.0004	0.01452	-1.6542	17.5651	(Reject)

Table 1: Descriptive statistics of the average daily return of the return matrix

The Jarque-Bera null hypothesis of normality is rejected as the skew is non-zero and the kurtosis is significantly higher than three. The data series covers two periods of high volatility, of which the last period (2008 stock market crash) impacted the American stock market severely.

5.2 Grouped Sparse LASSO

As the concept of the Group factor in the SGL aims to reduce potentially damaging effects from correlation between any individuals in the matrix, one has to find suitable grouping rules which would account for this. In this portfolio optimization framework we saw sector categories to be the most appropriate rule of grouping and used the Global Industry Classification Standard (GICS) taxonomy. Using the ten different sector categories where we have between three and eleven individuals in each group as the SGL does not require each group to be uniform in size.

Evaluating the various statistical elements used in the optimization was done in the most straight forward fashion by using MatLabs methods for evaluating a sample covariance matrix and a vector of expected return as the sample mean return. To subsequently perform the minimization objective we utilise the CVX package in MatLab, which is an excellent straight forward package for disciplined convex optimization. As of choosing the the λ, one could divide them as two separate and amplify each component differently. In this paper we decided that the weighing parameter α does this to an appropriate and sufficient extent. Trying out different types of scenario left us to chose $\lambda = 1$ as it did not make much of a difference untill one reached significantly higher values, which would be theoretically unappealing.

The tuning parameter α in equation 11, which is chosen such that $\alpha \in [0, 1]$ is set at 0.5 for simplicity so that the grouped vector of weights and the vector of individual weights are weighed equally.

5.3 Evaluation of the Gradient

As the objective is to maxmize the utility we created a versatile function in MatLab so that we could easily switch between the S-shaped and the Bilinear function. In equation 22 and 23 we present the mathematical method of evaluating the S-shaped utility and the Bilinear utility function respectively in accordance to the method presented by Hagströmer et al. (2008).

In the S-shaped utility function the x determines the location of the inflection point, while the other parameters (A, B, ψ_1, ψ_2) determines the curvature of the function.

$$u = \begin{cases} -A(x - r_p)^{\psi_1} & \text{if } x \geq r_p \\ B(r_p - x)^{\psi_2} & \text{if } x < r_p \end{cases} \tag{22}$$

In the Bilinear utility function presented in equation 23 the x is in this case the threshold value, or the so called kink, and the P constitutes the level of penalty of which the utility level is suffering during returns below the threshold value.

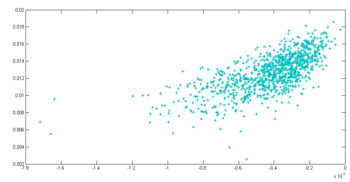
$$u = \begin{cases} ln(1 + r_p) & \text{if } r_p \geq x \\ P(r_p - x) + ln(1 + x) & \text{if } r_p < x, P > 0 \end{cases} \tag{23}$$

Kahneman and Tversky (1979) argues that these types of functions focuses on the utility in the return on an investment in a scenario where the agent has a different function for any return below a certain threshold value. Commonly this will be set to zero, so that all negative returns imply a steeper utility, meaning that losses will hurt more than gains pleases. This is the major idea in prospect theory, and prospect utility subsequently, and is the principle in which we aim to model a more realistic portfolio optimization which takes these insights into account.

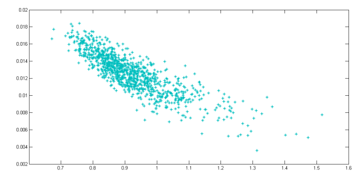
To justify our reasoning of the relation between the Expected utility and four first statistical moments we plotted these together in figure 4. In this simulation we used the S-shaped utility function and randomly generated one thousand different unit-sum weight vectors against our data set.

What we gather from this is that there is a very strong relation between the expected utility and the first two moments where a higher mean yields a higher utility and lower variance yields the same. The skewness also has a clear relationship although not as strong as the first two moments, the kurtosis is similar, but slightly weaker in correlation than the skewness. This is in line with the insights gathered from section 2.3 meaning that if one finds a way to optimize the utility in relation to the first four moments, one would find a better allocation that the standard MV technique.

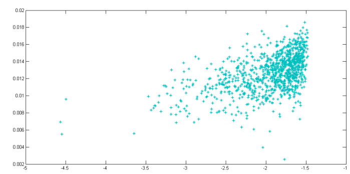
In order to evaluate the Gradient of the objective function as formulated in equation 15 we need to modify the weight vector ϕ. As we suggest in the



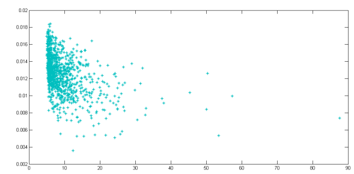
(a) Mean and Expected Utility



(b) Variance and Expected Utility



(c) Skewness and Expected Utility



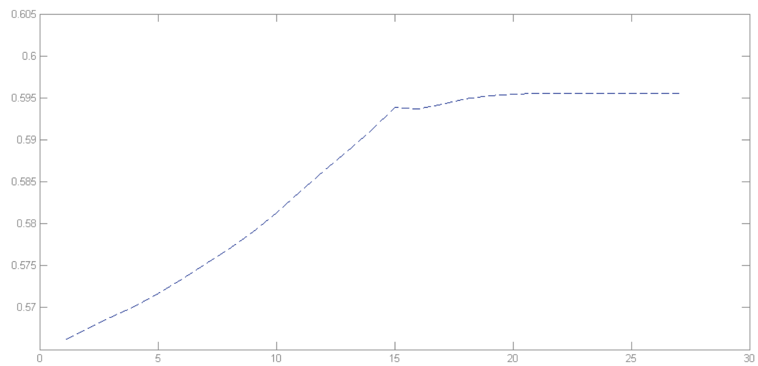
(d) Kurtosis and Expected Utility

Figure 4: The four moments plotted against a S-shaped Expected utility function

proposed algorithm 1, we look around for any adjacent improvement and evaluate the gradient in accordance to this proposed change and step in the direction where the gradient is the highest. We do this because we approximate the numerical gradient as the relative change in the expected utility over the relative change in statistical moments. Meaning that if our algorithm detects the direction which would increase our relative utility the most we would modify our weights in accordance to this direction. In order to keep a unit sum with non-negative weights due to the short selling constraint we made so that when increasing the ϕ_i we make a linear subtraction from all ϕ_j where $j \neq i$ and the $\phi_j \geq \dfrac{\eta}{n-1}$ where η denotes the step-size. This prevents any individual weights from being less than zero and keeps the sum of the weights at one. To illustrate, in the case of a step size at 0.01 and four equally weighted assets, one asset would gain the step size leading to the vector $(0.26, 0.25, 0.25, 0.25)$ which would give a sum of 1.01. Subtracting $\frac{0.01}{4-1}$ from all other weights yields the vector $(0.26, 0.246..67, 0.246..67, 0.246..67)$ which has a sum of one, and as long as a weight has above or equal value of 0.0033.. the weight will not be subtracted and will never go below zero.

The task at hand after formulating this process is to find the appropriate step-size. As explained, in our case this will be by how much we alter our weight vector in each step. We measure this path to convergence as the difference in improvement of the gradient, when it starts to diminish, we know that we are closing in on a local or global optima. We found that a good initial step-size was at 1%. What we did after that to ensure that there where no more room for improvement is that when our rule reached .01% we halved the step size, and repeated that until we reached a difference at .0001% and concluded that we had reached some converging point.

Figure 5: Figure presenting the level of convergence in the gradient based optimization technique where we can see how the expected utility converges at some point after about 20 steps



This convergence is illustrated in figure 5 where we performed the optimization over all three thousand days in the data set.

5.4 In-Sample Analysis

The in-sample analysis was performed in such a manner that we performed the optimization over the three thousand days in 'chunks' of three hundred and thirty respectively. This is to some what capture a concept of yearly and monthly re-weighing the portfolio.

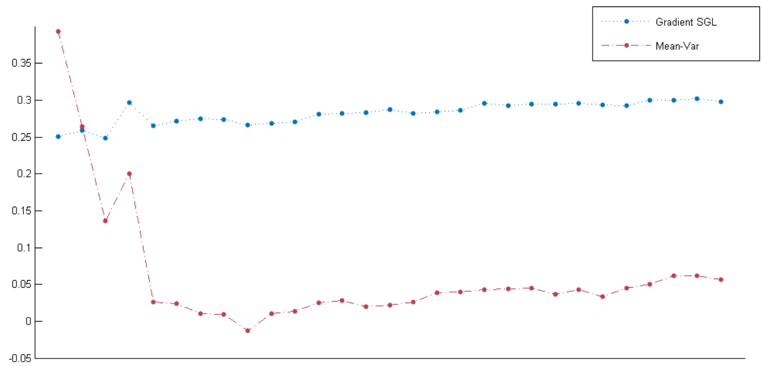
In figure 6 we present the resulting 29 Sharpe ratios from each so called 'monthly' re-weighing.

From this we can see that the Sharpe ratio for the Gradient SGL (Blue) is consistently higher than the Sharpe Ratio we got from the MV optimization.

In table 2 we averaged out all the monthly results for the Sharpe ratio as well as the third and fourth moment.

We see how the Sharpe Ratio for the Gradient SGL performs significantly better than the MV for both the S-shaped and Bilinear utility function. The results even yield a positive skewness for the Gradient SGL, also in both scenarios, while the skewness is negative for the MV. The kurtosis averaged out to a larger value for the Gradient SGL than for the MV.

Figure 6: Sharpe Ratio between the Gradient SGL (Blue) and the MV (Red) portfolio optimization over the 'monthly' in-sample test



	Gradient SGL (S-shaped)	Gradient SGL (Bilinear)	MV
Sharpe Ratio	0.2926	0.2835	0.0618
Skewness	0.8784	0.9804	-0.1509
Kurtosis	5.6185	5.9351	3.1662

Table 2: Monthly re-weight

In table 3 we did similarly as above where we averaged out the performance over all nine three hundred days iterations to get to following results.

	Gradient SGL (S-shaped)	Gradient SGL (Bilinear)	MV
Sharpe Ratio	0.1061	0.1059	0.0124
Skewness	0.0352	0.0400	-0.3622
Kurtosis	5.9535	5.9642	4.8846

Table 3: Yearly re-weight

Here we see similar results where the Sharpe ration is significantly higher in the case of both of the utility scenarios in the Gradient SGL method than the MV. And again a negative skewness for the MV with positive for the Gradient SGL. The kurtosis is still however lower in the MV in this case as well, but not by as much in this yearly scenario.

One reason which could cause this is the fact that the kurtosis does not have as much of a relationship with the expected utility as the other three moments.

Depending on the portfolio metric, we would say that this still would yield a better portfolio than the MV would as the kurtosis is still well below the value of the initial kurtosis of the entire data set with a simple average return for each asset. Measuring from a pure utility point of view one can see in our figure of the evolution of the convergence that the SGL-MV yields an initial ecpected utility of about 0.565, and with our Gradient method added converges with an expected utility at around 0.595.

5.5 Summary of Results & Implementation

The tests were done by a varying amount of settings on the various parameters, in the previous section we presented the implementation and results of the scenarios that produced the most theoretically interesting results. All of the theoretical models included in our method needed some form of simulation in order to find optimal presets before running it as a complete process. By using the very fast and precise optimization environment called CVX in Matlab we where able to obtain our SGL MV optima with ease. Subsequent formulation of the Gradient Ascent algorithm was done in a manner as straight forward as possible by the discrete approximation of the gradient vector of the portfolio.

The SGL method was very dependent on the selection of the parameter α. This was because that parameter adjusted how much the model should focus on finding the MV minima on a group level, or on an individual asset level. One could of course dive deeper into this depending on the agent performing the optimization. This is because some agents might prefer to have a portfolio optimization between the sector categories themselves, as well as a direct sparse MV allocation between all of the assets in total. That choice is however outside of the scope of this paper as we only aim to attain the beneficial properties of a sparse allocation which considers possible correlation between assets inside sector groups. We chose therefore to set the $\alpha = 0.5$ so that we would not discriminate too much between the two counter philosophies.

The calibration of the gradient optimization method was as well important for it to have the desired properties. This meant that we needed to find an appropriate step size. This is because a too large step size means it will fluctuate around the optima and a too small would take unnecessary time to reach the optima. We solved this by setting a condition that, when the change in the gradient decreased at a certain level, the algorithm would stop. Also, in order to ensure that the maxima would be reached, we formulated a condition which broke the step-size in half when the incremental improvement diminished in order to make sure that we could not improve the expected utility any further.

In the Bilinear utility framework we set the kink at zero. This is so that the undesirable dimension of the function is when the asset returns are negative. In the same manner, regarding the S-shaped utility function, we set the inflection point at zero as well. Choosing parameter settings on the use of the Bilinear and S-shaped utility function also required some simulation in order to achieve the desired theoretical properties. For example, when one used a lower penalty P for the Bilinear utility the gradient optimization routine did not find much better allocations than the MV. This has to do with the fact that as the penalty diminishes, the function will approximate into some simple function perfectly described by the first two moments, and defeat the purpose of performing the portfolio optimization in a prospect utility framework. To make sure that we would have the desired properties, we set the parameters $A = 20$, $B = 1$, $\psi_1 = 0.3$ and $\psi_2 = 0.55$ in the S-shaped utility framework. In the Bilinear framework we saw that a penalty level at $P = 2$ was sufficient to capture the desired characteristics with a kink at $x = 0$.

When in line with our theoretical assumptions of prospect utility we could however see how our optimization routine outperformed the standard MV portfolio in the first three moments, as well as by the portfolio metric Sharpe Ratio. In fact, the skewness of the MV portfolio was negative, while the Gradient Ascent SGL finds a portfolio with positive skewness. This tells us that the MV agent does not regard skewness as important, while the S-shaped or Bilinear Gradient Ascent SGL agent does. This could indicate that portfolio optimization in the latter method falls more in line with classical risk aversion utility theory. On the contrary, in our tests, we found the MV method yielded lower kurtosis than the Gradient Ascent SGL method, which as well is a desirable property within that theoretical framework.

We believe, in regards to the Gradient Ascent SGL optimized allocation having a lower kurtosis than the MV optimization, can be explained by the relatively low correlation that Kurtosis had with the expected utility as presented in figure 4. Hence, improvement in the expected utility was not by as much amplified by any improvement in kurtosis, when compared to the improvement in the mean, variance, and skewness.

Otherwise we see that this has yielded very interesting results as we could get significant improvements in regards to the Sharpe ratio, and portfolio skewness. This implies that there is room for improvement for the MV method, and in comparison to the FSO, our method could optimize over fifty assets on three thousand data points with ease. One could easily change the settings on the utility function to cater to any other utility function that we have not mentioned in this paper.

6 Conclusion

In this paper we have constructed a new method for portfolio optimization in order to properly take a more complex utility function into account. We used two examples of utility functions gathered from the concept of prospect utility, so that we could have a more realistic allocation in terms of investor preferences. This idea extends earlier findings by Hagströmer et al. (2008) and Adler and Kritzman (2006), but with a further extension with the aim of decreasing the computational load, so that in a similar manner to Jondeau and Rockinger (2006), and Athayde and Flores (2002) we focused on the changes relative to the first four moments. However, in this paper we diverged from their approach by formulating a Gradient Ascent method with a starting point gathered from a SGL MV allocation. This is assuming that such an allocation yields stable and sparse allocations which are optimized for the first two moments. This idea is an extension of previous research by Brodie et al. (2009), Fastrich et al. (2012), and Carrasco and Nérée (2011).

We saw some success in our method as it on average outperformed our benchmark MV method in the in-sample testing framework, in both a yearly and monthly re-weighing scheme. We managed to overcome the computational burden of the FSO, and provided an alternative on optimizing a portfolio relative to its gradient, while retaining full flexibility, and mathematical integrity in the usage of any utility function. However, as we saw that there was little effect to be gained from the kurtosis, we consider it reasonable to truncate the approximation at the fourth derivative, and not investigate any moments past the kurtosis.

Investigating the performance of this method on other types of asset classes could subject for future research, as the empirical part of this paper used S&P 500 equity returns, which has rather special distributional properties. Hence, looking at a portfolio of commodities, currencies, indexes or a mixed portfolio would be very interesting. Furthermore, as this paper mainly focused on motivating and formulating the method, and testing it in-sample, expanding the testing environment to an out-of-sample testing framework would as well be a necessity going forward.

References

Timothy Adler and Mark Kritzman. Mean-variance versus full-scale optimization: In and out of sample. *MIT Sloan Research Paper*, 4589-05, 2006.

Mengmeng Ao, Yingying Li, and Xinghua Zheng. Solving the high-dimensional markowitz optimization problem: When sparse regression meets random matrix theory. *Available at SSRN: 2699157*, 2015.

Kenneth Arrow. The theory of risk aversion. *Helsinki: Yrjo Jahnssonin Saatio. Reprinted in: Essays in the Theory of Risk Bearing*, 1965.

Gustavo M Athayde and Renato G Flores. *On Certain Geometric Aspects of Portfolio Optimisation with higher moments*. Multi-moment Asset Allocation and Pricing Models. John Wiley Sons, Inc., 2002.

Nicholas Barberis. Thirty years of prospect theory in economics. *Journal of Economic Perspectives*, 27:173–196, 2013.

Daniele Bernoulli. Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperiales Petropolitiane*, pages 175–192, 1782.

Joshua. Brodie, Ingrid Daubechies, Christine De Mol, Domenico Giannone, and Ignace Loris. Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences*, 106(30):12267–12272, 2009.

Marine Carrasco and Noumon Nérée. Optimal portfolio selection using regularization. *CIREQ working paper*, 2011.

Jan-Hein Cremers, Mark Kritzman, and Sebastien Page. Optimal hedge fund allocations - do higher moments matter? *Journal of Portfolio Management*, 3:70–81, 2005.

Louis Eeckhoudt. Beyond risk aversion: Why, how and what's next? *The Geneva Risk and Insurance Review*, 37:141–155, 2012.

Louis Eeckhoudt and Harris Schlesinger. *Higher-Order Risk Attitudes*. Handbook of Insurance. Springer New York. Pages: 41–57, 1995.

Eugene Fama. The behaviour of stock market prices. *Journal of Business*, 38 (1):4–105, 1965.

Björn Fastrich, Sandra Paterlini, and Peter Winker. Constructing optimal sparse portfolios using regularization methods. *Available at SSRN 2169062.*, 2012.

Jerome Friedman, Robert Tibshirani, and Trevor Hastie. A note on the group lasso and a sparse group lasso. *Preprint arXiv:1001:0736v1*, 2010.

Peter A. Frost and James E. Savarino. An empirical bayes approach to efficient portfolio selection. *The Journal of Financial and Quantitative Analysis*, 21 (3):293–305, 1986.

Christian Gollier and John W. Pratt. Risk vulnerability and the tempering effect of background risk. *Econometrica*, 64:1109–1124, 1996.

Björn Hagströmer, Richard Andersson, Jane Binner, Thomas Elger, and Birger Nilsson. Mean-variance vs. full-scale optimization: Broad evidence for the uk. *FRB of St. Louis Working Paper*, 2007-016D, 2008.

Campbell R. Harvey, John C. Liechty, and Merrill W. Liechty. Portfolio selection with higher moments. *Quantitative Finance*, 10(5):469–485, 2010.

Runar Heggelien Refsnæs. A brief introduction to the conjugate gradient method. *Lecture Notes NTNU*, 2009.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical learning.* Springer Texts in Statistics, 2015. ISBN 978-1-4614-7137-0.

Eric Jondeau and Michael Rockinger. Optimal portfolio allocation under higher moments. *European Financial Management*, 12(1):29–55, 2006.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47:263–291, 1979.

Miles S. Kimball. Precautionary savings in the small and in the large. *Econometrica*, page 53–73, 1990.

Miles S. Kimball. *Precautionary Motives for Holding Assets.* P. Newman, M. Milgate, and J. Falwell, eds. London: MacMillan., 1992.

Alan Kraus and Robert H Litzenberger. Skewness preference and the valuation of risk assets. *Journal of Finance*, 31:1085–100, 1976.

Yoram Kroll, Haim Levy, and Harry Markowitz. Mean-variance versus direct utility maximization. *Journal of Finance*, 39(1):47–61, 1984.

Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.

Harry Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.

John W. Pratt. Risk aversion in the small and in the large. *Econometrica*, page 122–136, 1964.

David Ruppert. *Statistics and Data Analysis for Financial Engineering*. Springer Texts in Statistics, 2010. ISBN 978-1-4614-2749-0.

Robert C Scott and Phillip A Horvath. On the direction of preference for moments of higher order than the variance. *Journal of Finance*, 35:915–919, 1980.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58:276–88, 1996.

Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society*, 68:49–67, 2007.