How Wall Street's "Dirty Little Secret" Affects Investor Portfolios

by

Liam Welsh

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Mathematical Finance

Department of Mathematical and Statistical Sciences
University of Alberta

Abstract

In the United States, investors of exchange-traded funds (ETFs) and mutual funds are required to pay tax on the capital gains that their funds have made throughout the year. However, ETFs are able to avoid making taxable capital gains by taking advantage of a legal loophole, subsequently benefiting ETF investors. This loophole is referred to as Wall Street's "dirty little secret". By contrast, mutual funds do not benefit from this loophole, and their investors must pay capital gains tax when the fund is selling appreciated assets. In this thesis, we explore the impact that the ETF tax loophole has on investor decisions and how an investor's asset allocation would change in the event the loophole is closed. However, investor biases have been observed between active and passive funds, so a classical model cannot be used. As such, we use a rank-dependent expected utility model allowing us to incorporate these biases. Given a particular setting, the investor's optimization problem can be solved explicitly. In a more general model, this problem is not explicitly solvable, but we can still obtain the approximate impact that the tax loophole has on an investor's portfolio. As the tax loophole allows ETF investors to defer capital gains tax, its impact increases with the investor's holding period. In the event the loophole is closed, we estimate that for holding periods between three and five years the investment in ETFs will decrease by approximately 2.75% to 7% of the total portfolio allocation, and the investment in mutual funds will increase by approximately 2.5% to 5.5% of the total portfolio allocation.

Acknowledgements

I would firstly like to thank my supervisor Dr. Christoph Frei for his consistent guidance, mentorship, and support throughout the past two years. Dr. Frei's aptitude for mathematical finance and overall encouragement were invaluable throughout my degree, and despite the global events of the past year he always provided a calm and welcoming environment in which to have academic and friendly discussions. I also thank Dr. Vishal Baloria, Dr. Abel Cadenillas, and Dr. Alexander Melnikov for their contributions to my defence committee. Next, I would like to thank my parents for allowing me to move back in with them (twice) during the COVID-19 pandemic and their constant support of my education. I would also like to thank all my friends and colleagues, both in Alberta and BC, of which there are far too many to personally name. Financial support by the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

Table of Contents

List of Tables

List of Figures

Abbreviations

CARA	constant absolute risk aversion
CDF	cumulative distribution function
CRRA	constant relative risk aversion
ETF	exchange-traded fund
LG	large growth
LV	large value
MF	mutual fund
PDF	probability density function
RDEU	rank-dependent expected utility

Chapter 1

Introduction

Investment in a fund provides an investor with instant diversification, reducing their inherent risk. There are numerous different types of funds that the investor can choose from, and these funds can be actively or passively managed. Active funds and passive funds have their own unique traits that can attract potential investors. An active fund, like most mutual funds (MFs), has a manager or a management team that make the investment decisions. These investment decisions made by the MF's management team can increase an investor's upside potential or minimize an investor's potential losses. By contrast, exchange-traded funds (ETFs) are often passive funds, which commonly mirror a specific index or security collection. As such, the ETF management team is not able to make specific investment decision to minimize an investor's potential investor loss or increase an investor's potential gain. ETFs do have their own benefits nonetheless, as they are more cost efficient. Additionally, ETFs in the United States are able to take advantage of a legal loophole that allows ETF investors to defer paying tax on an ETF's capital gains until they sell their position. ETFs are able to avoid making taxable capital gains by taking advantage of in-kind redemptions when performing a portfolio re-balance, given the aid of a friendly investor. Consequently, an ETF investor will pay capital gains tax only when selling their ETF holdings and not when the ETF realizes a capital gain. This legal loophole is not used by MFs because they rarely perform in-kind redemptions, as their investors typically prefer cash payouts. Due to this loophole, it is estimated by Colon [3] that investors in the top 25 ETFs on the US market benefit from roughly 60 billion USD of tax-free capital gains per year. These tax benefits aid in making ETFs an attractive investment opportunity to potential investors. In recent years, the debate surrounding this legal loophole has primarily focused on its mechanics and ethics considerations. We, however, intend to answer the question "how will investment in ETFs change in the event this loophole is closed?"

In this thesis, we will analyze how an investor's optimal portfolio, containing an MF and an ETF as the risky assets, is affected by the tax loophole and investigate how the optimal allocation would change in the event the loophole is closed. However, there is a catch. Prior work by Gruber [7] has shown there exists an investor preference for actively managed funds over their passively managed counterpart, despite the fact that actively managed funds have historically worse average performance than their passive fund equivalents. Therefore, in this thesis we use a model for portfolio optimization that takes into account investor bias. Polkovnichenko et al. [12] has shown that using a rank-dependent expected utility (RDEU) model for portfolio optimization is able to accurately represent this observed market behaviour for portfolios containing active and passive funds. The ability to accurately model how an investor will alter their portfolio in the event of the loophole closure will provide a deeper understanding as to how future policy changes will influence investments in a biased market. In doing so, valuable insight can be derived to help determine the implications of future policies.

To determine the impact the tax loophole has on investor portfolios, we will consider a one-period market model. In order to solve the portfolio optimization while incorporating investor bias, we utilize a Taylor approximation that will allow us to estimate the optimal strategy under a distorted probability model. With an estimation of the optimal portfolio strategy, we then will consider two investment scenarios: one where the loophole is open, and one where the loophole is closed. This comparison will allow us to determine the extent to which an investor benefits from the tax loophole. We expect that in the event the ETF tax loophole is closed, investment in ETFs will decrease and investment in MFs will increase. While this conclusion may seem obvious, it is our intention to provide a numerical result for the magnitude of these investment changes. The notion that the ETF tax loophole significantly affects investor portfolios is consistent with Moussawi et al. [11], who observe the importance of tax considerations in flows from MFs to ETFs for high-net-worth clients. Investment advisors for this clientele of highly tax-sensitive investors have a much larger portion of ETF holdings in their portfolios, compared to other investment advisors, as Moussawi et al. [11] report.

The structure of this thesis is as follows. First, we describe the necessary background information required to understand the contents of this thesis in Chapter 2, including active and passive funds, the ETF tax loophole, portfolio optimization techniques, and portfolio optimization under a distorted probability model. In Chapter 3, we explore a setting that allows us to solve the portfolio optimization under a distorted probability model explicitly and we analyze the solution's sensitivity to a shift in the expected return. In Chapter 4, we develop an approximation that will allow us to solve the optimal portfolio weights under a general distorted probability model. This approximation is used to estimate how an investor's strategy will change in the event that the ETF loophole is closed. We provide concluding remarks in Chapter 5.

Chapter 2

Necessary Background

In this chapter, we describe the background knowledge that is necessary to understand the remainder of this thesis. In Section 2.1, we provide an explanation of active funds and passive funds. We then introduce MFs and ETFs and discuss their differences. In Section 2.2, we discuss the legal loophole that allows an ETF to shield their investors from capital gains taxes. In Sections 2.3 and 2.4, we describe the fundamentals of portfolio optimization and expected utility, a common portfolio optimization technique. In Section 2.5, we discuss how expected utility can be extended to a distorted probability model that incorporates investor biases on tail events.

2.1 Active and Passive Funds

Creating a diverse portfolio significantly reduces the amount of inherent risk the portfolio has. A diverse portfolio is highly sought after among all kinds of investors. However, if the investor lacks the time or knowledge to properly research investment opportunities, they may not be able to achieve the amount of diversification they desire. Additionally, it is not cost efficient for an investor with a small amount of capital to invest directly into multiple stocks due to numerous transaction fees. To aid in diversification and reduce costs, the investor can choose to invest in a fund, in addition to other securities. A fund represents a collection of assets and usually follows an index, commodity, sector, or some other collection of securities. Investing in a fund has multiple benefits, as it provides exposure to a large amount of assets providing instant diversification, requires less knowledge and research for a time-strapped investor, and can be more cost efficient due to a smaller amount of transaction fees than investing in multiple stocks.

A fund can typically be placed into one of two categories: actively managed or passively managed. Both of these fund types have a fund manager. The active fund manager makes the final decision as to what the fund should invest in, which is a time consuming process. In addition to investment decisions, the manager also oversees performance reporting, portfolio balancing, and other duties related to the fund. For the manager's expertise and hands-on approach, an investor will typically pay a premium in order to invest their capital into an active fund. This premium takes the form of a management fee. A passively managed fund also requires a manager. The passive fund manager is required to re-balance the fund after a change in the underlying (e.g. an index, commodity, etc.), and to report performance among other duties. However, the passive fund manager does not make investment decisions like their active manager counterpart, as their fund mirrors some underlying. The passive fund manager typically has less work than that of the active fund manager. As such, the management fees that accompany investment into an active fund are significantly higher than those of a passive fund. While either an MF or an ETF can be actively or passively managed, it is more common for MFs to be actively managed and for ETFs to be passively managed.

It has been shown that actively managed funds (e.g. MFs) have historically worse average performance when compared to their passively managed (e.g. ETFs) counterparts; see Gruber [7] and Polkovnichenko et al. [12]. Despite this historically inferior performance and higher associated management fee, Gruber [7] found that investors prefer to invest in actively managed funds over their passive counterparts. This observed investor behaviour is referred to as the "mutual fund puzzle". Polkovnichenko et al. [12] proposed that an investor chooses to invest in an actively managed fund over its passively managed counterpart as an investor believes that the presence of a hands-on manager will protect them from significant losses and allow them to realize larger gains than those of a passive fund. This belief results in an investor bias between the two types of funds. For the assumed downside protection and upside potential, investors are willing to pay a relatively high management fee for active funds when compared to the management fee of the passive counterpart. This stands in contrast to what the standard mathematical models for portfolio optimization dictate, as the investor is giving up an expected higher average return due to a bias towards tail events. Therefore, we use a mathematical model that incorporates a tail-overweighting bias to describe the observed market behaviour between investment in actively managed and passively managed funds. Investor bias towards tail events is one of many possible explanations to the "mutual fund puzzle". In this thesis, we take an

MF to represent an actively managed fund and an ETF to represent the MF's passively managed counterpart. As such, when describing the observed market behaviour for portfolios containing these funds, we use a model that incorporates investor biases. To do so, we will employ an RDEU model, which distorts the underlying probability. This model is described in detail in Section 2.5.

2.2 The ETF Tax Loophole

Unlike MFs, ETFs were originally structured to have a high tax efficiency; see Gastineau [6]. The lower management fees and high tax efficiency of ETFs are partially responsible for driving a meteoric rise in their popularity among investors. In 2020 alone, Rosenbluth [16] found that 290 billion USD was invested into the top three ETF providers: BlackRock, State Street Global Advisors, and Vanguard. Together, these three firms account for 81% of the over 5 trillion USD invested in ETFs [16]. These firms, in addition to many others, have been found to be taking advantage of in-kind redemptions, which allow ETFs to avoid incurring taxable capital gains when selling appreciated assets. Subsequently, this practice allows ETF investors to defer paying capital gains taxes until they sell their ETF shares, as Mider et al. [10] report. While some view this as a smart tax strategy, Colon [3] and Hodaszy [8] consider it as a loophole in the tax code. We now describe the history of the tax loophole, how it works, and its consequences.

In the US, it is typical that tax must be paid on capital gains that have been realized on an asset when that asset is sold. Due to a 1969 tax law signed by then US President Nixon, ETF investors are able to avoid paying capital gains taxes when the fund is selling an appreciated asset. This tax law in fact precedes the first ETF by over 20 years. To explain how the loophole works, we present a similar scenario as that described in Mider et al. [10], where an ETF must undergo a re-balance.

First, imagine an ETF whose portfolio contains ten stocks labeled one through ten. This portfolio mirrors some market index that contains the same ten stocks. Further, we assume that the ETF holds this portfolio for one year. At the end of the year, stock five has been removed from the index, so the ETF must re-balance its portfolio due to this index adjustment. In order to re-balance the portfolio to match the underlying index, the ETF must sell its entire holding of stock five. If stock five has gained in value from a year ago and the ETF sells stock five at its current value, the ETF investors will incur a significant amount of capital gains tax. To avoid this, the ETF can take advantage of a loophole in the US tax code that allows for in-kind redemptions to withdrawing investors. To use this loophole, the ETF asks a friendly bank to invest using stocks that mirror the ETFs portfolio prior to re-balancing at the current value held of stock five. The next day, the bank withdraws its investment and to satisfy their withdrawal, the ETF transfers to the bank the appropriate amount of shares of stock five equal to the value of their original investment. This is called an in-kind redemption. The remaining value of stock five in the ETF's portfolio, leftover from the banks original investment, can be sold without incurring capital gains taxes, as the bank's withdrawal and stock sale typically takes place a day after the bank's original investment, so it is unlikely that a significant price change occurred. The ETF was able to successfully sell off an appreciated asset without incurring capital gains tax, thanks to a friendly bank and the ability to perform in-kind redemptions. If the ETF simply sold the appreciated asset in the traditional sense, then it would have incurred taxable capital gains taxes for its investors. The previously described transaction between the ETF and the bank is known as a "heartbeat trade" [10], as the large inflow and outflow of the ETF's capital when graphed resembles that of a heart rate monitor. These "heartbeat trades" will often occur when an ETF has to restructure its portfolio, and it allows an ETF to negate the majority of the taxable capital gains on an asset. Depending on the ETF issuer, it is generally known beforehand to external analysts when this kind of trade will occur due to the ETF's restructuring schedule. This is shown in Figure 2.1, where a spike in the inflow and outflow of capital of an ETF shows when restructuring occurs. In Mider et al. [10], this tax loophole is called "Wall Street's 'Dirty Little Secret'" following a comment by an ETF manager. The main driver of this loophole is an ETF's ability to perform tax-free in-kind redemptions.

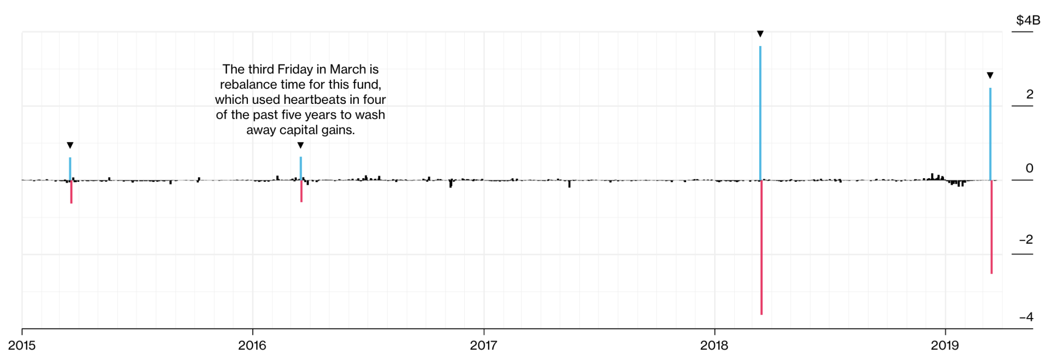

Figure 2.1: Regularly occurring "heartbeat" trades in an ETF due to portfolio restructuring, provided by Mider et al. [10]

The ability to perform in-kind redemptions is a highly desirable aspect of ETFs. These types of redemptions allow the ETF manager to defer capital gains at the fund level to taxable gains at the investor level to when the investor sells; compare to Colon [3]. This is advantageous for ETF investors because they can keep a large investment in the ETF and decide to realize their capital gains when it is optimal for them, taking tax considerations into account. If the ETF did not take advantage of in-kind redemptions when selling an appreciated asset, the incurred capital gains tax from the sale would be paid by the ETF investors in that tax year, reducing their overall return. MFs must typically sell securities in order to increase liquidity when there are large withdrawals by investors, resulting in the MF incurring a capital gains tax if the sold assets had appreciated. Meanwhile, when ETFs experience large volumes of outflows, they are able to take advantage of in-kind redemptions to significantly reduce the amount of capital gains tax that its investors must pay [16]. In fact, as the ETF is not actually selling the security, the investors do not incur capital gains tax. Further, due to the low turnover rate of their portfolio securities, ETFs often do not incur any taxable capital gains. It is estimated that a total of 94% of the ETFs managed by BlackRock, State Street Global Advisors, and Vanguard did not incur any capital gains tax for 2020 due to the ETF tax loophole and a low asset turnover rate [16]. By supposedly taking advantage of this legal loophole, State Street's S&P 500 ETF has not reported a taxable gain in 22 years as of 2019, while the corresponding MF run by Fidelity Investments that also tracks the S&P 500 reported a taxable gain in 10 of the past 22 years [10]. Using this tax loophole, State Street's ETF in 2018 was able to avoid paying taxes on approximately 4 billion USD of capital gains by declaring a 309 million USD loss to the IRS [10]. At the maximum tax rate for capital gains of 20% [8], that equates to 800 million USD of deferred taxes in 2018. Hodaszy [8] proposes a change in the tax code that would allow ETFs to remain an attractive investment among investors, while allowing the ETF investors to defer capital gains taxes from in-kind redemptions to a future year, where the amount of tax the ETF would pay is proportional to the amount of gain the fund realized on the in-kind redemption. It is interesting to note that ETFs are a popular investment also in countries that do not have this loophole or ability to perform tax-free in-kind redemptions [10].

This legal loophole is not a total removal of taxes on capital gains on ETF investors, however. When an ETF performs a "heartbeat trade" to avoid capital gains tax when selling an appreciated asset, the capital gains remain in the ETF, increasing its value. As this capital gain is not taxed, it increases the ETF's return by an amount larger than if taxes were applied to the gain. When an investor withdraws their investment, they are subsequently required to pay more in capital gains tax due to this higher return. A mathematical approach to determine the amount that an ETF's return is increased due to the loophole is discussed in Section 4.3.

2.3 Portfolio Optimization

When investing their capital, it is an investor's desire to maximize their expected return in accordance to their risk preferences. Additionally, if an investor decides to divide their wealth between numerous assets, then the natural question arises of how that wealth should be distributed in order to maximize their expected return given their risk preferences and the riskiness of the assets. This problem is the basis of portfolio optimization, and is one that retail investors and fund managers both share. Prior to exploring potential solutions to this problem, we first explore the underlying mathematics of portfolio optimization.

Since an investor does not know how their investment will behave over the course of a holding period, we use a random variable to represent the wealth at the terminus this period. The terminal wealth, denoted by W_T, is represented by the equation $W_T = w_0(1 + \theta^\top Y_T) = w_0(1 + \sum_{k=1}^{n} \theta_k Y_T^k)$. This random variable assumes a market model with n risky assets, where each component of the vector Y_T represents a risky asset's return. Each component of the vector θ, θ_k, represents the initial proportion of the investor's starting wealth allocated to the risky asset modelled by Y_T^k. Thus, we refer to θ as the weight vector. It is also sometimes referred to as a strategy. Each Y_T^k is itself a one-dimensional random variable, that has an associated expected return μ_k and standard deviation σ_k. The expected returns of each risky asset is compiled into an expected return vector denoted by $\mu = (\mu_1,\ldots,\mu_n)^\top$. The standard deviation represents the riskiness of an asset. If an asset's standard deviation is high, then the asset's return distribution is wider and there is a greater chance of achieving a large return or significant loss. As there is a higher chance of realizing an outlier event given a higher standard deviation, the asset is a riskier investment. If Y_T^k has a small σ_k, comparatively speaking, then the asset is much safer as we are more confident we know the range where the realized return will land in the return distribution. The components of Y_T may additionally be correlated with one another. These correlations, along with the standard deviations of the risky assets, can be used to form a covariance matrix, which we denote by Σ and whose elements are given by $\Sigma_{i,j} = \mathrm{Cov}(Y_T^i, Y_T^j) = \mathrm{corr}(Y_T^i, Y_T^j)\sigma_i\sigma_j$. Here, $\mathrm{Cov}(X,Y)$ and $\mathrm{corr}(X,Y)$ represent the covariance and correlation, respectively, between the random variables X and Y. The initial wealth is represented by w_0, and for simplicity we set $w_0 = 1$. With $w_0 = 1$, the wealth equation takes the form $W_T = 1 + \theta^\top Y_T$.

In our market model, there also exists a risk-free asset. Typically, this asset represents a bond or bank account, where the exact return is deterministic. We represent the wealth allocated to the risk-free asset as θ_0, and apply the constraint that $\sum_{k=0}^{n} \theta_k = 1$. In the case where $\sum_{k=1}^{n} \theta_k < 1$, the remainder of the investor's wealth is allocated to a risk-free asset, which has the risk-free interest rate of r_f. In the event where $\sum_{k=1}^{n} \theta_k > 1$, then θ_0 takes a value less than zero to represent a loan with interest rate r_f from the risk-free asset. For this thesis, we set $r_f = 0$, but all results can be extended to non-zero risk-free interest rates with simple adjustments.

Analyzing the equation for the terminal wealth, we see that it is dependent on the initial weight vector θ. The goal of portfolio optimization is to find the initial weight vector θ that will maximize the expected terminal wealth against some predefined risk preferences. The application of the risk preferences is dependent on the technique of portfolio optimization that is used. There are numerous techniques of portfolio optimization, and in this thesis we focus on an extension of the expected utility model. The expected utility model and its subsequent extension are described in Sections 2.4 and 2.5, respectively.

2.4 Expected Utility Portfolio Optimization

The technique of expected utility with regards to portfolio optimization is fairly straightforward. A utility function, denoted by u, is applied to an investor's terminal wealth. It is the goal of this technique of portfolio optimization to find the optimal portfolio weight vector, θ^*, such that the expectation of the terminal wealth's utility value is maximized. Mathematically, this optimization problem is expressed as

$$\underset{\theta}{\operatorname{argmax}}\, E[u(W_T)].$$

Recall that the terminal wealth W_T is dependent on weight vector θ, as discussed in Section 2.3. Depending on the utility function used and the distribution of the terminal wealth, this optimization cannot always be solved explicitly. In that case, a solution can be approximated numerically. With a continuously distributed W_T, the expectation takes the form of an integral.

The choice of utility function will represent an investor's risk preferences and how they perceive wealth. A concave utility function is used to represent an investor who is risk averse, and a convex function is used to represent an investor who is risk seeking. For example, if we are risk averse we perceive the difference between \$100 and \$150 to be much greater than the difference between \$1,000 and \$1,050, despite the absolute difference of \$50 being the same in both cases. This is because in the first case, the addition of \$50 is a much larger proportion of the wealth than in the second case, and the impact of a wealth change at a lower value has more influence over our decision. In order to represent this, we use concave utility functions. Additionally, a utility function should be increasing, as a higher utility value of the wealth is more desirable than a lower one. As the utility function is increasing and concave, it is effortless to find that $u(150) - u(100) > u(1{,}050) - u(1{,}000)$. Thus, an absolute gain or loss at a lower wealth level has more impact on a risk averse investor's decisions than the equivalent absolute gain or loss at a higher wealth level. To represent investors who are risk seeking a convex utility function is used, and an analogous explanation is given.

To further demonstrate the use a utility function for a risk averse investor, we explore a simple binomial model. Assume that a random variable W, representing the terminal wealth, can take either \$50 or \$0 with equal probability. It is simple to find that $E[W] = \$25$. With a concave utility function, we find $E[u(W)] = (u(\$50) + u(\$0))/2 < u(\$25) = u(E[W])$. This is also immediate from Jensen's inequality. This inequality appropriately represents a risk averse investor, as a negative outcome will have a larger effect on the investment decision more than a positive one. We can use a convex utility function if the investor is risk seeking, and the logic and inequality would be reversed. Due to these properties, utility functions are used to represent an investor's risk preferences. As an investor is typically risk averse, we will only focus on concave utility functions in this thesis.

There are various different forms of utility functions, and each form exhibits specific characteristics. An investor will choose a utility function in accordance to the function's characteristics that most accurately match their risk preferences. Furthermore, the choice of utility function is also dependent on the underlying random variable. Specifically, the domain of the utility function must include the range of the random variable. We analyze three possible utility functions below in Sections 2.4.1, 2.4.2, and 2.4.3, in addition to providing a visual representation of these in Figure 2.2.

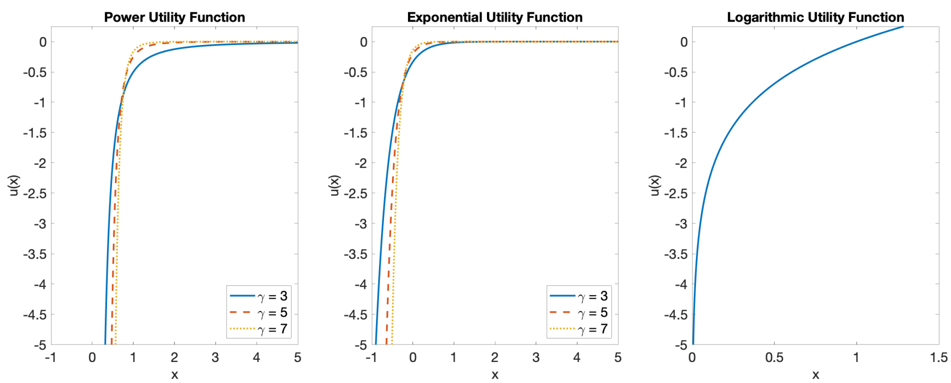

Figure 2.2: Various concave utility functions

2.4.1 Power Utility

A power utility function exhibits the property of constant relative risk aversion (CRRA). This property is represented by the ratio $-\frac{u''(x)}{u'(x)}x$. An example of a power utility function is $u(x) = x^{1-\gamma}/(1-\gamma)$, which is shown in the left panel of Figure 2.2 for different values of γ. The level of risk aversion in the case of power utility is dependent on the parameter γ, and γ is also equal to the CRRA measure. For $\gamma > 1$ and $0 < \gamma < 1$, we obtain a concave function. A larger value of γ increases the curvature, and thus represents an investor who is more risk averse. The limiting case of $\gamma \to 1$ is described in Section 2.4.3. A drawback of using a power utility function is that the domain is restricted to the positive domain. Consequently, the random variable it is being applied to is not able to take negative values.

2.4.2 Exponential Utility

An exponential utility function exhibits the property of constant absolute risk aversion (CARA). This is measured by the ratio $-\frac{u''(x)}{u'(x)}$. An example of an exponential utility function is $u(x) = -\mathrm{e}^{-\gamma x}/\gamma$, where γ represents the risk aversion parameter and the CARA measure is found to be equal to γ. We demonstrate the shape of exponential utility functions for different levels of risk aversion in the middle panel of Figure 2.2. For positive values of γ, we obtain a concave function that represents a risk averse investor. As in the case of power utility, a higher value of γ corresponds to an investor who is more risk averse. Unlike power utility functions, however, an exponential utility function is able to map values from the whole real line, so we are able to use any type of random variable in the optimization problem.

Using an exponential utility function and a random variable with a distribution from the exponential family will often allow one to explicitly solve for the optimal weight vector θ^* in the expected utility optimization problem. For normally distributed wealth and utility function $u(x) = -\mathrm{e}^{-\gamma x}/\gamma$, this solution takes the form of $\theta^* = \frac{1}{\gamma}\Sigma^{-1}\mu$, where Σ is the covariance matrix of the risky asset returns and μ is the expected return vector of the assets with the assumption that the risk-free interest rate is zero. Recall that our wealth equation is given by $W_T = 1+\theta^\top Y_T$ when the initial wealth is equal to one. We introduce a new random variable X_T, which is given by $X_T = W_T - 1 = \theta^\top Y_T$. This new random variable represents a shift in the initial wealth level. When using an exponential utility function for expected utility portfolio optimization, the optimal weight vector θ^* associated with W_T is the same optimal weight vector that is associated with the random variable X_T, as the optimal weights do not depend on the initial wealth level.

2.4.3 Logarithmic Utility

Recall that in the case of power utility, the function is $u(x) = x^{1-\gamma}/(1-\gamma)$. By taking the limit as $\gamma \to 1$ and applying L'Hôpital's rule, we obtain

$$\lim_{\gamma \to 1} \frac{x^{1-\gamma}}{1-\gamma} = \log(x),$$

where $\log(x)$ denotes the natural logarithm. The logarithmic utility function exhibits the CRRA property, with its CRRA value equal to one. We show the logarithmic utility function in the right panel of Figure 2.2. While both the power and exponential utility functions have horizontal asymptotes at zero when $\gamma > 1$, the logarithmic utility function has no horizontal asymptote. As is the case with the power utility function, the domain of the logarithmic utility function is restricted to positive values.

2.5 Rank-Dependent Expected Utility

When deciding what assets to invest their money into, an investor will sometimes overweight or underweight the probability of the extreme scenarios that an asset can realize. This phenomenon has been shown empirically numerous times, including by Camerer and Ho [2], and Tversky and Kahneman [17]. While a utility function can be used to represent an investor's risk aversion, it cannot properly represent an investor's bias of the tail probabilities. Therefore, we must extend the expected utility model to incorporate these investor's biases.

To better understand the phenomenon of investor bias, we offer some examples. Let us imagine that a certain asset's return follows a normal distribution with standard deviation σ. The probability that a fund has a very large gain ($> 2\sigma$) is very small and unlikely to occur. An eager investor would of course be ecstatic if their investment had such a high return, and will subsequently develop a bias. This investor will overweight the probabilities of the right tail events as they believe that they will be the lucky one who will realize this significant and unlikely return. Consequentially, this investor is more willing to invest their money into this asset, despite the real world probability of experiencing this significant gain being less than 2.5%. On the other side of the distribution we have the investor's losses. The probability of realizing a significant loss ($< -2\sigma$) is also incredibly small. Despite this, a hesitant investor will view this scenario and choose not to invest due to the possibility, not the probability, of a significant loss. This hesitant investor is overweighting the probability of the left tail events in the distribution. Meanwhile, a third and perhaps more relatable investor is likely to overweight the probability of both the left and right tail events when deciding to invest, as they not only believe they can achieve a significant gain, they are also hesitant due to the risk of a large loss. Despite certain events having an incredibly small probability of occurring, investors believe that these low probability events will happen to them. A more colloquial example of this phenomenon is someone who always buys a lottery ticket but refuses to hike in the Rockies out of fear of animal attacks. While both the probabilities of winning the lottery and being devoured by a hungry grizzly bear are essentially zero, this person believes that they are not only going to be the lucky winner of a lottery, but also that if they step foot in the mountains an unfortunate encounter with a grizzly bear is inevitable.

In order to represent these investor biases, a probability weighting function is used. A probability weighting function, which we will denote by G, satisfies certain properties that correspond to the aforementioned biases. The weighting function is monotonically increasing, continuous, has domain and range equal to [0,1], with $G(0) = 0$ and $G(1) = 1$, and is differentiable on (0,1). Additionally, G has an inverse S-shape which allows an overweighting to be applied to the probabilities of tail events. The exact curvature of the weighting function is typically parameter dependent. Examples of weighting function and the possible curvatures are explored later in Sections 3.2 and 4.2, and visual examples are provided in Figures 3.1 and 4.1. The curvature of a weighting function produces a non-linearity on a distribution's probabilities. Further analysis of a weighting function is given by Wu and Gonzalez [18]. The weighting function is applied to the cumulative distribution function (CDF) of a random variable in order to distort the probabilities. This allows the weighting function to model the investor's risk attitude for the probabilities of the ranked events [12].

When we use a probability weighting function, we are modifying the original probabilities. When we use the modified probabilities with the expected utility model, we obtain the RDEU model. The RDEU model is presented and thoroughly analyzed by Quiggin [15]. Using this new framework and a continuously distributed random variable, our classical expected utility integral for continuous random variables morphs into

$$\int_{-\infty}^{\infty} u(w)\, dG(F_{W_T}(w)) = \int_{-\infty}^{\infty} u(w) Z(F_{W_T}(w))\, dF_{W_T}(w)$$
$$= E[u(W_T) Z(F_{W_T}(W_T))],$$

where Z represents the derivative of our weighting function G, and F_{W_T} represents the CDF of the wealth W_T. With this new integral, our expected utility optimization problem under the RDEU framework now takes the form

$$\underset{\theta}{\operatorname{argmax}}\, E[u(W_T) Z(F_{W_T}(W_T))].$$

Analyzing this expectation, we see that an investor will be overweighting the probability of scenarios ω that satisfy $Z(F_{W_T}(W_T(\omega))) > 1$ compared to the real world probability. By contrast, the investor is underweighting the probability of scenarios ω with $Z(F_{W_T}(W_T(\omega))) < 1$. For scenarios ω with $Z(F_{W_T}(W_T(\omega))) = 1$, the investor has no bias when compared to the real world probabilities. Scenarios with overweighted probabilities will have a greater impact on the optimal solution, and those with underweighted probabilities will have less of an impact on the optimal solution. Depending on the form of the weighting function and its derivative, the utility function, and the distribution of the wealth, we may or may not be able to solve the optimization problem explicitly.

The RDEU model provides an explanation of the “mutual fund puzzle”, as it is able to mathematically represent an investor overweighting tail probabilities [12]. Additionally, Polkovnichenko et al. [12] found that the RDEU model allows one to represent the investor bias for active funds over passive funds due to the perceived downside protection and upside potential. In this thesis we use an RDEU model as we are interested in investor portfolios containing MFs and ETFs, where the MF represents an active fund and the ETF represents a passive fund. Given the observed market behaviour of investment in active and passive funds, it would be incorrect to use a classical model when describing investor portfolios due to the investor biases.

As weighting function G is applied to a CDF of a random variable, the rank orders of the outcomes for $Z(F_{W_T})$ and $Z(F_{X_T})$ are equivalent, where $W_T = 1 + \theta^\top \mu$ and $X_T = W_T - 1$. Depending on the utility function chosen, we are therefore able to find the same optimal strategy using either W_T or X_T, in the RDEU framework. If we also use an exponential utility function, discussed in Section 2.4.2, we will find the same optimal weight vector θ^* using either X_T or W_T under a distorted probability model.

Chapter 3

An Explicit Case

In this chapter, we will investigate a specific setting that will allow us to explicitly solve the RDEU optimization problem. We begin by specifying the market model in Section 3.1. In Section 3.2, we introduce a specific weighting function, analyze its properties, and solve for the optimal portfolio weights in the RDEU model using this specific weighting function. Finally, in Section 3.3, we analyze how the optimal portfolio weights vary given a theoretical change in the ETF's expected return.

3.1 The Market Model

We assume that our market has a risk-free asset and n risky assets, whose returns follow a multivariate normal distribution with mean vector μ and covariance matrix Σ. These parameters μ and Σ correspond to a single holding period. In a real-world setting, we can extract expected returns, volatilities, and correlations for risky assets from historical data, market projections, and options. The risky asset return random vector is represented by Y_T. We assume that Σ is invertible, therefore no arbitrage opportunities exist in the market. The optimal portfolio weights will be denoted by θ^*, and these will represent the proportion of initial capital that should be invested into the risky assets at the start of the holding period that will maximize the rank-dependent expected utility value. As stated in Section 2.3, we set the initial wealth w_0 equal to one and the risk-free interest rate equal to zero. Therefore, the equation for our terminal wealth is given by $W_T = 1 + \theta^\top Y_T$. Since our risky assets follow a multivariate normal model with mean vector μ and covariance matrix Σ, our wealth will have distribution $N(1 + \theta^\top \mu,\ \theta^\top \Sigma\theta)$. As there exists a possibility our wealth can realize a negative value, we will use an exponential utility function of the form $u(x) = -\exp(-\gamma x)/\gamma$, previously described in Section 2.4.2. As a consequence, we are able to use $X_T = W_T - 1 = \theta^\top Y_T$ in place of $W_T = 1 + \theta^\top Y_T$ to represent the terminal wealth. The random variable X_T will have the distribution $N(\theta^\top \mu, \theta^\top \Sigma\theta)$. For the remainder of this thesis, we will use the random variable X_T. The CDF of X_T we will denote by F_{X_T}, and its corresponding probability density function (PDF) by F'_{X_T}.

In our model, we assume that the gains are only realized at the end of the holding period. Therefore, we are only concerned with the distribution of the terminal wealth. Further, we assume no interaction can take place between the investor and their wealth after their initial investment until the end of the holding period, when they realize their gains. As we have no intertemporal consumption and are only interested at the beginning and end of the investment period, we do not use a continuous time stochastic model. Instead, we use a single-period model.

3.2 Derivation

In this section, we determine the optimal portfolio weights in the distorted probability model. As we are using an RDEU model, we require a weighting function that will allow us to explicitly solve the optimization problem from Section 2.5. To accomplish this, we will use the weighting function

$$G(P) = \Phi(\alpha\Phi^{-1}(P)),$$

where Φ represents the CDF of the standard normal distribution and α represents the level of distortion. This weighting function was presented by Hu et al. [9], however, they only consider the case where $\alpha = 1/\sqrt{2}$. We will now explore some key properties of this function. For $\alpha > 0$, this function satisfies the properties of a weighting function, which were previously described in Section 2.5. When $\alpha \in (0,1)$ we obtain an overweighting of both the tail probabilities that is symmetric about $P = 1/2$. This is shown in Figure 3.1 for various values of α. This means that an investor will apply equivalent overweightings to both left and right tail probabilities. For $\alpha = 1$, the weighting function becomes $G(P) = P$ and we have no distortion. As α decreases from one, the distortion increases symmetrically on both the right and left tails of P. The inflection point of this weighting function, for $\alpha \in (0,1)$, is equal to 1/2. For α in this range, the amount of tail distortion will increase as α decreases. We will refer to this weighting function as the *normal* weighting function.

In our RDEU optimization problem, recall that we are solving for the weight vector θ that maximizes the expected value $E[u(X_T)Z(F_{X_T}(X_T))]$,

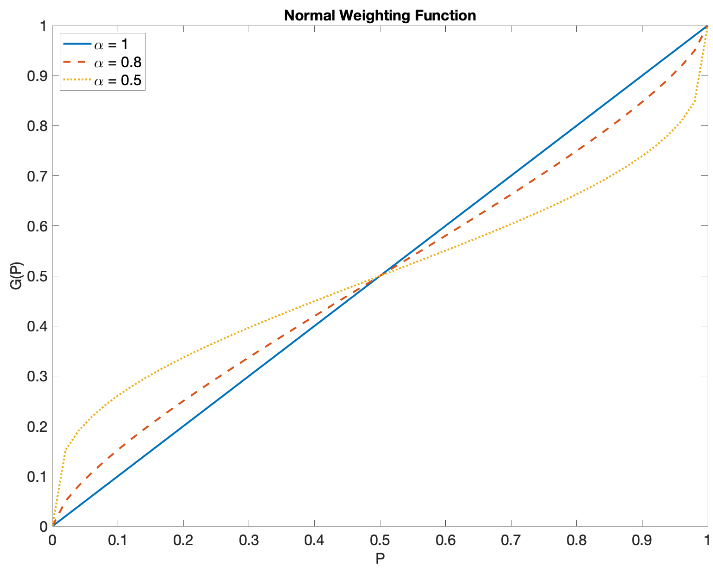

Figure 3.1: Effects of parameter α on normal weighting function curvature

where Z represent the derivative of a weighting function G. Taking the derivative of our normal weighting function, we obtain

$$Z(P) = \alpha \frac{\Phi'(\alpha \Phi^{-1}(P))}{\Phi'(\Phi^{-1}(P))}.$$

We substitute into the above equation the normal CDF of our terminal wealth X_T given by $F_{X_T}(X_T)$ to find

$$Z(F_{X_T}(X_T)) = \alpha \exp\left(\frac{1-\alpha^2}{2}\left(\frac{X_T - \theta^\top \mu}{\sqrt{\theta^\top \Sigma \theta}}\right)^2\right).$$

Given the framework described in Section 3.1, we are able to explicitly solve for the optimal weights given the distorted probability. We do so by writing

$$
\begin{aligned}
&E[u(X_T)Z(F_{X_T}(X_T))] \\
&= \int_{-\infty}^{\infty} u(x)Z(F_{X_T}(x))F'_{X_T}(x)\,dx \\
&= \frac{-\alpha}{\gamma\sqrt{2\pi\theta^\top\Sigma\theta}} \int_{-\infty}^{\infty} \exp\bigg(- \gamma x + \frac{1-\alpha^2}{2}\Big(\frac{x-\theta^\top\mu}{\sqrt{\theta^\top\Sigma\theta}}\Big)^2 - \frac{1}{2}\Big(\frac{x-\theta^\top\mu}{\sqrt{\theta^\top\Sigma\theta}}\Big)^2\bigg)\,dx \\
&= \frac{-\alpha}{\gamma\sqrt{2\pi\theta^\top\Sigma\theta}} \int_{-\infty}^{\infty} \exp\bigg(- \gamma x - \frac{\alpha^2}{2}\Big(\frac{x-\theta^\top\mu}{\sqrt{\theta^\top\Sigma\theta}}\Big)^2\bigg)\,dx \\
&= \frac{-\alpha}{\gamma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\bigg(- \gamma\big(\theta^\top\mu + \sqrt{\theta^\top\Sigma\theta}y\big) - \frac{\alpha^2}{2}y^2\bigg)\,dy \\
&= \frac{-1}{\gamma}\exp\bigg(- \gamma\theta^\top\mu + \frac{\gamma^2}{2\alpha^2}\theta^\top\Sigma\theta\bigg).
\end{aligned}
$$

From this point, we take the partial derivative of the above expression with respect to θ, utilizing properties from matrix calculus. As the variable we are optimizing over exists only in the argument of the exponential, it is enough to take the partial derivative of the argument itself. We set this partial derivative equal to zero, and find

$$
[0,\dots,0]^\top = \frac{\partial}{\partial\theta}\bigg(- \gamma\theta^\top\mu + \frac{\gamma^2}{2\alpha^2}\theta^\top\Sigma\theta\bigg) = -\gamma\mu + \frac{\gamma^2}{\alpha^2}\Sigma\theta,
$$

which yields the optimal solution for θ to be

$$
\theta^* = \frac{\alpha^2}{\gamma}\Sigma^{-1}\mu. \tag{3.1}
$$

This solution is obtained when using either W_T or X_T to represent our wealth. To verify that this indeed is a maximum, it is useful to check the second partial derivative with respect to θ^*. This is equal to the positive definite matrix $\frac{\gamma^2}{\alpha^2}\Sigma$, affirming that the solution in (3.1) is indeed the global optimizer. In the case of $\alpha = 1$, we recover the solution from classical expected utility with exponential utility, described in Section 2.4.2. For values $\alpha < 1$, the investor will simply allocate less of their initial wealth to their risky assets than they would in the case where no distortion is applied. This reduction is proportional across all risky assets. This is despite the fact that the investor is overweighting both the left and right tail probabilities equally. This reduction in risky asset investment can be explained by the fact that for a concave utility function, negative outcomes have a greater influence on the optimal weights than positive outcomes. This influence grows as more distortion is applied, as the weighted negative outcomes have more impact on the wealth's utility value than the weighted positive outcomes, due to the concavity of utility function. It follows that the investor will increase their investment in the risk-free asset due to the reduction in risky asset investment.

There are a few reasons why we were able to solve this optimization problem explicitly despite using a distorted probability model. In our expected value, the multiplication of our weighting function's derivative, PDF of the wealth distribution, and utility function resulted in a single exponential function. This exponential had an argument directly containing the variable of integration in a polynomial of degree less than two with a negative coefficient on the squared term. These traits allowed us to explicitly solve for the optimal weights, and will heavily influence our procedure in Chapter 4.

3.3 A Change in Expected Asset Returns

Now that we have an analytical result for the optimal portfolio weights under our specific RDEU framework, we are able to analyze how the optimal weights will change in a portfolio containing an MF and ETF in the event that the ETF's expected return shifts. We can obtain the optimal strategy for this bi-variate market model using (3.1). This optimal strategy will be denoted by $\theta^* = [\theta_1^*, \theta_2^*]^\top$, where θ_1^* corresponds to the initial proportion of wealth allocated to the MF and θ_2^* represents the proportion of wealth allocated to the ETF.

We represent the ETF's new expected return by shifting its original expected return, μ_2, by some value τ, so that the new return is given by $\tilde{\mu}_2 = \mu_2 - \tau$, while the expected return of the MF remains unchanged and equal to μ_1. We assume that $\tau > 0$, so we have $\tilde{\mu}_2 < \mu_2$. We denote $\tilde{\mu} = [\mu_1, \tilde{\mu}_2]^\top$ to be the expected return vector of the risky assets under this new scenario. The distribution of our terminal wealth with this shifted mean vector is $N(\theta^\top \tilde{\mu}, \theta^\top \Sigma \theta)$. For this new distribution, we can compute the optimal strategy using (3.1). We denote the optimal strategy under the shifted mean vector by $\tilde{\theta}^*$. We are interested in how the portfolio weights for the MF and ETF will change under this shifted mean vector, so we calculate

$$\begin{aligned} \tilde{\theta}^* - \theta^* &= \frac{\alpha^2}{\gamma}\Sigma^{-1}\tilde{\mu} - \frac{\alpha^2}{\gamma}\Sigma^{-1}\mu, \\ &= \frac{\alpha^2}{\gamma}\Sigma^{-1}(\tilde{\mu} - \mu). \end{aligned}$$

This formula holds for multiple risky assets, but since we are only interested in the relationship between an MF and an ETF, we consider the two-dimensional case. As a consequence, we can break this equation down into its components without much difficulty. We first note that $\tilde{\mu} - \mu = [0, -\tau]^\top$. As the ETF is the passive counterpart to the MF, we will assume a positive correlation, and we denote the correlation by ρ. Consequently, the off-diagonal components of Σ^{-1} will be negative. As Σ is a two-by-two matrix, we compute

$$\Sigma^{-1} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

Given this matrix inverse, we can directly calculate each component of $\tilde{\theta}^* - \theta^*$. These components, which we denote by $\Delta\theta_i^*$ for $i = 1, 2$, are given by

$$\Delta\theta_1^* = \frac{\alpha^2\rho}{\gamma\sigma_1\sigma_2(1-\rho^2)}\tau,$$
$$\Delta\theta_2^* = -\frac{\alpha^2}{\gamma\sigma_2^2(1-\rho^2)}\tau.$$

It is obvious from the above relations that the investment in the ETF reduces and investment in the MF grows when the ETF's expected return decreases, assuming positive correlation ρ. The magnitude of the change investment strategy is sensitive to changes τ, the size of the expected return shift, which is to be expected. Furthermore, we find that the absolute change in weights for both θ_1^* and θ_2^* decreases with α. This is because as α decreases, applying more distortion to the tail probabilities, the initial investment in both assets decreases, resulting in a smaller shift in the optimal strategy. For investors who are more risk averse, which is represented by a high γ value, their initial investment in both the ETF and MF is less than that of investors who have a lower risk aversion, so the change in strategy for highly risk averse investors is smaller as well. For $\Delta\theta_1^*$ and a non-zero correlation, we find that the investment in the MF will increase as the ETF's mean is decreased because the ETF becomes a less attractive investment, thus making the MF relatively more attractive to investors. This is an example of the substitution effect.

If the correlation between the returns of the two assets is zero, then the optimal weight of the MF will remain unchanged. That is, if $\rho = 0$, then any change in the mean of the ETF will not alter the investment in the MF. For positive and increasing correlation, the MF will become more sensitive to changes in the ETF, and vice versa. In this case, it will become easier to replace one asset by the other in the portfolio, due to their increasingly similar return structure. We do not consider the case where $\rho = 1$, as this will admit an arbitrage opportunity between the two risky assets. Interestingly, $\Delta\theta_1^*$ still contains a term of σ_2, so if the ETF has a high standard deviation, there will be a smaller change in the portfolio weight of the MF. While this is obvious mathematically, we also provide a financial explanation. Given positive correlation, an investor is less willing to invest more of their capital into the MF, as its return structure is still correlated to the ETF. If the ETF is a significantly risky investment, then covariance between the MF and ETF will be higher, leading to more volatility in the MF's return structure, resulting in the MF being a less attractive investment. Similar arguments can be made for the case where we increase the ETF's expected return, or in the case where we alter the MF's return.

We have analyzed using the normal weighting function how a change in the

ETF's expected return will result in a decrease in initial portfolio allocation to an ETF and an increase in initial portfolio allocation to the corresponding MF. The extent of the shift in optimal strategy is dependent on multiple factors, including the shift in the ETF's expected return, the distortion parameter of the weighting functions, and the covariance structure of the model. While this problem is relatively easy to solve given this chapter's explicit framework, we would like to consider this problem in a more general setting. We do so in the following chapter.

Chapter 4

An Approximation

In Chapter 3, we saw that if given a specific form of the weighting function's derivative, along with normally distributed wealth and an exponential utility function, we were able to solve explicitly for the optimal weights. With our normal weighting function from the previous chapter, shown in Figure 3.1, the distortion applied to the probabilities is symmetric about $P = 1/2$, and $P = 1/2$ is also the normal weighting function's inflection point. Keeping these properties in mind, we will form an approximation for the derivative of a generic weighting function in Section 4.1. Using this approximation, we will solve for the optimal portfolio weights using this in our RDEU model. In Section 4.2, we analyze the accuracy of our approximation with numerical results. Finally, in Section 4.3, using our approximation we answer our question of how the ETF tax loophole will affect investor portfolios in the event it is closed. As before, we only consider a single-period model.

4.1 Derivation

We form an approximation on the derivative of our probability weighting function, as in the RDEU model the calculation of the expected value requires the derivative of the weighting function. Firstly, we will assume for a general weighting function G, with derivative denoted by Z, that the inflection point is approximately $1/2$, so we have $G''(1/2) = Z'(1/2) \approx 0$. As before, we set the risk-free interest rate to zero and initial wealth to one, and we use X_T in place of W_T. We assume that the terminal wealth has an approximately normal distribution. In other words, the CDF of the wealth, F_{X_T}, can be approximated by a normal CDF. We maintain the notation previously introduced in Section 3.1, where μ represents the vector of expected returns, Σ represents the covariance matrix of the risky assets, and θ represents the vector of portfolio weights. We again assume that Σ is invertible, and therefore no arbitrage opportunities exist. Given our model parameters, for approximately normally distributed wealth we have the approximations $F_{X_T}(E[X_T]) \approx 1/2$, $F'_{X_T}(E[X_T]) \approx 1/\sqrt{2\pi\theta^\top\Sigma\theta}$, and $F''_{X_T}(E[X_T]) \approx 0$. We construct a second-order Taylor approximation on $Z(F_{X_T}(X_T))$ centered about $E[X_T]$, where in our setting we have $E[X_T] = \theta^\top\mu$. This approximation is formed by

$$
\begin{aligned}
&Z(F_{X_T}(X_T)) \\
&\approx Z(F_{X_T}(\theta^\top\mu)) + Z'(F_{X_T}(\theta^\top\mu))F'_{X_T}(\theta^\top\mu)(X_T - \theta^\top\mu) \\
&\quad + \frac{Z'(F_{X_T}(\theta^\top\mu))F''_{X_T}(\theta^\top\mu) + Z''(F_{X_T}(\theta^\top\mu))\big(F'_{X_T}(\theta^\top\mu)\big)^2}{2}(X_T - \theta^\top\mu)^2 \\
&\approx Z(1/2) + Z'(1/2)F'_{X_T}(\theta^\top\mu)(X_T - \theta^\top\mu)
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{Z'(1/2)F''_{X_T}(\theta^{*\top}\mu)+Z''(1/2)\big(F'_{X_T}(\theta^\top\mu)\big)^2}{2}(X_T-\theta^\top\mu)^2\\
&\approx Z(1/2)+\frac{Z''(1/2)}{4\pi\theta^\top\Sigma\theta}(X_T-\theta^\top\mu)^2\\
&=Z(1/2)\left(1+\frac{Z''(1/2)}{Z(1/2)\,4\pi\theta^\top\Sigma\theta}(X_T-\theta^\top\mu)^2\right)\\
&\approx Z(1/2)\exp\left(\frac{Z''(1/2)}{Z(1/2)\,4\pi\theta^\top\Sigma\theta}(X_T-\theta^\top\mu)^2\right)\\
&=c_1\exp\left(c_2\left(\frac{X_T-\theta^\top\mu}{\sqrt{\theta^\top\Sigma\theta}}\right)^2\right),
\end{aligned}
$$

where we have set $c_1 = Z(1/2)$, and $c_2 = Z''(1/2)/(4\pi Z(1/2))$. From this point, we solve for the optimal weights using our approximation and an exponential utility function by maximizing $E[u(X_T)Z(F_{X_T}(X_T))]$ with respect to θ. When computing this expectation, we perform a u-substitution to simplify the integral to obtain

$$
\begin{aligned}
&E[u(X_T)Z(F_{X_T}(X_T))]\\
&=\int_{-\infty}^{\infty}u(x)Z(F_{X_T}(x))F'_{X_T}(x)\,dx\\
&\approx\frac{-c_1}{\gamma\sqrt{2\pi\theta^\top\Sigma\theta}}\int_{-\infty}^{\infty}\exp\left(-\gamma x+c_2\left(\frac{x-\theta^\top\mu}{\sqrt{\theta^\top\Sigma\theta}}\right)^2-\frac{1}{2}\left(\frac{x-\theta^\top\mu}{\sqrt{\theta^\top\Sigma\theta}}\right)^2\right)dx\\
&=\frac{-c_1}{\gamma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\gamma\big(\theta^\top\mu+\sqrt{\theta^\top\Sigma\theta}u\big)+c_2u^2-\frac{1}{2}u^2\right)du\\
&=\frac{-c_1}{\gamma\sqrt{2\pi}}\frac{\sqrt{\pi}}{\sqrt{\frac{1}{2}-c_2}}\exp\left(\frac{\gamma^2\theta^\top\Sigma\theta}{2-4c_2}-\gamma\theta^\top\mu\right)\\
&=\frac{-c_1}{\gamma\sqrt{1-2c_2}}\exp\left(\frac{\gamma^2\theta^\top\Sigma\theta}{2-4c_2}-\gamma\theta^\top\mu\right).
\end{aligned}
$$

Similarly to Section 3.2, we take the partial derivative with respect to θ of the exponential argument and set it equal to zero to solve for the optimal weight vector. Again, by using properties from matrix calculus we are able to find

$$[0, \ldots, 0]^\top = \frac{\partial}{\partial \theta}\left(\frac{\gamma^2 \theta^\top \Sigma \theta}{2 - 4c_2} - \gamma \theta^\top \mu\right) = \frac{2\gamma^2}{2 - 4c_2} \Sigma \theta - \gamma \mu.$$

This provides the solution for the approximate optimal weight vector to be

$$\theta^* \approx \frac{1 - 2c_2}{\gamma} \Sigma^{-1} \mu. \tag{4.1}$$

As was the case in Section 3.2, we can easily confirm that this is the global solution by taking the second partial derivative. Using our approximation on the derivative of the weighting function, we have estimated the solution to the RDEU portfolio optimization problem as

$$\underset{\theta}{\operatorname{argmax}}\, E[u(X_T) Z(F_{X_T}(X_T))] \approx \frac{1 - 2c_2}{\gamma} \Sigma^{-1} \mu.$$

It is clear to see that this solution for the approximated optimal weights is the same when using either X_T or W_T to represent the wealth. If we were to use W_T, the approximation of the weighting function derivative is instead given by

$$Z(F_{W_T}(W_T)) \approx c_1 \exp\left(c_2 \left(\frac{W_T - (\theta^\top \mu + 1)}{\sqrt{\theta^\top \Sigma \theta}}\right)^2\right).$$

In fact, for any approximately normal random variable X, we find the approximation of a probability weighting function's derivative centered about the approximate inflection point of 1/2 to be

$$Z(F_X(X)) \approx c_1 \exp\left(c_2 \left(\frac{X - E[X]}{\sqrt{\mathrm{Var}(X)}}\right)^2\right).$$

To verify this approximation, we calculate the values for c_1 and c_2 for the case of the normal weighting function $G(P) = \Phi(\alpha\Phi^{-1}(P))$ from Chapter 3. If this approximation is valid, we hope to return the exact solution found in (3.1). We first compute c_1. Since $\Phi^{-1}(1/2) = 0$, this is a fairly easy computation, and we find

$$c_1 = Z(1/2) = \alpha\frac{\Phi'(\alpha\Phi^{-1}(1/2))}{\Phi'(\Phi^{-1}(1/2))} = \alpha\frac{\Phi'(0)}{\Phi'(0)} = \alpha.$$

So far, this aligns with the normal weighting function from Chapter 3. For c_2 we require the second derivative of Z, which for our normal weighting function we compute as

$$\begin{aligned}
Z(P) &= \frac{\alpha\Phi'(\alpha\Phi^{-1}(P))}{\Phi'(\Phi^{-1}(P))}, \\
Z'(P) &= \frac{\alpha^2\Phi''(\alpha\Phi^{-1}(P))}{\big(\Phi'(\Phi^{-1}(P))\big)^2} - \frac{\alpha\Phi''(\Phi^{-1}(P))\Phi'(\alpha\Phi^{-1}(P))}{\big(\Phi'(\Phi^{-1}(P))\big)^3}, \\
Z''(P) &= \frac{\alpha^3\Phi'''(\alpha\Phi^{-1}(P))}{\big(\Phi'(\Phi^{-1}(P))\big)^3} - \frac{2\alpha^2\Phi''(\alpha\Phi^{-1}(P))\Phi''(\Phi^{-1}(P))}{\big(\Phi'(\Phi^{-1}(P))\big)^4} \\
&\quad - \frac{\alpha\Phi'''(\Phi^{-1}(P))\Phi'(\alpha\Phi^{-1}(P)) + \alpha^2\Phi''(\alpha\Phi^{-1}(P))\Phi''(\Phi^{-1}(P))}{\big(\Phi'(\Phi^{-1}(P))\big)^4} \\
&\quad + \frac{3\Phi'(\Phi^{-1}(P))\Phi'(\alpha\Phi^{-1}(P))\big(\Phi''(\Phi^{-1}(P))\big)^2}{\big(\Phi'(\Phi^{-1}(P))\big)^6}.
\end{aligned}$$

Substituting in $P = 1/2$, and using the fact that $\Phi^{-1}(1/2) = 0$, $\Phi'(0) = 1/\sqrt{2\pi}$, $\Phi''(0) = 0$, and $\Phi'''(0) = -1/\sqrt{2\pi}$, we find that $Z''(1/2) = 2\pi\alpha - 2\pi\alpha^3$. Thus, we calculate c_2 as

$$c_2 = \frac{Z''(1/2)}{4\pi Z(1/2)} = \frac{2\pi\alpha - 2\pi\alpha^3}{4\pi\alpha} = \frac{1-\alpha^2}{2}.$$

We find that c_2 is equal to the coefficient of our normal weighting function from Chapter 3, which verifies our approximation. Furthermore, it is simple to see that $1 - 2c_2 = \alpha^2$, which agrees with the solution in (3.1).

In addition to approximating a general probability weighting function's derivative, we can also form an approximation on the utility function in a similar manner. For this approximation, we will take a first order Taylor approximation and centre it about zero, and find for wealth X_T that

$$\begin{aligned} u(X_T) &\approx u(0) + u'(0)X_T \\ &= u(0)\left(1 + \frac{u'(0)}{u(0)}X_T\right) \\ &\approx u(0)\exp\left(\frac{u'(0)}{u(0)}X_T\right) \\ &= u_1 \exp\big(u_2 X_T\big), \end{aligned}$$

where $u_1 = u(0)$ and $u_2 = u'(0)/u(0)$. With an exponential utility function, it is trivial to show that $u_1 = -\gamma^{-1}$, and $u_2 = -\gamma$ which agrees with the form of our exponential utility function from Section 2.4.2. With this additional approximation, our approximation for the optimal weight vector in the RDEU model now becomes

$$\theta^* \approx \frac{1 - 2c_2}{u_2}\Sigma^{-1}\mu, \tag{4.2}$$

where

$$c_2 = \frac{Z''(1/2)}{4\pi Z(1/2)},\ u_2 = \frac{u'(0)}{u(0)}.$$

The approximation of the utility function can also be used for classical expected utility optimization problems. However, for a more rigorous method of utility function approximation applied to portfolio optimization problems, we direct the reader to Fahrenwaldt and Sun [5], who provide a formula that decomposes the expected utility value using Taylor polynomials in continuous time and allows for the optimization to be solved explicitly.

4.2 Analysis of Approximation

To evaluate the accuracy of our approximation, we compare our approximated solution for the optimal weights against numerically solved optimal weights. For this analysis, we use the Prelec weighting function [14]. This weighting function takes the form

$$G(P) = \exp(-(-\beta \log(P))^{\alpha}).$$

For this specific weighting function, the shape of the curvature is dependent on the choice of parameters α and β. In the case where $\alpha = \beta = 1$, we have no distortion and we return the classic expected utility problem. We present how the parameters α and β affect the shape of the weighting function in Figure 4.1.

While an in-depth analysis of this weighting function was made by Polkovnichenko [12], we offer our own brief interpretation. Looking first at the curves in the left panel of Figure 4.1, where we set $\beta = 1$, we obtain the desired inverse S-shape of the weighting function. This shape becomes more pronounced for lower values of α, and we find that the parameter α directly contributes to

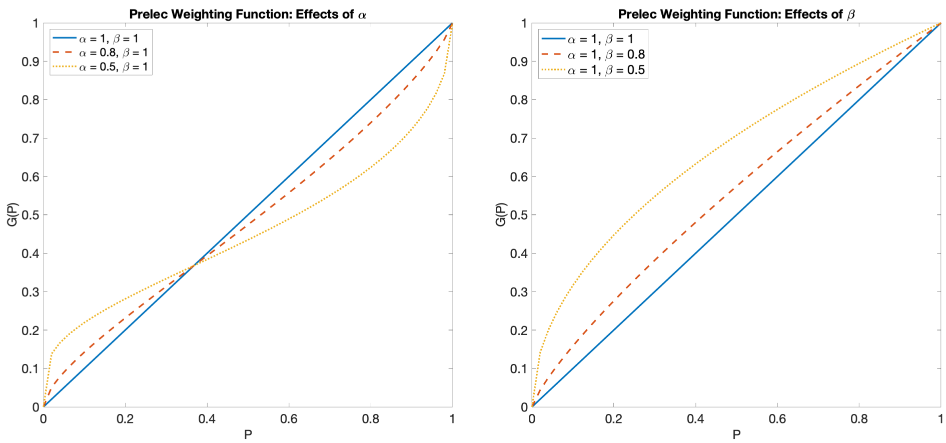

Figure 4.1: Effects of parameters α and β on the Prelec weighting function curvature

overweighting in both the left and right tails of the distribution. For $\beta = 1$, the inflection point of the Prelec weighting function is equal to $1/\mathrm{e}$. As a consequence, the distortion on the left and right tail is not symmetric.

Next, we briefly explore how changes in β affect the curvature of the weighting function, while holding $\alpha = 1$. This is shown in the right panel of Figure 4.1. We notice that the S-shape disappears, and the weighting function simplifies to $G(P) = P^{\beta}$. For $\beta < 1$, the weighting function is concave, and its curvature becomes more aggressive as we decrease β. We find that decreasing β will increase the weighting of the left tail probabilities. Consequentially, values of P close to one are weighted with less importance. In the case where $\beta \neq 1$, the inflection point of the Prelec weighting function is no longer guaranteed to be equal to $1/\mathrm{e}$, and for $\beta < 1$ and $\alpha = 1$ the inflection point does not exist.

4.2.1 Experimental Setup

To validate our approximation using the Prelec weighting function, we compare our approximate solution against a numerical solution for the optimal weights. As we are only concerned with portfolios containing an MF (active fund) and an ETF (passive fund), we validate the approximation using two risky assets from a bivariate normal distribution. We set the first risky asset in the model to represent the MF, and the second risky asset in our model to represent the ETF. For our distribution parameters, we set $\mu = [0.08, 0.085]^\top$ as our risky asset return vector, $\sigma = [0.20,0.25]^\top$ as the risky assets' standard deviations, and $\rho = 0.5$ as the correlation between the two risky assets. As before, we set the risk-free interest rate $r_f = 0$, and we use an exponential utility of the form $u(x) = -\exp(-\gamma x)/\gamma$ with $\gamma = 5$. We simulate 200,000 asset returns from a bivariate normal distribution, and with these simulated returns compute the numerical solution for the RDEU optimization problem. As we are simulating data, we expect there to be minor deviations between the simulated distribution parameters and the original parameters. Consequently, this will lead to deviations in the numerically computed optimal strategy and the analytically calculated optimal strategy. We compute the mean and standard deviation of our two simulated risky assets, as well as the correlation between them. We show this in Table 4.1. In this table we see deviations

	μ	σ	ρ
True Parameter	$[0.08, 0.085]^\top$	$[0.20,0.25]^\top$	0.5
Estimated Parameter	$[0.0799,0.0849]^\top$	$[0.1999,0.2497]^\top$	0.4987

Table 4.1: Comparison of true and estimated parameters from simulation

between the simulated distribution statistics and the original statistics. This will cause discrepancies between the numerically computed and analytical solutions for θ^*, even without probability weighting applied. As such, we can attribute a small amount in the difference of the numerical and approximated optimal weights to numerical and simulation error. This error will be clearly demonstrated in the case where $\alpha = 1$ and $\beta = 1$, when there is no probability distortion applied, as we are using an exponential utility function and know the true form of the optimal weights.

We will use (4.1) to approximate the optimal weights. However, we note that we would obtain the same estimated solution using (4.2), as the two approximations are equivalent given our choice of an exponential utility function. We compare the approximated optimal weights from (4.1) to the numerically computed optimal weights across multiple different parameters of the Prelec weighting function. For the Prelec weighting function parameters, we take $\alpha \in [0.8, 1]$ and $\beta \in [0.9, 1]$. We chose these ranges for the parameters in accordance to [13], where they found the approximate median values for α and β of the Prelec weighting function to be 0.9 and 0.95 respectively, so we take intervals such that these values are the centre of their respective interval. These median values were found to represent an accurate level of observed investor bias in the market. For values of α and β in their respective range, the coefficient $1 - 2c_2$ in (4.1) is positive and less than one, so as we found with (3.1), we expect to see a proportional decrease in the optimal weights using our approximation.

4.2.2 Analysis of Optimal Weights

First, we look directly at the approximated values compared to the numerical results for the optimal weights θ_1 and θ_2 in Figures 4.2 and 4.3.

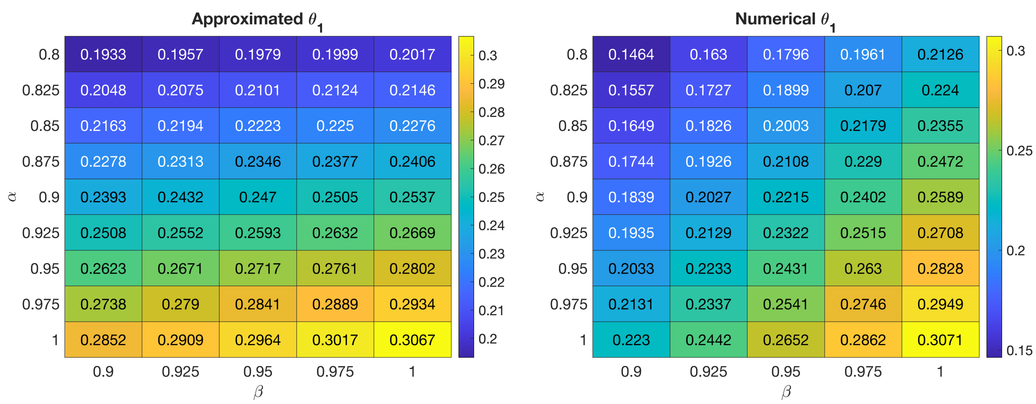

Figure 4.2: Approximated and numerical values for θ_1

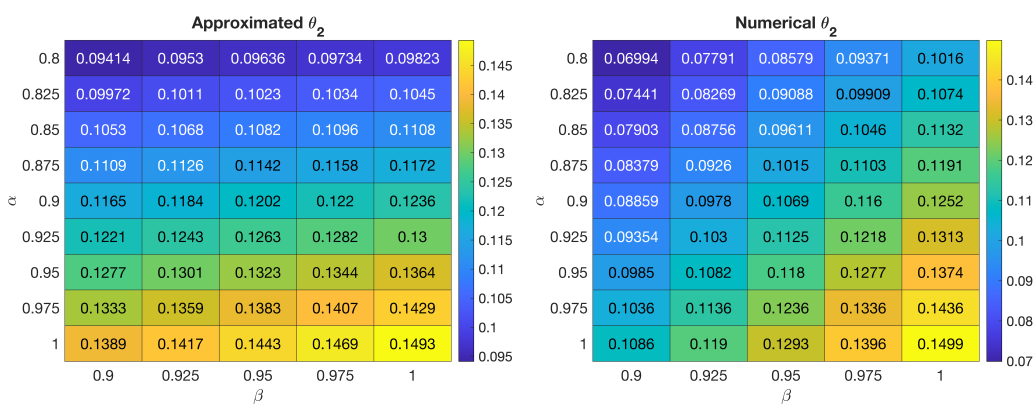

Figure 4.3: Approximated and numerical values for θ_2

Examining these two figures, we see that for high values of α and β (i.e. minor probability distortion) that approximation of the optimal weights is very close to that of the numerical solution for both risky assets. In the case where $\alpha = 1$ and $\beta = 1$, the approximation is more or less equivalent to the numerical solution, and the difference between these values can be attributed to deviations in the simulated distribution of the asset returns and numerical error. For additional comparison, we compute the difference between the approximated and numerical optimal weights as well as the percentage of numerical weight accounted for by the approximation, which are shown in Figures 4.4 and 4.5.

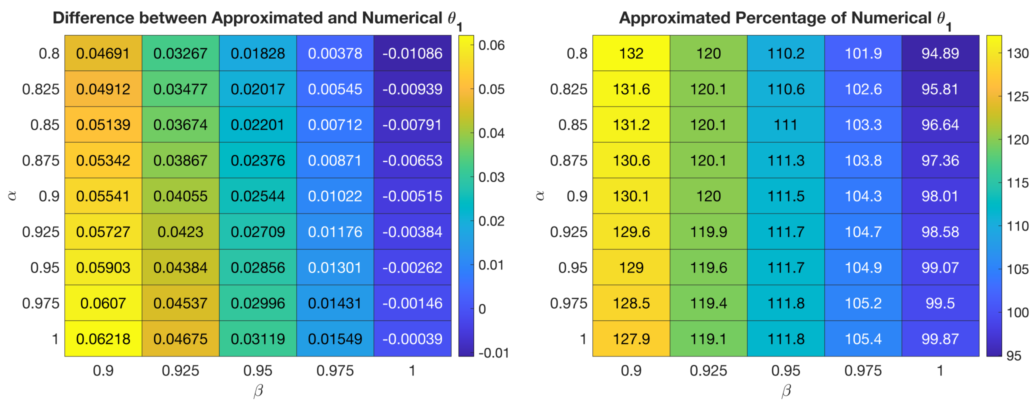

Figure 4.4: Comparison of approximated and numerical θ_1

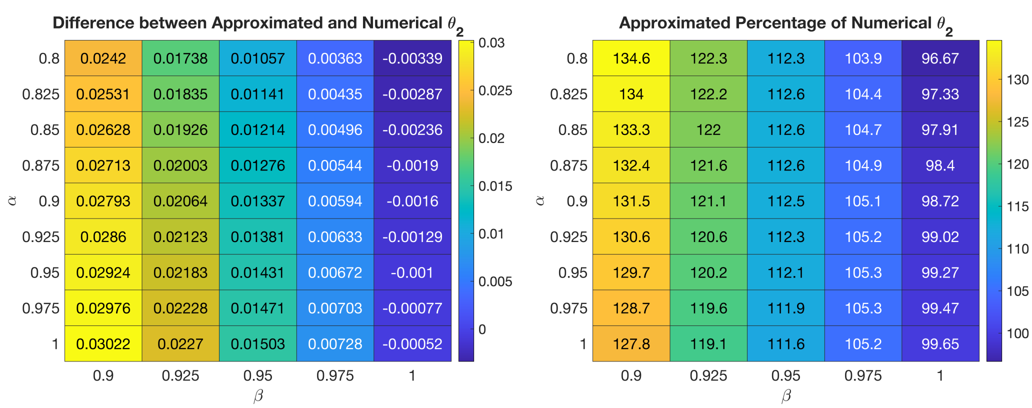

Figure 4.5: Comparison of approximated and numerical θ_2

Analyzing Figures 4.4 and 4.5, we notice a trend when either α or β are held equal to one. First, we will examine the trend when we hold $\alpha = 1$ and vary β. In this case, we overestimate the numerical weight. This is due to the fact that when $\alpha = 1$, the Prelec weighting function takes the form $G(P) = P^{\beta}$, and it subsequently does not have an inflection point. When this occurs, the Prelec weighting function overweights the left tail probabilities and underweights the right tail probabilities, as shown in the left panel of Figure 4.1. For $\alpha = 1$ and $\beta < 1$, the investor is more hesitant to invest into risky assets due to overweighting the left tail probabilities. Our approximation, however, assumes that the inflection point of the weighting function is approximately 1/2, and thus applies overweighting to both the left tail and right tail probabilities. As our approximation assumes the investor is also overweighting the right tail probabilities, we estimate larger optimal weights for the risky assets than if the right tail probabilities were underweighted. As such, for the case where $\alpha = 1$, our approximation overestimates the numerical solution.

In the case where $\beta = 1$ and we vary $\alpha < 1$, the Prelec weighting function has inflection point of 1/e, regardless of the value of the value α. This inflection point is below our approximation's assumed inflection point of 1/2. An inflection point less than 1/2 results in a non-symmetric distortion on the right and left tails, which is shown in the right panel of Figure 4.1. In fact, this causes more weighting being applied on the right tail probabilities. Consequentially, the investor is more willing to invest wealth into a risky asset than in the case of a symmetric weighting function, as they are overweighting the probabilities of realizing a large gain to a greater extent than with a symmetric weighting function. Thus, our approximation underestimates the optimal weights, as shown in Figures 4.4 and 4.5.

Finally, are interested in measuring the proportion of risky wealth allocated to the MF, the active fund, so we compute the ratio $\theta_A^* = \frac{\theta_1^*}{\theta_1^* + \theta_2^*}$. The computation of θ_A^* allows us to analyze the allocation of exclusively the risky wealth in the portfolio. Analyzing this ratio for both the approximated and numerical optimal weights will allow us to determine how the investor is allocating their wealth under different levels of probability distortion. Using either (4.1) or (4.2), we find that θ_A^* remains constant under different levels of distortion. In other words, the approximated value of θ_A^* is only dependent on μ and Σ. Using the market model parameters previously defined, we calculate the approximated value of θ_A^* to be equal to 0.6725. This means that roughly two thirds of the risky wealth is allocated to the MF, with the remaining third allocated to the ETF. We present the numerical values of θ_A^* in Figure 4.6. We see that the numerical value of θ_A^* remains approximately equal to two

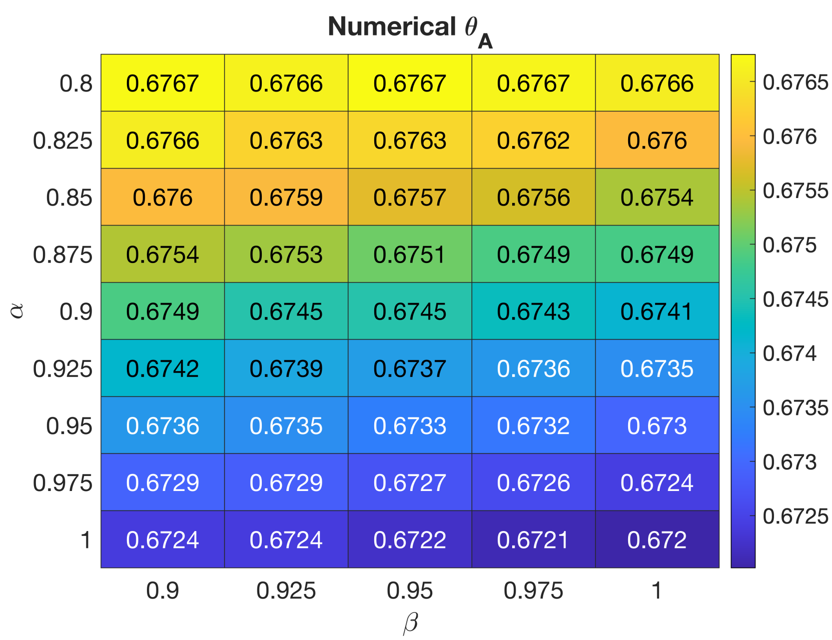

Figure 4.6: Numerical values for θ_A^*

thirds. Even in a numerical setting, the value of θ_A^* does not undergo significant deviation across different weighting function parameters. To determine the accuracy of the approximated θ_A^* against its numerical counterpart, we compute the percentage of the approximation with respect to the numerical value, which is displayed in Figure 4.7. We find that the approximation of the

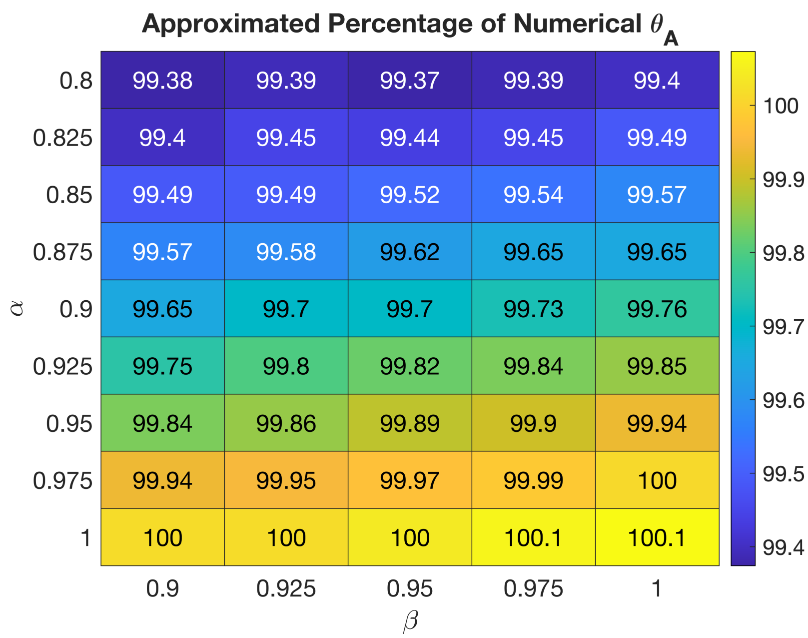

Figure 4.7: Percentage of numerical θ_A^* accounted for by approximation

active proportion of risky wealth is very accurate with respect to the numerical result, as for all tested levels of probability distortion we have less than a 0.7% error. Comparing with our previous numerical results, we find that the approximation of the risky wealth allocated to the active fund is more accurate than the approximation of the risky weights themselves.

4.2.3 Analysis of a Change in Expected Asset Returns

We now apply a shift in the expected return to certain assets. We follow in a similar fashion to that described in Section 3.3, where we defined a new return vector $\tilde{\mu}$ to represent the shifted expected returns of the assets. In addition to the distribution with parameters (μ, Σ), we now have a second distribution with model parameters given by $(\tilde{\mu}, \Sigma)$ and corresponding underlying random variable $\tilde{Y}_T$, where $\tilde{\mu}$ represents the shifted return vector. Now, we have a second wealth equation which is given by $\tilde{X}_T = \tilde{\theta}^\top \tilde{Y}_T$. As was the case in Section 3.3, Σ remains unchanged between the two distributions, since the variance structure of our risky assets is not affected by a shift in the expected returns. We wish to know how the optimal strategy changes between the return vectors μ and $\tilde{\mu}$, so we calculate $\tilde{\theta}^* - \theta^*$, where $\tilde{\theta}^*$ is the optimal strategy associated with the model parameters $(\tilde{\mu}, \Sigma)$. To determine what the change in strategy will be, we simply subtract the two optimal weight vectors, which can be approximated using (4.1) when using an exponential utility function, to obtain

$$\tilde{\theta}^* - \theta^* \approx \frac{1 - 2c_2}{\gamma} \Sigma^{-1} (\tilde{\mu} - \mu). \tag{4.3}$$

As was the case in Section 3.3, we see that the change in optimal portfolio weights is directly related to the magnitude of the shift in the expected return vector. For a market with two risky assets, the analysis of (4.3) is equivalent to the analysis performed in Section 3.3. Additionally, as we have $1 - 2c_2 < 1$ for the Prelec weighting function with α, $\beta < 1$, the shift in the optimal weights will decrease as the probability distortion is increased. The optimal weight decrease is proportional across all risky assets.

To verify the approximation for the change in optimal strategy, we use the same experimental set up that was described in Section 4.2.1, but now include a second simulated distribution with parameters $(\tilde{\mu}, \Sigma)$. For our new return vector $\tilde{\mu}$, we decrease the expected return of the second risky asset, the ETF, by 0.25%, so we set $\tilde{\mu} = [0.08, 0.0825]$. As was the case with our original simulated distribution, we again will have deviations in the statistics of the shifted simulated distribution, leading to an error between the approximated and numerical solution. This error will again be clear when analyzing the following figures for the case where $\alpha = \beta = 1$. We compute the numerical change in the optimal weights and compare it with the approximation given in (4.3). First, we analyze the approximated and numerical shift values for each component in Figures 4.8 and 4.9. For both the approximated and numerical

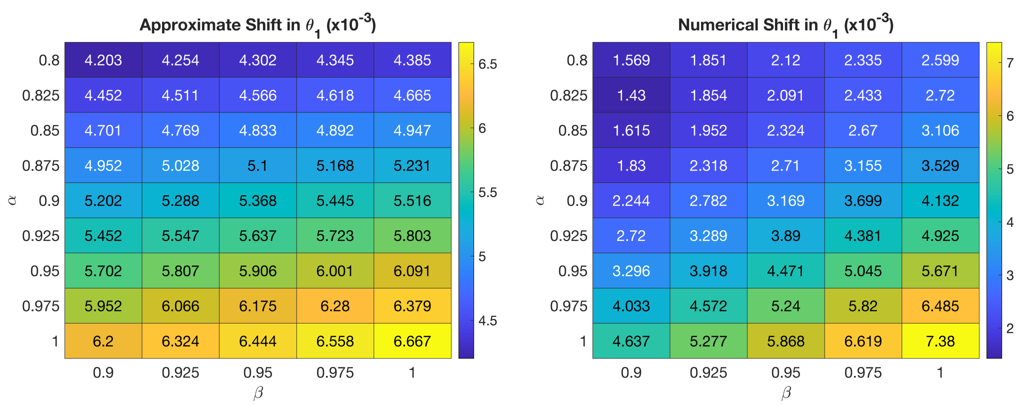

Figure 4.8: Approximated and numerical values for $\tilde{\theta}_1^* - \theta_1^*$

shifts, we see an increase in the optimal weight of the first risky asset, which did not undergo any change in expected return. Correspondingly, the optimal weight of the second risky asset decreases when its expected return decreases. The reasoning behind these findings is parallel to the analysis described in

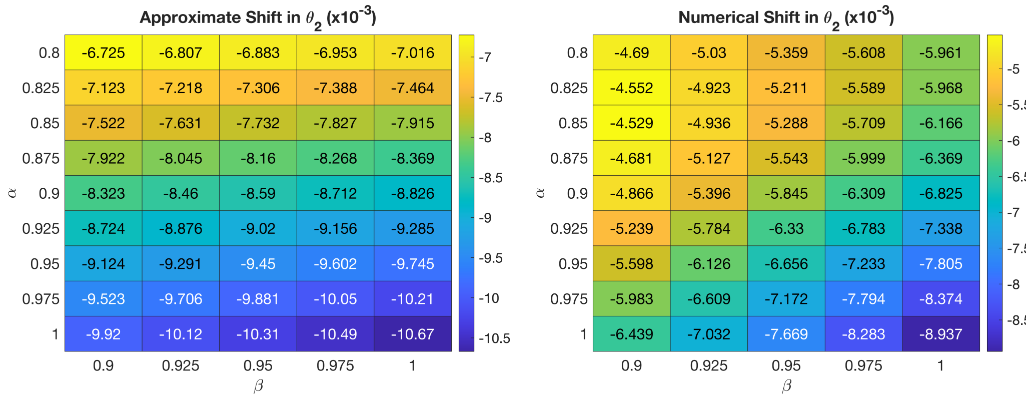

Figure 4.9: Approximated and numerical values for $\tilde{\theta}_2^* - \theta_2^*$

Section 3.3. For both risky assets, we see that the absolute change decreases as more distortion is applied for both α and β. For no distortion, that is when $\alpha = \beta = 1$, the largest shift in the optimal weights occur. This is clear from (4.3), where the coefficient $1 - 2c_2$ decreases the size of the shift as more distortion is applied. So we find that in the event an ETF's expected return is decreased, the proportion of an investor's portfolio allocated to the ETF will decrease, but to a lesser extent than if there were no distortion present.

To determine the accuracy of (4.3) against the numerical solution, we calculate the difference between the approximated shift in optimal strategy and the numerical shift optimal strategy, as well as the percentage of the numerical shift accounted for by the approximation. We display these comparisons in Figures 4.10 and 4.11. For the shift in both risky assets, we see that for almost all values of α and β, we are overestimating the numerically computed shift in the optimal weights. This result is expected, given the findings of Section 4.2.2 where the approximation typically overestimated the optimal weight, and is now amplified as we are approximating two sets of optimal weights. Despite

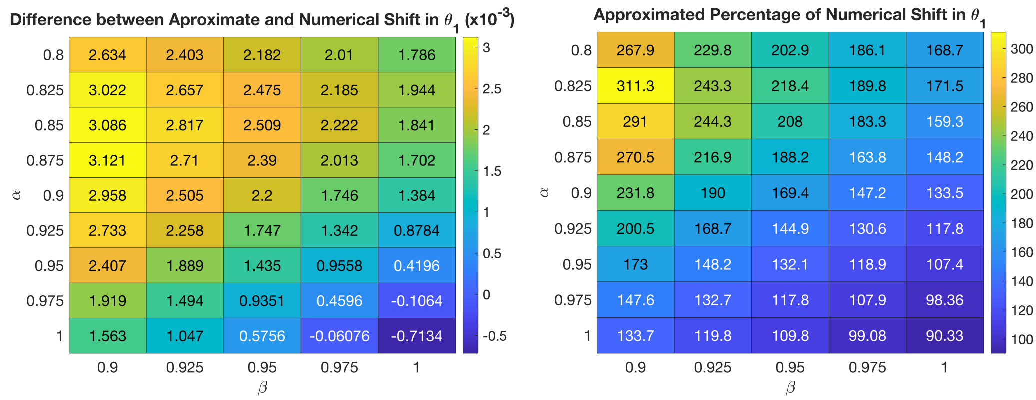

Figure 4.10: Comparison of approximated and numerical change in strategy $\tilde{\theta}_1^* - \theta_1^*$

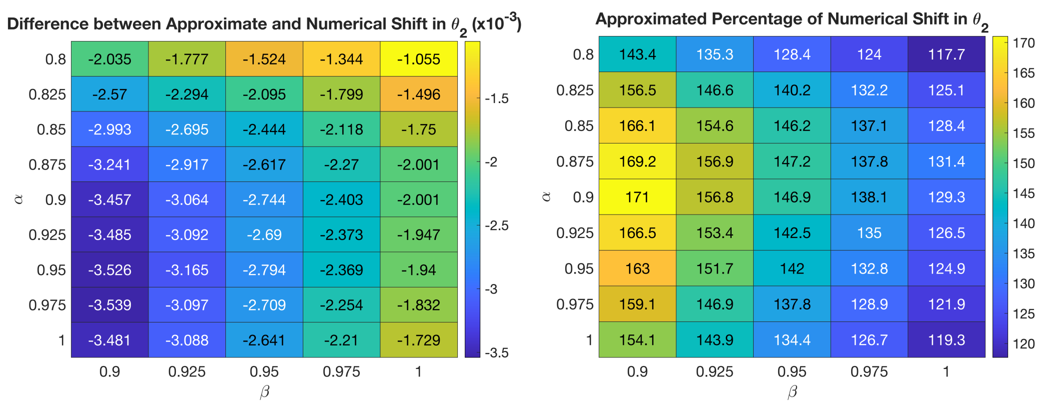

Figure 4.11: Comparison of approximated and numerical change in strategy $\tilde{\theta}_2^* - \theta_2^*$

losing some accuracy in our estimation, going back to Figures 4.8 and 4.9 we see that the approximation of the change in optimal weights encompasses the trend of how the numerically solved optimal strategy will change under various distortion parameters. We can further make the argument that if given a real world scenario where an ETF's return is decreased, the shift in the optimal strategies will be smaller than what is approximated by (4.3).

Finally, we compare how the active proportion of wealth between the two scenarios. We calculate $\tilde{\theta}_A^* = \frac{\tilde{\theta}_1^*}{\tilde{\theta}_1^* + \tilde{\theta}_2^*}$ and $\theta_A^* = \frac{\theta_1^*}{\theta_1^* + \theta_2^*}$ and analyze $\tilde{\theta}_A^* - \theta_A^*$ for numerically computed weights in Figure 4.12. As was the case in Section 4.2.2, the value $\tilde{\theta}_A^* - \theta_A^*$ for the approximation is constant with respect to the distortion parameter, as is evident with (4.3). The approximated value of $\tilde{\theta}_A^* - \theta_A^*$ is equal to 0.0207. Given the values from Figure 4.12, this result follows the

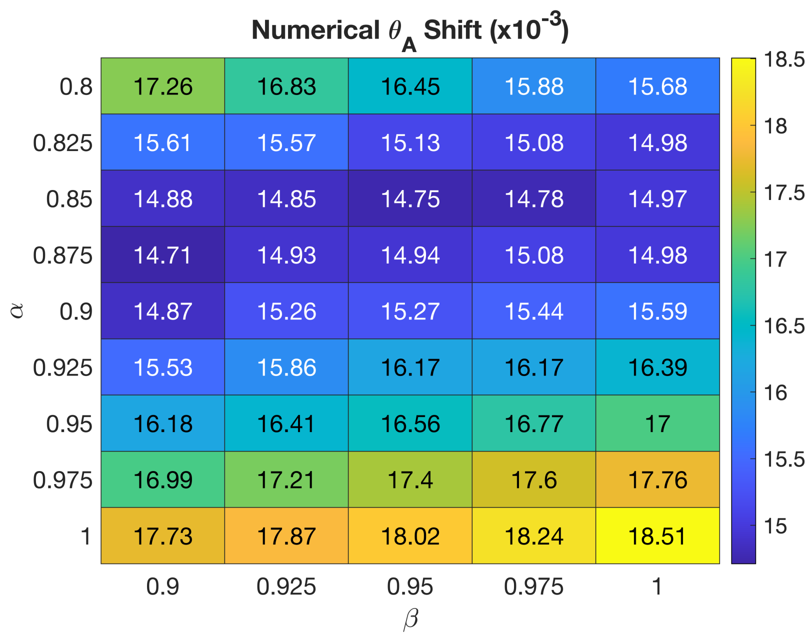

Figure 4.12: Numerical values for the shift in θ_A

theme of overestimating the numerical shift in active allocation that we have seen so far in our analysis. Additionally, viewing the numerical active weight allocation shift, we see this shift decrease as we increase both the α and β parameters since this reduces the amount of distortion in the model. We now present the percentage that our approximation accounts for of the numerical active allocation shift in Figure 4.13. The values from this figure clearly shows that we are over approximating the active allocation shift. It is noteworthy to point out that the approximation of the active allocation is much more accurate than the approximation of the change in weights themselves, an observation

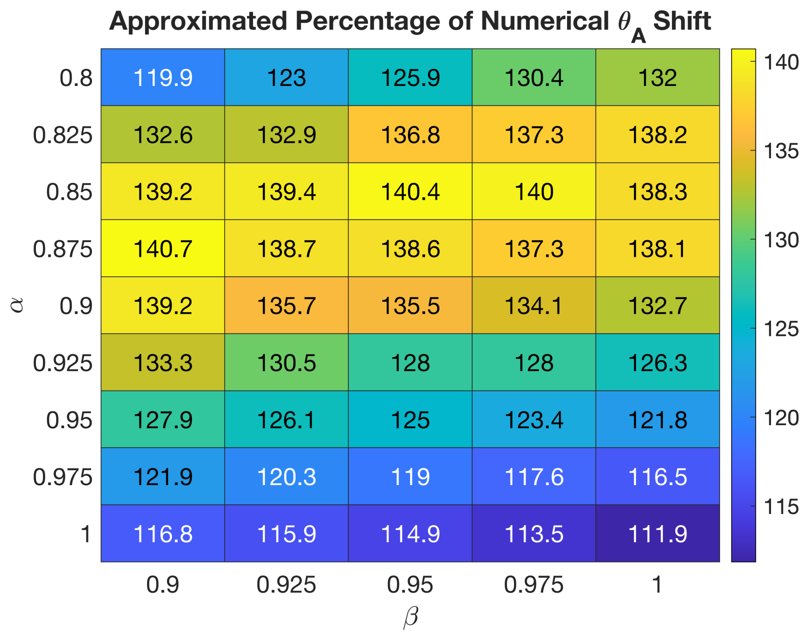

Figure 4.13: Percentage of numerical active proportion shift accounted for by approximation

that follows over from Section 4.2.2. As such, we have further evidence that in the distorted probability model, our approximation is overestimating the shift in the active allocation of the risky wealth.

4.3 Impact of Closing the ETF Tax Loophole on Investor Portfolios

In Section 2.2, we discussed how the ETF tax loophole is used to avoid having ETF investors pay capital gains tax on gains realized by the ETF selling assets. This allows the ETF to report a higher return. The tax is instead paid by withdrawing investors, as these investors will have realized a larger return due to the ETF's use of the tax loophole. In turn, they must pay more in capital gains tax due to their higher return. In this section, we formulate an expression to determine by how much an ETF's return is inflated due to the tax loophole. Using this expression, we then approximate how an investor's optimal portfolio containing only two risky assets, an MF and an ETF, will change in the event the loophole is closed.

Our analysis and its results will depend on a variety of assumptions on the fund returns, tax rates, the investor's preferences, and their holding periods. By nature, these are modelling assumptions that may not necessarily be satisfied in reality, which in turn would affect the results. Nonetheless, our analysis will allow for a quantitative estimation and will demonstrate the significance of the ETF tax loophole on investor portfolios.

First, we must determine the amount by which an ETF investor benefits when the loophole is used. To do so, we investigate a withdrawing investor's return under two scenarios: one where the loophole is freely used by the ETF and one where the loophole is closed. We assume that the ETF investment has yearly return μ_l when the ETF uses the loophole, and μ_c when the loophole is closed. The return of the ETF in the scenario where the ETF can no longer take advantage of the loophole, μ_c, is determined by applying a capital gains tax-rate to $\mu_l - d$, where d is the dividend rate of the ETF. We must subtract the dividend rate from μ_l as the dividend rate is included in the reported return of the ETF, and capital gains taxes do not apply to the dividends obtained from an investment. We denote this tax rate by r_t, and we find that $\mu_c = (1 - r_t)(\mu_l - d) + d$. Next, we assume that a withdrawing investor has held their investment for a period of T years. So in the event the loophole is used by the ETF, the investor realizes a return of $(1 + \mu_l)^T$, and in the event the loophole is closed the investor realizes a return of $(1+\mu_c)^T$. Note that since $\mu_c < \mu_l$, we have $(1+\mu_c)^T < (1+\mu_l)^T$, so we find that the benefits from the ETF tax loophole are compounding, and increase over the course of the investor's holding period T. In the scenario where the ETF paid no capital gains tax, we assume that the withdrawing investor must pay the tax rate of r_t, as this was the corresponding tax rate that would have been applied to the yearly capital gains distributions of the ETF to the investor. This capital gains tax is given by $((1+\mu_l-d)^T-1)r_t$. The annualized return for the withdrawing investor after capital gains tax has been applied is given by $((1+\mu_l)^T-((1+\mu_l-d)^T-1)r_t)^{\frac{1}{T}}-1$. With an analogous argument, we determine the annualized return after capital gains tax in the scenario where the loophole has been closed to be given by $((1+\mu_c)^T-((1+\mu_c-d)^T-1)r_c)^{\frac{1}{T}}-1$, where r_c represents the tax rate on capital gains in addition to the yearly capital gains tax. In order to compute the amount by which μ_l is inflated over μ_c, we compute the difference of the annualized capital gains tax adjusted returns from the two scenarios. We denote the annualized ETF return inflation by τ_T, and we compute τ_T as

$$\tau_T = ((1+\mu_l)^T-((1+\mu_l-d)^T-1)r_t)^{\frac{1}{T}}-((1+\mu_c)^T-((1+\mu_c-d)^T-1)r_c)^{\frac{1}{T}}. \quad (4.4)$$

Analyzing τ_T, we see that this value is dependent on how long the investor holds their investment prior to withdrawing. The value τ_T represents the amount by which the ETF's return is inflated due to the tax loophole. In order to represent the ETF's return after the loophole closure, we deduct the ETF's return by the value τ_T, as this shifted return will represent an ETF investor's return in the event capital gains taxes are deducted from the fund selling an appreciated asset.

For the expected returns and covariance structure of an ETF and an MF, we annualize the data from Polkovnichenko et al. [12], where they collected historical data from the period 1985 to 2008 for active funds and the market to generate bootstrapped samples of monthly returns. They form samples for two types of active funds, a large value (LV) fund and a large growth (LG) funds, as well as a sample for the market to represent a passive fund. For our purpose, we take the LV fund and LG fund to represent two generic MFs, and the market to represent a generic ETF. We use the annualized statistical moments generated for the net reported returns of the active funds and market. For the active funds, the net reported returns represent the returns after management fees are deducted. The sample parameters for the ETF do not change between the gross returns and net reported returns. Using the net reported returns allows us to better represent the investor's perspective, as an investor is required to pay noticeable management fees on their active fund investments. We use the provided funds' beta coefficient to the market to calculate the covariance and correlation of the active funds with respect to the market. As only the rounded monthly values were reported by Polkovnichenko et al. [12], we acknowledge a rounding error in our annualization. We summarize the relevant statistical moments in Table 4.2. We see that the annualized returns of all funds are equivalent, with varying standard deviations. The MFs are highly correlated to the ETF, so we expect that the substitution effect, briefly discussed in Section 3.3, will be very noticeable in this analysis. In other words, if we lower the expected return of the ETF, then investment in the ETF will decrease

Fund	μ	σ	Covariance with ETF	Correlation to ETF
LV	0.10	0.1455	0.020	0.882
LG	0.10	0.1836	0.026	0.908
ETF	0.10	0.1559	-	-

Table 4.2: Annualized sample parameters of MFs and ETF

and a greater investment into the MF will occur due to the highly correlated return structure of the two funds. Recall that in the investor's portfolio, the two risky assets that are held are a mutual fund, represented by either the LV fund or the LG fund, and an ETF.

In 2018, it is assumed that investors in 400 US equity ETFs avoided paying tax on more than 211 billion USD in gains due to the tax loophole, which is approximately 23 billion USD in taxes [10]. That \$23 billion, which was reinvested into the ETFs, equates to an approximate tax rate of 10.9%. We use this value as the capital gains tax rate r_t that would be applied to an ETF in the event the loophole is closed. Using the value r_t, (4.3), and (4.4), we will approximate how an investor's optimal portfolio will shift in the event the ETF loophole is closed. We adjust the expected return of the ETF from Table 4.2 in accordance to (4.4) for holding periods of 3, 5, 10, and 15 years, with $r_t = 10.9\%$. We assume the ETF has an annual dividend rate of 2.5%. In the event the loophole is closed, we assume that a withdrawing investor does not have to pay any capital gains taxes in addition to the yearly capital gains taxes, so we set $r_c = 0\%$. We calculate the value τ_T using (4.4) for various holding periods. These values are shown in Table 4.3. With this choice of r_c and a holding period of 1 year, we find that $\tau_1 = 0$, and the investor has no benefits from the tax loophole. Analyzing Table 4.3, we find that the tax

Holding Period (Years)	3	5	10	15
τ_T	0.00087	0.00162	0.003189	0.00436

Table 4.3: Amount of ETF return increase due to tax loophole for various holding periods

loophole benefits increase with the investor's holding period. This is due to the fact that the deferred tax amount, resulting from the tax loophole, is kept within the fund. For long term holding periods, the return on this amount compounds annually, providing a benefit to a holding investor. As previously mentioned, a withdrawing investor will have to pay more in capital gains tax at the end of the holding period, and this amount increases for longer holding periods.

To continue with our goal of estimating how an investor's portfolio will change in the event of the ETF loophole closure, we require a few more pieces to complete (4.3). We calculate the value $1 - 2c_2$ using the Prelec weighting function and the approximate median values of $\alpha = 0.9$ and $\beta = 0.95$ [13]. Polkovnichenko et al. [12] has shown that the Prelec weighting function can accurately portray the observed behaviour for preference of active funds over passive funds. Using the median values of α and β, we obtain $1 - 2c_2 = 0.805$. The choice of risk aversion parameter, γ, is dependent on the level of risk aversion the investor exhibits. Ang [1], after compiling multiple risk aversion studies and surveys, found the risk aversion parameter to be generally between one and ten. This relatively large range was due to multiple underlying factors of the participants in these studies, such as gender and socioeconomic status. We set our risk aversion parameter γ equal to five, the approximate mid-range value of this interval. This choice of γ will also appropriately represent a risk averse investor. Using the fund distribution parameters from Table 4.2 and the values τ_T from Table 4.3 to shift the mean of the ETF, we calculate the change in optimal strategy for a portfolio containing an MF, represented by the LV fund or LG fund, and an ETF using (4.3). We display the results in Table 4.4 and Table 4.5 for the portfolio containing the LV fund and LG fund, respectively. In addition to computing how the optimal weights will change, we also compute the change in the optimal active allocation of the risky wealth, given by $\tilde{\theta}^*_A - \theta^*_A$.

Holding Period (Years)	3	5	10	15
LV Fund Weight Change	0.0267	0.0505	0.0993	0.1358
ETF Weight Change	-0.0279	-0.0528	-0.1039	-0.1420
Change in Active Allocation	0.0360	0.0683	0.1347	0.1845

Table 4.4: Optimal weight changes for a portfolio containing an LV fund and an ETF

Holding Period (Years)	3	5	10	15
LG Fund Weight Change	0.0296	0.0560	0.1102	0.1506
ETF Weight Change	-0.0378	-0.0716	-0.1408	-0.1924
Change in Active Allocation	0.0387	0.0741	0.1490	0.2072

Table 4.5: Optimal weight changes for a portfolio containing an LG fund and an ETF

As expected, we see from Tables 4.4 and 4.5 that the optimal portfolio weight of the ETF decreases and a greater preference for mutual funds takes hold in the event the loophole is closed. This preference for the mutual fund increases as the holding period increases, as the benefits from the ETF tax loophole that were once compounding are now gone. As we observed with the numerical experiments from Section 4.2.3, these results are likely overestimating how the loophole closure will change an investor's portfolio. We can conclude, however, that in the event the loophole is closed, there will be a greater preference for mutual funds over ETFs for long-term investors. Due to the high amount of correlation between the two risky assets, the increase in the MF's allocation in an investor's portfolio will be similar to the decrease in the ETF's allocation. For investors who hold their portfolio for three to five years, we can expect investment in ETFs to decrease by approximately 2.75% and 7% of the total portfolio allocation and investment in MFs to increase between approximately 2.5% to 5.5% of the total portfolio allocation. Further, for holding periods between three and five years, the proportion of the risky wealth allocated to the MF is estimated to increase by approximately 3.5% and 7.5%. For long holding periods of 10 to 15 years, we find that the impact of the tax loophole closure will produce a significant shift within investor portfolios, with the ETF's allocation will decrease between approximately 10% and 19.5% while the MF's allocation will increase between approximately 10% and 15%. The proportion of risky wealth allocated to the MF will increase between approximately 13% and 21% for holding periods between 10 and 15 years. Our findings that the ETF tax loophole has a significant impact on investor portfolios are consistent with the findings from Moussawi et al. [11]. Recall that in 2020 alone, Rosenbluth [16] found that 290 billion USD was invested into the top three ETF providers, whose holdings are over 4 trillion USD. Given this fact, our estimation indicates that the closure of the ETF tax loophole could cause investors to shift billions of dollars from ETFs to MFs.

While we have explored the impacts that the closure of the ETF tax loophole would have on investor portfolios, it is also noteworthy to briefly discuss the proposed tax plan of US President Biden and its potential ramifications for ETFs and MFs. President Biden's administration is considering to raise the capital gains tax from 20% to 39.6% for investors whose household income is more than 1 million USD per year [4]. In the event the ETF tax loophole remains open and this tax plan becomes legislation, the comparative advantage ETFs have over MFs from using the tax loophole will increase, as MF investors affected by the tax plan would need to pay higher capital gains taxes in the relevant tax year while ETF investors could defer them, benefiting from higher compound returns and flexibility in making taxable capital gains. This would likely result in further inflows of capital into ETFs from MFs and other funds.

Chapter 5

Conclusion

This thesis adds to the increasing discussion surrounding the ETF tax loophole in the United States, by providing an approximate result on the impact that the ETF loophole and its possible closure would have on investor portfolios. We take no stance on whether benefiting from this legal loophole is ethical or not, but are fundamentally interested on the effects it has on portfolios containing MFs and ETFs. To properly represent observed investor behaviour for active and passive funds, we had to use a distorted probability model. While portfolio optimization using this distorted probability model typically does not allow for an explicit solution, we formed an approximation on the probability weighting function's derivative that allowed us to explicitly solve the RDEU portfolio optimization problem. Furthermore, we developed an equation to measure the impact of the ETF tax loophole on an investor's return, with respect to the holding period of the investment. Using our approximation and this equation, we were able to produce an estimate of how the closure of the ETF tax loophole would alter an investor's optimal portfolio allocation. We found that investment in ETFs would decrease and investment in MFs would correspondingly increase by a slightly smaller amount. The magnitude of the change in portfolio allocations crucially depended on the investor's holding period, as the tax benefits gained from the loophole's use compounded annually. For holding periods from three to five years, we found that an investor's portfolio allocation to ETFs would decrease between approximately 2.75% to 7% of the total allocation, while the allocation to MFs would increase between approximately 2.5% to 5% of the total portfolio allocation. As this legal loophole is in the US tax code, its closure would mainly impact the portfolios of American investors. Even so, as many Canadians choose to invest in American securities, and 50% of the total value of capital gains are taxable in Canada, the loophole's closure would also significantly influence the portfolio allocation of Canadian investors.

As this thesis only considered a single-period model, future work can extend our approximation of a general weighting function to a multi-period model and to a continuous-time model, potentially including stochastic volatility, with help from the prior work of Hu et al. [9]. Additionally, future work can include addressing how portfolio allocation will change under multiple closure scenarios of the ETF tax loophole, including a partial closure of the loophole, a restriction of "heartbeat" trades, or a capped amount of capital gains tax that the ETF can avoid yearly. The work and analysis performed in this thesis highlight the importance to consider effects on investor portfolios when determining the tax treatment of MFs and ETFs.

Bibliography

[1] Andrew Ang. *Asset Management: A Systematic Approach to Factor Investing*. Oxford University Press, 2014.

[2] Colin F. Camerer and Teck-Hua Ho. Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, 8(2):167–196, 1994.

[3] Jeffrey M. Colon. The great ETF tax swindle: The taxation of in-kind redemptions. *Penn State Law Review*, 122(1):1–68, 2017.

[4] Laura Davison. How capital gains are taxed and what Biden might do. *Available at https://www.bloomberg.com*, 2021.

[5] Matthias A. Fahrenwaldt and Chaofan Sun. Expected utility approximation and portfolio optimisation. *Insurance: Mathematics and Economics*, 93:301–314, 2020.

[6] Gary L. Gastineau. The benchmark index ETF performance problem. *The Journal of Portfolio Management*, 30(2):96–103, 2004.

[7] Martin J. Gruber. Another puzzle: The growth in actively managed mutual funds. *The Journal of Finance*, 51(3):783–810, 1996.

[8] Steven Z. Hodaszy. Tax-efficient structure or tax shelter: Curbing ETFs' use of section 852 (b)(6) for tax avoidance. *Tax Law*, 70:537, 2016.

[9] Ying Hu, Hanqing Jin, and Xun Yu Zhou. Consistent investment of sophisticated rank-dependent utility agents in continuous time. Forthcoming in *Mathematical Finance:* https://onlinelibrary.wiley.com/doi/abs/10.1111/mafi.12315, 2021.

[10] Zachary R. Mider, Rachel Evans, Carolina Wilson, and Christopher Cannon. The ETF tax dodge is Wall Street's 'Dirty Little Secret'. *Bloomberg Businessweek*, 2019.

[11] Rabih Moussawi, Ke Shen, and Raisa Velthuis. ETF heartbeat trades, tax efficiencies, and clienteles: The role of taxes in the flow migration from active mutual funds to ETFs. *SSRN preprint https://ssrn.com/abstract=3744519*, 2021.

[12] Valery Polkovnichenko, Kelsey D. Wei, and Feng Zhao. Cautious risk takers: Investor preferences and demand for active management. *The Journal of Finance*, 74(2):1025–1075, 2019.

[13] Valery Polkovnichenko and Feng Zhao. Probability weighting functions implied in options prices. *Journal of Financial Economics*, 107(3):580–609, 2013.

[14] Drazen Prelec. The probability weighting function. *Econometrica*, 66(3):497–527, 1998.

[15] John Quiggin. *Generalized Expected Utility Theory: The Rank-Dependent Model*. Springer Science & Business Media, 1993.

[16] Todd Rosenbluth. ETF tax benefits continued in 2020. *Available at https://www.etf.com*, 2020.

[17] Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

[18] George Wu and Richard Gonzalez. Curvature of the probability weighting function. *Management Science*, 42(12):1676–1690, 1996.