UNIVERSITY OF COPENHAGEN
FACULTY OF SOCIAL SCIENCES

Master's Thesis (F)

Dynamic trading strategies

With predictable (co)variances and quadratic transaction costs

Casper Nordblom Eneqvist (xfj248) & Niels Eriksen (dvp957)

Supervisor: Stefan Voigt

19th of December 2021

Keystrokes: Roughly 129500, normal pages: 54
Contract number: 129144

Abstract

This thesis considers dynamic multi-period minimum variance portfolio (MVP) optimization with an investor using DCC MGARCH models to forecast the conditional covariance of asset returns. We test three variance models and model the transaction costs as quadratic. We consider two strategies. In the first strategy, the investor seeks a minimum variance portfolio and ignores transaction costs associated with portfolio rebalancing (simple strategy). In the second strategy, the investor also seeks a minimum variance portfolio, but now the investor adjusts for the expected transaction costs (sophisticated strategy). We derive optimal dynamic solutions for both strategies with a closed-form solution for the simple strategy. The sophisticated strategy requires numerically solving for coefficients and additional restrictions. As new information arrives, the simple strategy immediately rebalances to the new MVP portfolio. In contrast, the sophisticated strategy partially trades towards an aim portfolio depending on the level of transaction costs. In gross returns, both strategies achieve a lower annualized standard deviation than an Equal-weight strategy and a Buy-and-hold strategy. However, both strategies only beat the Equal-weight strategy in terms of Sharpe Ratio. The simple strategy performs poorly in net returns as high transaction costs increase the portfolio's standard deviation drastically because of decreasing returns. After tuning the ex-ante transaction costs, the sophisticated strategy performs well, achieving lower annualized standard deviation than our benchmark strategies, even after subtracting transaction costs.

Division of Labor

Casper Nordblom Eneqvist (xfj248):
Section 2.1, 3, 4.1, 4.2.1, 5, 7.1, 8.3
Niels Eriksen (dvp957):
Section 2.2, 4.2.2, 6, 7.2, 8.1, 8.2, 9
Mutual Work:
Abstract, Section 1, Section 10

Contents

List of Figures

List of Tables

List of Algorithms

1 Introduction

The optimal allocation of funds, whether for retail investors, banks, or institutional clients, is a vast research topic. Given the compounding nature of returns, the difference in wealth achievable over a long horizon with an optimal portfolio compared to a sub-optimal portfolio is immense. Thus, it is theoretically possible to substantially improve the future living standards of savers or profits for firms by investing optimally. The insurance and pension sector is an excellent example of a place where better investment strategies can yield substantial welfare gains for millions of people.

Modern portfolio theory advanced the pursuit of optimal portfolios, which the Nobel prize-winning Harry Markowitz introduced in his seminal paper [Markowitz, 1952]. In this paper, Markowitz introduced the now widely applied mean-variance approach. Markowitz considers a risk-averse investor who maximizes expected returns and minimizes portfolio risk. While the theoretical work of [Markowitz, 1952] is the foundation of countless finance papers, his empirical implementation of the model is limited. Specifically, the approach is limited in accuracy by estimation uncertainty in the statistical moments of asset returns, mainly mean returns and covariances, a point we will expand upon later.

Another approach to the investor's problem was introduced by [Merton, 1969] where he considers multi-period continuous time models of varying time horizons for portfolio selection of two assets - famously computing the Merton fraction. This constant optimally determines what fraction of savings an investor should allocate to the single risky asset in the Merton model over an investor's lifetime. Surprisingly, the fraction is independent of time and wealth, which reduces the multi-period problem to a static solution. The static result to the multi-period problem is a consequence of Merton's simple setup where returns follow a Wiener process with a constant mean and variance. If either the mean or the variance were to be dynamic, this would makes the optimal wealth allocation dynamic as well.

While Markowitz's and Merton's approaches are theoretically appealing, three major problems occur when (naively) applying either model empirically. First, the future mean return and variance is time-varying and predicting them is the subject of much research, which implies a second problem. Namely, investors should change the allocation of wealth across time given new predictions of the future mean and variance. Dynamically changing the portfolio gives rise to a third problem; investors incur transaction costs to make portfolio changes, which means a target portfolio can take many days or weeks to achieve, rendering the static result unattainable.

[Gârleanu and Pedersen, 2013] mostly address the three issues from above, except the constant covariance matrix, by solving the multi-period problem for an investor with mean-variance preferences, facing quadratic transaction costs. They derive an analytical solution for returns given by a factor type model and a constant covariance matrix. The authors find that the optimal portfolio is a combination of the previous period's portfolio and an aim portfolio (or a target portfolio), depending on the level of transaction costs. It is analogous to how a missile homes onto its target by aiming for where it will be rather than where it currently is. Empirically, the authors find their dynamic approach is superior, even after subtracting transaction costs, emphasizing the importance of adjusting for transaction costs when deriving an optimal trading strategy for institutional investors.

[Gârleanu and Pedersen, 2013]'s focus on factor models for the mean of returns leaves an opportunity for us to focus on the conditional volatility and correlations of returns rather than the conditional mean. Literature like [Welch and Goyal, 2008] and [Jobson and Korkie, 1980] finds that forecasts of mean returns tent to be very noisy. [Jagannathan and Ma, 2003] writes:

> *"Tangency portfolios, whether constrained or not, do not perform as well as the global minimum variance portfolios in terms of the out-of-sample Sharpe ratio. This means that the estimates of the mean returns are so noisy that simply imposing the portfolio weight constraint is not enough..."*

Therefore, we consider a multi-period investment strategy similar to [Gârleanu and Pedersen, 2013] but focused on optimal minimum variance portfolios, which corresponds to investors with mean-variance preferences and extreme risk aversion. We model returns with a constant mean process and use a DCC MGARCH model for the conditional covariance matrix proposed by [Engle, 2002]. Engle decomposes the covariance matrix Ω_t into variance and correlation and models these separately with N univariate GARCH type models for the conditional variances and a single scalar BEKK model for the conditional correlation. This enables a two-stage estimation of the DCC MGARCH model, which significantly reduces the number of parameters to be estimated and allows us to model a larger number of assets.

The idea of using MGARCH models to model the conditional covariance matrix in minimum variance portfolios is not new as [Engle and Sheppard, 2001] consider a DCC MGARCH model as the data generating process of the conditional covariance matrix and the time-varying minimum variance weights resulting from using the conditional covariance matrix.

Thus, we want to find the optimal dynamic portfolio trading strategy for an investor with a long horizon where the covariance matrix is modeled as a DCC MGARCH process. A time-varying covariance matrix has been used in dynamic investing strategies before, like [Collin-Dufresne et al., 2015], who consider efficient portfolios with returns given by factors models. In their model, the residual covariance matrix, Σ, is constant, but the covariance matrix of the returns, $\Sigma_{t\rightarrow t+1}$, is time-varying by scaling it with time-varying factor exposures, $\beta_{i,t}$. The authors implement a time-varying covariance matrix into the result of [Gârleanu and Pedersen, 2013] and evaluate each period's Riccati equations. Although this adds extra computational demands, it significantly improves the outcome in terms of average terminal wealth and average utility across different levels of transaction costs.

Much literature estimates transaction costs for example [Robert et al., 2012], who find that trading 1.59% of the daily volume of an asset costs an average of 10 basis points on NYSE and NASDAQ. Additionally, the literature uses a wide range of transaction cost forms, from unrealistic yet straightforward forms like the rarely used constant transaction costs to extensively used forms like proportional transaction costs. An example is [Constantinides, 1986], who considers a two-asset case and shows that proportional transaction costs drastically decrease trading volume. [Mei et al., 2016] considers a problem similar to [Gârleanu and Pedersen, 2013] but with proportional and quadratic transaction costs. The difference between the two forms is that quadratic transaction costs exponentially punish portfolio changes compared to proportional transaction costs. We consider an investor facing quadratic transaction costs, which is generally more consistent with modeling large institutional investors.

To summarize, the research question of this thesis is *"What is the optimal minimum variance dynamic strategy when investors face quadratic transaction costs with returns modeled by a constant mean process and a DCC MGARCH model for the conditional covariance?"*.

This implies modeling a tiny subset of the efficient frontier with an infinitely risk-averse investor. Our reasoning behind this decision is the high noisiness of mean return estimates, [Jobson and Korkie, 1980] and [Jagannathan and Ma, 2003], and the inaccuracies of forecasts using factor models, [Welch and Goyal, 2008]. In contrast, it is somewhat possible to forecast conditional variances and correlations to find minimum variance portfolios.

The structure of this paper is as follows. Section 2 introduces Modern Portfolio Theory along with efficient and minimum variance portfolios (MVP) and arguments for focusing on the latter. Section 3 presents stylized facts of financial time series and how the GARCH type model captures these facts. We present both univariate and multivariate GARCH models, along with methods for forecasting the covariance matrix and how to use this forecast to find the optimal MVP portfolio. In section 4, we go through the dynamic programming techniques of [Bellman, 1966]. Then we use dynamic programming to solve the multi-period optimization problem both for an investor who ignores transaction costs and one who adjusts for them. Following a brief description of our sample data in section 5, we present the dynamics of both strategies and perform backtesting of the two strategies in section 7 and 8 to test their performance on historical data. This section also explores the differences between a static Markowitz portfolio using the covariance matrix from an empirical estimate and the covariance matrix from an MGARCH model to see if our strategy. Finally, we round off with a discussion and conclusion in section 9 and 10, respectively.

2 Modern portfolio theory

Modern portfolio theory was introduced in the seminal paper [Markowitz, 1952]. We give an overview of his approach, starting by characterizing the asset universe and the investor we model in section 2.1. Then we introduce Markowitz's mean-variance approach and review the optimal allocation of N assets in section 2.2. Furthermore, we discuss the investor's empirical challenges when using the mean-variance method in practice in section 2.2.1.

2.1 Characterization of the investor and the market

The market in this thesis is characterized by N risky assets with prices given as $P_{i,t}$. The risky assets may be stocks, ETFs, or other financial assets. Additionally, investors pay no taxes.

We model an institutional investor with an initial portfolio value of \$1 bn. Initially, the investor ignores transaction costs, but we later model an investor that adjusts for transaction costs. Transaction costs are brokerage fees, bid-ask spread costs, and price impact costs, but the latter is the most important for institutional investors. We model the investor as risk-averse, which is empirically in line with the vast majority of people being somewhat risk-averse.[1] Finally, the investor only cares about the return and variance of their portfolio. This type of preference is called *mean-variance* utility or *quadratic preferences*, which the following utility function can express:

$$U_t(v_{t-1}, r_t) = \mathbb{E}[v'_{t-1}r_t|\mathcal{F}_{t-1}] - \frac{\gamma}{2}\mathbb{V}[v'_{t-1}r_t|\mathcal{F}_{t-1}] = \mathbb{E}_{t-1}[v'_{t-1}r_t] - \frac{\gamma}{2}\mathbb{V}_{t-1}[v'_{t-1}r_t] \tag{1}$$

$\mathbb{E}_{t-1}[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{t-1}]$ denotes the conditional expectation with respect to the filtration, $\mathcal{F}_{t-1}$, and similar for $\mathbb{V}_{t-1}[\cdot]$ denoting the conditional variance. γ is the investor's level of risk aversion and $r_t = (r_{1,t}, r_{2,t}, ..., r_{N,t})'$ is the return of the financial assets given as $r_{i,t} = P_{i,t}/P_{i,t-1} - 1$. v_t is an $N \times 1$ vector of weights $v_t = (v_{1,t}, v_{2,t}, ...v_{N,t})'$ defining a portfolio of assets. The weights are normalized wrt. the total amount invested and are determined in the previous period. The portfolio has mean return and variance

$$\mathbb{E}_{t-1}[v'_{t-1}r_t] = v'_{t-1}\mathbb{E}_{t-1}[r_t] \qquad \text{and} \qquad \mathbb{V}_{t-1}[v'_{t-1}r_t] = v'_{t-1}\mathbb{V}_{t-1}[r_t]v_{t-1}$$

This utility function captures the trade-off between maximizing the expected returns of the portfolio, $\mathbb{E}_{t-1}[v'_{t-1}r_t]$, and minimizing the variance of the portfolio, $\mathbb{V}_{t-1}[v'_{t-1}r_t]$. For $\gamma = 0$, the investor is risk-neutral which means the only objective of the investor is to maximize the expected returns regardless of the risk it may pose, which shows as $\frac{\gamma}{2}\mathbb{V}_{t-1}[v'_{t-1}r_t]$ disappears from the utility function.

Real-world investors also care about intermediate consumption rather than just intermediate returns and volatility. Many authors have worked on this problem, like [Merton, 1969] who derived rules for intermediate consumption and wealth allocation. To limit scope of our thesis, we only consider an investor who cares about intermediate returns and volatility.

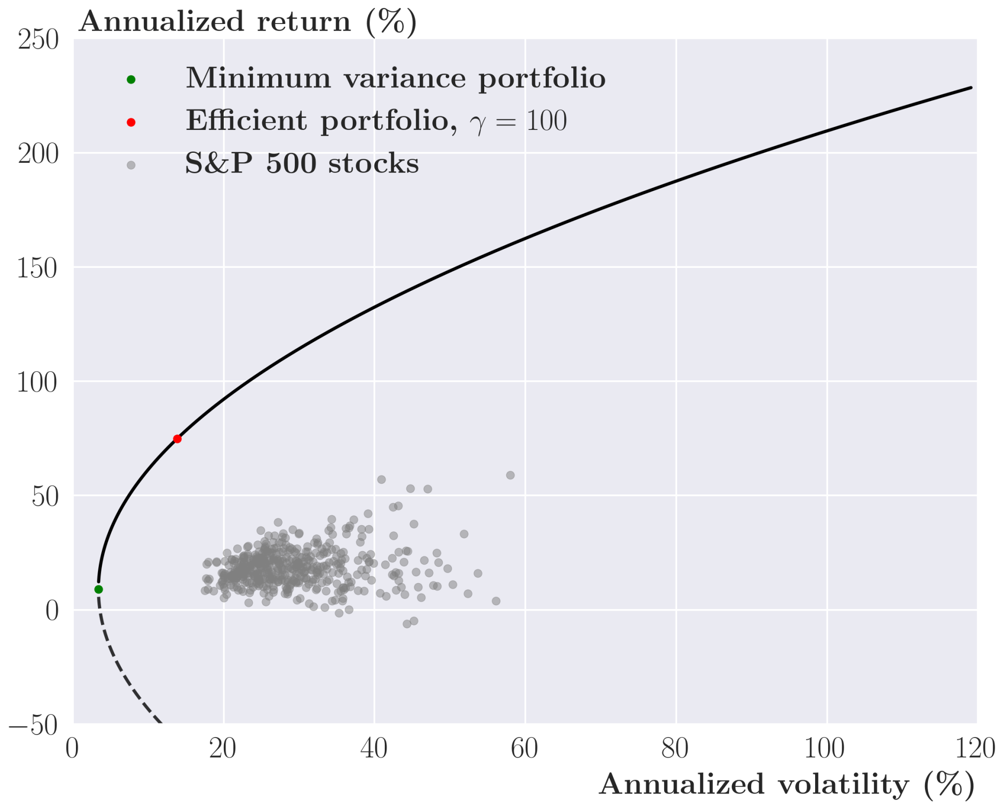
2.2 Mean-variance approach

The mean-variance approach is built upon [Markowitz, 1952]'s seminal paper that lays the foundation for the field of Modern Portfolio theory. Markowitz considers an investor with the objective of obtaining maximum future returns at the lowest possible risk. This corresponds to an investor with preferences as in equation (1). To apply this theory in practice, one must estimate mean asset returns and (co)variances. The simplest method is to use past returns to estimate mean returns and the sample averages for the (co)variances of returns.

[1] [Holt and Laury, 2002] observe that 81% of participants were risk-averse.

After estimating mean asset returns and the covariance of returns, the investor forms portfolios from N assets, and any set of mean and variance obtained by combining these N assets is so-called *feasible*. All feasible combinations are called *the feasible set*. The number of assets determines the size of the feasible set, which increases as the number of assets grows. In figure 1, the feasible set is contained within the black lines. Notice that the feasible set is much larger than any mean-variance combination obtainable from individual stocks. The

Figure 1: Estimated efficient frontier with minimum variance portfolio and efficient portfolio



Source: Yahoo Finance. Data from January 1^{st} *2013 to September* 1^{st} *2021*

investor can thus invest in any portfolio within the feasible set. However, a rational investor will only invest in a portfolio along *the efficient frontier*, which is the set of portfolios that either

- Has the highest expected return for a given variance
- Has the lowest variance for a given expected return

There are many efficient portfolios on the efficient frontier, which in figure 1 is all portfolios on the black line. The investor picks a portfolio from the efficient frontier that maximizes their utility with respect to their level of risk aversion, γ. The more risk-averse the investor is, the more to the left in the figure the investor prefers to be. An infinitely risk-averse investor, who only cares about minimizing risk, will prefer the *minimum variance portfolio* (the green dot) under our assumption of no outside option. In contrast, a moderately risk-averse investor will prefer the *efficient portfolio* (the red dot), a trade-off between expected return and risk.

Markowitz's mean-variance approach requires either Gaussian returns, $\mathcal{N}_N(\mu, \Sigma)$, or quadratic preferences given by equation (1). We use the second assumption that the investor has quadratic preferences even though we use Gaussian returns for the following derivations for simplicity. We begin by considering a standard mean-variance approach for N risky assets to see how this works in practice.

2.2.1 Solving the problem for multiple assets

Consider a market defined by N risky assets, the returns of which are given by a multivariate Gaussian distribution $\mathcal{N}_N(\mu, \Sigma)$. μ is an $N \times 1$ vector of mean returns, and Σ is the $N \times N$ covariance matrix of the N assets. v_t is a $N \times 1$ vector of weights. An investor with mean-variance preferences as given by equation (1) seeking to maximize their expected utility in period $t+1$ solves the following *static* optimization problem:

$$\max_{v_t} \left\{ \mathbb{E}_t[U_{t+1}(v_t, r_{t+1})] \right\} = \max_{v_t} \left\{ \mathbb{E}_t[v_t' r_{t+1}] - \frac{\gamma}{2} \mathbb{V}_t[v_t' r_{t+1}] \right\} =$$

$$\max_{v_t} \left\{ v_t'\mu - \frac{\gamma}{2} v_t' \Sigma v_t \right\} \qquad \text{s.t.} \quad v_t'\mathbf{1} = 1$$

The constraint $v_t'\mathbf{1} = 1$ means that the investor must invest all available funds into the risky assets of the market. The natural interpretation is that the investor has no outside option besides the N assets. Because the individual weights of the weight matrix, v_i, are not restricted to be positive, $v_i \geq 0$, the investor can short sell assets.

To solve the problem, we set up the Lagrangian with constraint $v_t'\mathbf{1} = 1$ with a Lagrangian multiplier λ

$$\mathcal{L}(v_t) = v_t'\mu - \frac{\gamma}{2} v_t' \Sigma v_t - \lambda(v_t'\mathbf{1} - 1)$$

Taking first-order conditions with respect to the weight, v_t

$$\frac{\partial \mathcal{L}}{\partial v_t} = \mu - \gamma \Sigma v_t - \lambda \mathbf{1} = 0$$

Solving for v_t yields

$$\lambda \mathbf{1} = \gamma \Sigma v_t \Leftrightarrow v_t = \frac{1}{\gamma} \Sigma^{-1}(\mu - \lambda \mathbf{1})$$

The constraint requires that $v_t'\mathbf{1} = 1 \Leftrightarrow \mathbf{1}' v_t = 1$, which can be used to solve for the Lagrangian multiplier λ

$$1 = \mathbf{1}' \left(\frac{1}{\gamma} \Sigma^{-1}(\mu - \lambda \mathbf{1}) \right) \Leftrightarrow \mathbf{1}'\Sigma^{-1}\mathbf{1}\lambda = \mathbf{1}'\Sigma^{-1}\mu - \gamma \Leftrightarrow \lambda = \frac{\mathbf{1}'\Sigma^{-1}\mu - \gamma}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$$

We insert the expression for λ into the weights, v_t

$$v_t = \gamma^{-1}\Sigma^{-1} \left(\mu - \left[\frac{\mathbf{1}'\Sigma^{-1}\mu - \gamma}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \right] \mathbf{1} \right) = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} + \frac{1}{\gamma} \left(\Sigma^{-1}\mu - \frac{\mathbf{1}'\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \Sigma^{-1}\mathbf{1} \right)$$

Resulting in

$$v_t^{\text{EFF}} = v_t^{\text{MVP}} + \frac{1}{\gamma}\underbrace{\left(\Sigma^{-1}\mu - \frac{\mathbf{1}'\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\Sigma^{-1}\mathbf{1}\right)}_{(i)} \tag{2}$$

Thus, the efficient portfolio, v_t^{EFF}, consists of the minimum variance portfolio v_t^{MVP} and a self-financing portfolio *(i)*[2].

An investor seeking to maximize their utility invests all available funds into the minimum variance portfolio, v_t^{MVP}, and creates a self-financing portfolio, *(i)*, that increases the expected return of the investor's portfolio but also increases the risk of the portfolio. In terms of figure 1, v_t^{EFF} corresponds to some portfolio on the efficient frontier, with the red dot as an example with $\gamma = 100$. As $\gamma \to \infty$ then $(i) \to 0$ and the investor only invests in the minimum variance portfolio, the green dot.

Intuitively, it makes sense to care about the expected return of the portfolio, μ. However, the investor does not know the true mean of returns, μ. It can be estimated by its empirical counterpart, $\hat{\mu}$. However, neither the true mean, μ, nor the empirical mean, $\hat{\mu}$, can predict the return of an asset tomorrow with any noticeable accuracy. Actually, [Jobson and Korkie, 1980] show that the sample mean from a wide range of stocks are unstable and biased or as [Jagannathan and Ma, 2003] puts it:

> *"The estimation error in the sample mean is so large that nothing much is lost in ignoring the mean altogether when no further information about the population mean is available"*

If the sample mean is used, the instabilities in $\hat{\mu}$ adds instabilities to the chosen weights, v_t. Therefore, some authors prefer to use a factor-like model that often includes factors like momentum, market-to-book ratio, etc., attempting to predict future returns. Famously [Fama and French, 1993], and [Carhart, 1997] presented their three and four-factor model with statistically significant parameters for the factors. However, [Welch and Goyal, 2008] showed that the estimates of many parameters of factor models are unstable and even prone to spurious results.

Another argument against factor models is the efficient market hypothesis, stating that no publicly available information today can predict returns tomorrow. Other market participants also have that information and should have traded the benefits of the information away. Thus, the explanatory power of the factors models should already be reflected in today's prices, thus giving the factor models no predictive power, theoretically.

In summary, no method exists to estimate future returns reliably. Even if the methods are successful, the literature tells us we can only expect to gain marginally helpful information from them. If we tried to estimate mean returns, the estimation uncertainty from these methods introduces two significant problems for the optimal portfolio weight, v_t^{EFF}.

[2]The portfolio is self financing as $\mathbf{1}'\left(\Sigma^{-1}\mu - \frac{\mathbf{1}'\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\Sigma^{-1}\mathbf{1}\right) = \mathbf{1}'\Sigma^{-1}\tilde{\mu} - \mathbf{1}'\Sigma^{-1}\tilde{\mu}\mathbf{1}'\Sigma^{-1}\mathbf{1}(\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1} = 0$.

Firstly, estimation uncertainty adds instability to the optimal weights. This leads to unnecessary rebalancing, which adds transaction costs without benefits since any deviation from the current portfolio is costly.

Secondly, not accounting for estimation uncertainty of the sample mean may cause extreme portfolio weights when naively implementing sample-based mean-variance portfolios. [DeMiguel et al., 2009] has a great example:

> *"Consider the following extreme two-asset example. Suppose that the true per annum mean and volatility of returns for both assets are the same, 8% and 20%, respectively, and that the correlation is 0.99. In this case, because the two assets are identical, the optimal mean-variance weights for the two assets would be 50%. If, on the other hand, the mean return on the first asset is not known and is estimated to be 9% instead of 8%, then the mean-variance model would recommend a weight of 635% in the first asset and -535% in the second. That is, the optimization tries to exploit even the smallest difference in the two assets by taking extreme long and short positions without taking into account that these differences in returns may be the result of estimation error."*

The tendency for extreme portfolio weights is problematic because [Jobson and Korkie, 1980] and [Jagannathan and Ma, 2003] show that the sample average mean is unstable and biased. Thus, the sample mean may be spurious, leading to erroneous portfolio allocations, which in the real market will badly hurt investors' returns. So to limit problems from estimation uncertainty regarding the sample mean, we narrow the scope of the thesis and focus solely on minimizing the volatility of the investor's portfolio.

Specifically, rather than maximizing expected utility, we minimize the variance of the portfolio.

$$\min_{v_t} \left\{ \frac{1}{2} v_t' \Sigma v_t \right\} \qquad \text{s.t.} \quad v_t' \mathbf{1} = 1$$

To solve the problem, we set up the Lagrangian with an identical weight constraint like before $v_t'\mathbf{1} = 1$, with a Lagrangian multiplier λ

$$\mathcal{L}(v_t) = \frac{1}{2} v_t' \Sigma v_t - \lambda (v_t' \mathbf{1} - 1)$$

Taking first-order conditions with respect to the weights, v_t

$$\frac{\partial \mathcal{L}}{\partial v_t} = \Sigma v_t - \lambda \mathbf{1} = 0$$

Solving for v_t yields

$$\lambda \mathbf{1} = \Sigma v_t \Leftrightarrow v_t = \Sigma^{-1} \mathbf{1} \lambda$$

The constraint requires that $v_t'\mathbf{1} = \mathbf{1}' v_t = 1$, which can be used to solve for the Lagrangian multiplier, λ

$$1 = \mathbf{1}' v_t = \mathbf{1}' \Sigma^{-1} \mathbf{1} \lambda \Leftrightarrow \lambda = \frac{1}{\mathbf{1}' \Sigma^{-1} \mathbf{1}}$$

Now, we insert the expression into the weights, v_t

$$v_t = \Sigma^{-1}\mathbf{1}\lambda = \Sigma^{-1}\mathbf{1}\frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \Leftrightarrow$$

$$v_t = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} = v_t^{\text{MVP}} \tag{3}$$

The Minimum Variance Portfolio (MVP) is the portfolio that has the lowest variance of all possible portfolios. To visualize this, consider figure 1, where the MVP (the green dot) has the lowest annualized volatility of all portfolios along the efficient frontier. Comparing the MVP portfolio to the efficient portfolio in figure 1, we see that the efficient portfolio (red dot) has a much higher expected return than the MVP portfolio (green dot) but recall that the sample mean is prone to estimation uncertainty.

We consider minimum variance portfolios not because we are entirely disinterested in the returns of the portfolio but because of the estimation uncertainty associated with the sample mean of returns. Additionally, [Jagannathan and Ma, 2003] have found that minimum variance portfolios have the highest out-of-sample Sharpe ratio. Thus, we have good reasons to believe that this more straightforward approach might be superior to an efficient portfolio in terms of risk-adjusted returns.

A significant drawback to the theoretical derivations above is that the assumed Gaussian returns do not mimic the empirical process of financial time series, which we elaborate on in the next section.

3 GARCH processes

Modeling Gaussian returns is mathematically convenient, but the subsequent sections use a Student's t-distributed returns in a GARCH type volatility model. In section 3.3.3, we show that changing the volatility process does not fundamentally change the results of the mean-variance approach. The following section explains why we are interested in deviating from conventional volatility modeling by exploring stylized facts about financial time-series and how the GARCH type models capture these features.

3.1 Stylized facts about returns

Numerous empirical studies of financial time-series like [Cont, 2001], [Fama, 1965] and [Mandelbrot, 1967] have revealed some stylized facts about returns. When modeling asset returns, it is important that the model of choice mimics these stylized facts. However, there is a fine line between a complete and rigorous model with high estimation uncertainty and a simple but sufficient model with low estimation uncertainty. Consider the following three stylized facts:

1. *The distribution of asset returns is non-Gaussian*

The empirical distribution of returns has higher kurtosis and fatter tails than a Gaussian distribution. Thus, financial returns are more likely to be centered around the mean and more likely to experience extreme events of either sign than returns drawn from a Gaussian distribution.

Distributions with a better fit to the empirical distribution of returns include the Generalized Error Distribution (GED) or the Student's t-distribution, both offering higher kurtosis and fatter tails than the Gaussian distribution. Both are good choices but we use the Student's t-distribution as this is most familiar to us.[3]

2. *There is almost no correlation between returns for different days*

Consider the sample autocorrelation function between period t and $t+\tau$ for $\tau > 0$ with a sample length of T

$$\hat{\rho}_{t,\tau} = \sum_{t=1}^{T-\tau}(r_t - \mathbb{E}[r_t])(r_{t+\tau} - \mathbb{E}[r_t]) \Big/ \sum_{t=1}^{T}(r_t - \mathbb{E}[r_t])^2 \tag{4}$$

which measures the autocorrelation, i.e., the average correlation between the values of a time series at different points in time. For almost all financial time series, the empirical autocorrelation, $\hat{\rho}_{t,\tau}$, is insignificant and thus close to zero.[4] Thus, an autoregressive (AR) model for returns would likely be a poor fit for financial time series, and past returns give little to no information about future returns. Therefore, we use a constant mean model for the mean process as it is simple and with mainly care about modeling the conditional covariance.

3. *There is a positive dependency between squared (or absolute) returns of nearby days*

Stylized fact 2 states that there is no autocorrelation in financial time series. Thus, one might be tempted to think that returns are *identical and independently distributed* or IID. However, this is not correct as many transformations of returns feature strong dependency across time. The most common of which is squared returns, the autocorrelation of which is rarely insignificant.[5] This also implies *volatility clustering*, which means that periods of high or low volatility tend to be clustered together across time or, as [Mandelbrot, 1967] puts it:

> *"...large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes."*

This latter part implies that returns are not homoskedastic, i.e., they do not have a constant variance across time.

To summarize, it would be inaccurate to model returns as IID or Gaussian. Thus, the next part of the thesis presents a model in which returns are neither Gaussian nor independent of each other across time. There are several possible approaches to achieve such a model. One of the most widely used is the Autoregressive Conditional Heteroskedasticity (ARCH) model with a non-Gaussian distribution, which is the first model we introduce.

[3] [Taylor, 2011], *Asset price dynamics, volatility, and prediction*, p. 70-76
[4] [Taylor, 2011], *Asset price dynamics, volatility, and prediction*, p. 76+77
[5] [Taylor, 2011], *Asset price dynamics, volatility, and prediction*, p. 82-86

3.2 Univariate GARCH models

[Engle, 1982] developed the precursor to the GARCH model to model and forecast variances more accurately, which should be very useful to an investor seeking to minimize the variance of a portfolio. The resulting Autoregressive Conditional Heteroskedasticity model, or ARCH for short, models the conditional variance as dependent on past shocks to the modeled time series. Consider a simple model for the mean of a time series with a constant mean

$$r_t = \mu + \epsilon_t \tag{5}$$

We introduce different GARCH models with different variance specifications, but they all share the constant mean specification in equation (5). Consider the simplest possible model given by an *ARCH(1)* where the conditional variance only depends on the last period's shock, ϵ_{t-1}.

$$\epsilon_t = \sigma_t z_t \qquad z_t \sim IID.D(0,1) \tag{6}$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 \qquad \omega > 0, \alpha \geq 0 \tag{7}$$

with initial values taken as given and $t = 0, 1, ...T$. Notice that the variance of this process is heteroskedastic. The parameter restrictions on α and ω are needed to ensure strictly positive variance for all t. $D(0,1)$ is some distribution with mean zero and unit variance. The distribution is often a Gaussian distribution for ease of computation, but other distributions with better fits to financial time series like a Generalized Error Distribution (GED) or a Student's t-distribution can be used. We utilize the latter, of which the probability density function is in the Appendix A.1. We elaborate on this choice in section 5.

ω is a constant that ensures strictly positive variance. The parameter α is the effect of past shocks on the conditional variance and can also be interpreted as short-run persistence of the conditional variance.

The ARCH model has an empirical weakness. The conditional variances from the model converge back to the minimum conditional variance after a large shock much quicker than empirical estimates indicate it should - even for large lag lengths.

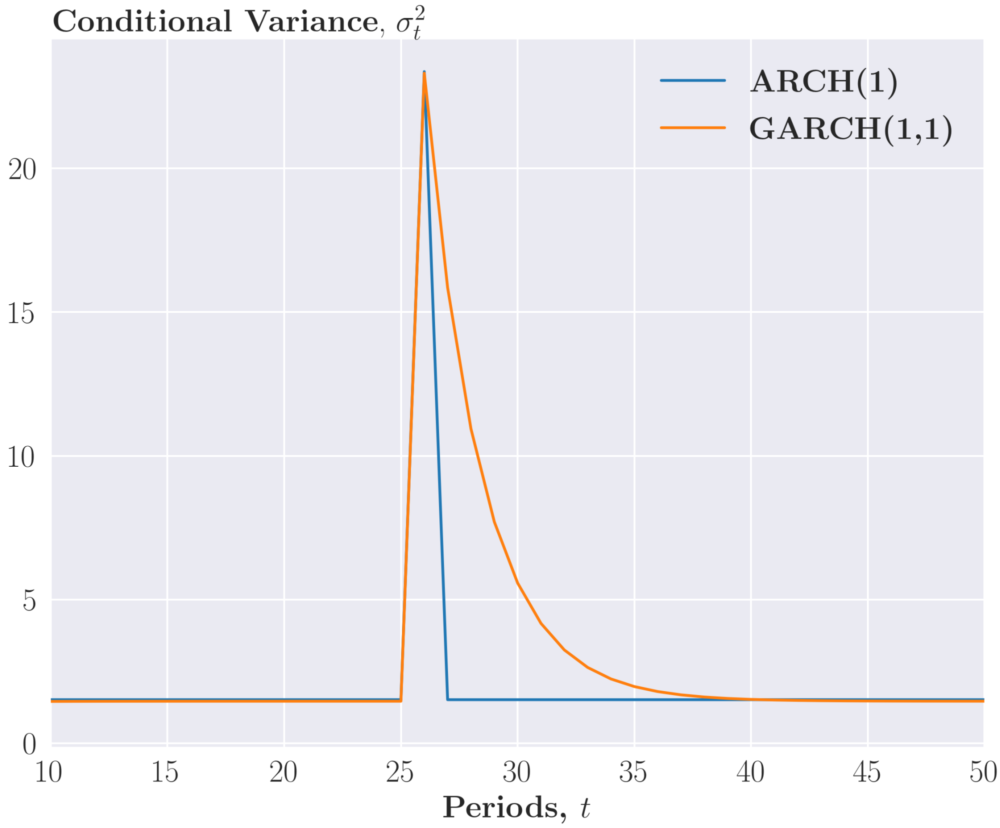
[Bollerslev, 1986] solved this problem with the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, which adds persistence between the individual measures of the conditional variance across time. It does so by adding $\beta\sigma_{t-1}^2$ to the equation for the conditional variance in equation (7), which results in the *GARCH(1,1)*.

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \qquad \omega > 0, \alpha, \beta \geq 0 \tag{8}$$

This reduces the speed with which the conditional variance decreases after a big shock as the parameter β is how persistent the conditional variance is and can be interpreted as the long-run persistence of volatility.

We illustrate the difference in convergence speed between the ARCH(1) and GARCH(1, 1) in figure 2, where we plot the conditional variance over 40 periods. In period $t = 25$, we give

Figure 2: Response of the conditional variance σ_t^2 to a shock



Note: In order to fairly compare the two models, parameters are chosen to give an identical unconditional variance for similar convergence.

the asset returns a large shock of 8%, which causes an increase in the conditional variance. For the ARCH(1) in eq. (7), we see a spike in the conditional variance in period $t = 26$, and an immediate return to the minimum conditional variance at $t = 27$ as the ARCH model has no persistence between measures of conditional variance. For the GARCH(1,1) in eq. (8), we see an identical spike in period $t = 26$, but in contrast to the ARCH(1), it exponentially convergences back to the minimum conditional variance. The exponential convergence is because the GARCH(1,1) has persistence between measures of conditional variance such that the shock affects the conditional variance for a longer time.

We see that adding a lagged conditional variance term to the ARCH model, $\beta\sigma_{t-1}^2$, causes the conditional variance to converge more slowly back to the *minimum conditional variance*, which is given as

$$\sigma_{\text{MIN}}^2 = \frac{\omega}{1-\beta} \tag{9}$$

and the *unconditional variance*, which, given weak stationarity, is given as

$$\sigma^2 = \frac{\omega}{1-\alpha-\beta},^{6} \tag{10}$$

This slower convergence is more in line with how the empirical variance behaves.

[6] [Taylor, 2011], *Asset price dynamics, volatility, and prediction*, p. 202

While the standard GARCH model does an admirable job of explaining the conditional variance of financial time series, econometricians have continuously worked to improve upon the work done by Engle and Bollerslev. [Glosten et al., 1993] presented the *GJR-GARCH(1,1)* model that allows for asymmetric effects of shocks where the conditional variance is given by

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \kappa\epsilon_{t-1}^2 I_{\{\epsilon_{t-1}<0\}} \qquad \omega > 0, \alpha, \beta, \kappa \geq 0 \tag{11}$$

such that the effect on the conditional variance is $\alpha + \kappa$ if the shock is negative, which is well documented to be accurate and named the "leveraged effect." For GARCH models in general, the number of lags is denoted GARCH(p,q), where p is the number of lagged residuals, and q is the number of lagged conditional variances terms.

These models can be estimated via maximum likelihood estimation (MLE). We define the parameters of the model as a vector, $\theta = (\mu, \omega, \alpha, \beta, \kappa, \nu)$, to be estimated by the maximum likelihood estimator (MLE), and the information set, $\mathcal{F}_{t-1} = (r_t, r_{t-1}, ..., r_T)$. Note that ν is the shape parameter for the Student's t-distribution.[7] The likelihood function is the joint probability of the observed data given the parameters of the model. It is more convenient to consider the log-likelihood function, which can be written as

$$L(\theta) = \sum_{t=1}^{T} \log l_t(\theta|\mathcal{F}_{t-1}) = \log \prod_{t=1}^{T} l_t(\theta|\mathcal{F}_{t-1}),$$

where

$$l_t(\theta) = F(r_i|\theta) \qquad i = 0, 1, 2, ..., T,$$

where F is the conditional probability density function for r_t given the parameters of the model, θ[8]. Then the maximum likelihood function with a Student's t density is given as

$$L(\theta) = \frac{\frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{\nu}{2})}}{\sqrt{(\nu-2)\pi\sigma_t^2(\theta)}}\left(1 + \frac{-(r_t-\mu)^2}{(\nu-2)\sigma_t^2(\theta)}\right).$$

with $\sigma_t^2(\theta) = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \kappa\epsilon_{t-1}^2 I_{\{\epsilon_{t-1}<0\}}$ and the gamma function $\Gamma(\cdot)$. We get the parameter estimates that make the data we observe most likely when maximizing the maximum likelihood function. The maximum of the maximum likelihood function cannot be found analytically. Instead, a consistent estimate, $\hat{\theta}$, can be obtained via numerical optimization given some initial guess of θ.

Under specific parameter conditions, the data of the GARCH type model can be shown to be weakly stationary. For GJR-GARCH(1,1), the data is weakly stationary when $\alpha+\beta+0.5\kappa < 1$ with the conditions from the ARCH(1) and GARCH(1,1) contained within as a special case of the GJR-GARCH(1,1)[9]. Suppose the model's data is weakly stationary given the stated conditions. In that case, the law of large numbers applies, which implies that the maximum likelihood estimator is consistent. The estimates, $\hat{\theta}$, then converge to the true value of the parameters, θ_0, i.e. $\hat{\theta} \rightarrow \theta_0$ in probability as $T \rightarrow \infty$.

[7] See appendix A.1 for details on the distribution

[8] [Bohn Nielsen, 2017]: Introduction to likelihood-based estimation and inference, Page 52-53

[9] [Taylor, 2011], *Asset price dynamics, volatility, and prediction*, p. 221

3.3 Multivariate GARCH models

In the univariate case, we consider one asset and the variance of this single asset. But an investor will almost always care about multiple assets and how their variances and covariances develop over time. Luckily, a natural extension of the univariate GARCH model exists into a multivariate GARCH model, MGARCH. Now, consider N assets with an $N \times 1$ vector of returns, r_t, given by a similar constant mean model to the univariate case, but now of vectors

$$r_t = \mu + \epsilon_t \tag{12}$$

with μ being a $N \times 1$ vector of the empirical mean of the individual assets and ϵ_t being an $N \times 1$ vector of error terms.

In the univariate case, ensuring positive variance only required restricting a few parameters. However, for MGARCH models, this is much more complicated. To ensure that the $N \times N$ covariance matrix, Ω_t, is indeed a covariance matrix, it must be positive definite for all t. The challenge is thus to parameterize the model such that Ω_t is positive definite for all t.

BEKK MGARCH

A mathematically simple MGARCH model that solves the parameterization problem is the BEKK GARCH by [Engle and Kroner, 1995] where the conditional covariance matrix resembles the univariate GARCH in form of a BEKK GARCH(1,1)

$$\epsilon_t = \Omega_t^{1/2} z_t \qquad z_t \sim IID.D(0, I_p) \tag{13}$$

$$\Omega_t(\theta) = \Omega + A\epsilon_{t-1}\epsilon_{t-1}' A' + B\Omega_{t-1}B' \tag{14}$$

with initial values taken as given and $t = 1, 2, ..., T$. Ω is positive definite, and A and B are $N \times N$-dimensional matrices. θ is a vector of parameters, where $\{\Omega, A, B\} \in \theta$ and D is some multidimensional distribution. We have chosen to use a multivariate Student's t-distribution for similar reasons as in section 3.2. The probability density function of the multivariate Student's t-distribution is in appendix A.1. Unfortunately, it is next to impossible to get a meaningful interpretation of the individual components of the matrix of parameters for A and B. Still, the overall interpretation of A is entirely analogous to α in the univariate case and the same for B to β.

A nice feature of the BEKK GARCH is that $\Omega_t(\theta)$ is positive definite for all t for any A and B. Thus, the BEKK GARCH easily solves the parameterization problem.[10]

However, the BEKK MGARCH is empirically impractical as the number of parameters explodes as N increases as the number of parameters is $N(N+1)/2 + 2N^2$. For $N = 4$, the number of parameters is 42, and for $N = 10$, it is 255. Thus this model is only practical when

[10][Engle and Kroner, 1995], *Multivariate Simultaneous Generalized Arch*, proposition 2.5

analyzing a small group of assets like $N < 10$. One way around this problem is to simplify the *scalar BEKK(1,1)* so A and B become scalars.

$$\Omega_t(\theta) = \Omega + \alpha\epsilon_{t-1}\epsilon_{t-1}' + \beta\Omega_{t-1} \qquad \alpha, \beta \geq 0 \tag{15}$$

This comes at a great loss to generality as all conditional (co)variances respond similarly to shocks and have similar persistence, which is a debatable assumption. The data of BEKK(1,1) is stationary and ergodic with $\mathbb{E}||X||^2 < \infty$ when

$$\varrho((A \otimes A) + (B \otimes B)) < 1 \tag{16}$$

where $\otimes$ is the tensor product, and $\varrho(\cdot)$ is the spectral radius. If the condition in equation (16) holds, then A and B can be estimated consistently by maximum likelihood. For the scalar-BEKK(1,1), the condition simplifies to $a + b < 1$.[11] This model can be estimated via maximum likelihood estimation where the maximum log-likelihood function has a multivariate Student's t density given as

$$\begin{aligned} L_{\text{BEKK}}(\theta) = \log\Gamma\left(\frac{\upsilon+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{N}{2}\log(\nu-2) \\ - \frac{N+\nu}{2}\log\left(1 + \frac{\epsilon_t'\Omega_t^{-1}(\theta)\epsilon_t}{\nu-2}\right) - \frac{1}{2}\log|\Omega_t(\theta)| \end{aligned}$$

where $\Omega_t(\theta)$ is given by equation (14) for the regular BEKK or (15) for the scalar BEKK and $\Gamma(\cdot)$ being the gamma function.[12] Similar to the univariate models, numerical optimization of the likelihood function is the only way to find the maximum. If the condition in equation (16) holds, the law of large numbers applies, and the maximum likelihood estimator is consistent.[13]

Similar to univariate GARCH models, multiple different Multivariate GARCH models exist, one of which is the Dynamical Conditional Correlation MGARCH or DCC MGARCH by [Engle, 2002], which have another way around the parameterization problem.

DCC MGARCH

As the regular BEKK is impractical, we use a *DCC MGARCH model* with a scalar BEKK for modeling conditional correlations. Consider the constant mean model given by equation (12), but the conditional covariance is now given by a DCC MGARCH model. This model exploits the fact that the $N \times N$ covariance matrix, Ω_t, in equation (13) can be decomposed into two variances matrices, Var_t and a correlation matrix, Γ_t,

$$\Omega_t = \text{Var}_t\Gamma_t\text{Var}_t \tag{17}$$

[11] [Engle and Kroner, 1995], *Multivariate Simultaneous Generalized Arch*, proposition 2.7

[12] [Rossi and Spazzini, 2010], *Model and distribution uncertainty in multivariate GARCH estimation: a Monte Carlo analysis*, p. 7

[13] [Engle and Kroner, 1995], *Multivariate Simultaneous Generalized Arch*, p. 138-139

Where

$$\text{Var}_t = \text{diag}\left(\sqrt{\sigma_{i,t}^2}\right) = \begin{pmatrix} \sigma_{1,t} & 0 & \cdots & 0 \\ 0 & \sigma_{2,t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{p,t} \end{pmatrix} \tag{18}$$

and each diagonal element is given by a univariate GARCH process. This GARCH process can be any univariate GARCH process like a GJR-GARCH(1,1) or a GARCH(1,1) as below.

$$\sigma_{i,t}^2 = \omega_i + \alpha_i \epsilon_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2 \quad \text{for} \quad i = 1, 2, ..., N \qquad \omega_i > 0, \alpha_i, \beta_i \geq 0 \tag{19}$$

These are weakly stationary given the conditions described in section 3.2. The correlation matrix is given as

$$\Gamma_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2} = \begin{pmatrix} 1 & \rho_{12,t} & \cdots & \rho_{1N,t} \\ \rho_{21,t} & 1 & \cdots & \rho_{2N,t} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N1,t} & \rho_{N2,t} & \cdots & 1 \end{pmatrix} \tag{20}$$

where Q_t is the pseudo correlation which follows a scalar BEKK(1,1) MGARCH process given by

$$Q_t = \bar{Q}(1 - a - b) + a\eta_{t-1}\eta_{t-1}' + bQ_{t-1} \qquad a, b \geq 0 \qquad \bar{Q} = \frac{1}{T}\sum_{t=1}^{T} \eta_t \eta_t' > 0 \tag{21}$$

where $\eta_t = \text{Var}_t^{-1} \epsilon_t$ is a $N \times 1$ matrix of standardized (or devolatized) disturbances of the correlations.

The ability to decompose the conditional covariance allows us to estimate this model in two stages. First, we estimate the parameters in the N univariate GARCH models for the N assets using the maximum likelihood estimation explained in section 3.2. Second, we estimate a multivariate scalar BEKK for the conditional correlation using the maximum likelihood estimation for the BEKK above, which reduces the number of parameters to estimate. However, by using a scalar BEKK(1,1) for the conditional correlations, we assume that the conditional correlations have identical responsiveness to shocks and persistence. This assumption is not as restrictive as one would think, because the conditional covariances still respond differently to shocks when the conditional variances are modeled individually. This multivariate specification of the conditional correlation given by equation (20) and (21) will remain the same throughout the thesis.

The DCC MGARCH model with univariate GARCH(1,1) for the conditional variances and a multivariate scalar BEKK(1,1) for the conditional correlations has $3N + 3$ parameters. Compared to the BEKK(1,1) for $N = 4$, the DCC MGARCH(1,1) has 12 parameters and 33 for $N = 10$. So the number of parameters for DCC MGARCH(1,1) still grows fast, and thus, to truly model a large number of assets, Eigenvalue MGARCH or λ-MGARCH models are preferable. Still, since this thesis does not explore $N \gg 10$, we stick to the simpler DCC MGARCH model.

3.3.1 One-period forecast of the conditional covariance, Ω_{t+1}

A very useful feature of GARCH type models is forecasting future (co)variance. Specifically, Ω_{t+1} can be forecasted with $\mathcal{F}_t$-measurable variables using equation (17)-(21).

Denote $X_{t+1|t}$ as the forecast of a variable X_{t+1} given $\mathcal{F}_t$. The forecast of Ω_{t+1} is then $\Omega_{t+1|t}$ and similarly for the variance matrix $\text{Var}_{t+1|t}$ and the correlation matrix $\Gamma_{t+1|t}$. Recall from (17) that the conditional covariance matrix, Ω_{t+1}, can be decomposed into the conditional variance and conditional correlation matrices such that the forecast, $\Omega_{t+1|t}$ can be written as

$$\Omega_{t+1|t} = \text{Var}_{t+1|t}\Gamma_{t+1|t}\text{Var}_{t+1|t}$$

Where:

$$\text{Var}_{t+1|t} = \text{diag}\begin{pmatrix} \sigma_{1,t+1|t} \\ \sigma_{2,t+1|t} \\ \vdots \\ \sigma_{N,t+1|t} \end{pmatrix} = \text{diag}\begin{pmatrix} \sqrt{\omega_1 + \alpha_1 \mathbb{E}_t[\epsilon_{1,t}^2] + \beta_1 \sigma_{1,t}^2} \\ \sqrt{\omega_2 + \alpha_2 \mathbb{E}_t[\epsilon_{2,t}^2] + \beta_2 \sigma_{2,t}^2} \\ \vdots \\ \sqrt{\omega_N + \alpha_N \mathbb{E}_t[\epsilon_{N,t}^2] + \beta_N \sigma_{N,t}^2} \end{pmatrix}$$

and $\mathbb{E}_t[\epsilon_{i,t}^2] = \epsilon_{i,t}^2$ as ϵ_t is known in period t. We can also decompose the forecast of the conditional correlation as:

$$\Gamma_{t+1|t} = \text{diag}(Q_{t+1|t})^{-1} Q_{t+1|t} \text{diag}(Q_{t+1|t})^{-1}$$

With $Q_{t+1|t}$ being forecast as

$$Q_{t+1|t} = \bar{Q}(1-a-b) + a\mathbb{E}_t[\eta_t\eta_t'] = \bar{Q}(1-a-b) + a\eta_t\eta_t' + bQ_t$$

where $\mathbb{E}_t[\eta_t\eta_t'] = \eta_t\eta_t$ as η_t is known in period t. Note that this is a forecast and thus subject to errors which will become highly relevant during portfolio optimization.

3.3.2 Shrinkage of the covariance matrix

Using the estimated covariance matrix, $\Omega_{t+1|t}$, but mainly the sample covariance matrix, Σ, introduces problems when used in portfolio optimization thoroughly explained in [Jobson and Korkie, 1980] and [Jagannathan and Ma, 2003]. To summarize, any estimation error may introduce extreme coefficients into the covariance matrix used in the mean-variance optimization. The extreme coefficients lead to exaggerated portfolio weights when maximizing utility for the investor. The portfolio weights are thus very unreliable given the estimation error causing inefficiencies, a phenomenon which [Michaud, 1989] dubbed "error-maximization."

A straightforward approach to reduce this problem is to "shrink" the estimates of the covariance towards the identity matrix I by some proportion δ as explained in [Ledoit and Wolf, 2004]

$$\hat{\Sigma}^{\text{Shrunk}} = \delta I + (1-\delta)\hat{\Sigma} \tag{22}$$

where we apply a shrinkage of 50% corresponding to $\delta = 0.5$ similar to [Gârleanu and Pedersen, 2013]. We don't shrink the entire covariance matrix, but only $\bar{Q}$ as this is the only part of the model estimated with a sample estimate corresponding to a pseudo sample correlation matrix such that

$$\bar{Q}^{\text{Shrunk}} = \delta I + (1-\delta)\bar{Q}$$

Which when used in $\Omega_{t+1|t}$ gives $\Omega_{t+1|t}^{\text{Shrunk}}$. Throughout the theoretical section and section 4 on Dynamic programming, we use $\Omega_{t+1|t}$ in the derivations to reduce awkward notation, but when we apply the theory, we use the shrunken version, $\hat{\Omega}_{t+1|t}^{\text{Shrunk}}$.

3.3.3 Multivariate GARCH models in Portfolio theory

Consider the investor's problem for N assets presented in section 2.2.1. Now, rather than the constant covariance matrix Σ, we use the time-varying covariance matrix, Ω_t, that comes from the constant mean DCC MGARCH(1,1) model from equation (12) and (17)-(21). The investor minimizes the variance of their portfolio given that the weights of the portfolio sum to 1.

$$\min_{v_t} \left\{ \frac{1}{2} v_t' \mathbb{E}_t[\Omega_{t+1}] v_t \right\} \qquad \text{s.t.} \quad v_t' \mathbf{1} = 1$$

As explained in section 3.3.1, the value of Ω_{t+1} can be forecast using $\mathcal{F}_t$-measurable variables such that $\mathbb{E}_t[\Omega_{t+1}] = \Omega_{t+1|t}$ i.e., the conditional expectation of the covariance matrix in period $t+1$ given period t is the forecast, $\Omega_{t+1|t}$. The Lagrangian is given as

$$\begin{aligned} \mathcal{L}(v_t) &= \frac{1}{2} v_t' \mathbb{E}_t[\Omega_{t+1}] v_t - \lambda(v_t' \mathbf{1} - 1) \\ &= \frac{1}{2} v_t' \Omega_{t+1|t} v_t - \lambda(v_t' \mathbf{1} - 1) \end{aligned}$$

Taking the first-order conditions wrt. v_t and solving for v_t and λ as in section 2.2.1 yields the minimum variance portfolio after some light algebra. The proof is in appendix A.3

$$v_t = \frac{\Omega_{t+1|t}^{-1} \mathbf{1}}{\mathbf{1}' \Omega_{t+1|t}^{-1} \mathbf{1}} = v_t^{\text{MVP}} \tag{23}$$

The result is intuitive as $\Omega_{t+1|t}$ simply replaces the constant Σ and, thus, it has the same interpretation. It is likewise easy to replace $\Omega_{t+1|t}$ with the shrunken version $\hat{\Omega}_{t+1|t}^{\text{Shrunk}}$.

4 Dynamic Trading Strategies

We have explored replacing Σ with the covariance matrix of an MGARCH model, $\Omega_{t+1|t}$ as the covariance of financial assets is time-varying. However, to truly use the possible benefit from the added complexity of MGARCH models, we need to consider a *multi-period problem* where the dynamics of the MGARCH model can be used to obtain an optimal dynamic trading strategy. We apply the *dynamic programming* framework introduced by [Bellman, 1966] to solve multi-period problems. We use notation similar to [Gan and Lu, 2014] and [Gârleanu and Pedersen, 2013].

4.1 Dynamic programming techniques

Consider an agent that seeks a policy that defines the optimal action the agent should take at time t in state s. The policy $\{v_t^*\}_{t=1}^{\infty}$ maximizes the present value of current rewards and future expected rewards, $f(v_t, s_t)$, discounted by $(1-\rho) \in (0,1]$ given some general constraint $g(v_t, s_t)$, that could be a budget constraint or a weight constraint.

$$\max_{\{v_t\}_{t=0}^{\infty}} \mathbb{E}_0\left[\sum_{t=0}^{\infty}(1-\rho)^{t+1}f(v_t,s_t)\right] \quad \text{s.t.} \quad g(v_t,s_t)=0 \tag{24}$$

$\mathbb{E}_0[\cdot]$ is the conditional expectation given the filtration in period 0, i.e., $\mathbb{E}_0[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_0]$. Since a maximization problem can be converted to a minimization problem,[14] equation (24) is equivalent to

$$\min_{\{v_t\}_{t=0}^{\infty}} -\mathbb{E}_0\left[\sum_{t=0}^{\infty}(1-\rho)^{t+1}f(v_t,s_t)\right] \quad \text{s.t.} \quad g(v_t,s_t)=0 \tag{25}$$

Consider an agent facing the minimization problem in (25). Dynamic programming decomposes this multi-period problem into a two-period problem, "now" and "later." The decomposition is accomplished by rewriting the sum by taking out the first period $t=0$.

$$\begin{aligned}
L_0 &= \min_{\{v_t\}_{t=0}^{\infty}} -\mathbb{E}_0\left[\sum_{t=0}^{\infty}(1-\rho)^{t+1}f(v_t,s_t)\right] \quad \text{s.t.} \quad g(v_t,s_t)=0\\
&= \min_{\{v_t\}_{t=0}^{\infty}} -\mathbb{E}_0\left[(1-\rho)\bigg(f(v_0,s_0)+\sum_{t=1}^{\infty}(1-\rho)^{t+1}f(v_t,s_t)\bigg)\right] \quad \text{s.t.} \quad g(v_t,s_t)=0\\
&= \min_{\{v_t\}_{t=0}^{\infty}} -\mathbb{E}_0\left[(1-\rho)\bigg(f(v_0,s_0)+\mathbb{E}_1\bigg\{\sum_{t=1}^{\infty}(1-\rho)^{t+1}f(v_t,s_t)\bigg\}\bigg)\right] \quad \text{s.t.} \quad g(v_t,s_t)=0\\
&= \min_{v_0} -\mathbb{E}_0\left[(1-\rho)\bigg(\underbrace{f(v_0,s_0)}_{(j)}+\underbrace{\min_{\{v_t\}_{t=1}^{\infty}} \mathbb{E}_1\bigg\{\sum_{t=1}^{\infty}(1-\rho)^{t+1}f(v_t,s_t)\bigg\}}_{(jj)}\bigg)\right] \quad \text{s.t.} \quad \underbrace{g(v_t,s_t)=0}_{(jjj)}
\end{aligned}$$

Here we have the immediate reward, *(j)*, and future reward,*(jj)*, under some constraint, *(jjj)*, which we transform to a Lagrangian type constraint with a time-varying Lagrangian multiplier, λ_t.[15]

$$L_0 = -\min_{v_t}\left[(1-\rho)f(v_0,s_0)+\min_{\{v_t\}_{t=1}^{\infty}} \mathbb{E}_0\bigg\{\sum_{t=1}^{\infty}(1-\rho)^{t+1}f(v_t,s_t)\bigg\}\right]-\lambda_t[g(v_t,s_t)]$$

This final part can be written as *the value function*, $V(s_0)$. We note that *(jj)* can be written as its own dynamic problem just one period into the future as $V(s_1)$ such that

$$V(s_0) = -\min_{v_0}\Big[(1-\rho)f(v_0,s_0)+\mathbb{E}_0[V(s_1)]\Big]-\lambda_t[g(v_t,s_t)]$$

[14] $\max_x\{f(x)\} = \min_x\{-f(x)\}$

[15] See [Gan and Lu, 2014], *General Setting and solution of Bellman equation in monetary theory*, page 2

Generalizing the value function to period t, we have

$$V(s_t) = -\min_{v_t}\Bigg[\underbrace{(1-\rho)f(v_t,s_t)}_{(j)} + \underbrace{\mathbb{E}_t[V(s_{t+1})]}_{(jj)}\Bigg] - \underbrace{\lambda_t[g(v_t,s_t)]}_{(jjj)}$$

The optimal policies for the agent are the solutions to the optimization problem contained within the value function, $V(s_t)$, for each period t. In our case, the policies are optimal portfolio weights.

4.2 Optimal dynamic strategies for multiple risky assets

Equipped with methods from dynamic programming, we proceed to derive optimal dynamic trading strategies. First, for an investor that ignores transaction costs, and lastly, for an investor that adjusts for transaction costs.

4.2.1 Derivation of the simple strategy (ignoring transaction costs)

Consider N assets with $N\times 1$ vector of returns, $r_t = (r_{1,t}, r_{2,t}, ..., r_{N,t})'$ given by a constant mean DCC MGARCH(1,1) model given by equation (12) and (17)-(21). The investor has mean-variance preferences given by equation (1) and seeks a dynamic trading strategy, $\{v_t\}_{t=0}^{\infty}$, that solves the following multi-period minimization problem

$$\max_{\{v_t\}_{t=0}^{\infty}} \mathbb{E}_0\left[\sum_{t=0}^{\infty}(1-\rho)^{t+1}\left(-\frac{1}{2}v_t'\Omega_{t+1}v_t\right)\right] \quad \text{s.t.} \quad v_t'\mathbf{1} = 1 \tag{26}$$

with $f(x_t,s_t) = -\frac{1}{2}v_t\Omega_{t+1}'$ and $g(x_t,s_t) = 0 \Rightarrow v_t'\mathbf{1} - 1 = 0 \Leftrightarrow v_t'\mathbf{1} = 1$. Multiplying (-1) into the minimization problem yields the investor's objective function

$$\min_{\{v_t\}_{t=0}^{\infty}} \mathbb{E}_0\left[\sum_{t=0}^{\infty}(1-\rho)^{t+1}\left(\frac{1}{2}v_t'\Omega_{t+1}v_t\right)\right] \quad \text{s.t.} \quad v_t'\mathbf{1} = 1 \tag{27}$$

The problem can be solved via dynamic programming using the method in section 4.1. Denote the minimization problem in (27) as denoted L_t^{GARCH}

$$L_t^{\text{GARCH}} = \min_{\{v_t\}_{t=0}^{\infty}} \mathbb{E}_0\left[\sum_{t=0}^{\infty}(1-\rho)^{t+1}\left(\frac{1}{2}v_t'\Omega_{t+1}v_t\right)\right] \quad \text{s.t.} \quad v_t'\mathbf{1} = 1$$

Following the same argument as in section 4.1 we get

$$= \min_{v_0}\mathbb{E}_0\left[(1-\rho)^1\left[\frac{1}{2}v_0'\Omega_{t+1}v_0\right] + \min_{\{v_t\}_{t=1}^{\infty}}\mathbb{E}_1\left\{\sum_{t=1}^{\infty}(1-\rho)^t\left(\frac{1}{2}v_t'\Omega_{t+1}v_t\right)\right]\right\}\right] - \lambda_t(v_t'\mathbf{1}-1)$$

$$= \min_{v_0}\left[(1-\rho)\frac{1}{2}v_0'\Omega_{t+1}v_0 + (1-\rho)\min_{\{v_t\}_{t=1}^{\infty}}\mathbb{E}_0\left\{\sum_{t=1}^{\infty}(1-\rho)^t\left(\frac{1}{2}v_t'\Omega_{t+1}v_t\right)\right\}\right] - \lambda_t(v_t'\mathbf{1}-1)$$

Generalizing to period $t-1$, we get that the value function, $V(v_{t-1})$ is given by

$$V(v_{t-1}) = \min_{v_t} \left[(1-\rho)\left(\underbrace{\frac{1}{2} v_t' \mathbb{E}_t[\Omega_{t+1}] v_t}_{(i)} + \underbrace{\mathbb{E}_t[V(v_t)]}_{(ii)} \right) \right] - \underbrace{\lambda(v_t'\mathbf{1} - 1)}_{(iii)}$$

The investor needs a measure of the next period's covariance, Ω_{t+1}, a random variable at period t. But given the MGARCH structure of the covariance. $\mathbb{E}_t[\Omega_{t+1}]$ can be given as the forecast value $\Omega_{t+1|t}$ (See section 3.3.1).

$$V(v_{t-1}) = \min_{v_t} \left[(1-\rho)\left(\underbrace{\frac{1}{2} v_t' \Omega_{t+1|t} v_t}_{(i)} + \underbrace{\mathbb{E}_t[V(v_t)]}_{(ii)} \right) \right] - \underbrace{\lambda(v_t'\mathbf{1} - 1)}_{(iii)} \tag{28}$$

The value function measures the value of the portfolio, in terms of utility for the investor, at time t with the weight v_{t-1} of the risky assets. Additionally, the value function measures utility for the current period *(i)* and future periods *(ii)* under the constraint that the weights sum to 1, *(iii)*.

Solving for the optimal weights requires solving the first-order conditions wrt. to the weights v_t. But first, we look into the conditional expectation of the value function in period t:

$$\mathbb{E}_t[V(v_t)] = \mathbb{E}_t\left(\min_{v_{t+1}} \left[(1-\rho)\left(\frac{1}{2} v_{t+1}' \Omega_{t+2} v_{t+1} + \mathbb{E}_{t+1}[V(v_{t+1})] \right) \right] - \lambda(v_{t+1}'\mathbf{1} - 1) \right)$$

Notice that none of the terms in $\mathbb{E}_t[V(v_t)]$ contain v_t. This turns the dynamic minimization problem into a series of one-period static problems. We can therefore proceed by solving the problem for one general period t, and reuse this result every period:

$$\frac{\partial V(v_{t-1})}{\partial v_t} = (1-\rho) v_t \Omega_{t+1|t} - \lambda_t \mathbf{1} = 0 \Leftrightarrow v_t = (1-\rho)^{-1} \Omega_{t+1|t}^{-1} \lambda_t \mathbf{1}$$

The constraint requires that $v_t'\mathbf{1} = 1$, which we can use to solve for the Lagrangian multiplier λ

$$1 = v_t'\mathbf{1} = \mathbf{1}' v_t = \mathbf{1}'(1-\rho)^{-1}\Omega_{t+1}^{-1}\lambda_t \mathbf{1} = (1-\rho)^{-1}\mathbf{1}'\Omega_{t+1|t}^{-1}\mathbf{1}\lambda_t \Leftrightarrow \lambda_t = \frac{1}{(1-\rho)^{-1}\mathbf{1}'\Omega_{t+1|t}^{-1}\mathbf{1}}$$

We insert the expression into the weights v_t

$$v_t = (1-\rho)^{-1}\Omega_{t+1|t}^{-1}\left[\frac{1}{(1-\rho)^{-1}\mathbf{1}'\Omega_{t+1|t}^{-1}\mathbf{1}} \right]\mathbf{1} = \frac{\Omega_{t+1|t}^{-1}\mathbf{1}}{\mathbf{1}'\Omega_{t+1|t}^{-1}\mathbf{1}} \equiv v_t^{\text{MVP}}$$

Because $\Omega_{t+1|t}$ is positive definite, $\mathbf{1}'\Omega_{t+1|t}^{-1}\mathbf{1} > 0$. Given that the covariance, $\Omega_{t+1|t}$ is time-varying, the weights, v_t, are likewise time-varying. The interpretation is that the investor forecasts the covariance of the asset returns in the feasible set every period using the information at time t. The investor can with some accuracy forecast the covariance within the near future or, at the very least, the investor believes he can forecast the covariance one period into the future.

This result is not surprising as, in the absence of trading costs, the investor can rebalance the portfolio each period to the new minimum variance portfolio at no charge. This is in line with [Gârleanu and Pedersen, 2013] and [Mei and Nogales, 2018]. They both find that the aim portfolio, i.e., the desired portfolio in the absence of transaction costs, is the Markowitz portfolio which corresponds to the minimum variance portfolio when the objective is to minimize volatility.

4.2.2 Derivation of the sophisticated strategy (adjusting for transaction costs)

Consider again N assets with returns, r_t, given by a constant mean DCC MGARCH(1,1) model from equation (12) and (17)-(21). An investor incurring transaction costs seeking to minimize portfolio risk faces the following problem when seeking a dynamic investment strategy:

$$\min_{\{v_t\}_{t=0}^{\infty}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} (1-\rho)^{t+1} \left(\frac{1}{2} v_t' \Omega_{t+1} v_t \right) + \frac{(1-\rho)^t}{2} \left(\frac{1}{2} (v_t - v_{t-1})' \Lambda_t (v_t - v_{t-1}) \right) \right] \quad \text{s.t.} \quad v_t' \mathbf{1} = 1 \tag{29}$$

with transaction costs taking the form $\frac{1}{2}(v_t - v_{t-1})'\Lambda_t(v_t - v_{t-1})$. The second term containing the transaction cost penalization is discounted in period t and not $t+1$ since the investor incurs transaction costs immediately. The investor now has to trade-off rebalancing their portfolio given new information with the cost of trading. We assume that the transaction costs, Λ_t, are given as

$$\Lambda_t = \Omega_{t+1|t} \gamma_D \tag{30}$$

meaning that transaction costs are time-varying as $\Omega_{t+1|t}$ is time-varying and γ_D is the risk-aversion parameter. This specification might seem odd at first glance, but similarly to [Gârleanu and Pedersen, 2013], one can think of it as a dealer taking the opposite side of the trade Δv_t that our investor makes. The dealer holds this position and then sells it back to the market. During this period, the dealer's risk is equivalent to $\Delta v_t' \Omega_{t+1|t} \Delta v_t$. The transaction costs can thus be interpreted as compensation for the dealer's risk, the level of which depends on their risk-aversion, γ_D.

We solve this problem by using the method in section 4.1. Denote the minimization problem in (29) as L_t^{TC}

$$L_t^{\text{TC}} = \min_{\{v_t\}_{t=0}^{\infty}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} (1-\rho)^{t+1} \left(\frac{1}{2} v_t' \Omega_{t+1} v_t \right) + \frac{(1-\rho)^t}{2} (v_t - v_{t-1})' \Lambda_t (v_t - v_{t-1}) \right] - \lambda_t (v_t' \mathbf{1} - 1)$$

Following the same argument as in section 4.1 we get

$$
\begin{aligned}
&= \min_{v_0} \mathbb{E}_0 \Bigg[(1-\rho)^1 \left(\frac{1}{2} v_0' \Omega_{t+1} v_0 \right) + \frac{(1-\rho)^0}{2} (v_0 - v_{-1})' \Lambda (v_0 - v_{-1}) \\
&+ \min_{\{v_t\}_{t=1}^{\infty}} \mathbb{E}_1 \Bigg\{ \sum_{t=1}^{\infty} (1-\rho)^{t+1} \left(\frac{1}{2} v_t' \Omega_{t+1} v_t \right) + \frac{(1-\rho)^t}{2} (v_t - v_{t-1})' \Lambda_t (v_t - v_{t-1}) \Bigg\} \Bigg] - \lambda_t (v_t' \mathbf{1} - 1) \\
&= \min_{v_0} \Bigg[\frac{1}{2} (v_0 - v_{-1})' \Lambda_t (v_0 - v_{-1}) + (1-\rho) \left(\frac{1}{2} v_0' \Omega_1 v_0 \right) \\
&+ \min_{\{v_t\}_{t=1}^{\infty}} \mathbb{E}_0 \Bigg\{ \sum_{t=1}^{\infty} (1-\rho)^{t+1} \left(\frac{1}{2} v_t' \Omega_{t+1} v_t \right) + \frac{(1-\rho)^t}{2} (v_t - v_{t-1})' \Lambda_t (v_t - v_{t-1}) \Bigg\} \Bigg] - \lambda_t (v_t' \mathbf{1} - 1)
\end{aligned}
$$

Where v_{-1} is the weight of the investor's initial portfolio which sums to 1. Generalizing to period t, we get that the value function, $V(v_{t-1})$ is given by

$$
V(v_{t-1}) = \min_{v_t} \left[\frac{1}{2} (v_t - v_{t-1})' \Lambda_t (v_t - v_{t-1}) + (1-\rho) \left(\frac{1}{2} v_t' \mathbb{E}_t [\Omega_{t+1}] v_t + \mathbb{E}_t [V(v_t)] \right) \right] - \lambda_t (v_t' \mathbf{1} - 1) \tag{31}
$$

Similar to section 4.2.1, the investor needs a measure of the next period's covariance, Ω_{t+1}, a random variable at period t. $\mathbb{E}_t[\Omega_{t+1}]$ can be given as the forecast value $\Omega_{t+1|t}$ (See section 3.3.1) with an analogous interpretation as in section 4.2.1.

In contrast to the problems without transaction costs in section 4.2.1, $\mathbb{E}_t[V(v_t)]$ also becomes relevant as transaction costs add persistence between periods, i.e., v_t now also affects the value function of the next period. Thus, an expression for the next period's $V(v_t)$ is needed. Following the method of [Gârleanu and Pedersen, 2013], we apply the 'guess and verify' method, which can be divided into 6 steps:

1. Make an ansatz of the form of the value function
2. Set up the Bellman equation of the guessed value function
3. Find the first-order conditions and solve for the optimal policy (the weights)
4. Insert the optimal policy (the weights) into the value function
5. Compare the new value function with the ansatz and verify it solves the problem
6. Solve for coefficients

Our ansatz for $V(v_t)$ is

$$
V(v_t) = \frac{1}{2} v_t' A_{vv} v_t - v_t' A_{v1} \mathbf{1} - \frac{1}{2} \mathbf{1}' A_{11} \mathbf{1}
$$

Very tedious algebra can show that this is indeed a solution. The proof is in appendix A.4 along with expressions for A_{vv}, A_{v1}, and A_{11}. We solve for the optimal weights by taking the partial derivative with respect to v_{t-1} as the optimal solutions for v_t are already embedded within the guessed value function and its parameters, A_{vv}, A_{v1}, and A_{11}. Thus, the investor essentially finds the optimal weights in period $t-1$ given what the investor believes about the optimal weights in period t. Differentiating the Bellman equation (31) with respect to v_{t-1} yields:

$$\begin{aligned} A_{vv}v_{t-1} - A_{v1}\mathbf{1} &= \Lambda_t(v_t - v_{t-1}) \\ \Lambda_t v_t &= \Lambda_t v_{t-1} + A_{vv}v_{t-1} - A_{v1}\mathbf{1} \\ v_t &= v_{t-1} + \Lambda_t^{-1}A_{vv}\left[v_{t-1} - A_{vv}^{-1}A_{v1}\mathbf{1}\right] \end{aligned}$$

Define $A_{vv}^{-1}A_{v1}\mathbf{1}$ as *the aim portfolio*, aim_t and insert $\Lambda_t = \gamma_D\Omega_{t+1|t}$

$$v_t = v_{t-1} + (\gamma_D\Omega_{t+1|t})^{-1}A_{vv}\left[v_{t-1} - \text{aim}_t\right] \tag{32}$$

This result has an intuitive interpretation similar to [Gârleanu and Pedersen, 2013]. The investor starts with the previous period's weights, v_{t-1} and then estimates the aim portfolio, aim_t, the investor's preferred portfolio given no transaction costs. Then given how costly it is to trade in the period, $\gamma_D\Omega_{t+1|t}$, the investor changes their portfolio to a combination of v_{t-1} and aim_t resulting in v_t.

Finally, we check if the weights v_t sum to 1 using $v_t'1 = 1$

$$\begin{aligned} \mathbf{1}'\left(v_{t-1} + \Lambda_t^{-1}A_{vv}\left[v_{t-1} - A_{vv}^{-1}A_{v1}\mathbf{1}\right]\right) &\stackrel{?}{=} 1 \\ \mathbf{1}'v_{t-1} + \mathbf{1}'\Lambda_t^{-1}A_{vv}v_{t-1} - \mathbf{1}'\Lambda_t^{-1}A_{v1}\mathbf{1} &\neq 1 \end{aligned}$$

They clearly don't sum to 1, which is undoubtedly a problem.[16] To be able to move on, we enforce that the aim portfolio and the next period's portfolio weights sum to 1, so equation 32 becomes:

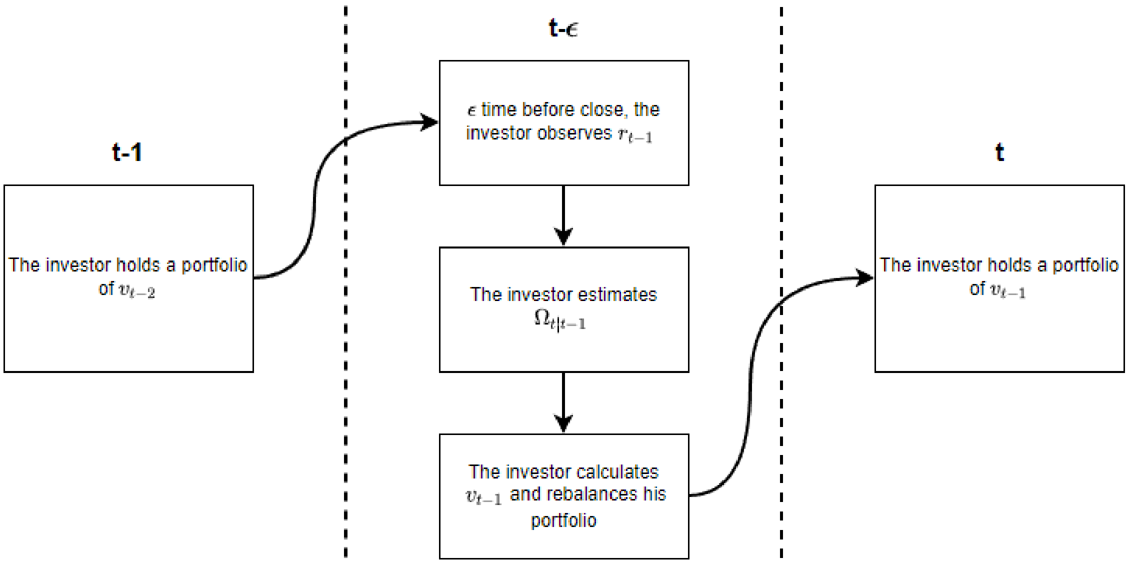
$$v_t = \frac{v_{t-1} + (\gamma_D\Omega_{t+1|t})^{-1}A_{vv}\left[v_{t-1} - \frac{\text{aim}_t}{\mathbf{1}'\text{aim}_t}\right]}{\mathbf{1}'(v_{t-1} + (\gamma_D\Omega_{t+1|t})^{-1}A_{vv}\left[v_{t-1} - \frac{\text{aim}_t}{\mathbf{1}'\text{aim}_t}\right])} \tag{33}$$

The additional restriction to the aim portfolio might seem odd, but we later show empirically that the normalized aim portfolio is very similar to the simple strategy. This result intuitively makes sense, leading us to speculate that a mathematically correct solution should be possible. However, this is a topic for further research as the time constraints of this project force us to move on. Notice that the normalized portfolio requires $\mathbf{1}'(v_{t-1} + (\gamma_D\Omega_{t+1|t})^{-1}A_{vv}\left[v_{t-1} - \frac{\text{aim}_t}{\mathbf{1}'\text{aim}_t}\right]) \neq 0$ and $\mathbf{1}'\text{aim}_t \neq 0$ which is not necessarily fulfilled. However, this has not led to any problems in our implementation of the theory.

Below in figure 3, we summarize how the investor gains new information and updates their portfolio. Here, ϵ is a small time-step just before market close.

[16] This means we must have made a mathematical error in A.4 or used an incomplete approach to begin with, for example, with too few constraints. However, we have made substantial progress in unifying MGARCH models with dynamic trading strategies

Figure 3: The trading cycle of the investor



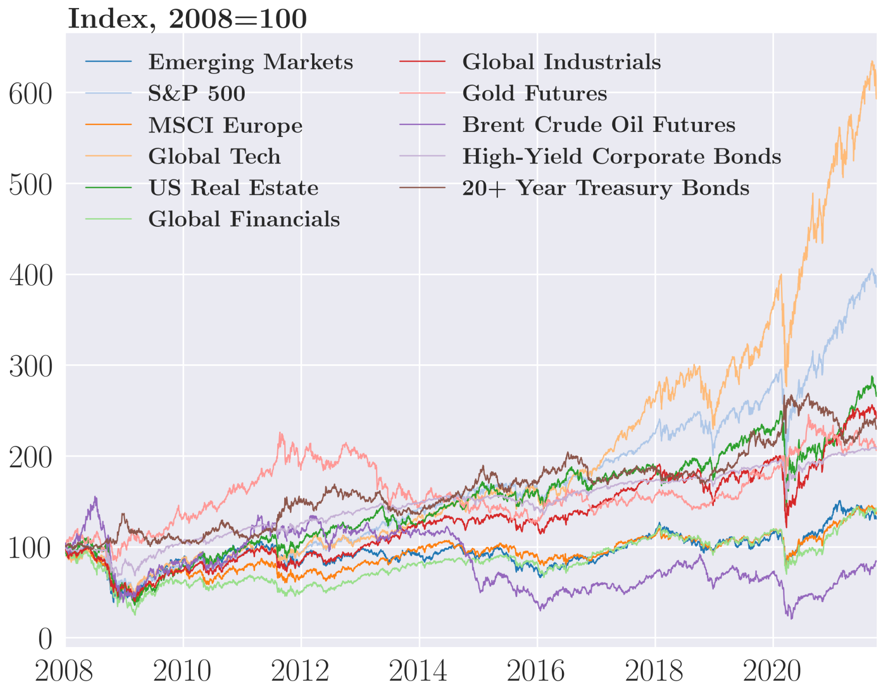
To summarize, we have derived two optimal minimum variance dynamic strategies when the investor faces quadratic transaction costs with returns modeled by a constant mean process and a DCC MGARCH model for the conditional covariance. Before proceeding to case studies using historical backtesting, we briefly describe the data we use.

5 Overview of assets used in backtesting

We have chosen multiple different assets to test the dynamic strategies. Our source is Yahoo Finance, where the data is freely available. We are primarily interested in stock ETFs, but we also try the strategy on gold and crude oil and two bond ETFs. The mix of assets covers a broad range of the financial markets without the need for thousands of individual assets. We purposefully chose diversified assets to minimize the risk of spurious empirical results because of the development in a specific sector or region. We use data from January 1^{st} 2008 to October 1^{st} 2021. The indexed time series of these assets are displayed in figure 4. Here, we see that traditionally safer assets like gold and bonds steadily increase and generally fair well around crises like 20+ year treasuries did around 2008 and 2020. In contrast, stocks typically suffer substantial drops in value around the same market crashes, but in turn, achieve higher average returns. This pattern is not surprising as finance theory predicts that exposing yourself to higher risks, in general, is rewarded with higher returns. This pattern becomes even more apparent when looking at table 2, where the stock ETFs generally have the highest annualized returns at the cost of the highest risk. In contrast, bonds typically have lower returns at much lower risk. From table 1, we note that the commodities are primarily uncorrelated with the remaining assets.

Specifically, gold (GC=F) is very close to being uncorrelated with any of the assets except

Figure 4: Indexed prices of our asset sample



Brent crude oil (BZ=F), which itself has a low positive correlation to most assets. In sharp contrast, all the stock ETFs, EEM to EXI, are highly positively correlated with a correlation coefficient above 0.7 in all cases. Finally, the 20 year+ treasuries (TLT) negatively correlate with all assets except Brent crude oil (BZ=F). The correlations are essential in forming the minimum variance portfolio because they create diversification and hedging opportunities, which increase as assets get more negatively correlated. As mentioned in the theoretical

Table 1: Correlation matrix of selected asset returns

	EEM	IVV	IEV	IXN	IYR	IXG	EXI	GC=F	BZ=F	HYG	TLT
EEM	1.00	0.85	0.86	0.82	0.70	0.82	0.84	0.10	0.36	0.63	-0.36
IVV	0.85	1.00	0.89	0.92	0.79	0.90	0.92	0.01	0.35	0.66	-0.43
IEV	0.86	0.89	1.00	0.82	0.70	0.89	0.92	0.09	0.38	0.66	-0.41
IXN	0.82	0.92	0.82	1.00	0.68	0.80	0.84	0.03	0.32	0.61	-0.38
IYR	0.70	0.79	0.70	0.68	1.00	0.78	0.73	0.03	0.21	0.53	-0.25
IXG	0.82	0.90	0.89	0.80	0.78	1.00	0.89	-0.01	0.34	0.67	-0.45
EXI	0.84	0.92	0.92	0.84	0.73	0.89	1.00	0.06	0.37	0.67	-0.43
GC=F	0.10	0.01	0.09	0.03	0.03	-0.01	0.06	1.00	0.18	0.03	0.13
BZ=F	0.36	0.35	0.38	0.32	0.21	0.34	0.37	0.18	1.00	0.32	-0.23
HYG	0.63	0.66	0.66	0.61	0.53	0.67	0.67	0.03	0.32	1.00	-0.27
TLT	-0.36	-0.43	-0.41	-0.38	-0.25	-0.45	-0.43	0.13	-0.23	-0.27	1.00

section, returns of financial time-series are not Gaussian, a point thoroughly proven by the p-values of the Jarque-Bera test of normality being 0% for every asset in our sample, thus, rejecting the null of normality. In figure 5 (a)-(c), the fitted Gaussian densities are also far from the shape of the histogram of the actual data. The Generalized Error Distribution (GED) or a non-central Student's t-distribution is a far better fit. The latter takes the slight edge as the GED often has excess kurtosis. We provide the parameters for the fitted non-central Student's t-distribution for each asset in table 2.[17]

Table 2: Descriptive statistics, tests, and distribution fit

	Mean	Std. Dev	Normality	DF	Non central
	Annualized			NC Student's t parameters	
Emerging Markets (EEM)	6.394%	29.919%	0.0 %	2.793	-0.199
S&P 500 (IVV)	11.921%	20.353%	0.0 %	2.154	-0.156
Europe (IEV)	5.333%	24.651%	0.0 %	2.473	-0.188
Global Tech (IXN)	15.511%	22.825%	0.0 %	2.612	-0.195
Real estate (IYR)	12.097%	31.307%	0.0 %	1.815	-0.127
Global financials (IXG)	6.659%	28.924%	0.0 %	2.105	-0.165
Global Industrials (EXI)	8.997%	22.549%	0.0 %	2.498	-0.178
Gold (GC=F)	6.880%	18.071%	0.0 %	3.067	-0.092
Brent crude oil (BZ=F)	5.992%	38.229%	0.0 %	2.484	-0.081
High-yield bonds (HYG)	5.998%	11.827%	0.0 %	1.732	-0.064
20+ year treasuries (TLT)	7.140%	15.017%	0.0 %	5.705	-0.146

Source: Yahoo Finance, *Note*: Data from January 1st 2008 to October 1st 2021

For backtesting purposes, the distribution is of minor importance because the parameters for the MGARCH model can be consistently estimated via quasi maximum likelihood using the Gaussian distribution or using maximum likelihood using the Student's t-distribution as explained in section 3.3.

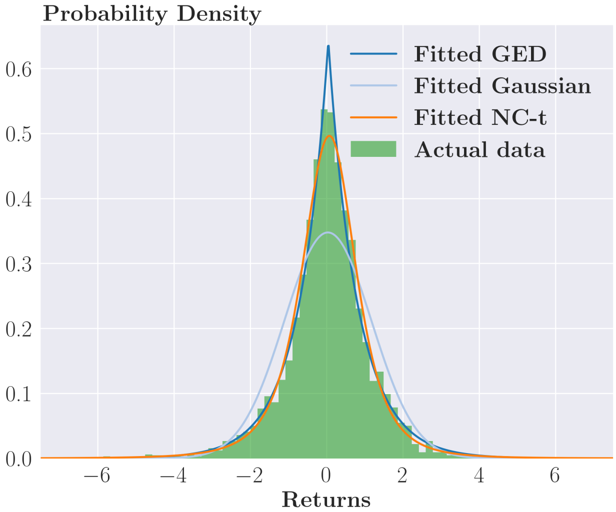
Lastly, we consider the dependency structure provided by a GARCH type model. Recall from the section 3.1 on stylized facts that daily returns are dependent across time; thus, drawing returns IID would be inaccurate. From figure 5 (d)-(f), we see that one can fit GJR-GARCH(1,1) models to the time-series for gold, S&P500, and high-yield bonds such that the dependency structure of the simulated data mimics the historical time series data.

The ARCH(1) and GARCH(1,1) models can also fit the time series of many different asset classes in the same way, so we only display the GJR-GARCH(1,1) for the sake of brevity. Notice how the shocks of the simulated returns are of a similar size and how the shocks of the returns are clustered together, similar to the historical data. Keep in mind the location of the shocks for the simulated data is random as the data is simulated.

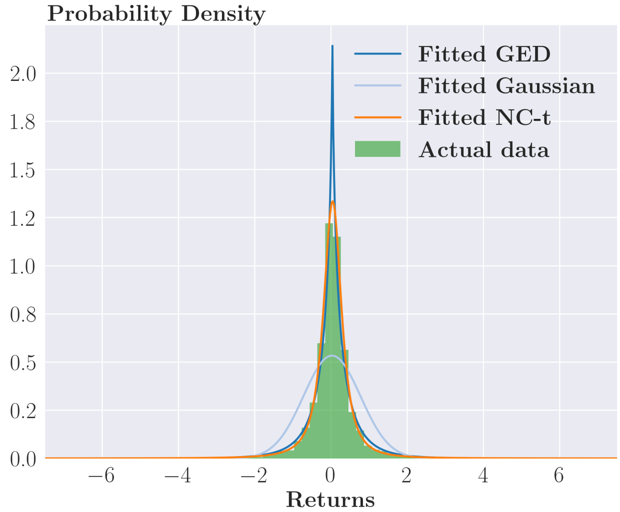
[17] See appendix A.2 for the exact density function for the non-central Student's t-distribution

Figure 5: Asset return densities and time series plots

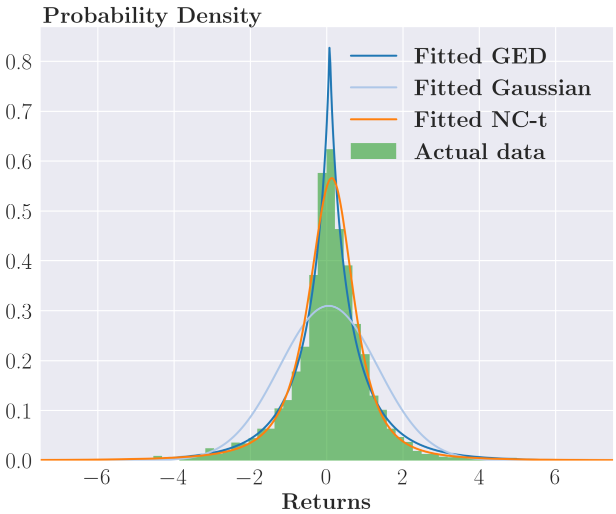
(a) Gold densities



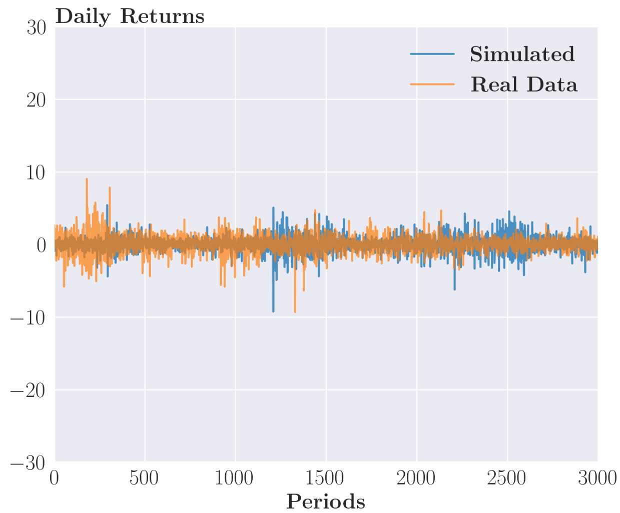
(b) HYG bond densities



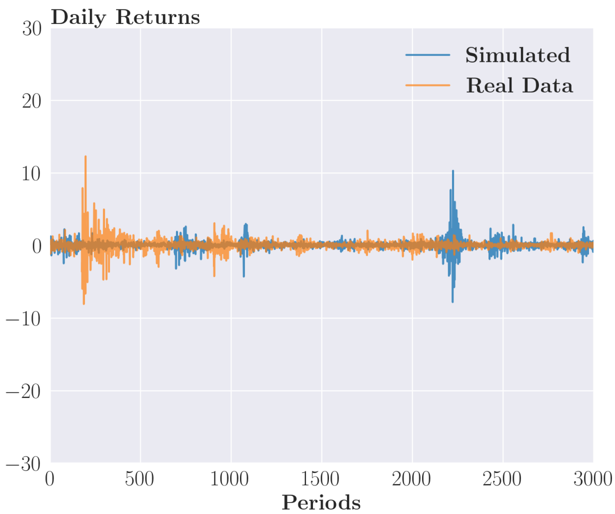
(c) S&P500 densities



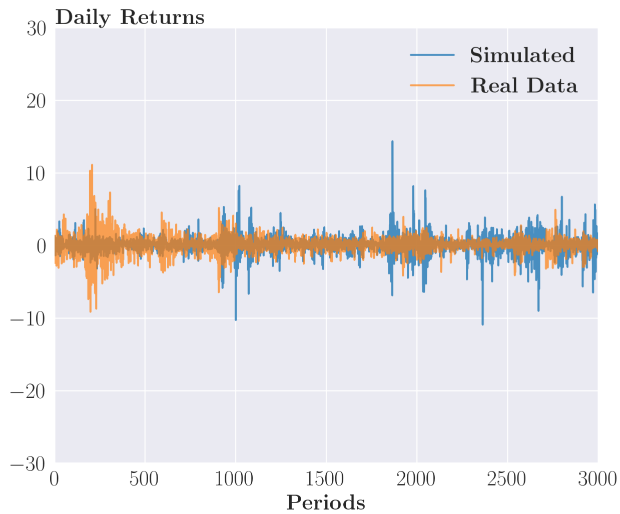
(d) Gold GJR-GARCH(1,1) time-series



(e) HYG bond GJR-GARCH(1,1) time-series



(f) S&P500 GJR-GARCH(1,1) time-series



Note: The model for the simulated data is fitted on the historical data in the plots

Combining the non-central Student's t-distribution with a GARCH type process captures the stylized facts about returns. The returns simulated from the process will be non-Gaussian because they follow the above-mentioned distribution. They will also be uncorrelated as the conditional mean has no correlation structure. Furthermore, they will be time dependent as the conditional variance is correlated across time, again in line with the stylized facts from section 3.1. All of this is to say that our chosen (G)ARCH models are flexible enough to match all statistical properties of the time series in our data sample. In the next section, we test whether this promising result can lead to superior portfolio risk management on historical data.

6 Calibration, model fit, and the Buy-and-hold strategy

In this section, we apply our theoretical results to real-world data to assess the performance and characteristics of the theoretically derived dynamic minimum variance portfolios. We compare our strategies with two benchmark strategies: an Equal-weight strategy holding $1/N$ in each asset with daily rebalancing and a Buy-and-hold strategy with the same initial portfolio as the simple and sophisticated strategies but with no rebalancing. This initial portfolio is the minimum variance portfolio given by the sample covariance using in-sample data with the correlations shrunk 50%, elaborated upon in section 6.1. We consider the cases for which we have theoretical results

- An investor ignoring transaction costs, 'the simple strategy'
- An investor adjusting for transaction costs, 'the sophisticated strategy'

By 'ignoring transaction costs,' we mean whether the investor's objective function penalizes transaction costs or not, corresponding to minimizing either equation (29) or (26).

We test the simple and sophisticated strategies for three univariate GARCH specifications. 'Simple' or 'sophisticated' refers to the type of dynamic investment strategy used, whereas ARCH(1), GARCH(1,1), or GJR-GARCH(1,1) refers to the kind of variance model used to forecast conditional variance in the DCC MGARCH model. A combination of a dynamic strategy and a univariate variance model yields a single investment strategy with a total of six strategies.

For example, the simple ARCH(1) strategy is a dynamic minimum variance portfolio where the investor ignores transaction costs. The investor uses a forecast of the conditional covariance matrix, $\Omega_{t+1|t}$ given by a DCC MGARCH model where the univariate variance models are ARCH(1) models. Similarly, for the sophisticated GARCH(1,1), which is a dynamic minimum variance portfolio where the investor adjusts for transaction costs. The investor uses a forecast of the conditional covariance matrix, $\Omega_{t+1|t}$ given by a DCC MGARCH model where the univariate variance models are GARCH(1,1) models. Notice that the structure of the multivariate model is the same, but the structure of the univariate models changes.

We choose the Equal-weight strategy as one of our two benchmark strategies as it is a simple strategy that often has solid out-of-sample performance, as shown by [DeMiguel et al., 2009]. The daily rebalancing implies that the Equal-weight strategy sells wining assets and buys losing assets, a kind of anti-momentum strategy. We also benchmark against a Buy-and-hold strategy. This strategy starts with the same portfolio as the simple and sophisticated strategies but never rebalance. Therefore, it can show whether the simple and sophisticated strategies add value by rebalancing and incorporating new information.

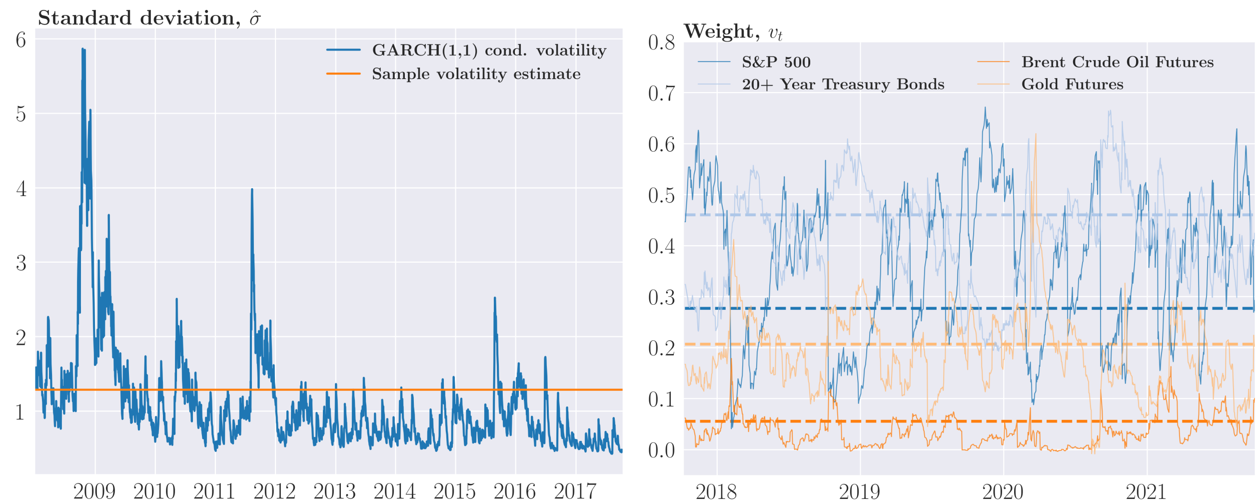
First, we test whether a GARCH estimate of the conditional variance can improve portfolio performance compared to a myopic approach with a static sample covariance estimate. Second, we fit the DCC MGARCH models and calibrate the transaction costs parameter, γ_D.

6.1 Motivating the Buy-and-hold strategy as a benchmark

This section tests if a GARCH model can make a difference in volatility forecasting compared to a myopic approach. Specifically, we compare the conditional covariance forecast from a GARCH(1,1) model to the sample covariance estimate from our out-of-sample dataset. First, an overview of a typical GARCH conditional variance estimate of the S&P 500 compared to the myopic estimate seen in figure 6 panel (a). We only show S&P 500 for the sake of brevity, but the conditional volatility estimates for the other assets are similar in nature.

Figure 6: Comparison of sample covariance estimate vs. GARCH conditional estimate

(a) Volatility estimates σ_t... **(b)** ...resulting in portfolio weights, v_t^{MVP}



Note: In panel(b), the dotted lines are the MVP weights given a sample covariance estimate, and the solid lines are the weights given by the MGARCH covariance estimate.

In panel (a), we observe that both estimates forecast roughly the same average volatility levels. However, the GARCH estimate is dynamic and sensitive to shocks compared to the static estimate. The dynamic nature of the GARCH estimate will later prove to lead to costly rebalancing for the GARCH model strategies. In panel (b), we backtest the two volatility estimation methods using four assets[18] using the strategy from equation (28) with a 50% regularization applied to the correlations.

Table 3: Annualized performance in gross return

Strategy	Std. deviation	Return	Sharpe ratio
GARCH(1,1)	0.0661	0.0537	0.8125
Sample estimate	0.0775	0.0823	1.0615

Unsurprisingly, the fixed covariance estimates result in a portfolio with constant weights (the dashed lines), which we dub the Buy-and-hold strategy. The simple and sophisticated strategies we test in the subsequent sections also have these weights as their initial portfolio. In the following sections, we will use the Buy-and-hold benchmark strategy along with an Equal-weight strategy that holds $1/N$ in each asset for comparison with our dynamic strategies.

Even though this is just one case, we see that at least for these assets and the GARCH(1,1) model, there seem to be significant potential improvements from using a GARCH conditional covariance estimate compared to the Buy-and-hold approach. However, the Buy-and-hold approach results in a strategy with no transaction costs, which could offset the worse performance in gross returns. We will explore this trade-off for more assets and other volatility models in section 7 and 8.

6.2 Calibration and model fit

Before we apply our theoretical results, we calibrate the transaction cost parameter, γ_D, from equation (30), interpreted as the dealer's risk preference. We define a portfolio, ψ, as a time-series of weights, v_t. The daily return $r_{\psi,t}$ of the portfolio is the product of the lagged weights and the asset returns: $r_{\psi,t} = v'_{t-1} r_t$. We lag the weights since the investor rebalances the portfolio at the end of each trading day. The standard deviation of the portfolio returns $\hat{\sigma}_\psi$ is measured as the sample standard deviation:

$$\hat{\sigma}_\psi \equiv \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(r_{\psi,t} - \hat{\mu}_\psi)^2} \tag{34}$$

which is an essential measure when evaluating minimum variance portfolios. With inspiration from [Hautsch and Voigt, 2019], we measure the daily turnover of the portfolio in percentage points as:

$$TO_{\psi,t} = v_t - \frac{v_{t-1} \circ (1 + r_t)}{1 + v'_{t-1} r_t} \tag{35}$$

This specification includes intraday returns of the weights before rebalancing instead of assuming flat intraday returns. Notice that this is only measurable from period $t = 2$.

[18] S&P 500, 20+ Year Treasury Bonds, Brent Crude Oil Futures, and Gold Futures

We calibrate γ_D like [Gârleanu and Pedersen, 2013], even though our returns are measured in relative terms and not in dollar terms. We can convert the dollar-cost back into relative costs with $\Delta x_t = TO_t \circ \bar{P}\psi_v$, where Δx_t is a $N \times 1$ matrix of the number of shares traded, ψ_v is the total portfolio of initially \$1bn and $\bar{P}$ is a $N \times 1$ vector of average asset prices. We do this to use the result of [Robert et al., 2012] who find costs of 10 basis points (bp) of the average price per asset when trading 1.59% of the daily volume. Combining the turnover equation (35) with quadratic transaction costs, we obtain the daily transaction costs in dollar terms:

$$TC_{\psi,t} = \Delta x_t' \Lambda_t \Delta x_t \tag{36}$$

Now, we calibrate γ_D from our specification of transaction costs where $\Lambda_t = \Omega_{t+1|t}\gamma_D$. In the following, we make a simplifying assumption and use the average daily volume of shares traded in the market $\bar{\mathcal{V}}_i$ for asset $i = 1, 2, ...N$.
We know from [Robert et al., 2012] that

$$\frac{\Delta x_i}{\bar{\mathcal{V}}_i} = 1.59\% \Leftrightarrow \Delta x_i = \bar{\mathcal{V}}_i \cdot 1.59\%$$

and

$$TC_i = 10bp \cdot \bar{P}_i$$

Combining these results yields:

$$\begin{aligned} 10bp \cdot \bar{P}_i = \bar{\mathcal{V}}_i \cdot 1.59\% \cdot \Lambda_i = \bar{\mathcal{V}}_i \cdot 1.59\% \cdot \sigma_i^2 \gamma_D \\ \gamma_D = 10bp \cdot \bar{P}_i (1.59\% \cdot \sigma_i^2 \bar{\mathcal{V}}_i)^{-1} \end{aligned}$$

We estimate σ_i using the sample standard deviation of asset returns for the in-sample period. As an example, we calibrate γ_D for IVV (S&P 500) where we estimate the following empirical values:

$$\sigma_{\text{IVV}} = 1.7389 \qquad \bar{\mathcal{V}}_{\text{IVV}} = 4,190,495 \qquad \bar{P}_{\text{IVV}} = 136.82$$

This yields γ_D for IVV:

$$\gamma_D = 1.7389^2 (136.82 \cdot 4190495 \cdot 0.0159) {-1} \cdot 0.001 \cdot 1\text{e}9 = 1.41\text{e}{-5}$$

Repeating this for all assets, we get a mean γ_D of 3.37e−5 and a median of 1.52e−6. To get the relative loss from transaction costs in our backtesting, we divide the dollar transaction costs we have just formulated with the total portfolio value. Algorithm 3 in the appendix explains the procedure in detail.

Moving on to the fit of the MGARCH model, where we estimate the parameters via maximum likelihood. As mentioned in section 3.3, a DCC MGARCH model comprises N univariate GARCH type models, one for each asset, and a single multivariate model for the conditional correlation.

In table 4, we present the estimates of the $N = 11$ univariate GARCH(1,1) models. The estimates all lie within the expected range with β around 0.8-0.9, α around 0.1, and $\alpha + \beta < 1$

Table 4: Estimates of a DCC MGARCH(1,1) with univariate GARCH(1,1) - t_ν errorterms

	Univariate GARCH				
Asset	μ	ω	α	β	ν
Emerging Markets (EEM)	0.062 (0.019)	0.038 (0.01)	0.102 (0.014)	0.883 (0.015)	9.458 (1.262)
S&P 500 (IVV)	0.100 (0.011)	0.020 (0.004)	0.169 (0.018)	0.830 (0.016)	5.286 (0.452)
Europe (IEV)	0.068 (0.015)	0.021 (0.006)	0.121 (0.017)	0.876 (0.016)	5.697 (0.511)
Global Tech (IXN)	0.127 (0.015)	0.029 (0.007)	0.122 (0.013)	0.868 (0.013)	5.900 (0.555)
Real estate (IYR)	0.086 (0.014)	0.016 (0.004)	0.119 (0.017)	0.875 (0.016)	7.812 (0.904)
Global financials (IXG)	0.081 (0.015)	0.026 (0.007)	0.125 (0.017)	0.869 (0.016)	5.996 (0.555)
Global Industrials (EXI)	0.083 (0.013)	0.018 (0.005)	0.122 (0.016)	0.873 (0.016)	6.508 (0.656)
Gold (GC=F)	0.040 (0.014)	0.005 (0.001)	0.033 (0.002)	0.963 (0.010)	4.206 (0.306)
Brent crude oil (BZ=F)	0.057 (0.025)	0.041 (0.015)	0.082 (0.012)	0.915 (0.012)	4.599 (0.389)
High-yield bonds (HYG)	0.036 (0.004)	0.003 (0.010)	0.168 (0.021)	0.830 (0.020)	4.776 (0.341)
20+ year treasuries (TLT)	0.026 (0.013)	0.015 (0.006)	0.066 (0.014)	0.915 (0.019)	14.94 (3.582)
	Multivariate GARCH				
	a	b	ν		
Scalar-BEKK(1,1)	0.0156 (0.003)	0.973 (0.003)	9.465 (0.349)		

Note: Estimated via MLE using data from January 1st 2008 to October 11th 2017. Robust standard errors in $(\cdot)$.

such that the parameters are estimated consistently. Additionally, we see that $\alpha + \beta \simeq 1$, meaning that the GARCH models are very persistent and close to being integrated. Similar to table 2, there is a wide range of estimates of ν for the Student's t-distribution. The higher the ν, the closer the asset is to a Gaussian distribution with 20+ year treasuries (TLT) being the closest though still quite far from Gaussian. The least Gaussian asset returns are the commodities and High yield bonds (HYG).

Estimates for the ARCH(1) and GJR-GARCH(1,1) univariate models are in table 9 and 10 in the appendix.

7 Simple strategy (ignoring transaction costs)

We start with the simple strategy where the investor ignores transaction costs in his objective function corresponding to equation (26). First, we delve into the dynamics of the model to understand how shocks impact this equation and, thereby, the weights. Then, we show performance before (gross return) and after subtracting transaction costs (net return) to highlight the importance of adjusting for transaction costs.

7.1 Dynamics of the simple strategy

As explained in section 3, the GARCH type models capture the persistent variances and covariances of asset returns. For an investor seeking to minimize portfolio variance, the effects of past shocks and conditional (co)variances on future (co)variances are vital to understanding how these affect the optimal weights.

To illustrate the dynamics of the simple strategy, we model three of the assets for 10.000 periods from $t = -5000$ to $t = 5000$ and give a single shock of -2% to S&P500 ETF at $t = 0$, and plot how the weights of the different assets change and converge back to their long-run equilibrium. We model all correlations with a scalar-BEKK(1,1) from equation (12) and (17)-(21) and only vary the variance model for the individual asset returns. Keep in mind that this is a very artificial setup with only one shock in 10.000 periods which does not replicate real time-series.

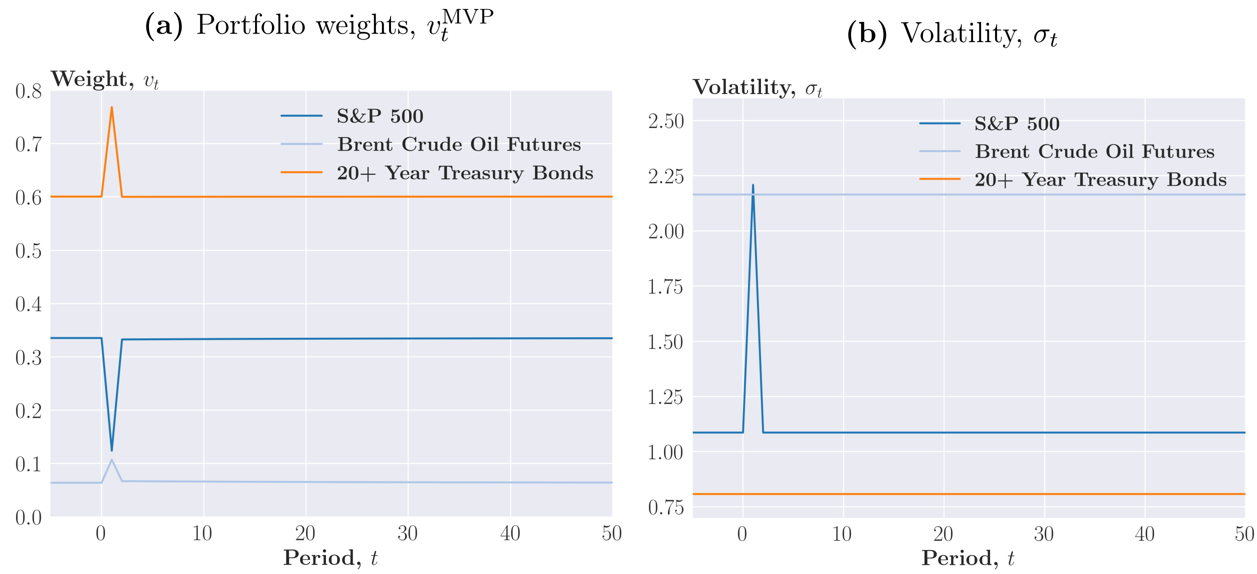
The weights, that the simple strategies (and sophisticated strategies) diverge from and converge back to after a shock hits, are given by the minimum conditional covariance matrix, which is a combination of the minimum conditional variance, σ^2_{MIN}, from equation (9) and the minimum conditional correlations. The constant variance term, ω, has a significant effect on σ^2_{MIN}, profoundly affecting the weights. High-yield bonds have the lowest minimum conditional variance, with Brent crude oil futures having the highest. We see the result in the IRF plots, where the investor prefers High-yield bonds the most and Brent crude oil the least. This is most profound for the simple ARCH(1), where the weights continuously oscillate around the weights from the minimum conditional covariance estimate.

For the simple dynamic strategy, there are two primary dynamic effects to consider: how the conditional variance responds to shocks and how persistent the impact of the shock is for the conditional variance. How the conditional variance responds to a shock is dictated by α and also κ for the GJR-GARCH model. The higher α and κ is, the more the conditional variance of the assets increases in the period following a shock. β models the persistence of the shock for the conditional variance - the closer β is to one, the more persistent the variance process is, and by extension, the slower the weights converge back.

Consider the simple ARCH(1) in figure 7 where we plot the optimal weights and the conditional volatility estimates.

In figure 7 panel (b), the volatility of the S&P500 ETF spikes from 1.1 to 2.2 from period $t = 0$ to $t = 1$ as the shock of -2 hits. Then in period $t = 2$, the volatility returns to 1.1 as the shock exits the model. Thus, we see the lack of volatility persistence from the ARCH(1) model. Looking at the weight of the S&P500 ETF, the allocation drops at period $t = 1$ because the investor now perceives the S&P500 as riskier than treasuries and crude

Figure 7: Impulse response functions of v_t^{MVP} for the simple ARCH(1) strategy



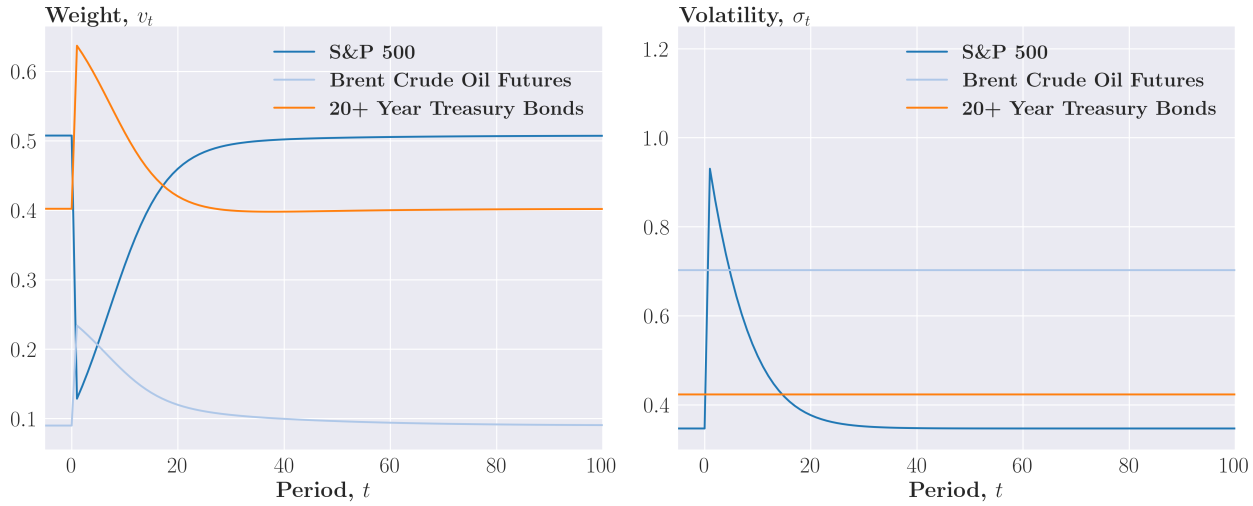
oil. Still, at period $t = 2$, the allocation goes back very close to the pre-shock value as the shock is now out of the model, but the model still shows some persistence as the weights fully converge after around 100 periods. This persistence, although barely visible, comes from the conditional correlations between the asset returns, which the shock also affects at period $t = 1$. The changes in the conditional correlations are minimal as $a \approx 0.01$ and very persistent as $b \approx 0.96$ in the scalar-BEKK model for the conditional correlations. Recall from section 3.2, that the ARCH(1) converged back too quickly to its unconditional variance compared to empirical estimates, which also shows in the weights, which converge back very soon after the shock. So, the volatility estimates are by far the most important factor for the investor's asset allocation in the simple ARCH(1) strategy, with the correlations accounting for a tiny fraction. Recall from section 3.2, that the ARCH(1) converged back quicker than empirical estimates indicated, which was fixed mainly by the GARCH(1,1) model from equation 8. Will the simple GARCH(1,1) weight dynamics show persistence as the conditional variance in figure 2?

In figure 8, we see the conditional volatility and optimal weights for the simple GARCH(1,1) strategy after an identical shock of -2%. Here we see a different dynamic to the simple ARCH(1) strategy. In figure 8 panel (b), the volatility of the S&P500 ETF increases much less, courtesy of lower α estimates. However, the volatility is now persistent and decays exponentially to the minimum conditional variance after roughly 30 periods at $t = 30$. The sticky volatility profoundly affects the weights where we observe a 'two-stage' convergence. In the first stage, both the conditional correlations and conditional variances converge back, roughly from period $t = 1$ to $t = 30$. At this point, the conditional variance has fully converged, and the remaining dynamics are due to the conditional correlations, which converge very slowly. This effect is visible in panel (a), where the weight of the S&P500 ETF converges quickly in period $t = 1$ to $t = 30$ and then very slowly for the remaining periods.

In conclusion, using a GARCH(1,1) for the conditional variance adds more persistence to

Figure 8: Impulse response functions of v_t^{MVP} for the simple GARCH(1,1) strategy

(a) Portfolio weights, v_t^{MVP} **(b)** Volatility, σ_t

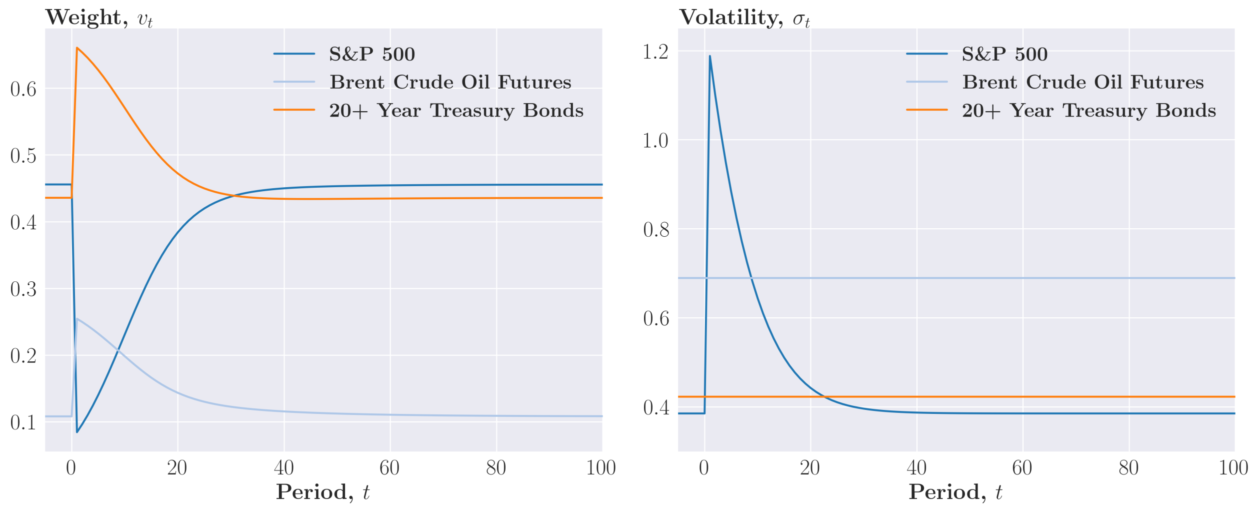


the weights. Recall the "leverage effect" from section 3.2, where negative shocks, parameterized with $\alpha + \kappa$, generally have a more significant effect on the variance than positive shocks, α. How will the model that captures the leverage effect impact the weight dynamics?

The univariate GJR-GARCH(1,1) captures the leverage effect with the α parameter to positive shocks and $\alpha + \kappa$ parameters to negative shocks. In figure 9, we plot the conditional variance and optimal weights for the simple GJR-GARCH(1,1). In panel (b), the volatility of

Figure 9: Impulse response functions of v_t^{MVP} for the simple GJR-GARCH(1,1) strategy

(a) Portfolio weights, v_t^{MVP} **(b)** Volatility, σ_t



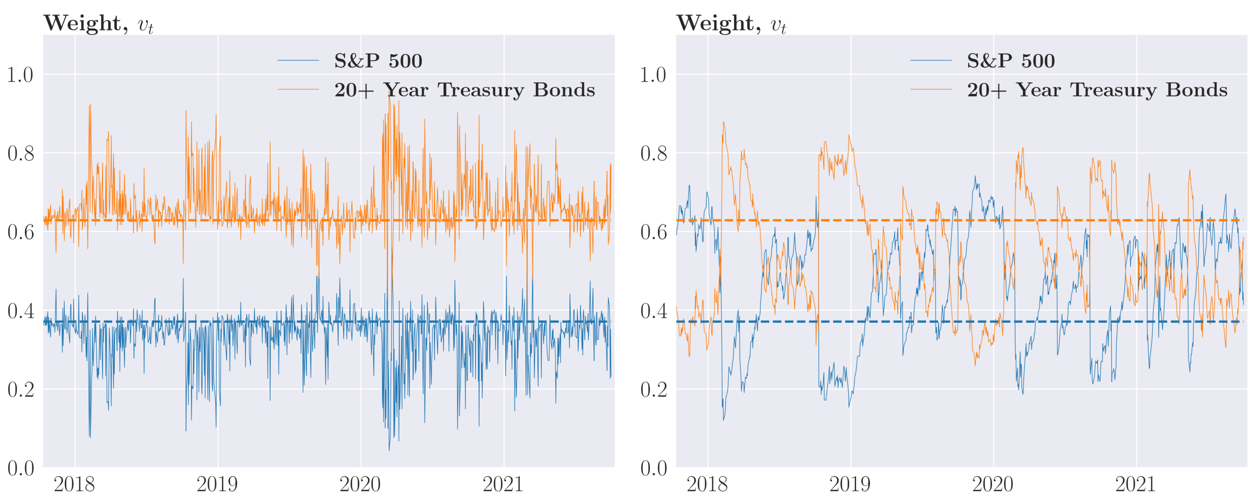
the S&P500 ETF increases to 1.2 in period $t = 1$, which is larger than the GARCH(1,1) model increase of 0.9, thanks to the leverage effect. The higher volatility results in a slightly larger weight adjustment of 1%-point for the simple GJR-GARCH(1,1) than the simple GARCH(1,1). Thus, there is a non-linear decreasing relationship between variance and weights. Overall, the simple GJR-GARCH(1,1) strategy exhibits a similar 'two-stage' convergence to the simple GARCH(1,1) strategy, resulting in identical weight dynamics in panel (a). The differences in the long-run weights is due to different minimum conditional variances of the two models. In the GJR-GARCH models, estimates of α are often around 0 and estimates of κ around 0.2. Therefore, positive shocks do not affect the conditional variance and weights by extension. We see the lack of response to positive shocks in figure 24 in the appendix, where we give the S&P500 ETF a positive shock of 2%. The shock does not increase the conditional variance and thus results in negligible effects on the weights, with the only dynamics coming from the conditional correlations. Therefore, figure 24, shows the pure effects from the conditional correlations, which are tiny compared to the impact of the conditional covariance.

To summarize the IRF plots, the conditional covariance and the weights exhibit dynamics that strongly depend on the univariate model used. Keep in mind that the conditional covariance never fully converges back to the minimum conditional covariance in the real world as the asset returns are hit by shocks of varying sizes daily. However, the overall dynamics that the investor allocates less capital to assets perceived as riskier immediately after a shock hits and then slowly changes the allocation back as the volatility subsides will remain moving to real-world data.

Another feature of the simple and sophisticated strategies is that portfolio weights oscillate around the estimated unconditional sample covariance weights. To illustrate this, we plot the weights for two assets from the unconditional covariance (the dotted lines) and the weights from the (G)ARCH model's conditional covariance (the solid lines).

Figure 10: Portfolio weights from the unconditional and conditional covariance

(a) Simple ARCH(1) **(b)** Simple GARCH(1,1)



Note: The dotted lines are the weights from the unconditional covariance

Looking at figure 10, we see that the conditional weights roughly oscillate around the unconditional weights. The effect is evident in the simple ARCH(1) strategy, where the weights jump back to the unconditional value when a shock ends. The oscillation is more persistent for the simple GARCH(1,1) and the simple GJR-GARCH(1,1) strategy, and the conditional weights have extended periods below and above their unconditional counterpart.

With a thorough understanding of the dynamics of the model, we move from the artificial setup to historical backtesting to examine the performance of the simple strategy with real-world data, answering the question: "What would the approximate performance of our strategies have been during the last four years?"

7.2 Backtesting the simple strategy

We test the dynamic trading strategies using historical prices by fitting the models on a sample period and testing them out-of-sample. Our out-of-sample period is 1.000 days and begins October 12th 2017 and ends October 2nd 2021. The model is thus only fitted with data from before this period to avoid polluting the optimization process with "future" data. We calculate the optimal portfolio weights for each period using the simple strategies from section 4.2.1 where the investor ignores transaction costs, i.e., an investor with the objective function of equation (26). We assume that rebalancing is feasible at closing prices. See algorithm 1 in the appendix for a detailed explanation of the backtesting procedure. We measure performance before (gross return) and after subtracting transaction costs (net returns).

7.2.1 Backtesting the simple strategy with all assets

We start by backtesting the simple strategy for all 11 assets. First, we look at the minimum variance portfolio weights of the simple ARCH(1) strategy in figure 11 panel (a).

Historical dynamic portfolio weights

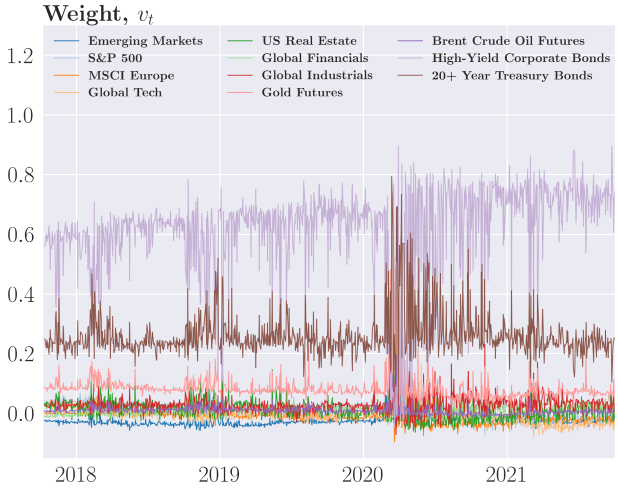
Here, we see that the weights fluctuate heavily around the weights given by the unconditional covariance estimate, as shown in section 7.1. Recall from the section on dynamics that the ARCH(1) shows little to no persistence in the weights, which is present for all assets but easily spotted for 20+ year government bonds and High-yield bonds. Thus, every time a shock hits, the investor rebalances depending on the size of the shock, which will later turn out to be very costly in transaction costs.

Moving on to the simple GARCH(1,1) strategy, we plot the minimum variance portfolio weights in figure 11 panel (b). In contrast to the weights of the simple ARCH(1) strategy in panel (a), the weights given by the simple GARCH(1,1) strategy are more persistent and slowly converge to back as explained in section 7.1 rather than jump back as in the simple ARCH(1) strategy. The main difference between the IRF plots and this is the number of shocks that hit the model in the real world such that the weights never fully converge back.

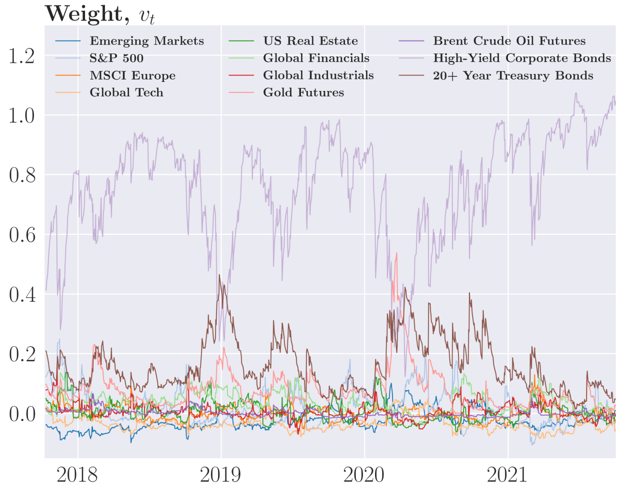
The weights of the GJR-GARCH(1,1) strategy shown in figure 26 in the appendix are similar to those given by the simple GARCH(1,1) strategy. The main difference between the two is that the drops in the weights given by the simple GJR-GARCH(1,1) strategy are slightly larger than the simple GARCH(1,1) strategy caused by the asymmetric responses to shocks

Figure 11: Portfolio weights of the simple strategy, all assets

(a) ARCH(1) weights, v_t^{MVP}



(b) GARCH(1,1) weights, v_t^{MVP}



in the GJR-GARCH(1,1). An important detail is that the extra adjustment of the weights for the simple GJR-GARCH(1,1) strategy compared to the simple GARCH(1,1) is the most costly as transaction costs are quadratic.

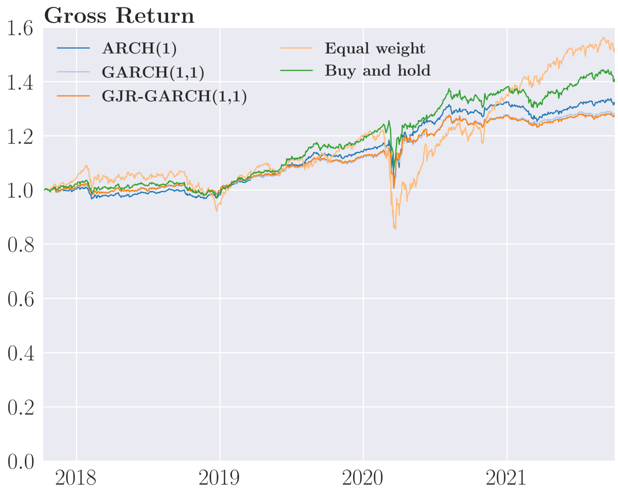
For the simple strategies, it is evident that significant portfolio changes happened during the market crash caused by the COVID-19 pandemic of March and April 2020, especially for the ARCH(1). Additionally, the bond ETFs unsurprisingly dominate the portfolio regardless of the choice of variance model because the bond ETFs have the lowest estimated volatility of all assets, which is attractive for minimum variance investors.

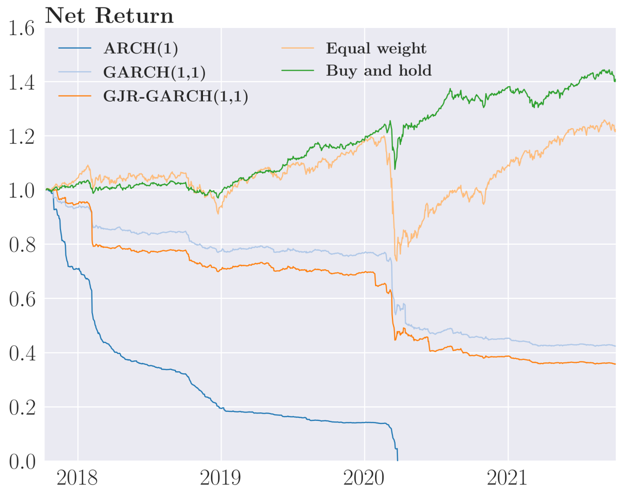
Historical performance

Recall that we are interested in minimum variance portfolios, which makes the standard deviation a vital performance measure. Still, we also report the average return and risk-adjusted return in the form of Sharpe ratios because a low-risk strategy that loses all its money is not desirable.

Looking at gross returns in the top of table 5, the simple GARCH(1,1) has the lowest annualized standard deviation of 6.3%, closely followed by the GJR-GARCH(1,1) with a standard deviation of 6.5%. Even when the variance is modeled by the very simple ARCH(1,1), the annualized standard deviation is 7.8%. Compare this to the Equal-weight and Buy-and-hold with an annualized standard deviation of 15.7% and 7.8%. Thus, the simple strategies have lower or similar volatility than the benchmark strategies before transaction costs. Henceforth, we refer to every annualized measure without the "annualized" prefix and write volatility instead of standard deviation. In terms of the Sharpe ratio, the Buy-and-hold strategy has the highest Sharpe ratio of 1.15, closely followed by the simple GARCH(1,1) strategy with a Sharpe ratio of 1.03 and the Equal-weight as the worst-performing with a Sharpe ratio of 0.71. The simple strategies perform well in gross returns as they beat the Equal-weight and

Figure 12: Cumulative returns of the simple strategy, all assets





perform similar to the Buy-and-hold. But can the simple strategies replicate this success in net returns given the high trading volume present in figure 11?

The short answer: No. After transaction costs, the simple strategies unsurprisingly do far worse. Especially the simple ARCH(1) strategy as its trading volume is extremely high, as seen in both figure 11 panel(a) and table 5. The simple ARCH(1) loses 100% of its portfolio value to transaction costs and becomes worthless after 2.5 years. Transaction costs cause drops in portfolio value which has an unfortunate knock-on effect on portfolio volatility. With quadratic transaction costs, the rebalancing of the simple ARCH(1) strategy is costly. It greatly hurts performance measured in net returns where it earns -100% as the investor goes bankrupt with a volatility of 68.7%.[19]

In contrast to the simple ARCH(1) strategy, the simple GARCH(1,1) strategy does not suffer as much from transaction costs. The persistent covariance estimates of the simple GARCH(1,1) imply less trading, so it annually loses 24.1% of its portfolio value to transaction costs compared to the 100% of the simple ARCH(1) strategy. The lower transaction costs improve net returns to -19.3% and decrease volatility to 11%, which is still atrocious performance in absolute terms. The main reason the simple GARCH(1,1) strategy performs better than the simple ARCH(1) strategy is its dynamics. Firstly, the simple GARCH(1,1) strategy reacts less to shocks courtesy of lower α estimates than the simple ARCH(1) strategy. Secondly, the simple GARCH(1,1) converges back exponentially after a shock leaves the model. The simple ARCH(1,1) strategy converges back immediately. The two effects make the simple ARCH(1) strategy perform a violent adjustment when a shock hits and an often equally significant adjustment back as the shock leaves. The simple GARCH(1,1) strategy makes one moderate

[19]Note that the interpretation of the Sharpe ratio becomes rather difficult with negative returns. The simple ARCH(1) strategy has a better Sharpe ratio than the other GARCH strategies even though the simple ARCH(1) performs worse in all aspects, with much lower net returns and higher volatility.

Table 5: Annualized performance of the simple strategy, all assets

Strategy	Std. deviation	Return	Sharpe ratio	Transaction costs
	Before transaction costs			
ARCH(1)	0.0788	0.0724	0.9186	
GARCH(1,1)	0.0628	0.0644	1.0246	
GJR-GARCH(1,1)	0.0650	0.0626	0.9632	
Equal-weight	0.1565	0.1105	0.7062	
Buy-and-hold	0.0780	0.0894	1.1462	
	After transaction costs			
ARCH(1)	0.6869	-1.0000	-1.4559	100.00%
GARCH(1,1)	0.1098	-0.1928	-1.7553	24.144%
GJR-GARCH(1,1)	0.1227	-0.2263	-1.8442	27.137%
Equal-weight	0.1626	0.0517	0.3182	5.2374%
Buy-and-hold	0.0780	0.0894	1.1462	0%

Note: Transaction costs are the relative annual share of the portfolio lost to transaction costs

adjustment when the shock hits and several minor adjustments afterward. The latter is far less costly and explains part of the difference in net return, the other reason being the more accurate volatility predictions of the simple GARCH(1,1).

The simple GJR-GARCH(1,1) strategy is very similar to the simple GARCH(1,1) strategy but performs a bit worse. The strategy earns a net return of -22.6% with a volatility of 12.3%. The poor performance is likely due to asymmetry of the GJR-GARCH(1,1) variance model such that when a negative shock hits the simple GJR-GARCH(1,1) adjusts slightly more than the simple GARCH(1,1) strategy. Since the model have quadratic transaction costs, a slight extra adjustment is the most expensive. This is evident from table 5 where the simple GJR-GARCH(1,1) strategy annually loses 27.1% of its portfolio value to transaction costs, which is three percentage points more than the simple GARCH(1,1) strategy, which adds up over four years.

The Buy-and-hold strategy is unaffected by transaction costs as it does not rebalance and thus takes the lead with the lowest volatility after transaction costs. The Equal-weight strategy is only moderately affected by transaction costs as it only needs to make minor adjustments every period. It earns net returns of 5.2% and has a slightly higher volatility of 16.3%.

To summarize, the simple strategy can reduce portfolio risk when measuring gross returns compared to the Buy-and-hold and Equal-weight strategies. Gross Sharpe ratios are higher than the Equal-weight strategy but slightly lower than the Buy-and-hold strategy. However, after transaction costs, the simple strategy performs poorly as costly rebalancing chisels away at the value of the portfolio.

The weights in figure 11 show that the minimum variance portfolios heavily favor non-equity ETFs. In gross return, the GARCH strategy is successful at reducing risk. Is this due to the simple strategies favoring bonds, or does it also apply when restricting it to only equity ETFs?

7.2.2 Backtesting the simple strategy with stocks only

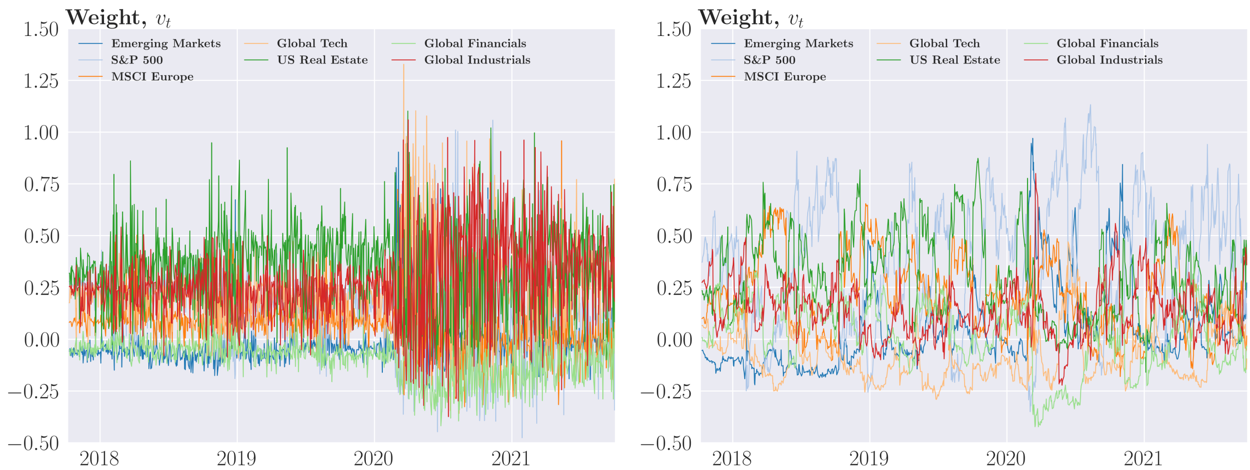
From figure 11 panel (a)-(b), it is evident that the simple strategies mainly consist of 20+ years treasuries and high-yield corporate bonds. But can the forecast of Ω_{t+1}, $\Omega_{t+1|t}$, help reduce portfolio risk when the investor does not have access to safe assets with low correlations to market risk? To answer this, we restrict the asset universe to stock ETFs, leaving out commodities and bonds.

Historical dynamic portfolio weights

From figure 13 panel (a)-(b),[20] we see that the weight dynamics of the simple strategies do not change from the case with all assets. However, a greater span in portfolio weights from -50%

Figure 13: Portfolio weights of the simple strategy, stocks

(a) ARCH(1) weights, v_t^{MVP} **(b)** GARCH(1,1) weights, v_t^{MVP}



to 125% indicates that the optimization takes on more extreme positions to obtain the lowest possible risk. Consider the S&P 500 index, which changes from weights of around 125% to as low as -25% for the simple ARCH(1) strategy. After the Covid-19 pandemic market crash of March and April 2020, we see even more significant weight fluctuations in the simple strategies. However, it is most visible for the simple ARCH(1) strategy in panel(a).

Historical performance

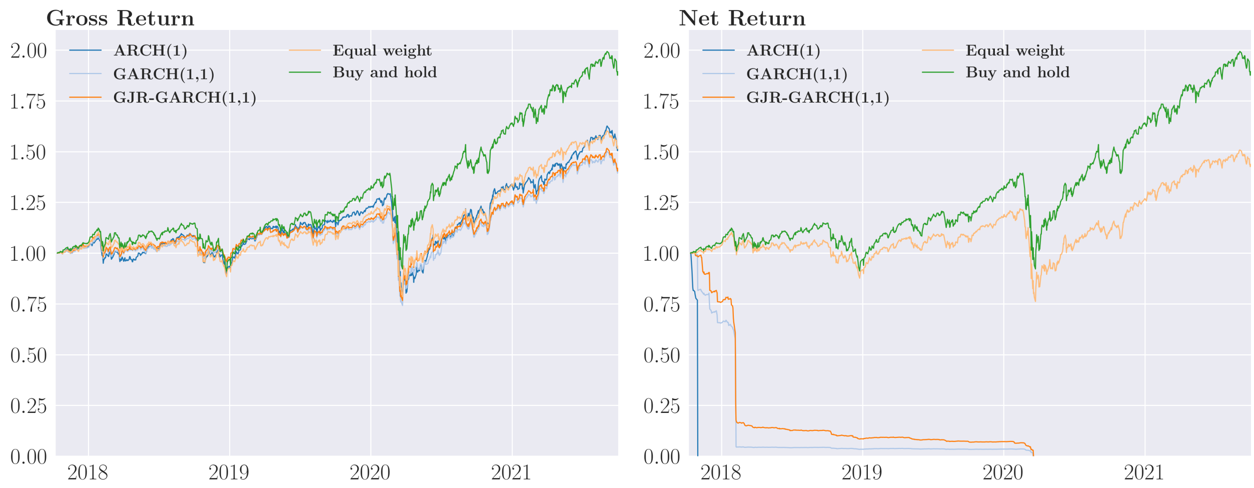
The significant, frequent changes are very costly, evident from the differences between gross returns in panel (a) and net returns in panel (b) of figure 14. It takes 14 days for the simple ARCH(1) strategy to lose 100% of the portfolio value almost exclusively from transaction

[20] with the weight of the simple GJR-GARCH(1,1) in the appendix in figure 26 panel (b)

costs. For the simple GARCH(1,1) and GJR-GARCH(1,1) strategies, it takes roughly two years for both strategies to lose everything because of transaction costs, with both of them annually losing 84%.

Figure 14: Cumulative performance of the simple strategy, stocks

(a) Gross return **(b)** Net return



The impact of transaction costs can also be seen in table 6, where the simple ARCH(1) strategy earns a gross return of 11%. In net terms, this translates to an abysmal return of -100%. The simple GARCH(1,1) and GJR-GARCH(1,1) earn gross returns of 8.9% and 9.1%, respectively, and a net return of -100%. The failure of these strategies undoubtedly shows why investors should adjust for transaction costs for these strategies to be viable.

To answer whether the simple strategy can reduce risk when only considering equity ETFs, the simple GARCH(1,1) and GJR-GARCH(1,1) strategies can reduce risk - even in this restricted asset universe. The simple GARCH(1,1) strategy has the lowest volatility of 18.7%, closely followed by the simple GJR-GARCH(1,1) strategy of 18.9% compared to 22% for the Buy-and-hold and 20.8% for the Equal-weight strategy. The picture is very different when looking at net returns, where the performance of the simple strategies suffers heavily from very high transaction costs. Additionally, the Equal-weight and Buy-and-hold strategies obtain better Sharpe Ratios in both gross and net returns.

The main takeaway is that it is vital to adjust for transaction costs when setting up an investment strategy. As seen in the performance table above, a good strategy before transaction costs can be a downright terrible strategy after transaction costs if the portfolio changes are sufficiently large. Thus, can we improve the performance of our dynamic strategies after transaction costs by making the investor anticipate and adjust for transaction costs when rebalancing?

Table 6: Annualized performance of the simple strategy, stocks

Strategy	Std. deviation	Return	Sharpe ratio	Transaction costs
	Before transaction costs			
ARCH(1)	0.2082	0.1105	0.5306	
GARCH(1,1)	0.1876	0.0892	0.4755	
GJR-GARCH(1,1)	0.1898	0.0912	0.4808	
Equal-weight	0.2082	0.1117	0.5366	
Buy-and-hold	0.2197	0.1732	0.7882	
	After transaction costs			
ARCH(1)	4.3101	-1.0000	-0.2320	100.00%
GARCH(1,1)	0.9179	-1.0000	-1.0894	83.949%
GJR-GARCH(1,1)	0.8117	-1.0000	-1.2320	83.779%
Equal-weight	0.2083	0.0951	0.4565	1.4989%
Buy-and-hold	0.2197	0.1732	0.7882	0%

Note: Transaction costs are the relative annual share of the portfolio lost to transaction costs

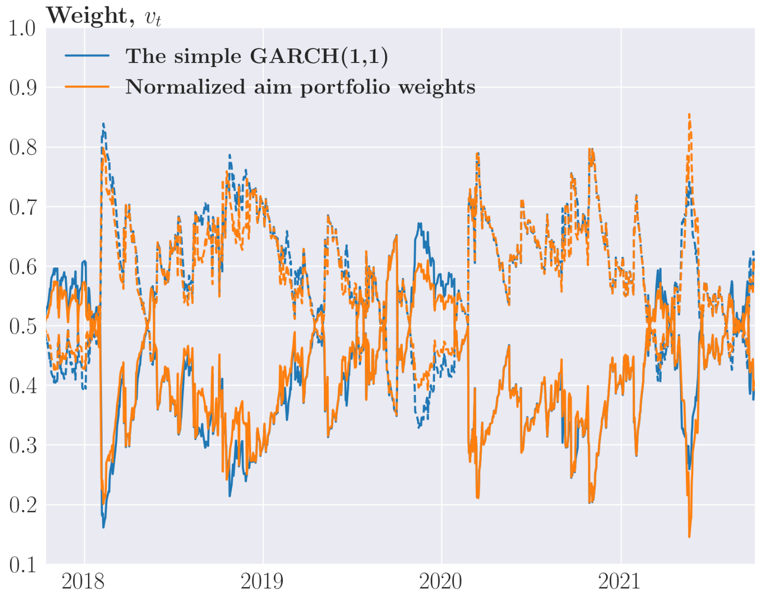
8 Sophisticated strategy (adjusting for transaction costs)

We continue with the more complex case where the investor adjusts for transaction costs (the sophisticated strategy). Specifically, the investor now adjusts for transaction costs in his objective function (see equation 29) by being penalized for portfolio changes. We also explore how transaction costs affect the dynamics of the model and, thereby, the optimal weights v_t.

8.1 The simple strategy and the normalized aim portfolio

First, we address the point in section 4.2.2 about the normalized aim portfolio being almost equivalent to the simple strategy from equation (28), i.e., a time-varying minimum variance portfolio. As an example, we plot the weights for two random assets of both portfolios over an out-of-sample period of 1000 days in figure 15. We observe very similar weights in both portfolios, except for some noise partly due to instability when normalizing the weights of the aim portfolio when the aim portfolio nearly sums to 0. The similarity is promising for our otherwise incomplete theoretical results. The workaround is not ideal, but we have to leave it as a topic for further research and move on. However, it is comforting that the normalized aim portfolio closely resembles the simple strategy because it is the minimum variance portfolio that ignores transaction costs. So, we interpret the sophisticated strategy as having the same target as the simple strategy but with an additional element to optimally account for transaction costs.

Figure 15: Similarities of the normalized aim portfolio and the simple strategy



Note: The two assets are S&P500 (solid line) and 20+ year Treasuries (dotted line)

8.2 The dynamics of the sophisticated strategy

Similar to the simple strategy, we illustrate the dynamics of the sophisticated strategy by modeling three assets for 10.000 periods from $t = -5000$ to $t = 5000$. We give a single shock of -2 to S&P500 ETF at $t = 0$ and plot how the weights of the different assets change and converge back to their long-run equilibrium. How does penalizing the investor for trading change the dynamics of the strategy?

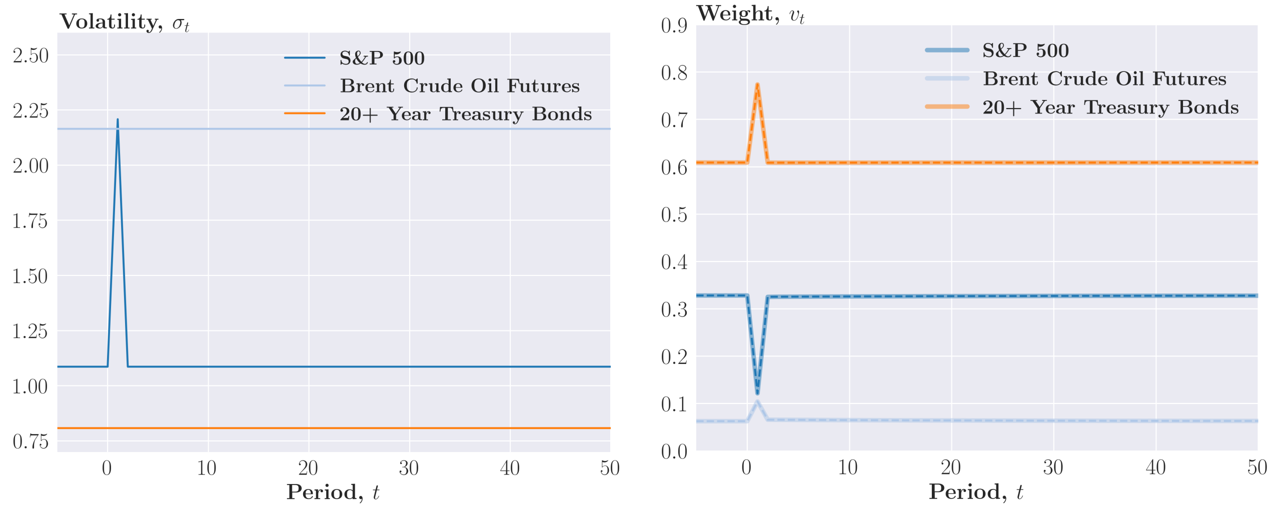
Consider the sophisticated ARCH(1) strategy with $\gamma_D = 1.52\text{e}{-6}$ which is the empirically calibrated level. Figure 16 panel (a) shows that the volatility spikes to an identical level as the simple strategy one period after the shock hits. The remaining dynamics of the sophisticated strategies depend on two factors: Firstly, the increased volatility results in adjustments to the investor's aim portfolio, which strongly resembles the simple strategy in 7.1. Secondly, the investor's actual adjustment amount depends negatively on transaction costs.

In figure 16 panel (b), we consider the sophisticated ARCH(1) strategy with $\gamma_D = 1.52\text{e}{-6}$. At this level of ex-ante transaction costs, the investor almost completely adjusts his portfolio (solid line) to the aim portfolio (dotted line). We interpret this as the investor believing that the future risk reduction from changing his portfolio outweigh the transaction costs associated with rebalancing. Thus, the dynamics are very similar to the simple strategy in section (7.1). Unfortunately, it seems like our sophisticated strategy does not make a noticeable difference to the weight dynamics. How can we alleviate this problem?

We implement an idea similar in spirit to [Hautsch and Voigt, 2019] where we increase the parameter γ_D above the empirically calibrated value to punish portfolio changes more. Note that this does not increase the actual cost of transactions, but the investor thinks it

Figure 16: Impulse response functions of v_t^{MVP} for the sophisticated ARCH(1)

(a) Volatility, σ_t **(b)** Portfolio weights, v_t^{MVP}, $\gamma_D = 1.52\text{e}{-6}$



Note: The aim portfolio is the dotted lines, the optimal portfolio weights are the solid lines

does ex-ante. We implement transaction costs tuning by modifying equation (33) which now becomes:

$$v_t = \frac{v_{t-1} + (\tilde{\gamma}_D \Omega_{t+1|t})^{-1} A_{vv} \left[v_{t-1} - \frac{\text{aim}_t}{\mathbf{1}'\text{aim}_t} \right]}{\mathbf{1}'(v_{t-1} + (\tilde{\gamma}_D \Omega_{t+1|t})^{-1} A_{vv} \left[v_{t-1} - \frac{\text{aim}_t}{\mathbf{1}'\text{aim}_t} \right])}$$

Where $\tilde{\gamma}_D$ is the parameter we tune. Intuitively, the bigger $\tilde{\gamma}_D$ is, the less the investor trades as transaction costs increase. The interpretation is that the investor perceives transaction costs as higher than what he actually incurs. The convergence towards the aim portfolio happens quickly for the calibrated value of γ_D. We only show the sophisticated ARCH(1) strategy, but the convergence is quick in all the sophisticated strategies. The fast convergence turns out to be very costly with quadratic transaction costs where large portfolio changes are punished heavily. Increasing $\tilde{\gamma}_D$ increases the penalty of trading such that

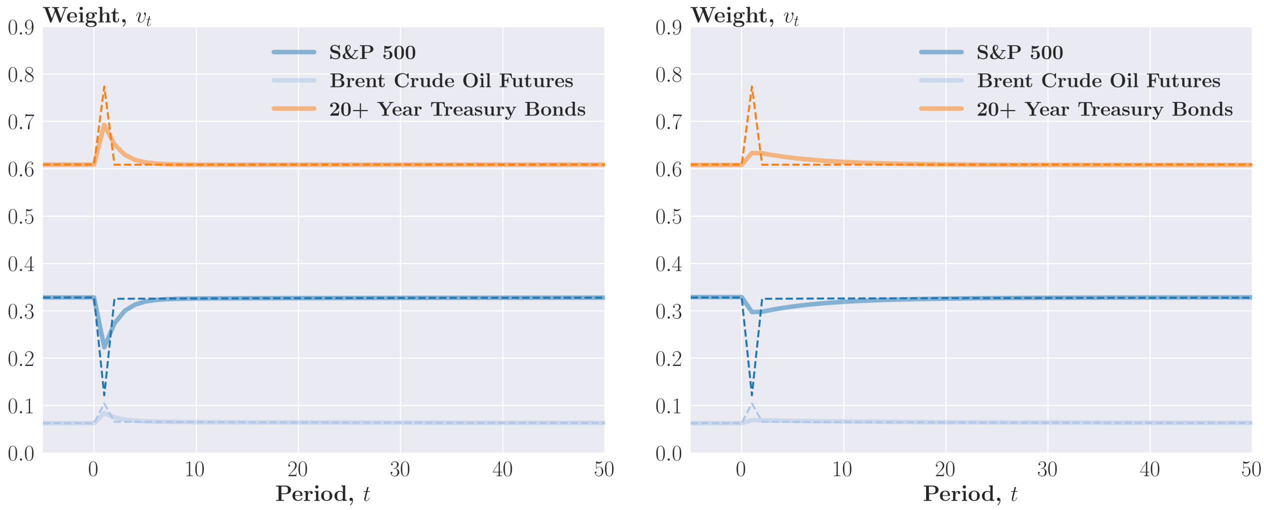
$$\lim_{\tilde{\gamma}_D \to +\infty} v_t = \frac{v_{t-1}}{\mathbf{1}' v_{t-1}}$$

the investor will not rebalance. To illustrate the effects of tuning, we plot the dynamics for the sophisticated strategies for $\tilde{\gamma_D} \in \{1.52\text{e}{-6}, 3\text{e}{-6}, 1\text{e}{-5}\}$, where $1.52\text{e}{-6}$ is the empirically calibrated value. Increasing the ex-ante transaction costs highlights the investor's trade-off between lowering transaction costs and obtaining a portfolio with expected lower risk.

Consider ex-ante transaction costs with $\tilde{\gamma}_D = 3\text{e}{-6}$, roughly double the empirical estimate. Figure 17 panel (a) shows that the investor adjusts halfway towards the aim portfolio. After the shock subsides, the aim portfolio immediately returns to the pre-shock weights given the dynamics of the simple ARCH(1) strategy. However, the investor's portfolio only slowly returns to the pre-shock weight, as changing the portfolio is now perceived as costlier, with the convergence to the aim portfolio taking roughly five periods.

Figure 17: Impulse response functions of v_t^{MVP} for the sophisticated ARCH(1)

(a) Portfolio weights, v_t^{MVP}, $\tilde{\gamma}_d = 3\text{e}{-6}$ **(b)** Portfolio weights, v_t^{MVP}, $\tilde{\gamma}_d = 1\text{e}{-5}$



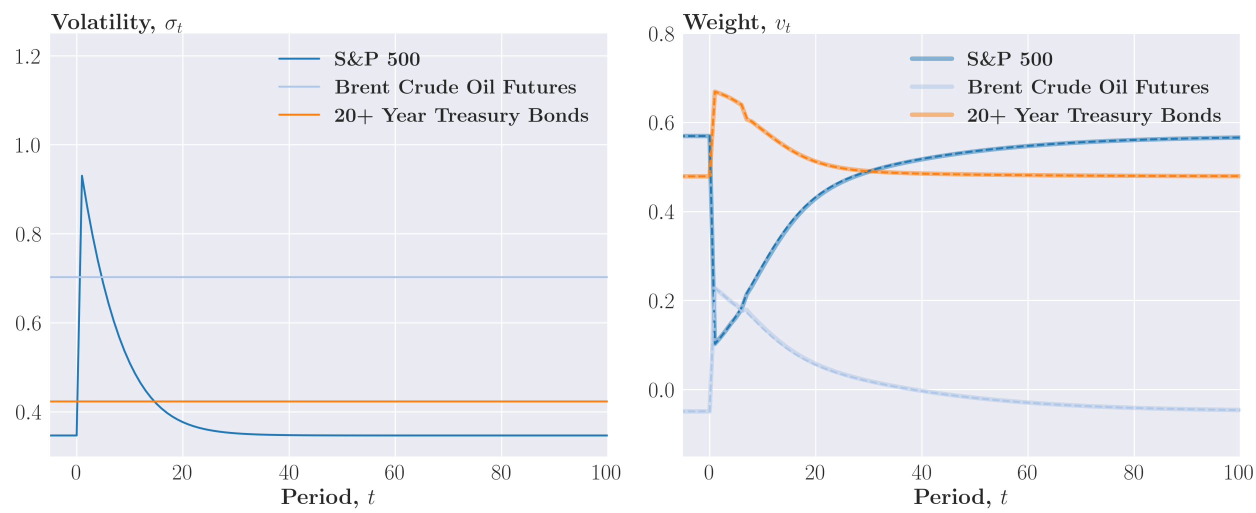
Note: The aim portfolio is the dotted lines, the optimal portfolio weights are the solid lines

In figure 17 panel (b), we see that further increasing $\tilde{\gamma}_D$ to 1e−5, results in even smaller adjustment steps towards the aim portfolio and an even slower adjustment back to aim portfolio after the shock subsides. Despite the small reaction to the shock, the investor's portfolio takes roughly 15 periods to converge to the aim portfolio. Note that the aim portfolio is identical across different values of $\tilde{\gamma}_D$, only the adjustment of the investor's actual portfolio changes. In summary, tuning $\tilde{\gamma}_D$ seems to fix the quick convergence of the empirically calibrated γ_D

We plot the impulse response functions of the GARCH(1,1) case in figure 18. Like the

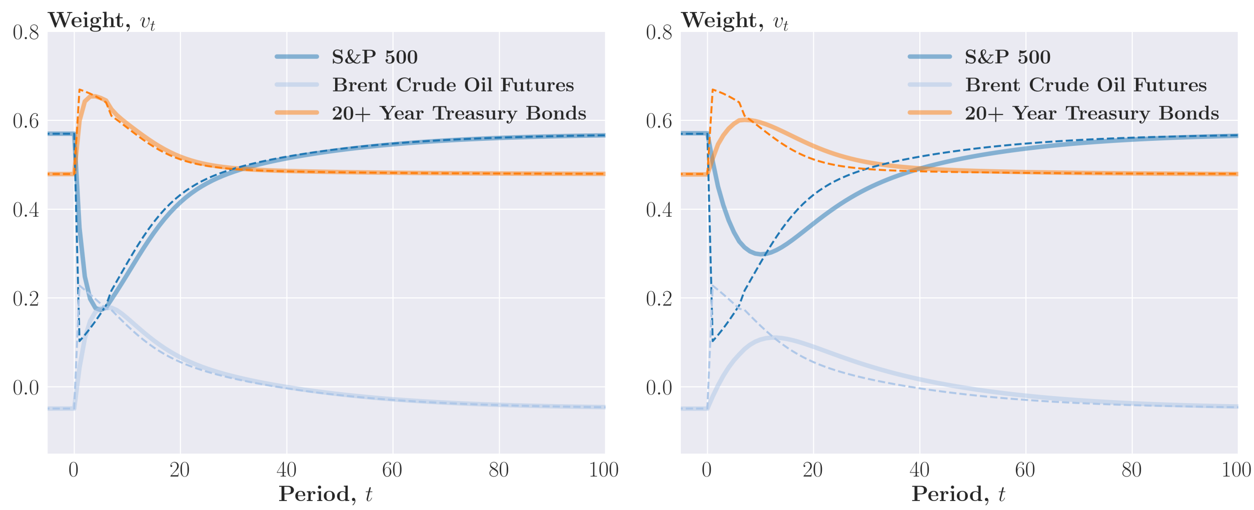
Figure 18: Impulse response functions of v_t^{MVP} for the sophisticated GARCH(1,1)

(a) Volatility, σ_t **(b)** Portfolio weights, v_t^{MVP}, $\tilde{\gamma}_d = 1.52\text{e}{-6}$



simple strategy, the conditional volatility plotted in panel (a) spikes and then exponentially

(c) Portfolio weights, v_t^{MVP}, $\tilde{\gamma}_d = 3\text{e}{-6}$ **(d)** Portfolio weights, v_t^{MVP}, $\tilde{\gamma}_d = 1\text{e}{-5}$



Note: The aim portfolio is the dotted lines, the optimal portfolio weights are the solid lines

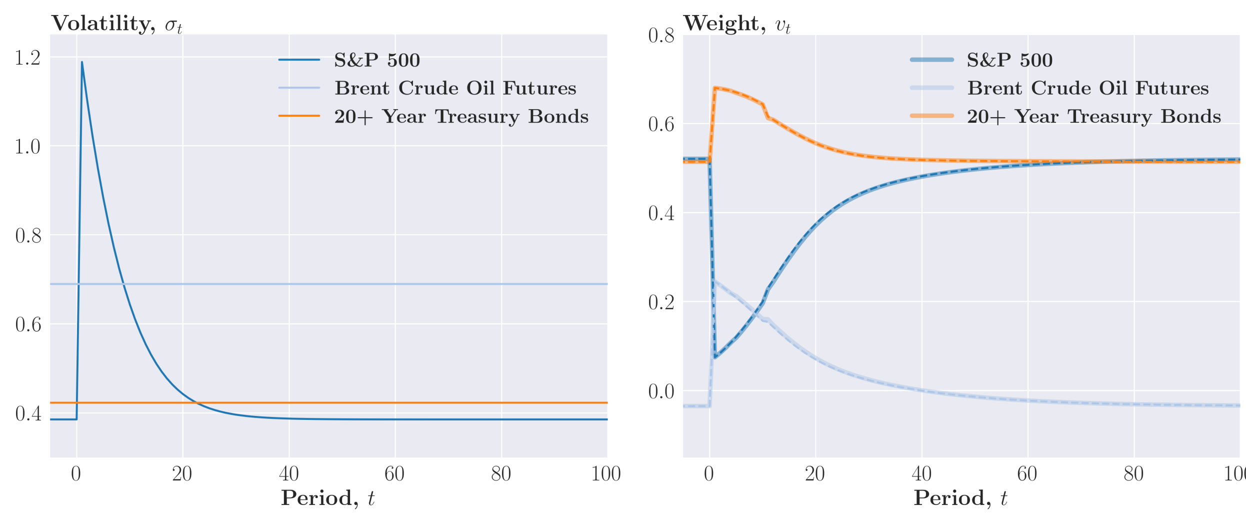
converges to the equilibrium value. In panel (b) where $\tilde{\gamma}_D = 1.52\text{e}{-6}$, we see the investor immediately adjusts to the aim portfolio, behaving almost identically to the simple strategy.

In panels (c) and (d), we plot the cases with increased ex-ante transaction costs. The investor's portfolio first undershoots and then overshoots the aim portfolio. This is most clear in the case where $\tilde{\gamma}_D = 1\text{e}{-5}$, where the aim portfolio spikes one period after the shock, which the investor slowly adjusts to as trading is costly. Then, as the volatility converges back to equilibrium, the aim portfolio converges faster than the weights, which adjust rather slowly and are thus overtaken by the aim portfolio.

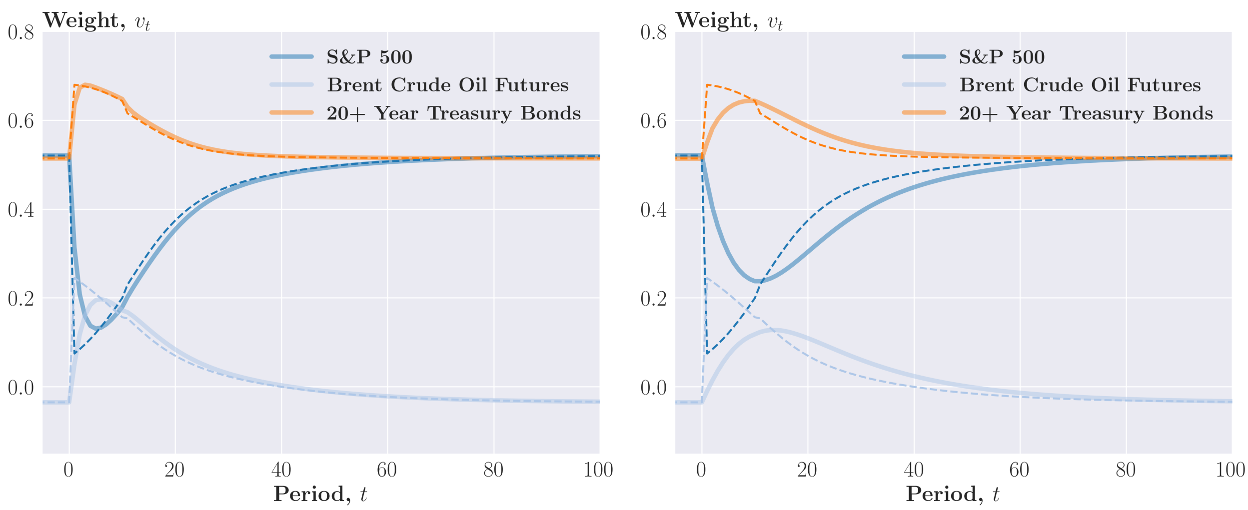
The sophisticated GJR-GARCH(1,1) in figure 19 is very similar to the GARCH(1,1) case but with a more significant effect from the shock given the asymmetry of the GJR-

Figure 19: Impulse response functions of v_t^{MVP} for the sophisticated GJR-GARCH(1,1)

(a) Volatility, σ_t **(b)** Portfolio weights, v_t^{MVP}, $\tilde{\gamma}_d = 1.52\text{e}{-6}$



(c) Portfolio weights, v_t^{MVP}, $\tilde{\gamma}_d = 3\text{e}{-6}$ **(d)** Portfolio weights, v_t^{MVP}, $\tilde{\gamma}_d = 1\text{e}{-5}$



Note: The aim portfolio is the dotted lines, the optimal portfolio weights are the solid lines

GARCH(1,1).Thus, we see the investor react more for the GARCH(1,1) when $\tilde{\gamma}_D$ is equal to 3e−6 and 1e−5. As transaction costs are quadratic, the extra adjustment is very costly for the investor.

Across volatility models, we see trading volume decreases as $\tilde{\gamma}_D$ increase. This is both unsurprising as the investor perceives trading as more costly and is similar to the result of [Gârleanu and Pedersen, 2013]'s proposition 2.

As mentioned before, the forecast of Ω_{t+1} is not necessarily accurate. Thus, it might be beneficial to consider backtesting for a higher $\tilde{\gamma}_D$ than the empirically calibrated $\gamma_D = 1.52\text{e}{-6}$ to limit extreme and costly portfolio movements. However, limiting rebalancing too much prevents the investor from adjusting his portfolio when new information becomes available. Thus, a trade-off exists between trading a lot to gain lower expected future portfolio risk and transaction costs. To find the optimal parameter value, we plot the volatility and Sharpe ratio for different $\tilde{\gamma}_D$ for the sophisticated strategies and the benchmark strategies.

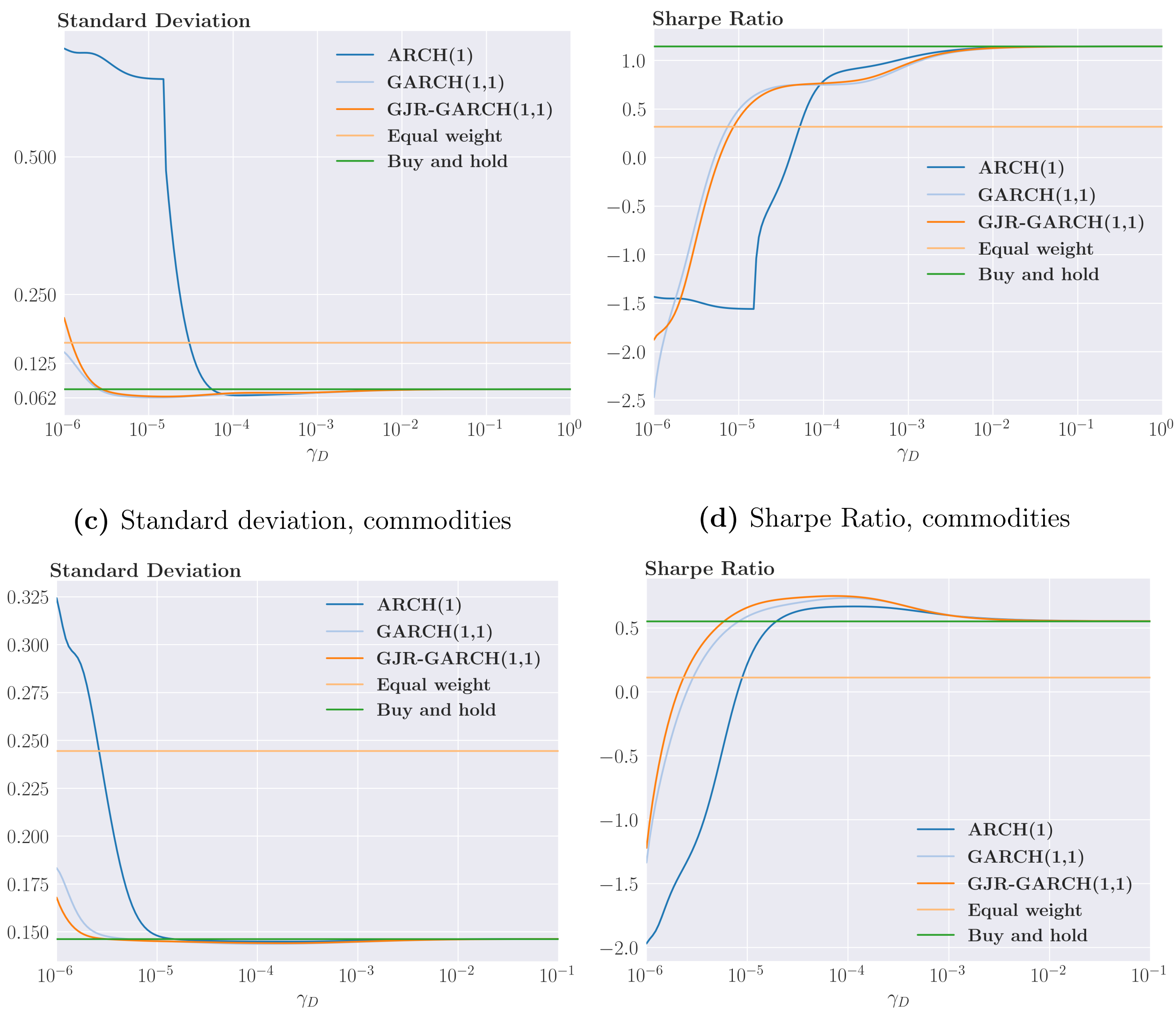
Looking at figure 20 panel (a) and (b), we see that for a portfolio of all 11 assets, the minimum volatility is reached for a $\tilde{\gamma}_D$ of around 1.3e−5 with the Sharpe ratio converging to the Buy-and-hold portfolio as γ increases. Thus, a clear trade-off between lowering risk and increasing the Sharpe ratio is evident for the case with all 11 assets. The sophisticated ARCH(1) still performs poorly for this $\tilde{\gamma}_D$ so we also consider 1e−4. The sophisticated ARCH(1) reaches its minimum volatility at this value. $\tilde{\gamma}_D$ = 1e−4 also yields a higher Sharpe ratio for the other sophisticated strategies.

Suppose the investor's asset universe is limited to the two commodities, which we plot in figure 20 panel (c) and (d). In that case, we obtain a higher Sharpe ratio and lower volatility than the Buy-and-hold and Equal-weight strategy with a single optimal $\tilde{\gamma}_D$ of around 1e−4.[21]

Unsurprisingly, as $\tilde{\gamma}_D$ increases, the performance of all the sophisticated strategies con-

[21] The subset using all stocks and all bonds are in the appendix 25

Figure 20: Annualized performance for different values of $\tilde{\gamma}_D$ for GARCH(1,1)



(a) Standard deviation, all assets

(b) Sharpe Ratio, all assets

(c) Standard deviation, commodities

(d) Sharpe Ratio, commodities

verges towards the performance of the Buy-and-hold portfolio as trading decreases when $\tilde{\gamma}_D$ increases. We will use the optimal parameter values in the following backtesting section.

8.3 Backtesting the sophisticated strategy

We proceed with historical backtesting for an investor who minimizes portfolio risk and is penalized for trading corresponding to an investor with an objective function given by equation (29). The exact backtesting process can be seen in algorithm 2 in the appendix. As mentioned in 8.2 section above, different values of ex-ante transaction costs may greatly benefit the performance of the investor's portfolio. Therefore, we backtest for multiple values of $\tilde{\gamma}_D$.

8.3.1 Backtesting the sophisticated strategy with all assets

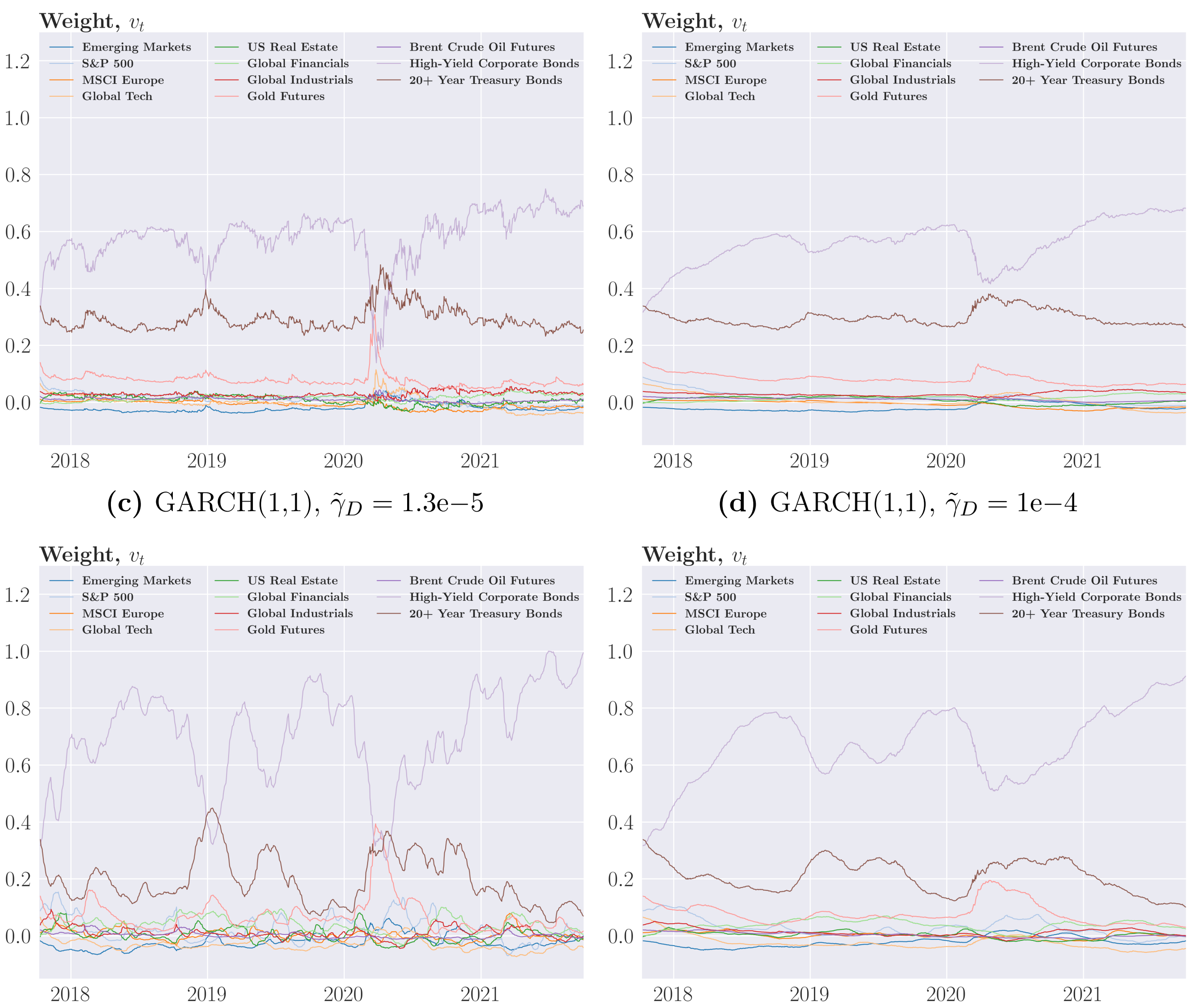
We consider the same 11 assets as for the simple strategy. We will first consider the effects of adjusting for transaction costs on the real-world weights.

Historical dynamic portfolio weights

As for the sophisticated strategy, we first plot the portfolio weights for each variance model for $\tilde{\gamma}_D$ equal to 1.52e−6 (Calibrated value), 1.3e−5 (Lowest volatility) and 1e−4 (Higher Sharpe ratio). The weights for the case where $\tilde{\gamma}_D = 1.52\text{e}{-6}$ are almost identical to the case where the investor ignores transaction costs, and thus, we have placed these plots in the appendix.[22]

First, consider the sophisticated ARCH(1) strategy in figure 21 panel (a)+(b). Recall from section 7.2 that the optimal weights fluctuated heavily in the simple ARCH(1). In figure 21

Figure 21: Portfolio weights of the sophisticated strategy, all assets



(a) ARCH(1), $\tilde{\gamma}_D = 1.3\text{e}{-5}$

(b) ARCH(1),$\tilde{\gamma}_D = 1\text{e}{-4}$

(c) GARCH(1,1), $\tilde{\gamma}_D = 1.3\text{e}{-5}$

(d) GARCH(1,1), $\tilde{\gamma}_D = 1\text{e}{-4}$

Note: The weights for the sophisticated GJR-GARCH(1,1) are in figure 28 in the appendix

panel (a)+(b) with $\tilde{\gamma}_D$ of 1.3e−5 and 1e−4, the weights are clearly more persistent, and any changes to the weights happen over a far greater period than in the simple strategy. The main differences between $\tilde{\gamma}_D$ of 1.3e−5 and 1e−4 is that the changes of the weights are smaller for

[22]See figure 27

$\tilde{\gamma}_D = 1e{-}4$ and that the small fluctuations present in the case with 1.3e−5 are almost gone for $\tilde{\gamma}_D = 1e{-}4$ as the weights here are closer to a smooth line. The reason for this smoothing is that the higher ex-ante transaction costs act as a dampener on the fluctuating aim portfolio as explained in section 8.2. This can be interpreted as the investor adjusting more for transaction costs and that large adjustments require multiple shocks over an extended period like in March and April 2020.

The optimal weights from the GARCH(1,1) model are depicted in figure 21 panel (c) and (d). Compared to the simple ARCH(1) portfolio, the portfolio changes become smoother and slower, thus, adding an extra layer of persistence. The sophisticated GARCH(1,1) is analogous to the sophisticated GJR-GARCH(1,1) in figure 28 and has almost identical dynamics and very similar weights. The portfolio weights of the three different models become more and more identical as $\tilde{\gamma}_D$ increases as they all converge to the Buy-and-hold strategy as trading decreases.

Performance measures

We now evaluate the performance measurements for the sophisticated strategies. First, we see that the gross return in figure 22 panel (a) is almost identical to the simple strategy in figure 12 panel (a). Therefore, the sophisticated GARCH strategies have a very similar performance to the simple strategies in terms of gross return. For example, the sophisticated GARCH(1,1) has gross returns of 6.5% and volatility of 6.3% compared to 6.4%, and volatility of 6.3% in the simple GARCH(1, 1).

After transaction costs, the sophisticated GARCH strategies only perform mildly better than their simple counterparts. Using the empirically calibrated γ_D only causes a minor dampening of the investor's trading activity and thus only slightly improves the performance, which ideally shouldn't be the case. Recall that the simple GARCH(1,1) strategy earns a net return of -19.3% at 11% volatility, compared to the sophisticated version, which yields a net return of -18.4% at 11% volatility, a rather insignificant improvement. A similar pattern holds for the GJR-GARCH(1,1) strategy. The performance of the sophisticated strategies is far worse than the Equal-weight and Buy-and-hold. Thus, the empirically calibrated γ does not significantly improve performance for the sophisticated strategies. This is further backed by the minimal drop in transaction costs in table 7 where the sophisticated ARCH(1) strategy loses 100% annually, identical to the simple ARCH(1). The sophisticated GARCH(1,1) strategy and the sophisticated GJR-GARCH(1,1) strategy loses 23.5% and 26.1% respectively, which is noticeably lower than their simple counterparts but still bad in absolute terms.[23]

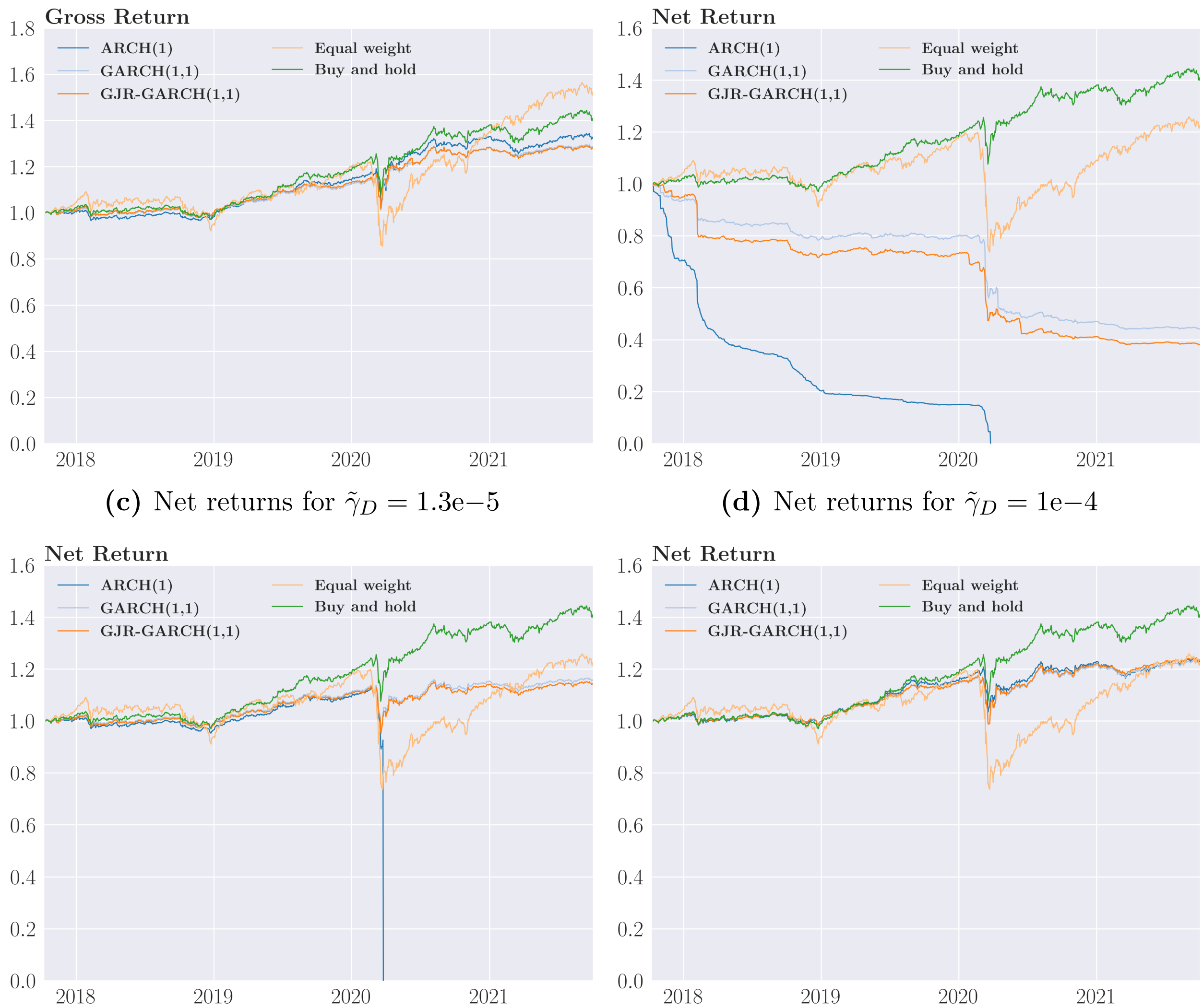
This naturally begs the question as to why this penalization does not improve performance? The primary reason is probably forecast errors as the investor believes that the changes he

[23]We acknowledge the absurdity of having a multiyear strategy with annualized transaction costs of 100%. But sometimes, the ARCH(1) makes extreme changes to its portfolio such that the transactions costs exceed the portfolio value. This can cause annual transaction costs of +100%. We cap the cost at 100%, assuming the investor cannot lose more than everything.

Figure 22: Cumulative performance of the sophisticated strategy, all assets

(a) Gross returns **(b)** Net returns for $\tilde{\gamma}_D = 1.52\text{e}{-6}$

(c) Net returns for $\tilde{\gamma}_D = 1.3\text{e}{-5}$ **(d)** Net returns for $\tilde{\gamma}_D = 1\text{e}{-4}$



makes to his portfolio will reduce the risk of the portfolio in the next period to such a degree that the significant transaction costs the investor incurs are made up for in lower future expected risk. But as $\Omega_{t+1|t}$ is a forecast and not the true covariance, the changes to the weights might end up not being as beneficial as the investor believes. Thus, further reducing the trading volume may be beneficial as it reduces transaction costs and limits the investor's extreme positions. This can be interpreted as the investor only changing the portfolio when deemed absolutely necessary.

Increasing the ex-ante transaction costs $\tilde{\gamma}_D$ to 1.3e−5 causes the investor to reduce trading volume, and by extension, the transaction costs the investor incurs. This drastically improves the performance of the sophisticated GARCH(1,1) and sophisticated GJR-GARCH(1,1) strategies. In table 7, the sophisticated ARCH(1) strategy does not improve as annualized transaction costs of 100% still wipe it out. In contrast, the sophisticated GARCH(1,1) and GJR-GARCH(1,1) strategies perform admirably, earning a net return of 3.8% at 3.4% with

Table 7: Annualized performance of the sophisticated strategy, all assets

Strategy	Std. deviation	Return	Sharpe ratio	Transaction costs
	Before transaction costs			
ARCH(1)	0.0794	0.0735	0.9257	
GARCH(1,1)	0.0632	0.0651	1.0299	
GJR-GARCH(1,1)	0.0654	0.0640	0.9788	
Equal-weight	0.1565	0.1105	0.7062	
Buy-and-hold	0.0780	0.0894	1.1462	
	After transaction costs $\tilde{\gamma}_D = \gamma_D = 1.52\text{e}{-6}$			
ARCH(1)	0.6890	-1.0000	-1.4514	100.00%
GARCH(1,1)	0.1114	-0.1846	-1.6569	23.553%
GJR-GARCH(1,1)	0.1257	-0.2136	-1.7000	26.174%
Equal-weight	0.1626	0.0517	0.3182	5.0201%
Buy-and-hold	0.0780	0.0894	1.1462	0%
	After transaction costs $\tilde{\gamma}_D = 1.3\text{e}{-5}$			
ARCH(1)	0.6415	-1.0000	-1.5588	100.00%
GARCH(1,1)	0.0630	0.0375	0.5943	2.0706%
GJR-GARCH(1,1)	0.0648	0.0343	0.5291	2.4152%
	After transaction costs $\tilde{\gamma}_D = 1\text{e}{-4}$			
ARCH(1)	0.0670	0.0529	0.7890	0.1038%
GARCH(1,1)	0.0696	0.0522	0.7507	0.01630%
GJR-GARCH(1,1)	0.0709	0.0543	0.7660	0.01902%

Note: Transaction costs are the relative annual share of the portfolio lost to transaction costs. The actual transaction costs the investor pays are identical for each $\tilde{\gamma}$.

volatility at 6.3% and 6.5%, respectively. In this case, the GARCH(1,1) and GJR-GARCH(1,1) have lower volatility than the Equal-weight and Buy-and-hold but at a lower net return with the Buy-and-hold yields the highest Sharpe ratio of 1.14. The increased ex-ante transaction costs cause lower trading volume for the sophisticated GARCH(1,1) and GJR-GARCH(1,1) strategies and thus lower transaction costs of 2.1% and 2.4% annually, which is a drastic reduction. The lower transaction costs boost net returns and lower volatility compared to the empirically calibrated value. The sophisticated ARCH(1) strategy still performs poorly for $\tilde{\gamma}_D = 1.3\text{e}{-5}$. However, for the sophisticated GARCH(1,1) and GJR-GARCH(1,1) strategies, $\tilde{\gamma}_D = 1.3\text{e}{-5}$ is the optimal $\tilde{\gamma}_D$ in terms of reducing volatility, i.e., the optimal trade-off between rebalancing in the face of new information and transaction costs from implementing these changes.

The investor still loses a moderate sum in transaction costs, especially the sophisticated ARCH(1), so we further increase the ex-ante transaction costs to $\tilde{\gamma}_D = 1\text{e}{-4}$. Figure 22 panel(d) shows that all the sophisticated strategies perform very similarly and closely follow each other throughout the four years. This is due to the considerable ex-ante transaction costs such that the differences between the covariance predictions of the models are significantly reduced, which results in the sophisticated ARCH(1), GARCH(1, 1), and GJR-GARCH(1,1) strategies only having annual transaction costs of 1%, 0.2%, and 0.2% respectively. The consequence is that the sophisticated strategies have nearly identical performance with Sharpe ratios of around 0.76. This thoroughly beats the Equal-weight strategies in both volatility and net return. The Buy-and-hold strategy is also beat in terms of volatility, but it has a higher Sharpe Ratio than the sophisticated strategies. The reason for the very similar performance is as $\tilde{\gamma}_D \to \infty$ the sophisticated strategies converge to a Buy-and-hold. Thus, the different covariance forecasts of the models affect the weights less as the rebalancing from these forecasts is seen as much more costly.

In summary, for the empirically calibrated γ_D, the sophisticated strategies do not sufficiently adjust for transaction costs resulting in negative net returns or bankruptcy for the sophisticated ARCH(1). We see significant performance improvements when the ex-ante transaction costs increase, especially for lowering volatility. There is a clear trade-off between decreasing risk and increasing net returns as $\tilde{\gamma}_D = 1\text{e}{-4}$ has the highest net return of the three cases but with higher volatility than $\tilde{\gamma}_D = 1.3\text{e}{-5}$. Either the investor can trade more to reduce risk but at the cost of higher transaction costs or trade less to get higher net return and volatility with lower transaction costs.

8.3.2 Backtesting the sophisticated strategy with commodities only

As shown in figure 20, the sophisticated strategies outperform the Buy-and-hold and Equal-weight when only considering the commodities gold and oil. Additionally, gold and oil are also used by [Gârleanu and Pedersen, 2013], and since our strategies build on their work, it is interesting to look into the same assets.

Historical weight dynamics

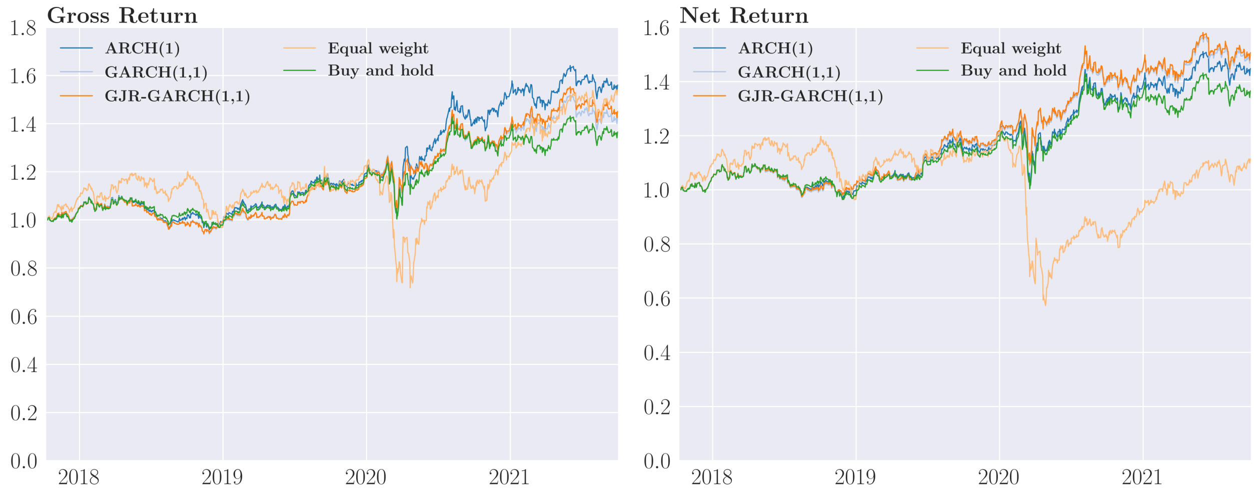
We will not focus on the portfolio weights for this case. They are plotted in figure 29, 30, and 31 in the appendix. They exhibit identical dynamics as the case with all assets and display the same reduction in trading volume as $\tilde{\gamma}_D$ increases.

Historical performance

In terms of gross returns in figure 23 panel (a), the sophisticated strategies have similar volatility to the Buy-and-hold of around 14.6% and easily beat the Equal-weight, which has volatility of 24.1%. But the sophisticated strategies have the highest Sharpe Ratio, with the sophisticated ARCH(1) achieving a ratio of 0.74. It seems that the high trading volume of the ARCH(1) gives it an edge over the remaining strategies in terms of gross returns. Carefully comparing our performance to [Gârleanu and Pedersen, 2013] even though we use different sample periods, their dynamic portfolio achieves a Sharpe ratio of 0.62 in gross returns, we see very similar performance.

Figure 23: Cumulative performance of the sophisticated strategy, commodities

(a) Gross returns **(b)** Net returns for $\tilde{\gamma}_D = 1e{-}4$



When considering net returns given the empirically calibrated γ_D in figure 23 panel (b), we get slightly higher volatility for the sophisticated GARCH(1,1) and GJR-GARCH(1,1) strategies earning net returns of -5.3 % and -8.8%, respectively as a result of high transaction costs of 16.7% and 13.8%. The sophisticated ARCH(1) strategy performs poorly given the large trading volume with transaction costs of 55.4%, resulting in net returns of -50% at a volatility of 29.5%. Carefully comparing to [Gârleanu and Pedersen, 2013] where their dynamic strategy earned a Sharpe ratio of 0.58 in net returns, we see our sophisticated strategies have negative Sharpe ratios for the same reasons as the sophisticated strategies that considered all assets. Thus, we increase the ex-ante transaction costs to $\tilde{\gamma}_D = 1e{-}4$.

Increasing $\tilde{\gamma}_D$ yields a substantial increase in performance partly in slightly lower volatility but mainly in higher net returns. For the sophisticated GARCH(1,1) and GJR-GARCH(1,1) strategies, this results in Sharpe ratios of 0.74 and 0.75, whereas the sophisticated ARCH(1) earns 0.69. This performance comfortably beats the Equal-weight and the Buy-and-hold strategies. Additionally, the Sharpe ratios of our sophisticated strategies are slightly higher than the ratio of 0.58 that [Gârleanu and Pedersen, 2013] found, although we cannot make a fair and direct comparison.

We briefly consider the sophisticated strategies for completeness for the remaining two asset classes, bonds and stocks. Consider figure 25 panel (a) and (b) for stocks along with table 11. We see that the sophisticated strategies outperform the Equal-weight in term of Sharpe ratios for high values of ex-ante transaction costs but never beats the Buy-and-hold. For lower values of $\tilde{\gamma}_D$, the sophisticated GARCH(1,1) and GJR-GARCH(1,1) strategies have lower volatility than the Equal-weight and the Buy-and-hold for a majority of values. Looking at figure 25 for bonds in panel (c) + (d) and table 12, the sophisticated strategies, especially

Table 8: Annualized performance of the sophisticated strategy, commodities

Strategy	Std. deviation	Return	Sharpe ratio	Transaction costs
	Before transaction costs			
ARCH(1)	0.1590	0.1176	0.7398	
GARCH(1,1)	0.1461	0.0929	0.6355	
GJR-GARCH(1,1)	0.1446	0.0976	0.6747	
Equal-weight	0.2412	0.1141	0.4733	
Buy-and-hold	0.1463	0.0809	0.5531	
	After transaction costs $\tilde{\gamma}_D = \gamma_D = 1.52\text{e}{-6}$			
ARCH(1)	0.2951	-0.4988	-1.6905	55.432%
GARCH(1,1)	0.1619	-0.0882	-0.5448	16.695%
GJR-GARCH(1,1)	0.1517	-0.0526	-0.3466	13.751%
Equal-weight	0.2446	0.0275	0.1125	5.2374%
Buy-and-hold	0.1463	0.0809	0.5531	0%
	After transaction costs $\tilde{\gamma}_D = 1\text{e}{-4}$			
ARCH(1)	0.1450	0.0969	0.6682	0.0794%
GARCH(1,1)	0.1437	0.1056	0.7347	0.0417%
GJR-GARCH(1,1)	0.1442	0.1078	0.7477	0.0419%

Note: Transaction costs are the relative annual share of the portfolio lost to transaction costs. The actual transaction costs the investor pays is identical for each $\tilde{\gamma}$

the GJR-GARCH(1,1), have lower volatility than both the Equal-weight and Buy-and-hold for all ex-ante transaction costs displayed. For high values of $\tilde{\gamma}_D$, the sophisticated strategies outperform the Equal-weight strategy in Sharpe ratios. The Buy-and-hold is also outperformed but for a narrower range of ex-ante transaction costs. The main takeaway is that the choice of asset class(es) is crucial for the performance of the sophisticated strategies.

To summarize, investors need to adjust for transaction costs when trading is costly. An optimal low-risk strategy in gross returns can be an atrocious strategy in net returns, such as the simple strategies and some low $\tilde{\gamma}_D$ versions of the sophisticated strategies. It is very beneficial for the investor to consider optimization under higher transaction costs than what the investor incurs. Increasing transaction costs ex-ante dampens trading volume and reduces large trades to reduce future risk that might be slightly spurious given estimation uncertainty. Overall, the different sophisticated strategies perform well in gross returns with lower annualized portfolio standard deviations and higher annualized Sharpe Ratios than the Equal-weight and slightly lower than the Buy-and-hold strategy. For $\tilde{\gamma}_D \in \{2.3\text{e}{-5}, 1\text{e}{-4}\}$, the sophisticated strategies likewise outperformed the Equal-weight across all measures and yielded a lower annualized volatility than the Buy-and-hold, which had a higher annualized Sharpe ratio than the sophisticated strategies. The sophisticated strategy for high $\tilde{\gamma}_D$ shows very similar performances to the dynamic trading strategies of [Gârleanu and Pedersen, 2013] when considering the same assets.

9 Discussion

No one completes a thesis without impactful decisions and tough compromises. Thus, we will attempt to address some of these choices and explain our thought process behind them.

An obvious question to ask two researchers looking at minimum variance portfolio is why we haven't included a risk-less asset. The answer to this is threefold. Firstly, the minimum variance-seeking investor would always invest 100% into the risk-less asset because it reduces portfolio risk the most, which is not the most exciting result. Secondly, a genuinely risk-less asset is a theoretical construction that one could argue doesn't exist in the real world though often approximated by US treasuries or a bank account.

Why do we report annualized returns and Sharpe Ratios when we model a minimum variance investor? We look at minimum variance portfolios because authors like [Jagannathan and Ma, 2003] have shown that minimum variance portfolios have better out-of-sample Sharpe Ratios than tangency portfolios. The out-performance of minimum variance portfolios is because sample estimates of mean returns are a lousy predictor of future returns as shown [Jobson and Korkie, 1980] resulting in extreme portfolio weights, see [DeMiguel et al., 2009]. But, if we are interested in the return of our portfolio, why did we not implement a factor-type model or another more complex mean prediction model rather than discard them altogether? First and foremost, it is beyond the scope of this paper to unite factor models and multivariate GARCH models. Secondly, as mentioned in section 2.2.1, [Welch and Goyal, 2008] finds that the estimates of many factor models are unstable, insignificant, and yield spurious results. Thus, we decided to focus on modeling the conditional covariance.

Regarding modeling the covariance, one might question why we don't update the model parameters with the new information that the investor receives while going through the out-of-sample period. Firstly, continuously updating the model parameters increases the computational effort needed to solve the model dramatically with little to no benefits in return because the model parameters are unlikely to change significantly with the continuous addition of new information. Secondly, adding this extra dimension will likely cause mild instability given the estimation uncertainty, which makes interpreting the results more difficult as there would be an extra dimension to consider. One could argue that re-estimating the model every year, for example, might be a good idea to enrich the parameter estimates with new information. We will, however, leave this question for future research. Another interesting future research venture would be to add more sophisticated univariate GARCH models to the DCC MGARCH model. One could consider the EGARCH model by [Nelson, 1991] or with the more frequent use of high-frequency data, a Realized GARCH model presented in [Hansen et al., 2012]. Especially, the latter could increase the forecast accuracy of the conditional variance and thus improve the forecast of the covariance, decreasing estimation uncertainty and increasing investor utility.

Speaking of uncertainty, we currently shrink the correlations of $\bar{Q}$ by 50%. This number is similar to [Gârleanu and Pedersen, 2013] and is our reason for using it. But it is possible to estimate the optimal amount of shrinkage intensity, δ, derived by [Ledoit and Wolf, 2004]. We could have used an estimate for δ rather than simply setting it to 50%, which might have yielded better performance for the simple and sophisticated strategies, but this is beyond the scope of this paper, and we leave it for future research.

We only test real-world data in our current empirical setup, i.e., historical backtesting, which will always be a case study of a particular period for certain assets. Thus, the results will always depend on the sample period, choice of assets, number of assets, etc. This is quite evident in our backtesting, as the performance of the simple and sophisticated strategies vary substantially for different asset classes. Another drawback of historical backtesting is that past performance is not necessarily an indicator of future performance. Thus, good performance in historical backtesting does not mean that the investment strategy will perform well in the future. Another way to test performance is to use Monte Carlo simulation. We could have used simulated returns from some distribution to allow us to change the data generating process of the returns and investigate how this impacts our results and whether the dynamic strategies are robust to changes to the data generating process. However, we will leave Monte Carlo approaches to future research.

Another point of valid critique is why we use ETFs when an institutional investor can feasibly replicate the index themselves and avoid the administration fees of the ETFs. While this will most likely be true for institutional investors, it would be computationally highly challenging to estimate multivariate GARCH models for hundreds or thousands of assets with our current setup. The curse of dimensionality would make accurately estimating the covariance of these thousands of assets difficult with the amount of data available. Therefore, modeling the individual assets that compromise the ETFs we use is beyond the scope of this paper but could be a possible benefit to the investor with more diverse investment opportunities, i.e., increasing the investor's feasible set.

Lastly, we model an investor facing quadratic transaction costs similar to [Gârleanu and Pedersen, 2013]. While these seem the best choice for large institutional investors, proportional transaction costs might be better for smaller investors. [Mei et al., 2016] considers a version of [Gârleanu and Pedersen, 2013]'s model with proportional transaction costs. As such, this change is not new to the literature, and thus, sticking more closely to the model made by [Gârleanu and Pedersen, 2013] made comparisons easier. We only use the median value of γ_D of all the assets rather than the individual asset calibrations. Using the individual values would be more realistic but would also add an extra dimension that would make interpretation more difficult. Sticking to the setup of [Gârleanu and Pedersen, 2013] also allows easier comparisons.

10 Conclusion

This thesis derives optimal multi-period dynamic minimum variance trading strategies for an investor that either ignores (simple strategy) or adjusts for quadratic transaction costs (sophisticated strategy). We use multivariate DCC GARCH models to forecast the conditional covariance for both strategies. These optimal strategies are what we set out to find in our research question. The simple strategy has a closed-form solution, while the sophisticated strategy requires numerically solving the coefficient matrices. Additionally, the solution to the sophisticated strategy is slightly incomplete as it requires numerically imposing the weights to sum to 1.

The two solutions have simple dynamics. In the simple strategy, the investor ignores transaction costs and completely rebalances to the new minimum variance portfolio each period. Thus, the investor responds immediately to shocks with the convergence speed depending on the exact specification of the DCC MGARCH model. In the sophisticated strategy, the investor partially rebalances the portfolio toward an aim portfolio, which closely resembles the next period's minimum variance portfolio. The amount of rebalancing towards the aim depends on transaction costs; the higher the transaction costs, the lower the amount of trading.

We backtest the two strategies for 1.000 trading days (4 calendar years) using 11 different EFTs from different sections, regions, and assets types. In terms of gross returns, we find that the simple strategy achieves a lower standard deviation and a higher Sharpe ratio when benchmarked against a naive Equal-weight and a Buy-and-hold strategy. Of the three volatility models tested, the GARCH(1,1) fares best as it slightly outperforms the GJR-GARCH(1,1) and ARCH(1). However, we find abysmal performance in net returns (after transaction costs) as the investor trades heavily, especially for the simple ARCH(1). Large amounts of trading are very costly given quadratic transaction costs. We calibrate the transaction cost parameter, γ_D, to match empirical transaction costs using a result from [Robert et al., 2012]. Using the calibrated γ_D in the sophisticated strategy barely improves the performance over the simple strategy in net returns. The interpretation is that the investor believes rebalancing almost completely toward the aim portfolio reduces the expected future portfolio risk to such a degree that the high rebalancing is worth the high transaction costs. Increasing the ex-ante (or perceived) transaction costs, $\tilde{\gamma}_D$, increases net performance significantly. When the investor perceives trading as more costly, the investor decreases trading volumes. Applying optimal $\tilde{\gamma}_D$ values to the sophisticated GARCH(1,1) and GJR-GARCH(1,1), we obtain lower standard deviations than the Equal-weight and Buy-and-hold strategies with higher net Sharpe ratios than the Equal-weight but lower than the But-and-hold. Considering only gold and oil futures like [Gârleanu and Pedersen, 2013], we find similar net Sharpe ratios, and our dynamic strategies outperform the Equal-weight and Buy-and-hold benchmark across all measures when using the optimal $\tilde{\gamma}_D$.

Our contribution to portfolio theory is thus the foundations of a framework for solving dynamic multi-period with an investor seeking to minimize portfolio variance with quadratic transaction costs and incorporating time-varying covariance estimates model by multivariate GARCH models. Future research with offset in this thesis should first and foremost focus on deriving a complete theoretical solution to the sophisticated strategy by finding optimal weights that sum to 1. Additionally, one could find an analytical solution or an approximation to the coefficient matrix, A_{vv}. Afterward, one could incorporate more complex univariate GARCH models like the Realized GARCH or EGARCH for better covariance forecasts. Furthermore, it might be possible to combine the focus of [Gârleanu and Pedersen, 2013] on mean prediction models and our emphasis on covariance prediction into a complete model that incorporates both mean and covariance predictions.

References

[Bellman, 1966] Bellman, R. (1966). Dynamic programming. *Science*, 153(3731):34–37.

[Bohn Nielsen, 2017] Bohn Nielsen, H. (2017). *Introduction to likelihood-based estimation and inference*. Hans Reitzels, 3 edition.

[Bollerslev, 1986] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

[Carhart, 1997] Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82.

[Collin-Dufresne et al., 2015] Collin-Dufresne, P., Daniel, K. D., Moallemi, C. C., and Saglam, M. (2015). Dynamic asset allocation with predictable returns and transaction costs. *Available at SSRN 2618910*.

[Constantinides, 1986] Constantinides, G. M. (1986). Capital market equilibrium with transaction costs. *Journal of political Economy*, 94(4):842–862.

[Cont, 2001] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance*, 1(2):223.

[DeMiguel et al., 2009] DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953.

[Engle, 2002] Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

[Engle, 1982] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

[Engle and Kroner, 1995] Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric theory*, 11(1):122–150.

[Engle and Sheppard, 2001] Engle, R. F. and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate garch. *Department of Economics, UC San Diego, University of California at San Diego, Economics Working Paper Series*.

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

[Fama and French, 1993] Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

[Gan and Lu, 2014] Gan, X. and Lu, W. (2014). A general setting and solution of bellman equation in monetary theory. *Journal of Applied Mathematics*, 2014.

[Gârleanu and Pedersen, 2013] Gârleanu, N. and Pedersen, L. H. (2013). Dynamic trading with predictable returns and transaction costs. *The Journal of Finance*, 68(6):2309–2340.

[Glosten et al., 1993] Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801.

[Hansen et al., 2012] Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

[Hautsch and Voigt, 2019] Hautsch, N. and Voigt, S. (2019). Large-scale portfolio allocation under transaction costs and model uncertainty. *Journal of Econometrics*, 212(1):221–240.

[Holt and Laury, 2002] Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655.

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

[Jobson and Korkie, 1980] Jobson, J. D. and Korkie, B. (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554.

[Ledoit and Wolf, 2004] Ledoit, O. and Wolf, M. (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119.

[Li and Nadarajah, 2020] Li, R. and Nadarajah, S. (2020). A review of student'st distribution and its generalizations. *Empirical Economics*, 58(3):1461–1490.

[Mandelbrot, 1967] Mandelbrot, B. (1967). The variation of some other speculative prices. *The Journal of Business*, 40(4):393–413.

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

[Mei et al., 2016] Mei, X., DeMiguel, V., and Nogales, F. J. (2016). Multiperiod portfolio optimization with multiple risky assets and general transaction costs. *Journal of Banking & Finance*, 69:108–120.

[Mei and Nogales, 2018] Mei, X. and Nogales, F. J. (2018). Portfolio selection with proportional transaction costs and predictability. *Journal of Banking & Finance*, 94:131–151.

[Merton, 1969] Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, pages 247–257.

[Michaud, 1989] Michaud, R. O. (1989). The markowitz optimization enigma: Is 'optimized' optimal? *Financial analysts journal*, 45(1):31–42.

[Nelson, 1991] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370.

[Robert et al., 2012] Robert, E., Robert, F., and Jeffrey, R. (2012). Measuring and modeling execution cost and risk. *The Journal of Portfolio Management*, 38(2):14–28.

[Rossi and Spazzini, 2010] Rossi, E. and Spazzini, F. (2010). Model and distribution uncertainty in multivariate garch estimation: a monte carlo analysis. *Computational statistics & data analysis*, 54(11):2786–2800.

[Taylor, 2011] Taylor, S. J. (2011). *Asset price dynamics, volatility, and prediction.* Princeton university press.

[Virtanen et al., 2020] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

[Welch and Goyal, 2008] Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.

A Appendix

A.1 Student's t-distribution

Consider Z and Y as an independent random variable, where $Z \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2(\nu)$. The Student's t random variable can be defined by

$$X = \frac{Z}{\sqrt{Y/\nu}}$$

with the probability density function (pdf) given as

$$f(x|\nu) = \frac{\frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{\nu}{2})}}{\sqrt{\nu\pi}} \left(1 + \frac{x^2}{\nu}\right)^{\frac{-\nu+1}{2}}$$

with ν degrees of freedom, which determines the fatness of the tails and the number of moments which are finite as such $\nu > 2$ for the variance to be finite and $\Gamma(\cdot)$ is the Gamma function. Note that for $\nu \to \infty$, the student t distribution converges to a normal distribution (See [Li and Nadarajah, 2020] for more).

This distribution is a special case of the multivariate Student's t-distribution. Consider yet again Z and Y now as independent random vectors, where Z is a multivariate standard normal and $Y \sim \chi^2(\nu)$. The multivariate Student's t random variable can be defined by

$$X = \frac{Z}{\sqrt{Y/\nu}}$$

with the multivariate probability density function (pdf) given as

$$f(x|\nu) = \frac{\Gamma\big((\nu+p)/2\big)}{\Gamma\big(\nu/2\big)\nu^p p/2\pi^{p/2}|\Sigma|^{1/2}} \left(1 + \frac{1}{v}x'\Sigma^{-1}x\right)^{-(v+p)/2}$$

Note here that Σ is the covariance matrix of the multivariate normal distribution. ν is degrees of freedom that determine the fatness of the tails and the number of finite moments. p is the number of dimensions.

A.2 Non-Central Student's t-distribution

Consider Z and Y as independent random variable, where $Z \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2(\nu)$. The non-central Student's t random variable with non-centrality parameter nc can be defined by

$$X = \frac{Z + nc}{\sqrt{Y/\nu}}$$

with the probability density function (pdf) given as

$$f(x|\nu, nc) = \frac{e^{-nc^2/2}\nu^{\nu/2}}{\sqrt{\pi}(\nu + x^2)^{(\nu+1)/2}\Gamma(\nu/2)} \sum_{k=0}^{+\infty} \frac{\Gamma\left(\frac{v+k+1}{2}\right)nc^k 2^{k/2} x^k}{\Gamma(k+1)(\nu + x^2)^{k/2}}$$

with $I_x(a,b)$ being the incomplete beta function ratio. Where nc dictates which direction the distribution is moved. ν is the degrees of freedom that determine the fatness of the tails and the number of finite moments such $\nu > 2$ for the variance to be finite and $\Gamma(\cdot)$ is the Gamma function. Note that for $\nu \to \infty$, the student t distribution converges to a non-central normal distribution (See [Li and Nadarajah, 2020] for more).

A.3 Multivariate GARCH models in Portfolio theory

Taking first order conditions with respect to the weight, v_t

$$\frac{\partial \mathcal{L}}{\partial v_t} = \Omega_{t+1|t} v_t - \lambda_t \mathbf{1} = 0$$

Solving for v_t yields

$$\lambda \mathbf{1} = \Omega_{t+1|t} v_t \Leftrightarrow v_t \Omega_{t+1|t} \mathbf{1} \lambda_t$$

The constraint requires that $v_t'\mathbf{1} = 1$, which can be used to solve for the Lagrangian multiplier λ_t

$$\begin{aligned} 1 =& v_t'\mathbf{1} = \mathbf{1}'v_t \\ 1 =& \mathbf{1}'\Omega_{t+1|t}^{-1}\mathbf{1}\lambda_t \\ \lambda_t =& \frac{1}{\mathbf{1}'\Omega_{t+1|t}^{-1}\mathbf{1}} \end{aligned}$$

Now, we insert the expression into the weights v_t

$$v_t = \Omega_{t+1|t}^{-1}\mathbf{1}\lambda_t = \Omega_{t+1|t}^{-1}\mathbf{1}\frac{1}{\mathbf{1}'\Omega_{t+1|t}^{-1}\mathbf{1}}$$

$$v_t = \frac{\Omega_{t+1|t}^{-1}\mathbf{1}}{\mathbf{1}'\Omega_{t+1|t}^{-1}\mathbf{1}} = v_t^{\text{MVP}}$$

A.4 The sophisticated strategy (Adjusting for transaction costs)

We evaluate the conditional expectation of the guessed value function with $\mathbb{E}_t[\Omega_{t+1}] = \Omega_{t+1|t}$

$$\mathbb{E}_t[V(v_t)] = \frac{1}{2}v_t'A_{vv}v_t - v_t'A_{v1}\mathbf{1} - \frac{1}{2}\mathbf{1}'A_{11}\mathbf{1}$$

Inserting this into our value function in equation (31) yields the problem:

$$\begin{aligned} V(v_{t-1}) = \min_{v_t} \Bigg[\frac{1}{2}(v_t - v_{t-1})'\Lambda_t(v_t - v_{t-1}) + (1-\rho)\bigg(\frac{1}{2}v_t'\Omega_{t+1|t}v_t + \frac{1}{2}v_t'A_{vv}v_t \\ -v_t'A_{v1}\mathbf{1} - \frac{1}{2}\mathbf{1}'A_{11}\mathbf{1}\bigg)\Bigg] - \lambda_t(v_t'\mathbf{1} - 1) \end{aligned}$$

We redefine terms and see the investor minimizes the following quadratic problem:

$$V(v_{t-1}) = \min_{v_t} \left[(1-\rho)\left(\frac{1}{2}v_t' J_t v_t - v_t' j_t - d_t\right)\right] - \lambda_t(v_t'\mathbf{1} - 1) \tag{37}$$

with

$$\begin{aligned} J_t &= \Omega_{t+1|t} + A_{vv} + \bar{\Lambda}_t \\ j_t &= \bar{\Lambda}_t v_{t-1} + A_{v1}\mathbf{1} \\ d_t &= \frac{1}{2}v_{t-1}'\bar{\Lambda}_t v_{t-1} + \frac{1}{2}\mathbf{1}' A_{11}\mathbf{1} \end{aligned}$$

Where we define $\bar{\Lambda}_t = (1-\rho)^{-1}\Lambda_t$. Minimize the reformulated problem wrt. to v_t:

$$\begin{aligned} \frac{\partial V(v_{t-1})}{\partial v_t} &= (1-\rho)(J_t v_t - j_t) - \lambda_t\mathbf{1} = 0 \Leftrightarrow (1-\rho)J_t v_t = (1-\rho)j_t + \lambda_t\mathbf{1} \Leftrightarrow \\ v_t &= J_t^{-1}(j_t + (1-\rho)^{-1}\lambda_t\mathbf{1}) \end{aligned}$$

Solve for the Lagrangian multiplier λ_t using the constraint $v_t'\mathbf{1} = 1$

$$1 = \mathbf{1}'\big[J_t^{-1}(j_t + (1-\rho)^{-1}\lambda_t\mathbf{1})\big] = \mathbf{1}'J_t^{-1}j_t + \mathbf{1}'J_t^{-1}\mathbf{1}(1-\rho)^{-1}\lambda_t \Leftrightarrow \lambda_t = \frac{1-\mathbf{1}'J_t^{-1}j_t}{(1-\rho)^{-1}\mathbf{1}'J_t^{-1}\mathbf{1}}$$

Inserting the Lagrangian multiplier back into the problem yields

$$v_t = J_t^{-1}(j_t + (1-\rho)^{-1}\lambda_t\mathbf{1}) = J_t^{-1}\left(j_t + \frac{1-\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}\right)$$

Detour a bit to check that the weights sum to one by inserting v_t into the constraint, $v_t'\mathbf{1} = \mathbf{1}'v_t = 1$

$$\begin{aligned} 1 &= \mathbf{1}'\left[J_t^{-1}\left(j_t + \frac{1-\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}\right)\right] = \mathbf{1}'J_t^{-1}j_t + \frac{\mathbf{1}'J_t^{-1}\mathbf{1} - \mathbf{1}'J_t^{-1}\mathbf{1}\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}} \\ 1 &= \mathbf{1}'J_t^{-1}j_t + \frac{\mathbf{1}'J_t^{-1}\mathbf{1} - \mathbf{1}'J_t^{-1}\mathbf{1}\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}} = \mathbf{1}'J_t^{-1}j_t + 1 - \mathbf{1}'J_t^{-1}j_t = 1 \end{aligned}$$

Recall the rewritten version of the quadratic problem $V(v_{t-1})$ in equation (37) and define $\bar{\lambda}_t = (1-\rho)^{-1}\lambda_t$, then we insert the solution of v_t into equation 37 (Ignore λ_t for the time being)

$$\begin{aligned} V(v_{t-1}) &= (1-\rho)\left\{\frac{1}{2}\big[J_t^{-1}(j_t + \bar{\lambda}_t\mathbf{1})\big]' J_t\big[J_t^{-1}(j_t + \bar{\lambda}_t\mathbf{1})\big] - \big[J_t^{-1}(j_t + \bar{\lambda}_t\mathbf{1})\big]' j_t - d_t\right\} - \lambda_t(v_t'\mathbf{1} - 1) \\ &= (1-\rho)\left\{\frac{1}{2}\big[J_t^{-1}(j_t + \bar{\lambda}_t\mathbf{1})\big]'(j_t + \bar{\lambda}_t\mathbf{1}) - \big[J_t^{-1}(j_t + \bar{\lambda}_t\mathbf{1})\big]' j_t - d_t\right\} - \lambda_t(v_t'\mathbf{1} - 1) \\ &= (1-\rho)\left\{\big[J_t^{-1}(j_t + \bar{\lambda}_t\mathbf{1})\big]'\big[\frac{1}{2}(j_t + \bar{\lambda}_t\mathbf{1}) - j_t\big] - d_t\right\} - \lambda_t(v_t'\mathbf{1} - 1) \\ &= (1-\rho)\left\{\big[J_t^{-1}(j_t + \bar{\lambda}_t\mathbf{1})\big]'\big[\frac{1}{2}(\bar{\lambda}_t\mathbf{1} - j_t)\big] - d_t\right\} - \lambda_t(v_t'\mathbf{1} - 1) \end{aligned}$$

All elements of J_t can be chosen as symmetric and because a sum of symmetric matrices is also symmetric J_t is symmetric, meaning $(J_t^{-1})' = J_t^{-1}$.

$$\begin{aligned}
V(v_{t-1}) &= (1-\rho)\left\{\frac{1}{2}(j_t + \bar{\lambda}_t\mathbf{1})'J_t^{-1}\left[\bar{\lambda}_t\mathbf{1} - j_t\right] - d_t\right\} - \lambda_t(v_t'\mathbf{1} - 1) \\
&= (1-\rho)\left\{\frac{1}{2}(\bar{\lambda}_t\mathbf{1})'J_t^{-1}(\bar{\lambda}_t\mathbf{1}) - \frac{1}{2}j_t'J_t^{-1}j_t + \frac{1}{2}j_t'J_t^{-1}(\bar{\lambda}_t\mathbf{1}) - \frac{1}{2}(\bar{\lambda}_t\mathbf{1})'J_t^{-1}j_t - d_t\right\} - \lambda_t(v_t'\mathbf{1} - 1) \\
&= (1-\rho)\left\{\frac{1}{2}(\bar{\lambda}_t\mathbf{1})'J_t^{-1}(\bar{\lambda}_t\mathbf{1}) - \frac{1}{2}j_t'J_t^{-1}j_t + \frac{1}{2}j_t'J_t^{-1}(\bar{\lambda}_t\mathbf{1}) - \frac{1}{2}j_t'J_t^{-1}(\bar{\lambda}_t\mathbf{1}) - d_t\right\} - \lambda_t(v_t'\mathbf{1} - 1)
\end{aligned}$$

Now insert for v_t in the constraint $\lambda_t(v_t'\mathbf{1} - 1)$

$$\begin{aligned}
V(v_{t-1}) &= (1-\rho)\left\{\frac{1}{2}(\bar{\lambda}_t\mathbf{1})'J_t^{-1}(\bar{\lambda}_t\mathbf{1}) - \frac{1}{2}j_t'J_t^{-1}j_t - d_t\right\} - \lambda_t([J_t^{-1}(j_t + \bar{\lambda}_t\mathbf{1})]'\mathbf{1} - 1) \\
&= (1-\rho)\left\{\frac{1}{2}(\bar{\lambda}_t\mathbf{1})'J_t^{-1}(\bar{\lambda}_t\mathbf{1}) - \frac{1}{2}j_t'J_t^{-1}j_t - d_t - \bar{\lambda}_t(j_t'J_t^{-1}\mathbf{1} + \bar{\lambda}_t\mathbf{1}'J_t^{-1}\mathbf{1} - 1)\right\} \\
&= (1-\rho)\left\{\frac{1}{2}(\bar{\lambda}_t\mathbf{1})'J_t^{-1}(\bar{\lambda}_t\mathbf{1}) - \frac{1}{2}j_t'J_t^{-1}j_t - d_t - \bar{\lambda}_t j_t'J_t^{-1}\mathbf{1} - (\bar{\lambda}_t\mathbf{1})'J_t^{-1}(\mathbf{1}\bar{\lambda}_t) + \bar{\lambda}_t\right\} \\
&= (1-\rho)\left\{-\frac{1}{2}(\bar{\lambda}_t\mathbf{1})'J_t^{-1}(\bar{\lambda}_t\mathbf{1}) - \frac{1}{2}j_t'J_t^{-1}j_t - d_t - \bar{\lambda}_t j_t'J_t^{-1}\mathbf{1} + \bar{\lambda}_t\right\}
\end{aligned}$$

Now, insert for $\bar{\lambda}_t$

$$\begin{aligned}
V(v_{t-1}) &= (1-\rho)\left\{-\frac{1}{2}\left((1-\rho)^{-1}\frac{1-\mathbf{1}'J_t^{-1}j_t}{(1-\rho)^{-1}\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}\right)'J_t^{-1}\left((1-\rho)^{-1}\frac{1-\mathbf{1}'J_t^{-1}j_t}{(1-\rho)^{-1}\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}\right)\right. \\
&\quad\left. -\frac{1}{2}j_t'J_t^{-1}j_t - d_t - (1-\rho)^{-1}\frac{1-\mathbf{1}'J_t^{-1}j_t}{(1-\rho)^{-1}\mathbf{1}'J_t^{-1}\mathbf{1}}j_t'J_t^{-1}\mathbf{1} + (1-\rho)^{-1}\frac{1-\mathbf{1}'J_t^{-1}j_t}{(1-\rho)^{-1}\mathbf{1}'J_t^{-1}\mathbf{1}}\right\} \\
&= (1-\rho)\left\{-\frac{1}{2}\left(\frac{1-\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}\right)'J_t^{-1}\left(\frac{1-\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}\right) - \frac{1}{2}j_t'J_t^{-1}j_t - d_t\right. \\
&\quad\left. - \frac{1-\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}}j_t'J_t^{-1}\mathbf{1} + \frac{1-\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}}\right\} \\
&= (1-\rho)\left\{\left[1 - j_t'J_t^{-1}\mathbf{1} - \frac{1}{2}\left(\frac{1-\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}\right)'J_t^{-1}\mathbf{1}\right]\left(\frac{1-\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}}\right) - \frac{1}{2}j_t'J_t^{-1}j_t - d_t\right\} \\
&= (1-\rho)\left\{\left[1 - j_t'J_t^{-1}\mathbf{1} - \frac{1}{2}\frac{1-\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}'J_t^{-1}\mathbf{1}\right]\left(\frac{1-\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}}\right) - \frac{1}{2}j_t'J_t^{-1}j_t - d_t\right\}
\end{aligned}$$

notice that $\lambda_t = (1-\mathbf{1}'J_t^{-1}j_t)/(\mathbf{1}'J_t^{-1}\mathbf{1})$ is a scalar such that $(\lambda_t\mathbf{1})' = \lambda_t\mathbf{1}'$

$$\begin{aligned}
V(v_{t-1}) &= (1-\rho)\left\{\left[1 - j_t'J_t^{-1}\mathbf{1} - \frac{1}{2} + \frac{1}{2}\mathbf{1}'J_t^{-1}j_t\right]\left(\frac{1-\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}}\right) - \frac{1}{2}j_t'J_t^{-1}j_t - d_t\right\} \\
&= (1-\rho)\left\{\left[\frac{1}{2} - \frac{1}{2}\mathbf{1}'J_t^{-1}j_t\right]\left(\frac{1-\mathbf{1}'J_t^{-1}j_t}{\mathbf{1}'J_t^{-1}\mathbf{1}}\right) - \frac{1}{2}j_t'J_t^{-1}j_t - d_t\right\}
\end{aligned}$$

now, we insert for j_t and d_t

$$
\begin{aligned}
V(v_{t-1}) &= (1-\rho)\Bigg\{\left[\frac{1}{2}-\frac{1}{2}\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]\right]\left(\frac{1-\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]}{\mathbf{1}'J_t^{-1}\mathbf{1}}\right)\\
&\quad -\frac{1}{2}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]-\frac{1}{2}v_{t-1}'\bar{\Lambda}_t v_{t-1}-\frac{1}{2}\mathbf{1}'A_{11}\mathbf{1}\Bigg\}\\
&= (1-\rho)\Bigg\{\left[\frac{1}{2}-\frac{1}{2}\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]\right]\left(\frac{1-\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]}{\mathbf{1}'J_t^{-1}\mathbf{1}}\right)\\
&\quad -\frac{1}{2}v_{t-1}'\bar{\Lambda}_t J_t^{-1}\bar{\Lambda}_t v_{t-1}-\frac{1}{2}\mathbf{1}'A_{v1}J_t^{-1}A_{v1}\mathbf{1}-v_{t-1}'\bar{\Lambda}J_t^{-1}A_{v1}\mathbf{1}-\frac{1}{2}v_{t-1}'\bar{\Lambda}_t v_{t-1}-\frac{1}{2}\mathbf{1}'A_{11}\mathbf{1}\Bigg\}\\
&= (1-\rho)\frac{1}{2}\Bigg\{\underbrace{\left[1-\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]\right]\left(\frac{1-\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]}{\mathbf{1}'J_t^{-1}\mathbf{1}}\right)}_{(\mathbb{I})}\\
&\quad -\underbrace{v_{t-1}'(\bar{\Lambda}_t J_t^{-1}\bar{\Lambda}_t+\bar{\Lambda}_t)v_{t-1}-\mathbf{1}'(A_{v1}J_t^{-1}A_{v1}+A_{11})\mathbf{1}-2v_{t-1}'\bar{\Lambda}J_t^{-1}A_{v1}\mathbf{1}}_{(\mathbb{II})}\Bigg\}
\end{aligned}
$$

Consider ($\mathbb{I}$)

$$
\begin{aligned}
&\frac{1-\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]}{\mathbf{1}'J_t^{-1}\mathbf{1}}-\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]\frac{1-\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]}{\mathbf{1}'J_t^{-1}\mathbf{1}}\\
&=\underbrace{\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}-2\frac{\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]}{\mathbf{1}'J_t^{-1}\mathbf{1}}}_{(\mathbb{III})}+\underbrace{\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]\left(\frac{\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]}{\mathbf{1}'J_t^{-1}\mathbf{1}}\right)}_{(\mathbb{IV})}
\end{aligned}
$$

Starting with ($\mathbb{III}$)

$$
\begin{aligned}
&=\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}-2\frac{\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]}{\mathbf{1}'J_t^{-1}\mathbf{1}}=\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}-2\frac{\mathbf{1}'J_t^{-1}\bar{\Lambda}_t v_{t-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}-2\frac{\mathbf{1}'J_t^{-1}A_{v1}\mathbf{1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\\
&=\mathbf{1}'\frac{J_t^{-1}}{(\mathbf{1}'J_t^{-1}\mathbf{1})(\mathbf{1}'J_t^{-1}\mathbf{1})}\mathbf{1}-v_{t-1}'\frac{2\bar{\Lambda}_t J_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}-\mathbf{1}'\frac{2J_t^{-1}A_{v1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}
\end{aligned}
$$

Continuing with ($\mathbb{IV}$)

$$
\begin{aligned}
&=\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]\left(\frac{\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]}{\mathbf{1}'J_t^{-1}\mathbf{1}}\right)\\
&=\mathbf{1}'J_t^{-1}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}[\bar{\Lambda}_t v_{t-1}+A_{v1}\mathbf{1}]'J_t^{-1}\mathbf{1}\\
&=\mathbf{1}'J_t^{-1}\bar{\Lambda}_t v_{t-1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}[\bar{\Lambda}_t v_{t-1}]'J_t^{-1}\mathbf{1}+\mathbf{1}'J_t^{-1}A_{v1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}[A_{v1}\mathbf{1}]'J_t^{-1}\mathbf{1}\\
&\qquad +\mathbf{1}'J_t^{-1}\bar{\Lambda}_t v_{t-1}\frac{2}{\mathbf{1}'J_t^{-1}\mathbf{1}}[A_{v1}\mathbf{1}]'J_t^{-1}\mathbf{1}\\
&=v_{t-1}'\bar{\Lambda}_t J_t^{-1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}'J_t^{-1}\bar{\Lambda}_t v_{t-1}+\mathbf{1}'J_t^{-1}A_{v1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}[A_{v1}\mathbf{1}]'J_t^{-1}\mathbf{1}\\
&\qquad +v_{t-1}'\bar{\Lambda}_t J_t^{-1}\mathbf{1}\frac{2}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}'J_t^{-1}A_{v1}\mathbf{1}
\end{aligned}
$$

Now insert for (𝕀𝕀𝕀) and (𝕀𝕍) into (𝕀)

$$= \mathbf{1}'\frac{J_t^{-1}}{(\mathbf{1}'J_t^{-1}\mathbf{1})(\mathbf{1}'J_t^{-1}\mathbf{1})}\mathbf{1} - v_{t-1}'\frac{2\bar{\Lambda}_t J_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1} - \mathbf{1}'\frac{2J_t^{-1}A_{v1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1} + v_{t-1}'\bar{\Lambda}_t J_t^{-1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}'J_t^{-1}\bar{\Lambda}_t v_{t-1}$$
$$+ \mathbf{1}'J_t^{-1}A_{v1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}[A_{v1}\mathbf{1}]'J_t^{-1}\mathbf{1} + v_{t-1}'\bar{\Lambda}_t J_t^{-1}\mathbf{1}\frac{2}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}'J_t^{-1}A_{v1}\mathbf{1}$$
$$= v_{t-1}'\left(\bar{\Lambda}_t J_t^{-1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}'J_t^{-1}\bar{\Lambda}_t\right)v_{t-1} - v_{t-1}'\left(\frac{2\bar{\Lambda}_t J_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\left[1 - \mathbf{1}\mathbf{1}'J_t^{-1}A_{v1}\right]\right)\mathbf{1}$$
$$+ \mathbf{1}'\left(J_t^{-1}A_{v1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}[A_{v1}\mathbf{1}]'J_t^{-1} + \frac{J_t^{-1}}{(\mathbf{1}'J_t^{-1}\mathbf{1})(\mathbf{1}'J_t^{-1}\mathbf{1})} - \frac{2J_t^{-1}A_{v1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\right)\mathbf{1}$$

Returning to $V(v_{t-1})$ with the calculated expressions for (𝕀) and (𝕀𝕀𝕀)

$$V(v_{t-1}) = (1-\rho)\frac{1}{2}\Bigg\{\underbrace{-v_{t-1}'(\bar{\Lambda}_t J_t^{-1}\bar{\Lambda}_t + \bar{\Lambda}_t)v_{t-1} - \mathbf{1}'(A_{v1}J_t^{-1}A_{v1} + A_{11})\mathbf{1} - 2v_{t-1}'\bar{\Lambda}_t J_t^{-1}A_{v1}\mathbf{1}}_{(\mathbb{III})}$$
$$+ \underbrace{v_{t-1}'\left(\bar{\Lambda}_t J_t^{-1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}'J_t^{-1}\bar{\Lambda}_t\right)v_{t-1} - v_{t-1}'\left(\frac{2\bar{\Lambda}_t J_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\left[1 - \mathbf{1}\mathbf{1}'J_t^{-1}A_{v1}\right]\right)\mathbf{1}}_{(\mathbb{I})}$$
$$+ \underbrace{\mathbf{1}'\left(A_{v1}J_t^{-1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}'J_t^{-1}A_{v1} + \frac{J_t^{-1}}{(\mathbf{1}'J_t^{-1}\mathbf{1})(\mathbf{1}'J_t^{-1}\mathbf{1})} - \frac{2J_t^{-1}A_{v1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\right)\mathbf{1}}_{(\mathbb{I})\text{ continued}}\Bigg\}$$

Combining terms with $v_{t-1}'(\,\cdot\,)v_{t-1}$, $v_{t-1}'(\,\cdot\,)\mathbf{1}$ and $\mathbf{1}'(\,\cdot\,)\mathbf{1}$

$$V(v_{t-1}) = (1-\rho)\Bigg\{\frac{1}{2}v_{t-1}'\Bigg(\underbrace{\bar{\Lambda}_t J_t^{-1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}'J_t^{-1}\bar{\Lambda}_t - \bar{\Lambda}_t J_t^{-1}\bar{\Lambda}_t - \bar{\Lambda}_t}_{A_{vv}}\Bigg)v_{t-1}$$
$$- v_{t-1}'\Bigg(\underbrace{\frac{\bar{\Lambda}_t J_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\left[1 - \mathbf{1}\mathbf{1}'J_t^{-1}A_{v1}\right] + \bar{\Lambda}_t J_t^{-1}A_{v1}}_{A_{v1}}\Bigg)\mathbf{1}$$
$$+ \frac{1}{2}\mathbf{1}'\Bigg(\underbrace{A_{v1}J_t^{-1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}'J_t^{-1}A_{v1} + \frac{J_t^{-1}}{(\mathbf{1}'J_t^{-1}\mathbf{1})(\mathbf{1}'J_t^{-1}\mathbf{1})} - \frac{2J_t^{-1}A_{v1}}{\mathbf{1}'J_t^{-1}\mathbf{1}} - A_{v1}J_t^{-1}A_{v1} - A_{11}}_{A_{11}}\Bigg)\mathbf{1}\Bigg\}$$

Compare this to the guessed value function

$$V(v_{t-1}) = \frac{1}{2}v_t'A_{vv}v_t - v_t'A_{v1}\mathbf{1} - \frac{1}{2}\mathbf{1}'A_{11}\mathbf{1}$$

We see that this is indeed a solution. This implies that the following restriction on the coefficient matrices must hold

$$(1-\rho)^{-1}A_{vv} = \bar{\Lambda}_t J_t^{-1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}'J_t^{-1}\bar{\Lambda}_t - \bar{\Lambda}_t J_t^{-1}\bar{\Lambda}_t - \bar{\Lambda}_t$$
$$(1-\rho)^{-1}A_{v1} = \frac{\bar{\Lambda}_t J_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\left[1 - \mathbf{1}\mathbf{1}'J_t^{-1}A_{v1}\right] + \bar{\Lambda}_t J_t^{-1}A_{v1}$$
$$(1-\rho)^{-1}A_{11} = A_{v1}J_t^{-1}\mathbf{1}\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}'J_t^{-1}A_{v1} + \frac{J_t^{-1}}{(\mathbf{1}'J_t^{-1}\mathbf{1})(\mathbf{1}'J_t^{-1}\mathbf{1})} - \frac{2J_t^{-1}A_{v1}}{\mathbf{1}'J_t^{-1}\mathbf{1}} - A_{v1}J_t^{-1}A_{v1} - A_{11}$$

Now, we proceed to solve for the coefficient matrices A_{vv}, A_{v1} and A_{11}. Starting with A_{vv} and begin with inserting J_t

$$\begin{aligned}(1-\rho)^{-1}A_{vv} = & \\ & \bar{\Lambda}_t\big[\Omega_{t+1|t}+A_{vv}+\bar{\Lambda}_t\big]^{-1}\mathbf{1}\frac{1}{\mathbf{1}'\big[\Omega_{t+1|t}+A_{vv}+\bar{\Lambda}_t\big]^{-1}\mathbf{1}}\mathbf{1}'\big[\Omega_{t+1|t}+A_{vv}+\bar{\Lambda}_t\big]^{-1}\bar{\Lambda}_t \\ & -\bar{\Lambda}_t\big[\Omega_{t+1|t}+A_{vv}+\bar{\Lambda}_t\big]^{-1}\bar{\Lambda}_t+\bar{\Lambda}_t\end{aligned}$$

Here, we face a major problem because we are not able to solve for A_{vv} analytically as we cannot see a way to to isolate A_{vv} in $\mathbf{1}'\big[\Omega_{t+1|t}+A_{vv}+\bar{\Lambda}_t\big]^{-1}\mathbf{1}$. We see only two approaches. Either one would have to take the inverse of $\mathbf{1}$ and $\mathbf{1}'$, which is not defined as these are not square matrices. Even more exotic inverses like a left- and right-inverse requires either full column or row rank, which a matrix of 1's does not have. Alternatively, one could take the inverse of the entire scalar $\mathbf{1}'\big[\Omega_{t+1|t}+A_{vv}+\bar{\Lambda}_t\big]^{-1}\mathbf{1}$ which simply moves the problem to one of the other terms. We, therefore, see no way to solve this problem analytically.

However, we can numerically solve for Avv by using the two sides of the equation (LHS and RHS, respectively). We define the objective function as the sum of squared differences element-wise between the matrix $\text{LHS}_{l,j}(A_{vv})$ and $\text{RHS}_{l,j}(A_{vv})$. Denote row as l and column as j

$$\arg\min_{A_{vv}}\sum_{l=1}^{N}\sum_{j=1}^{N}\big(\text{LHS}_{l,j}(A_{vv})-\text{RHS}_{l,j}(A_{vv})\big)^2 \tag{38}$$

We implement the minimization problem in Python using the Scipy package of [Virtanen et al., 2020], with the SLSQP solver. We constrain A_{vv} to be symmetric to fulfill the requirements from our derivations. We achieve numerical convergence after about 20 iterations on average.

In contrast to A_{vv}, it is possible to solve for A_{v1} analytically

$$\begin{aligned}(1-\rho)^{-1}A_{v1} &= \frac{\bar{\Lambda}_tJ_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\Big[1-\mathbf{1}\mathbf{1}'J_t^{-1}A_{v1}\Big]+\bar{\Lambda}J_t^{-1}A_{v1} \\ (1-\rho)^{-1}A_{v1}-\bar{\Lambda}J_t^{-1}A_{v1}+\frac{\bar{\Lambda}_tJ_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}\mathbf{1}'J_t^{-1}A_{v1} &= \frac{\bar{\Lambda}_tJ_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}} \\ \left((1-\rho)^{-1}-\bar{\Lambda}J_t^{-1}+\frac{\bar{\Lambda}_tJ_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}\mathbf{1}'J_t^{-1}\right)A_{v1} &= \frac{\bar{\Lambda}_tJ_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\end{aligned}$$

resulting in

$$A_{v1}=\left((1-\rho)^{-1}-\bar{\Lambda}_tJ_t^{-1}+\frac{\bar{\Lambda}_tJ_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}\mathbf{1}\mathbf{1}'J_t^{-1}\right)^{-1}\frac{\bar{\Lambda}_tJ_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}$$

Now insert for J_t and $\bar{\Lambda}_t=\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}$

$$\begin{aligned}A_{v1}=\Bigg((1-\rho)^{-1}-&\frac{\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}]^{-1}}{\mathbf{1}'[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}\mathbf{1}}\mathbf{1}'\left[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}\right]^{-1} \\ &+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}\left[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}\right]^{-1}\Bigg)^{-1}\frac{\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}]^{-1}}{\mathbf{1}'[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}]^{-1}\mathbf{1}}\end{aligned} \tag{39}$$

Note while we can analytically solve for A_{v1}, the solution is not analytical as it uses A_{vv}, which we solved for numerically, which causes A_{v1} to be a numerical solution.
Lastly, we solve for A_{11} which, similarly to A_{v1} can be solved analytically but requires both A_{vv} and A_{v1} and is thus not an analytical solution but numerical.

$$\begin{aligned}(1-\rho)^{-1}A_{11} &= \frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\left[A_{v1}J_t^{-1}\mathbf{1}\mathbf{1}'J_t^{-1}A_{v1}+\frac{J_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}-2J_t^{-1}A_{v1}\right]-A_{v1}J_t^{-1}A_{v1}-A_{11}\\ \left(1+\frac{1}{1-\rho}\right)A_{11} &= \frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\left[A_{v1}J_t^{-1}\mathbf{1}\mathbf{1}'J_t^{-1}A_{v1}+\frac{J_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}-2J_t^{-1}A_{v1}\right]-A_{v1}J_t^{-1}A_{v1}\\ \left(\frac{1-\rho+1}{1-\rho}\right)A_{11} &= \frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\left[A_{v1}J_t^{-1}\mathbf{1}\mathbf{1}'J_t^{-1}A_{v1}+\frac{J_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}-2J_t^{-1}A_{v1}\right]-A_{v1}J_t^{-1}A_{v1}\end{aligned}$$

Multiplying the term on A_{11} over yields

$$\begin{aligned}A_{11} &= \left(\frac{1-\rho+1}{1-\rho}\right)^{-1}\left\{\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\left[A_{v1}J_t^{-1}\mathbf{1}\mathbf{1}'J_t^{-1}A_{v1}+\frac{J_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}-2J_t^{-1}A_{v1}\right]-A_{v1}J_t^{-1}A_{v1}\right\}\\ A_{11} &= \left(\frac{1-\rho}{1-\rho+1}\right)\left\{\frac{1}{\mathbf{1}'J_t^{-1}\mathbf{1}}\left[A_{v1}J_t^{-1}\mathbf{1}\mathbf{1}'J_t^{-1}A_{v1}+\frac{J_t^{-1}}{\mathbf{1}'J_t^{-1}\mathbf{1}}-2J_t^{-1}A_{v1}\right]-A_{v1}J_t^{-1}A_{v1}\right\}\end{aligned}$$

Now insert for J_t and $\bar{\Lambda}_t = \frac{\Omega_{t+1|t}\gamma_D}{1-\rho}$

$$\begin{aligned}A_{11} =&\left(\frac{1-\rho}{1-\rho+1}\right)\left\{\frac{1}{\mathbf{1}'[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}]^{-1}\mathbf{1}}\left(\frac{[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}]^{-1}}{\mathbf{1}'[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}]^{-1}\mathbf{1}}\right.\right.\\ &+A_{v1}\left[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}\right]^{-1}\mathbf{1}\mathbf{1}'\left[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}\right]^{-1}A_{v1}\\ &\left.\left.-2\left[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}\right]^{-1}A_{v1}\right)-A_{v1}\left[\Omega_{t+1|t}+A_{vv}+\frac{\Omega_{t+1|t}\gamma_D}{1-\rho}\right]^{-1}A_{v1}\right\}\end{aligned} \tag{40}$$

A.5 Empirical

A.5.1 MGARCH model estimates

Table 9: Estimates of a DCC MGARCH(1,1) with univariate ARCH(1) - t_ν error terms

	Univariate GARCH			
Asset	μ	ω	α	ν
Emerging Markets (EEM)	0.062 (0.026)	2.674 (0.26)	0.408 (0.074)	3.171 (0.196)
S&P 500 (IVV)	0.089 (0.014)	1.482 (0.275)	0.816 (0.182)	2.518 (0.139)
Europe (IEV)	0.073 (0.021)	2.072 (0.246)	0.576 (0.102)	2.89 (0.163)
Global Tech (IXN)	0.107 (0.018)	1.599 (0.176)	0.418 (0.076)	2.901 (0.166)
Real estate (IYR)	0.108 (0.019)	1.399 (0.132)	0.999 (0.09)	2.95 (0.107)
Global financials (IXG)	0.078 (0.021)	2.557 (0.374)	0.903 (0.16)	2.653 (0.143)
Global Industrials (EXI)	0.087 (0.018)	1.547 (0.179)	0.636 (0.108)	2.93 (0.176)
Gold (GC=F)	0.042 (0.019)	1.408 (0.102)	0.112 (0.029)	3.58 (0.274)
Brent crude oil (BZ=F)	0.016 (0.031)	4.192 (0.442)	0.467 (0.079)	3.034 (0.206)
High-yield bonds (HYG)	0.042 (0.006)	0.222 (0.025)	0.999 (0.106)	2.756 (0.1)
20+ year treasuries (TLT)	0.036 (0.0006)	0.738 (0.033)	0.994 (0.0015)	4.00 (0.087)
	Multivariate GARCH			
	a	b	ν	
Scalar-BEKK(1,1)	0.004 (0.003)	0.973 (0.003)	9.465 (0.349)	

Note: Estimated via MLE using data from January 1$^{\text{st}}$ 2008 to October 11$^{\text{th}}$ 2017. Robust standard errors in $(\cdot)$.

Table 10: Estimates of a DCC MGARCH(1,1) with univariate GJR-GARCH(1,1) - t_ν error terms

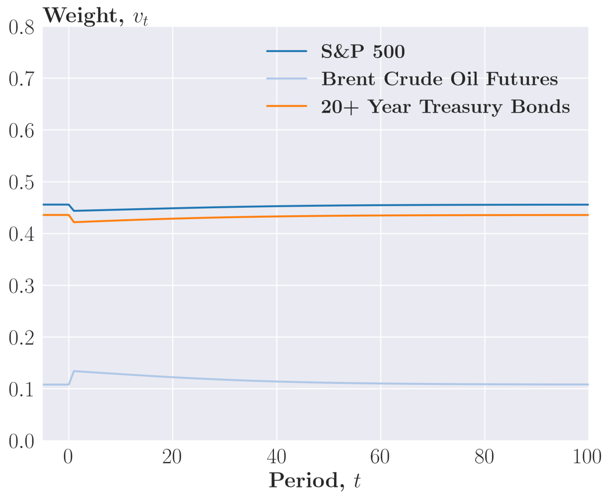
	Univariate GARCH					
Asset	μ	ω	α	β	κ	ν
Emerging Markets (EEM)	0.021 (0.025)	0.016 (0.009)	0.001 (0.012)	0.933 (0.02)	0.116 (0.021)	10.096 (1.796)
S&P 500 (IVV)	0.061 (0.013)	0.018 (0.004)	0.000 (0.015)	0.859 (0.021)	0.268 (0.045)	5.587 (0.631)
Europe (IEV)	0.044 (0.019)	0.019 (0.008)	0.019 (0.011)	0.900 (0.02)	0.146 (0.034)	5.978 (0.708)
Global Tech (IXN)	0.082 (0.017)	0.021 (0.007)	0.000 (0.010)	0.902 (0.019)	0.161 (0.033)	6.143 (0.738)
Real estate (IYR)	0.068 (0.018)	0.011 (0.004)	0.053 (0.018)	0.896 (0.02)	0.091 (0.024)	7.764 (1.056)
Global financials (IXG)	0.049 (0.02)	0.019 (0.008)	0.019 (0.012)	0.903 (0.021)	0.139 (0.031)	6.75 (0.875)
Global Industrials (EXI)	0.052 (0.017)	0.01 (0.006)	0.000 (0.019)	0.923 (0.033)	0.136 (0.033)	7.913 (1.26)
Gold (GC=F)	0.037 (0.018)	0.007 (0.002)	0.038 (0.006)	0.961 (0.001)	-0.008 (0.010)	4.692 (0.456)
Brent crude oil (BZ=F)	0.000 (0.029)	0.008 (0.004)	0.018 (0.005)	0.956 (0.001)	0.049 (0.011)	6.091 (0.718)
High-yield bonds (HYG)	0.031 (0.006)	0.003 (0.001)	0.041 (0.014)	0.861 (0.023)	0.192 (0.032)	4.84 (0.436)
20+ year treasuries (TLT)	0.037 (0.017)	0.006 (0.002)	0.055 (0.007)	0.953 (0.002)	-0.03 (0.011)	16.048 (5.143)
	Multivariate GARCH					
	a	b	ν			
Scalar-BEKK(1,1)	0.0149 (0.001)	0.972 (0.003)	10.04 (0.469)			

Note: Estimated via MLE using data from January 1st 2008 to October 11th 2017. Robust standard errors in $(\cdot)$.

A.5.2 Excess figures

Figure 24: Impulse response function of v_t^{MVP} for GJR-GARCH(1,1) with a shock of 2

(a) Weights, v_t^{MVP}



(b) Volatility, σ_t

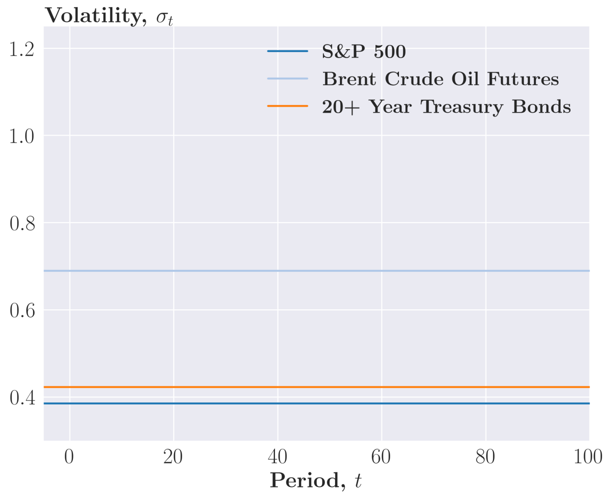
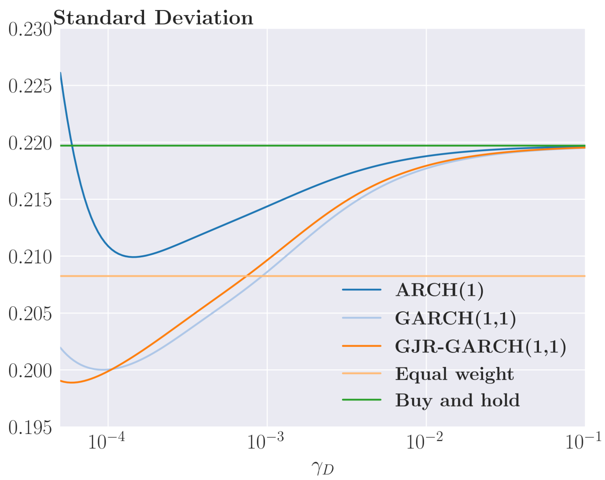
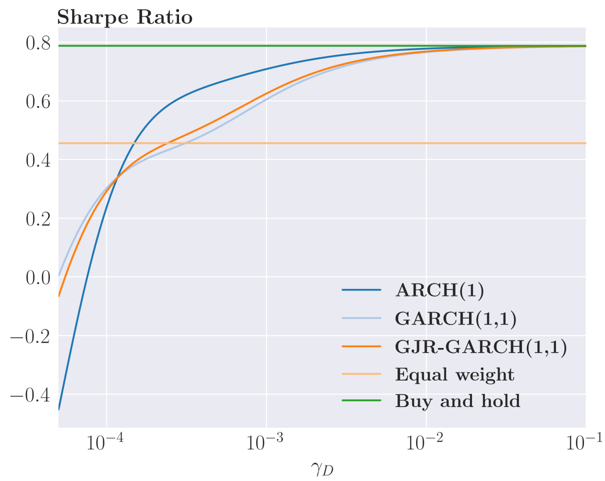


Figure 25: Annualized performance measures across different values of $\tilde{\gamma}_D$

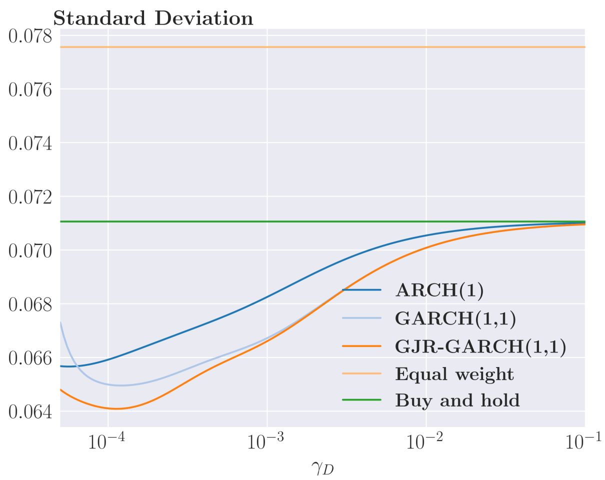
(a) Standard deviation, Stocks



(b) Sharpe Ratio, Stocks



(c) Standard deviation, bonds



(d) Sharpe Ratio, bonds

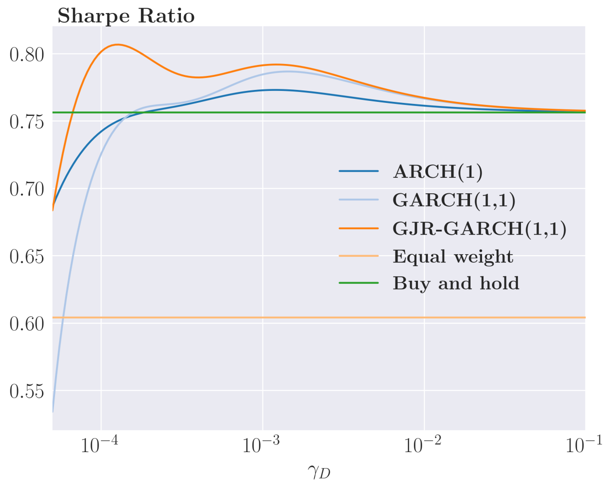
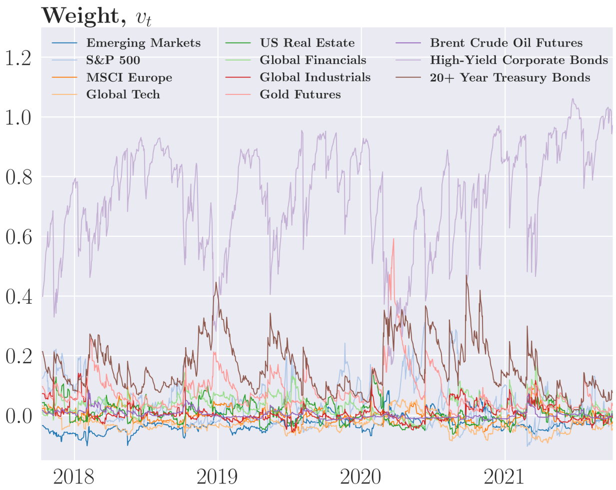


Figure 26: Simple GJR-GARCH(1,1) portfolio weights

(a) All assets



(b) Stock ETFs

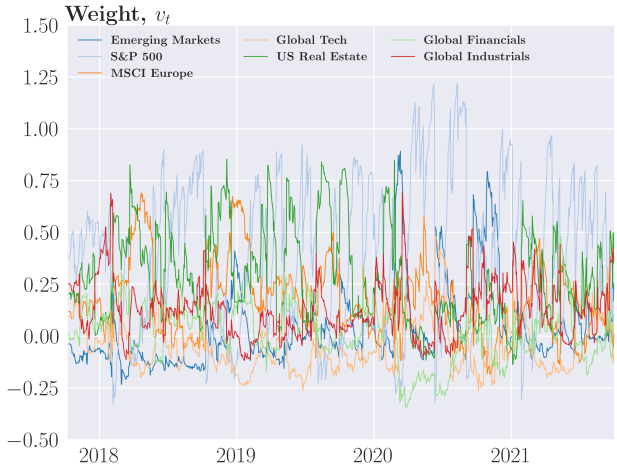
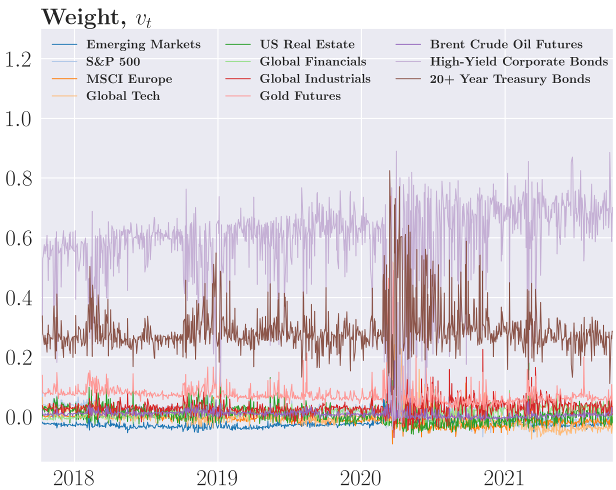
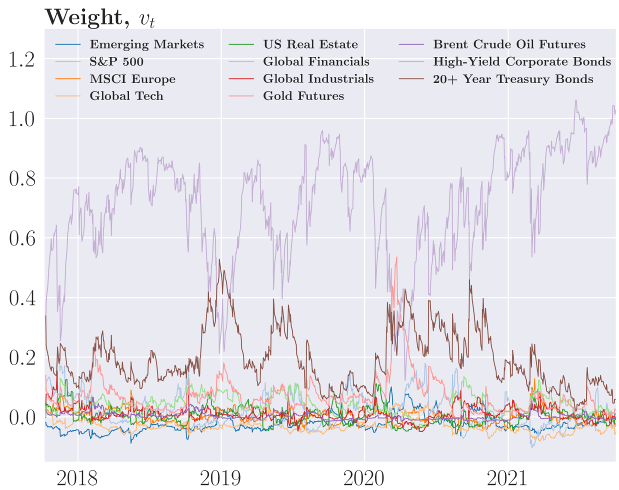


Figure 27: Weights of different sophisticated GARCH type strategies for $\tilde{\gamma}_D = 1.52\text{e}{-6}$

(a) v_t^{MVP}, ARCH(1)



(b) v_t^{MVP}, GARCH(1,1)



(c) v_t^{MVP}, GJR-GARCH(1,1)

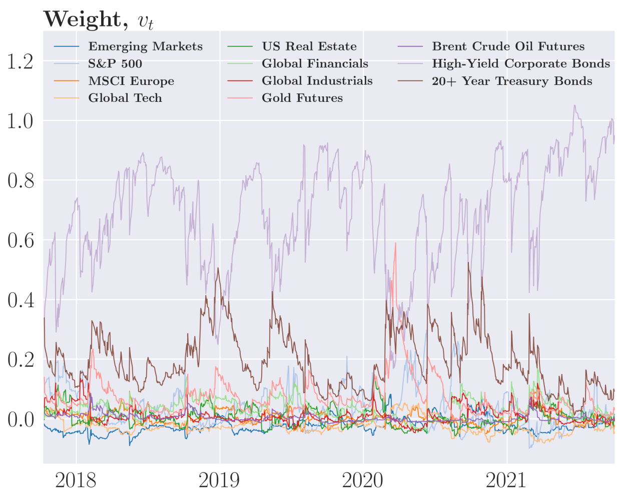
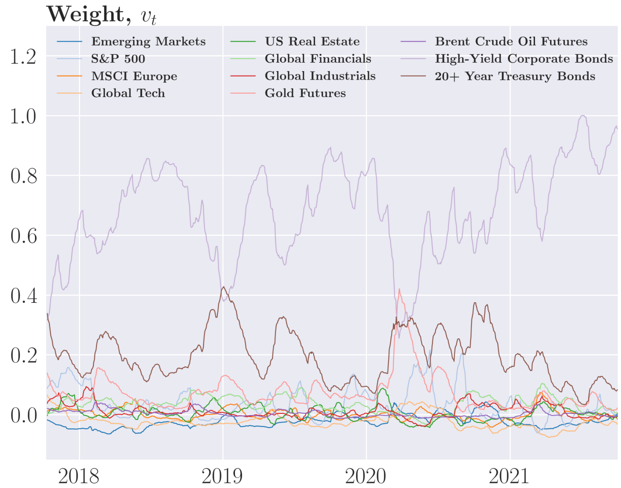


Figure 28: Sophisticated GJR-GARCH(1,1) portfolio weights

(a) v_t^{MVP} for $\tilde{\gamma}_D = 2.3\text{e}{-5}$



(b) v_t^{MVP} for $\tilde{\gamma}_D = 1\text{e}{-4}$

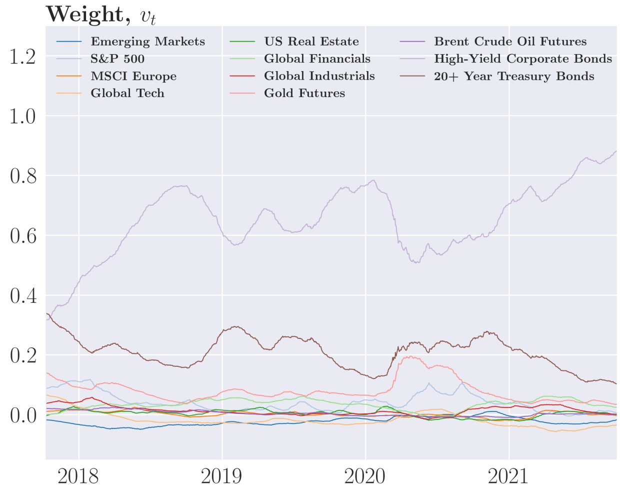
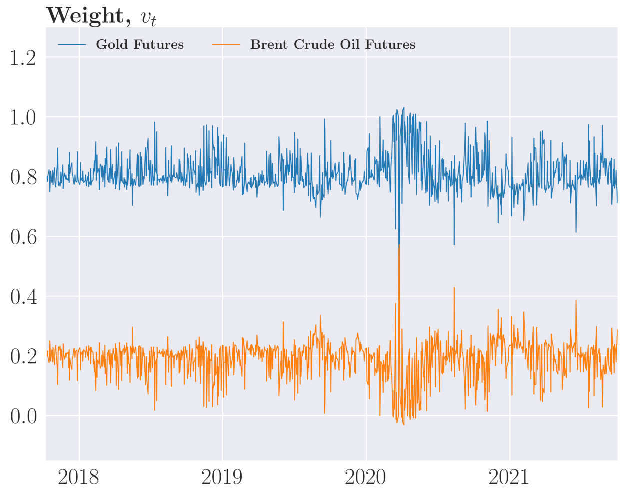


Figure 29: ARCH(1) MVP weights for different $\tilde{\gamma}_D$, commodities

(a) $\tilde{\gamma}_D = 1.52\text{e}{-6}$



(b) $\tilde{\gamma}_D = 1\text{e}{-4}$

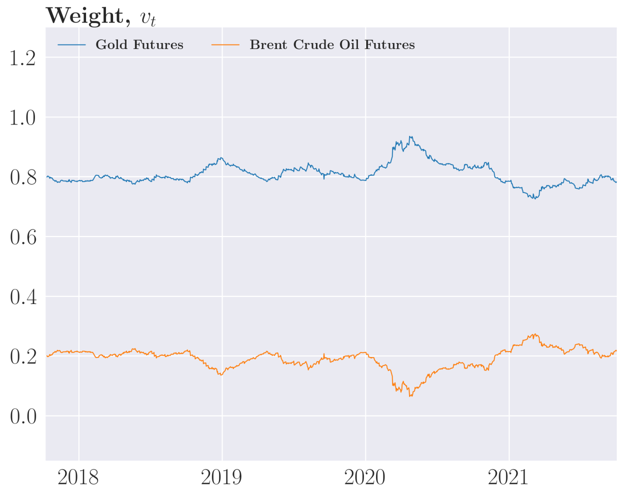
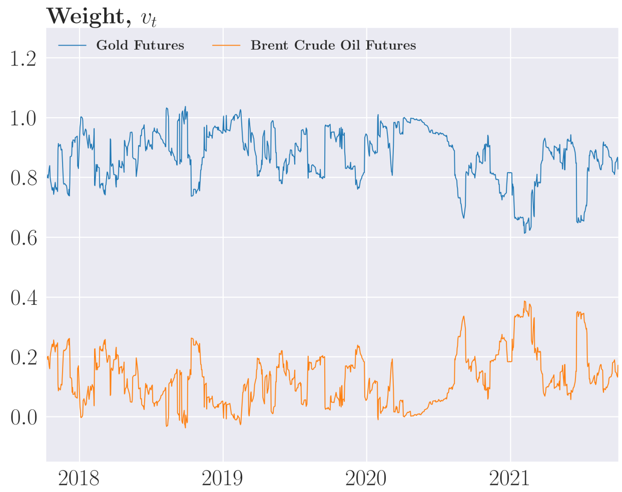


Figure 30: GARCH(1,1) MVP weights for different $\tilde{\gamma}_D$, commodities

(a) $\tilde{\gamma}_D = 1.52\text{e}{-6}$



(b) $\tilde{\gamma}_D = 1\text{e}{-4}$

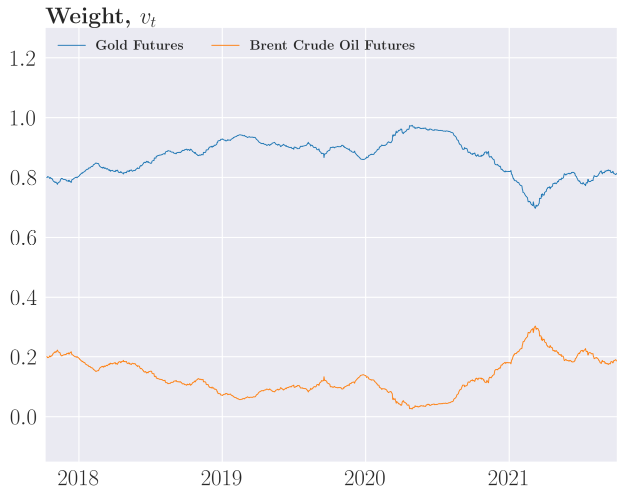
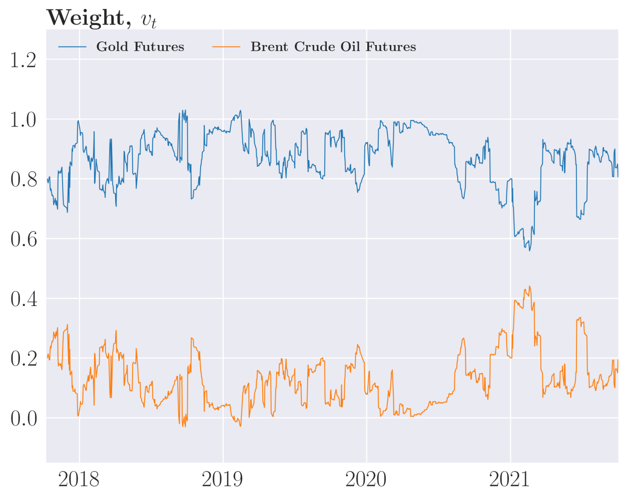
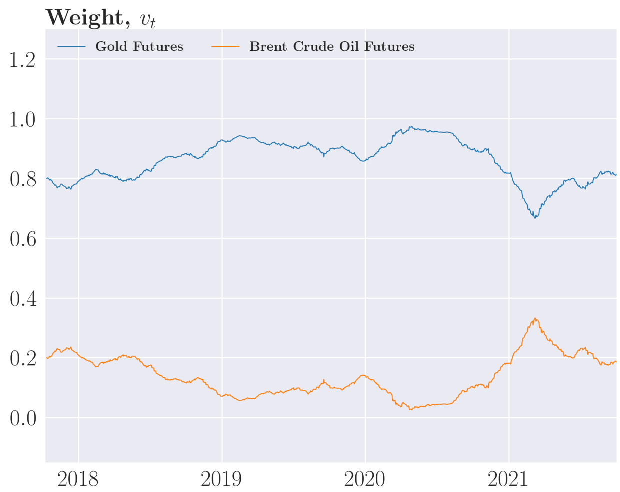


Figure 31: GJR-GARCH(1,1) MVP weights for different $\tilde{\gamma}_D$, commodities

(a) $\tilde{\gamma}_D = 1.52\text{e}{-6}$



(b) $\tilde{\gamma}_D = 1\text{e}{-4}$



A.5.3 Algorithms

The algorithms are all formulated as the general GJR-GARCH(1,1) case as the ARCH(1), and GARCH(1,1) are special cases of it.

Algorithm 1 Optimal weights using a DCC MGARCH model without transaction costs

Data: Asset returns of N assets until period T. Out of sample period is $[T+1, M]$
Result: Array of the weights of v_t of size $(M - [T+1]) \times N$
Model fit:
    Use in-sample data from period 0 to T
    Fit a Dynamic Conditional Correlation Multivariate GARCH model based on R
Univariate part:
for Asset $i \in N$ **do**
    Estimate a univariate GJR-GARCH(1,1) model
    Receive the parameters of the model $\alpha_i, \beta_i, \omega_i$,κ_i
    Receive variables from the model, σ_i, ϵ_i
end for
Multivariate part:
    Estimate a scalar BEKK GARCH(1,1) model for the pseudo correlation, Q_t
    Receive the parameters of the model a, b
One-period out of sample forecast of Ω_t
while $T \leq M$ **do**
    Get variables from current period for all N assets: $\epsilon_t^2, \sigma_t^2, r_t$
    $\text{Var}_t = \text{diag}(\sigma_{1,t}, \sigma_{2,t}, ..., \sigma_{N,t})$
    for Asset $i \in N$ **do**
        $\epsilon_{i,t} = r_{i,t} - \mu$
        $\sigma^2_{i,t+1|t} = \omega_i + \alpha_i \epsilon^2_{i,t} + \beta_i \sigma^2_{i,t} + \kappa_i \epsilon_t^2 I_{\{\epsilon_t < 0\}}$
    end for
    $\eta_t = \text{Var}_t^{-1} \epsilon_t$
    Forecast Ω_{t+1} from $\mathcal{F}_t$-measurable variables, denoted $\Omega_{t+1|t}$
    $\text{Var}_{t+1|t} = \text{diag}(\sigma_{1,t+1|t}, \sigma_{2,t+1|t}, ..., \sigma_{N,t+1|t})$
    $Q_{t+1|t} = \bar{Q}(1 - a - b) + a\eta_t\eta_t' + bQ_t$ where $\bar{Q} = \frac{1}{T}\sum_{t=1}^{T} \eta_t \eta_t'$ and shrunk
    $\Gamma_{t+1|t} = \text{diag}(Q_{t+1|t})^{-1/2} Q_{t+1|t} \text{diag}(Q_{t+1|t})^{-1/2}$
    $\Omega_{t+1|t} = \text{Var}_{t+1|t} \Gamma_{t+1|t} \text{Var}_{t+1|t}$
end while
Portfolio Optimization:
Initial weights are the benchmark Buy-and-hold MVP: $v_{-1} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$
while $T \leq M$ **do**
    Use result from section 4.2.1, equation (28) to find the optimal weights: $v_t = \frac{\Omega^{-1}_{t+1|t}\mathbf{1}}{\mathbf{1}'\Omega^{-1}_{t+1|t}\mathbf{1}}$
end while

Algorithm 2 Optimal weights using a DCC MGARCH model with transaction costs

Data: Asset returns of N assets until period T. Out of sample period is $[T+1, M]$
Result: Array of the weights of v_t of size $(M - [T+1]) \times N$
Model fit:
    Use in-sample data from period 0 to T
    Fit a Dynamic Conditional Correlation Multivariate GARCH model based on R
Univariate part:
for Asset $i \in N$ **do**
    Estimate a univariate GJR-GARCH(1,1) model
    Receive the parameters of the model $\alpha_i, \beta_i, \omega_i\ , \kappa_i$
    Receive variables from the model, σ_i, ϵ_i
end for
Multivariate part:
    Estimate a scalar BEKK GARCH(1,1) model for the pseudo correlation, Q_t
    Receive the parameters of the model a, b
One-period out of sample forecast of Ω_t
while $T \leq M$ **do**
    Get variables from current period for all N assets: $\epsilon_t^2, \sigma_t^2, r_t$
    $\text{Var}_t = \text{diag}(\sigma_{1,t}, \sigma_{2,t}, ..., \sigma_{N,t})$
    for Asset $i \in N$ **do**
        $\epsilon_{i,t} = r_{i,t} - \mu$
        $\sigma_{i,t+1|t}^2 = \omega_i + \alpha_i \epsilon_{i,t}^2 + \beta_i \sigma_{i,t}^2 + \kappa_i \epsilon_t^2 I_{\{\epsilon_t < 0\}}$
    end for
    $\eta_t = \text{Var}_t^{-1} \epsilon_t$
    Forecast Ω_{t+1} from $\mathcal{F}_t$-measurable variables, denoted $\Omega_{t+1|t}$
    $\text{Var}_{t+1|t} = \text{diag}(\sigma_{1,t+1|t}, \sigma_{2,t+1|t}, ..., \sigma_{N,t+1|t})$
    $Q_{t+1|t} = \bar{Q}(1 - a - b) + a\eta_t \eta_t' + bQ_t$ where $\bar{Q} = \frac{1}{T}\sum_{t=1}^{T} \eta_t \eta_t'$ and shrunk
    $\Gamma_{t+1|t} = \text{diag}(Q_{t+1|t})^{-1/2} Q_{t+1|t} \text{diag}(Q_{t+1|t})^{-1/2}$
    $\Omega_{t+1|t} = \text{Var}_{t+1|t} \Gamma_{t+1|t} \text{Var}_{t+1|t}$
    Calculate A_{vv}, A_{v1} and A_{11} using equation (38), (39) and (40)
end while
Portfolio Optimization:
Initial weights are the benchmark Buy-and-hold MVP: $v_{-1} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$
while $T \leq M$ **do**
    Find optimal weights using equation (32): $v_t = v_{t-1} + (\gamma_D \Omega_{t+1|t})^{-1} A_{vv} \left[v_{t-1} - \text{aim}_t\right]$
end while

Algorithm 3 Calculation of transaction costs

Data: Out of sample returns from period $[T+1, M]$ of the N assets
Result: Array of net return size $(M - [T+1]) \times N$
Calculation:
while $T \leq M$ **do**
    Calculate turnover (TO) using equation (35): $TO_{\psi,t} = v_t - \frac{v_{t-1} \circ (1+r_t)}{1+v'_{t-1} r_t}$
    Convert relative change in weight into number of stocks traded: $\Delta x_i = TO_i \psi_v P_i^{-1}$
    Calculate the dollar transaction cost: $TC_{\psi,t} = \Delta x_t \Omega_{t+1|t} \gamma_D \Delta x_t$
    Calculate the relative transaction cost: $\%TC_{\psi,t} = TC_{\psi,t}/\psi_v$
    Calculate net returns: $\hat{r}_{\psi,t} = v'_{t-1} r_t - \%TC_{\psi,t}$
end while

A.5.4 Excess tables

Table 11: Annualized performance, sophisticated strategy, stocks

Strategy	Std. deviation	Return	Sharpe ratio	Transaction costs
	Before transaction costs			
ARCH(1)	0.2085	0.1105	0.5297	
GARCH(1,1)	0.1878	0.0865	0.4609	
GJR-GARCH(1,1)	0.1899	0.0884	0.4656	
Equal weight	0.2082	0.1117	0.5366	
Buy and hold	0.2197	0.1732	0.7882	
	After transaction costs $\tilde{\gamma}_D = \gamma_D = 1.52\text{e}{-6}$			
ARCH(1)	1.4366	-1.0000	-0.6961	100.00%
GARCH(1,1)	0.9056	-1.0000	-1.1042	83.362%
GJR-GARCH(1,1)	0.8124	-1.0000	-1.2309	84.315%
Equal weight	0.2083	0.0951	0.4565	1.4981%
Buy and hold	0.2197	0.1732	0.7882	0.0000%
	After transaction costs $\tilde{\gamma}_D = 1.03\text{e}{-4}$			
ARCH(1)	0.2107	0.0547	0.2597	5.3285%
GARCH(1,1)	0.1995	0.0592	0.2968	2.2943%
GJR-GARCH(1,1)	0.1994	0.0577	0.2891	2.7637%
	After transaction costs $\tilde{\gamma}_D = 1.03\text{e}{-3}$			
ARCH(1)	0.2100	0.1379	0.6565	2.1754%
GARCH(1,1)	0.2039	0.1050	0.5148	1.0545%
GJR-GARCH(1,1)	0.2048	0.1114	0.5443	1.2475%

Note: The actual transaction costs the investor pays is identical for each $\tilde{\gamma}$

Table 12: Annualized performance, sophisticated strategy, bonds

Strategy	Std. deviation	Return	Sharpe ratio	Transaction costs
	Before transaction costs			
ARCH(1)	0.0807	0.0649	0.8047	
GARCH(1,1)	0.0703	0.0508	0.7224	
GJR-GARCH(1,1)	0.0777	0.0324	0.4174	
Equal weight	0.0756	0.0590	0.7806	
Buy and hold	0.0711	0.0538	0.7565	
	After transaction costs $\tilde{\gamma}_D = \gamma_D = 1.52\text{e}{-6}$			
ARCH(1)	0.3829	-0.5652	-1.4761	59.078%
GARCH(1,1)	0.8087	-1.0000	-1.2366	100.02%
GJR-GARCH(1,1)	0.8087	-1.0000	-1.2366	48.585%
Equal weight	0.0776	0.0469	0.6043	1.1341%
Buy and hold	0.0711	0.0538	0.7565	0.0000%
	After transaction costs $\tilde{\gamma}_D = 1.3\text{e}{-4}$			
ARCH(1)	0.0661	0.0497	0.7506	0.0655%
GARCH(1,1)	0.0650	0.0486	0.7483	0.1112%
GJR-GARCH(1,1)	0.0641	0.0517	0.8067	0.0927%
	After transaction costs $\tilde{\gamma}_D = 2.13\text{e}{-4}$			
ARCH(1)	0.0666	0.0505	0.7580	0.0251%
GARCH(1,1)	0.0652	0.0497	0.7615	0.0395%
GJR-GARCH(1,1)	0.0645	0.0512	0.7941	0.0356%

Note: The actual transaction costs the investor pays is identical for each $\tilde{\gamma}$

A.6 Python Code

All the code used for backtesting and plots is available at `https://github.com/neriksen/thesis`