2 Modern portfolio theory

Modern portfolio theory was introduced by [Markowitz, 1952] in his Nobel pricing wining work. We will give an overview of his approach, starting by characterize the asset market and the investor in section 2.1. Then we will introduce the mean-variance approach of Markowitz showing examples of allocation between first two and then p assets in section 2.2. Furthermore, we will discuss the empirical problem facing the mean-variance approach in practice in section 2.2.2.

2.1 Characterization of the investor and the asset markets

The market in this paper is characterized by p risky assets, $S_t = (S_1, S_2, ..., S_p)'$ with prices given as $P_{i,t}$. The risky assets may be stocks, EFTs or other derivatives. Additionally, investors pays no taxes and markets are perfectly liquid. Initially, the investor faces no trading costs, though this assumption will be relaxed later.

Finally, investors care only about the return and variance of their portfolio returns. This is empirically in line with the vast majority of investors being risk averse (reference maybe?). This type of preferences is called mean-variance utility which can be given by the following utility function

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

$\mathbb{E}_{t-1}[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{t-1}]$ denotes the conditional expectation with respect to the filtration, $\mathcal{F}_{t-1}$, and similar for $\mathbb{V}_{t-1}[\cdot]$. γ is the level of risk aversion and $r_t = (r_{1,t}, r_{2,t}, ..., r_{p,t})'$ is the return where $r_{i,t} = P_{i,t}/P_{i,t-1} - 1$ and v_t is a $p \times 1$ vector of weights $v_t = (v_{1,t}, v_{2,t}, ...v_{p,t})'$ defining a portfolio of assets. The weight are normalized wrt. the total amount invested and are determined in the previous period. This portfolio will have 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 essentially captures the trade-off of the risk-averse investor between maximizing returns $\mathbb{E}_{t-1}[v'_{t-1}r_t]$ and minimizing the variance $\mathbb{V}_{t-1}[v'_{t-1}r_t]$. Note that for $\gamma = 0$, $\frac{\gamma}{2}\mathbb{V}_{t-1}[v'_{t-1}r_t]$ disappears and thus the investor is risk neutral and will only care about maximizing his expected return regardless of the risk it may pose.

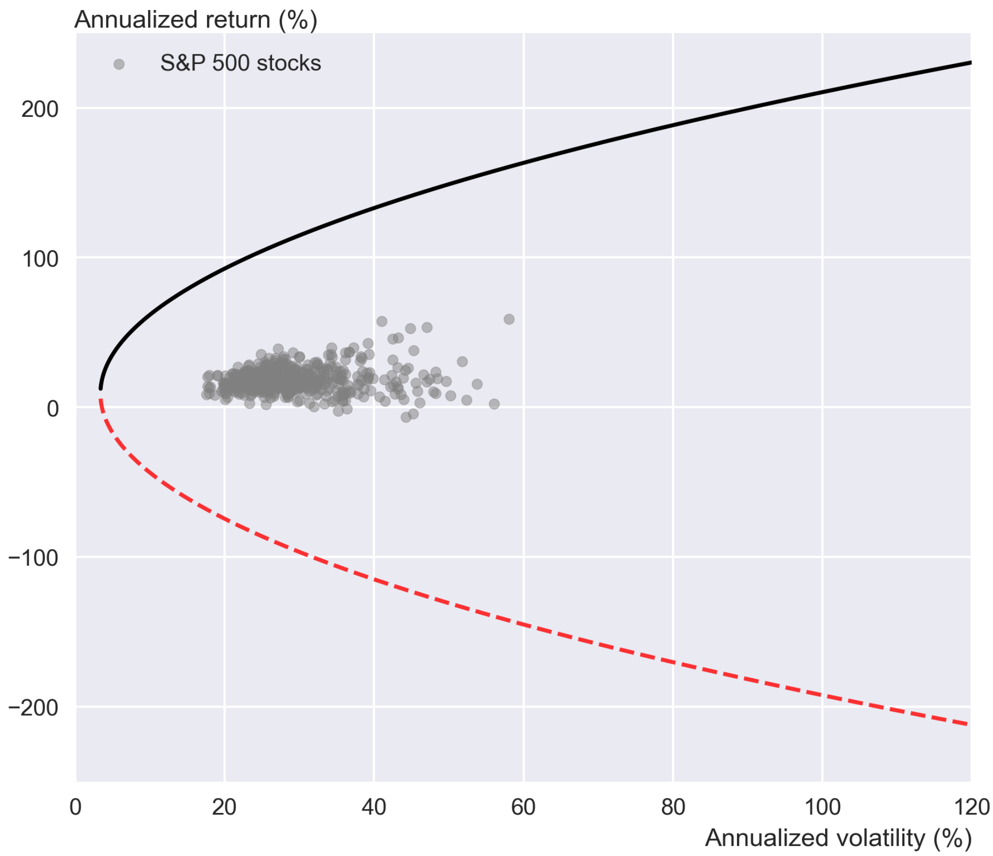
In the real world, investors also care about intermediate consumption rather than only intermediate returns and volatility. This problem has been considered by many authors in the past like [Merton, 1969] who set up rules for both intermediate consumption and allocation of wealth. While this problem is certainly a interesting one, we will only consider a investor who cares about intermediate returns and volatility, thus focusing on portfolio allocation.

2.2 Mean-Variance Approach

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

The investor can form portfolios from p assets and any set of mean and variance possible by combining these p assets is called *feasible* and all feasible combinations is called *the feasible set*. The number of assets determines the size of the feasible set, increasing as the number of assets grow. From figure 1, we see that the feasible set is any portfolio with mean and variance between the black and red striped line. Notice that the feasible set is much larger than any mean variance combination given by individual stocks.

Figure 1: Feasible set of portfolios of S&P500 stocks



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

The investor can, thus, invest in any possible portfolio within the feasible set. However, a rational investor will only invest 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 will pick a portfolio from the efficient frontier that maximizes his utility with respect to his level of risk aversion, γ. The more risk averse, the investor is, the more to the left, the investor will prefer to be, thus, having a lower risk at the cost of lower expected return. In the other end, a risk neutral investor, who does not care about risk, will simply pick the portfolio from the efficient frontier with the highest expected return regardless of the risk it carries. This translate to an investor prefer to be as far up as possible.

To see how this works in practice, lets consider a simple case with a risk averse investor allocating wealth into two risky asset.

2.2.1 Simplest case with two risky assets

Consider two risky assets, S_1 and S_2, the returns of which, $r_{1,t}$ and $r_{2,t}$, are assumed by the investor to be given by a bivariate Gaussian $N_2(\mu, \Sigma)$ distribution. With the constraint that the allocation must sum to 1 which is automatically fulfilled in this simple case as the allocation into asset 2, v_2 is given as $v_2 = 1 - v_t$.
The optimization problem for an investor maximizing his expected utility in period $t+1$

$$\max_{v_t}\{\mathbb{E}_t[U_{t+1}(v_{t-1}, r_t)]\} = \max_{v_t}\{\mathbb{E}_t[v'_{t-1}r_t] - \frac{\gamma}{2}\mathbb{V}_t[v'_{t-1}r_t]\} \tag{2}$$

with the portfolio return and variance given as

$$\mathbb{E}_t[r_{t+1}] = v_1\mu_1 + (1-v_1)\mu_2 \qquad \text{and} \qquad \mathbb{V}_t[r_{t+1}] = v_1^2\sigma_1^2 + (1-v_1)^2\sigma_2^2 + 2v_1(1-v_1)\sigma_1\sigma_2\rho$$

Setting up the Lagrangian

$$\mathcal{L}(v_1) = v_1\mu_1 + (1-v_1)\mu_2 - \frac{\gamma}{2}\big(v_1^2\sigma_1^2 + (1-v_1)^2\sigma_2^2 + 2v_1(1-v_1)\sigma_1\sigma_2\rho\big)$$

Taking first order conditions with respect to the weight v_1

$$\frac{\partial\mathcal{L}}{\partial v_1} = \mu_1 - \mu_2 - \frac{\gamma}{2}\big(2v_1\sigma_1^2 - 2(1-v_1)\sigma_2^2 + (2-4v_1)\sigma_1\sigma_2\rho\big) = 0$$

Solving for v_1 by isolating v_1 on one side and factorizing

$$\mu_1 - \mu_2 + \gamma\sigma_2^2 - \gamma\sigma_1\sigma_2\rho = \gamma v_1\big(\sigma_2^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho\big) \Leftrightarrow v_1^* = \frac{\mu_1 - \mu_2 + \gamma\sigma_2^2 - \gamma\sigma_1\sigma_2\rho}{\gamma\big(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho\big)}$$

Though this problem is trivial to solve, the intuition behind the choice of the investor is clearer than with p assets. An investor with mean-variance preferences choosing between two assets faces two trade-offs.

Firstly, the investor weighs the expected returns of both assets against each other, $\mu_1 - \mu_2$. Simply put, when the expected return of asset 1 increases, the investor increases his exposure to asset 1 and decreases his exposure to asset 2.

Secondly, the investors weighs the variance of asset 2, $\gamma\sigma_2^2$ and covariance, $\gamma\sigma_1\sigma_2\rho$ against the combined portfolio variance $\gamma\big(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho\big)$. So, as the variance of asset 2 increases it becomes relatively more risky compared to asset 1 and thus the investor will increase the weight in asset 1. This trade-off logic is built in to all the following optimal solutions.

No rational investor would only consider a portfolio of two assets as the investor would be ignoring obvious possible diversification benefits. Thus, we now consider a problem with p assets for the investor to choose from.

2.2.2 Multiple Assets Problem

Now, the market is defined by p risky assets, the returns of which is given by a multivariate Gaussian distribution $N_p(\mu, \Sigma)$. μ is a $p \times 1$ vector of mean returns and Σ is the $p \times p$ covariance matrix of the p assets. v_t is now a $p \times 1$ vector of weights instead of a scalar. The optimization problem for an investor maximizing his expected utility in period $t+1$ is:

$$\max_{v_t}\{\mathbb{E}_t[U_{t+1}(v_{t-1}, r_t)]\} = \max_{v_t}\{\mathbb{E}_t[v'_{t-1}r_t] - \frac{\gamma}{2}\mathbb{V}_t[v'_{t-1}r_t]\} =$$

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

The constraint $v'_t\mathbf{1} = 1$ means that the investor must investor all available funds into the risky assets of the market. This means that there is no outside option besides the p assets and given that we do not restrict weights to be positive, $v_i \geq 0$ that investors can short sell individual 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 $v'_t\mathbf{1} = 1$, which can be used to solve for the Lagrangian multiplier λ

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

$$\mathbf{1}'\Sigma^{-1}\mathbf{1}\lambda = \mathbf{1}'\Sigma^{-1}\mu - \gamma \Leftrightarrow$$

$$\lambda = \frac{\mathbf{1}'\Sigma^{-1}\mu - \gamma}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$$

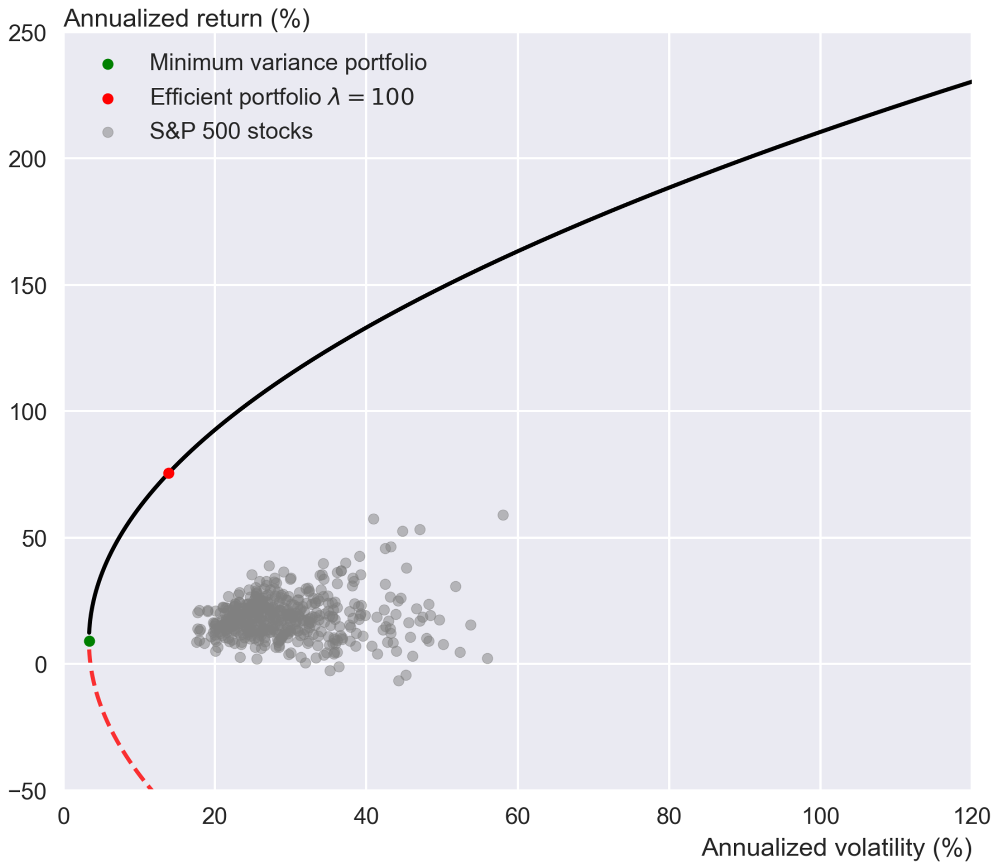
Now, we insert the expression into the weights v_t

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

$$v_t^{\text{EFF}} = v_t^{\text{MVP}} + \frac{1}{\gamma}\underbrace{\left(\Sigma^{-1}\mu - \frac{\mathbf{1}'\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\Sigma^{-1}\mathbf{1}\right)}_{(i)} \tag{3}$$

Thus, the efficient portfolio, v_t^{EFF}, consists of the minimum variance portfolio v_t^{MVP} and a self-financing portfolio *(i)*[1]

Figure 2: Efficient frontier with minimum variance portfolio and efficient portfolio



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

An investor seeking to maximize his utility will invest all available funds into the minimum variance portfolio and create a self-financing portfolio that increases the expected return of

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

the investor's portfolio and increase the risk of the portfolio too. In terms of figure 2, the less risk-averse an investor is, the further up along the efficient frontier, the investor prefer to be. Essentially preferring higher return at the cost of higher risk. An example could be an investor with $\gamma = 100$ preferring the portfolio that yield the mean variance combination of the red dot.

Intuitively it makes sense to care about the expected return of the portfolio μ. However, we do not know the true mean of returns, μ. We can attempt to estimate it with its empirical counterpart, $\hat{\mu}$. However, the neither the true mean, μ, or the empirical mean, $\hat{\mu}$, can predict the return of an asset tomorrow with any noticeable accuracy. Actually, [Jobson and Korkie, 1980] shows that the sample average 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 then the instabilities in $\hat{\mu}$ will add instabilities to the chosen weights, v_t.

Therefor, some authors prefer to use a factor-like model with factor that attempts to predict the future return of a stock. Factors like momentum, market-to-book ratio etc. While many others have used factor models famously, [Fama and French, 1993] and [Carhart, 1997] with their three and four factors models, citing statistically significant parameters for the factors, [Welch and Goyal, 2008] showed that many estimates of the parameters for the factor are mostly unstable and even prone to spurious results. Additionally, one could consider the efficient market hypothesis, stating that no information today can be used to predict returns tomorrow as other players also have that information and have traded the benefits of the information away. Thus, the explanatory power of the factors models is already taken into account by the prices of today thus given the factor models no predictive power.

All in all, no method exists to reliably estimate future returns, and even if they are successful we can only expect to gain marginally useful information from them. However, the estimation uncertainty from these methods introduce two major problems for the optimal portfolio weight, v^*. Firstly, estimation uncertainty adds instability to the optimal weights which in the real world with trading costs adds cost to the investor without adding any benefits. The reason for this is that changes to the estimated optimal weights may arise from uncertainty rather than changes in the market.

Secondly, estimation uncertainty may cause extreme long or short positions in the investors portfolio. To explain why consider two assets, i and j, where the sample mean of asset i is higher than j then the investor will long asset i and short j, all else equal. But as much literature has shown, the sample average is unstable and biased such that the difference in sample mean may be spurious or the other way around, which in the real market will hurt the investors returns badly. For these reasons we narrow the scope of the thesis and just focus on

volatility moving forwards.
Specifically, rather than maximizing expected utility, we now minimize the variance of the portfolio.

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

To solve the problem, we set up the Lagrangian where the investor is limited by the normalized weights need to sum to 1 which is the constraint $v_t'\mathbf{1} = 1$ 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 weight, 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 $v_t'\mathbf{1} = 1$, which can be used to solve for the Lagrangian multiplier λ

$$\begin{aligned} 1 =& v_t'\mathbf{1} = \mathbf{1}'v_t \Leftrightarrow \\ 1 =& \mathbf{1}'\Sigma^{-1}\mathbf{1}\lambda \Leftrightarrow \\ \lambda =& \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \end{aligned}$$

Now, we insert the expression into the weights v_t

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

$$v_t = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} = v_t^{\text{MVP}} \tag{4}$$

The Minimum Variance Portfolio (MVP) is the portfolio that has the lowest variance of all possible portfolios. To visualize this consider figure 2, where the Minimum Variance Portfolio (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 2, 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.

The mean-variance approach of Markowitz requires either normality of returns or quadratic preference of the investor to hold. Where we will use the second assumption of quadratic preferences even though we have Gaussian return for the derivations above. However, IID Gaussian returns do not mimic the process for financial time series, which we will elaborate on in the next section.

3 GARCH processes

Modelling Gaussian returns is mathematically convenient and preserves the major point of intuition behind the central results of the mean-variance approach, but it is not empirically correct. This section explores stylized facts about financial time-series and the GARCH type models that captures these features.

3.1 Stylized facts about returns

Numerous empirical studies of returns of financial time-series have revealed some stylized facts about returns. When modelling return it is important that the model of choice mimics these stylized facts.

1. *The distribution of return is non Gaussian*

The distribution of returns does not follow a Gaussian distribution as it has a higher kurtosis and fatter tails than a Gaussian distribution allows for. Thus, returns are more likely to be centered around the mean than the Gaussian distribution and more likely to be extreme events of either sign than the Gaussian distribution.

Better fits to the empirical distribution of returns includes the Generalized Error Distribution (GED) or the Student's t-distribution, both offering higher kurtosis and fatter tails than the Gaussian distribution.[2]

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

Consider the sample autocorrelation function between period period t and $t + \tau$ for $\tau > 0$

$$\hat{\rho}_{t,\tau} = \sum_{t=1}^{T-\tau}(r_t - \mathbb{E}[r_t])(r_{t+\tau} - \mathbb{E}[r_t]) \Bigg/ \sum_{t=1}^{n}(r_t - \mathbb{E}[r_t])^2 \tag{5}$$

which measure the autocorrelation i.e. the average correlation between the values of a time series in different point in time. For almost all financial timeseries, the empirical autocorrelation, $\hat{\rho}_{t,\tau}$, is insignificant and thus close to zero.[3] Thus, an autoregressive (AR) model for returns would likely be a poor fit for financial time series and past returns gives little to no information about future returns.

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

Previously, we consider autocorrelation of returns where we found no correlation, thus, one might be tempted to think that returns are *identical and independently distributed* or IID. However, this is not correct as many transformation of returns feature strong dependency across time. The most common of which is squared returns, the autocorrelation of which is

[2] [Taylor, 2011], *Asset price dynamics, volatility, and prediction*, p. 70-76

[3] [Taylor, 2011], *Asset price dynamics, volatility, and prediction*, p. 76+77

rarely insignificant.[4] This also implies *volatility clustering* of returns which means that period of higher or low volatility tend to be clustered together across time or as [Mandelbrot, 1967] simply 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, we want to built a model in which returns are not Gaussian and where returns are not modelled as independent of each other across time. There are several approach to achieve such a model, but one of the most widely used is Autoregressive Conditional Heteroskedacity or ARCH model with a non normal distribution.

3.2 Univariate GARCH models

The precursor to the GARCH model was developed by [Engle, 1982] to model and forecast variances more accurately. The resulting Autoregressive Conditional Heteroskedacity model, or ARCH for short, models the conditional variance as depending on past shocks to the time series being modelled. Consider a simple case with the return of an asset with constant mean

$$r_t = \mu + \epsilon_t \tag{6}$$

Now consider the the variance of this process is heteorskedastic and depends on the lagged shocks to the process, $\sum_{i=0}^{t-1} \epsilon_i$. The example below being the simple ARCH(1) where the conditional variance only depends on last periods shock, ϵ_{t-1}.

$$\epsilon_t = \sigma_t z_t \qquad z_t \sim IID.D(0,1) \tag{7}$$

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

with initial values taken as given and $t = 0, 1, ...T$. The restrictions on the parameters is needed to ensure strictly positive variance for all t. $D(0,1)$ is some distribution with mean zero and unit variance. This is often a Gaussian distribution for easy of computation, but other choices like a Generalized Error Distribution (GED) or a Student's t-distribution can be used with better fits to financial time series. We will use a Student's t-distribution, the probability density function of which can be found in the Appendix A.1. We will elaborate on this choice in section 5.

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

[4] [Taylor, 2011], *Asset price dynamics, volatility, and prediction*, p. 82-86

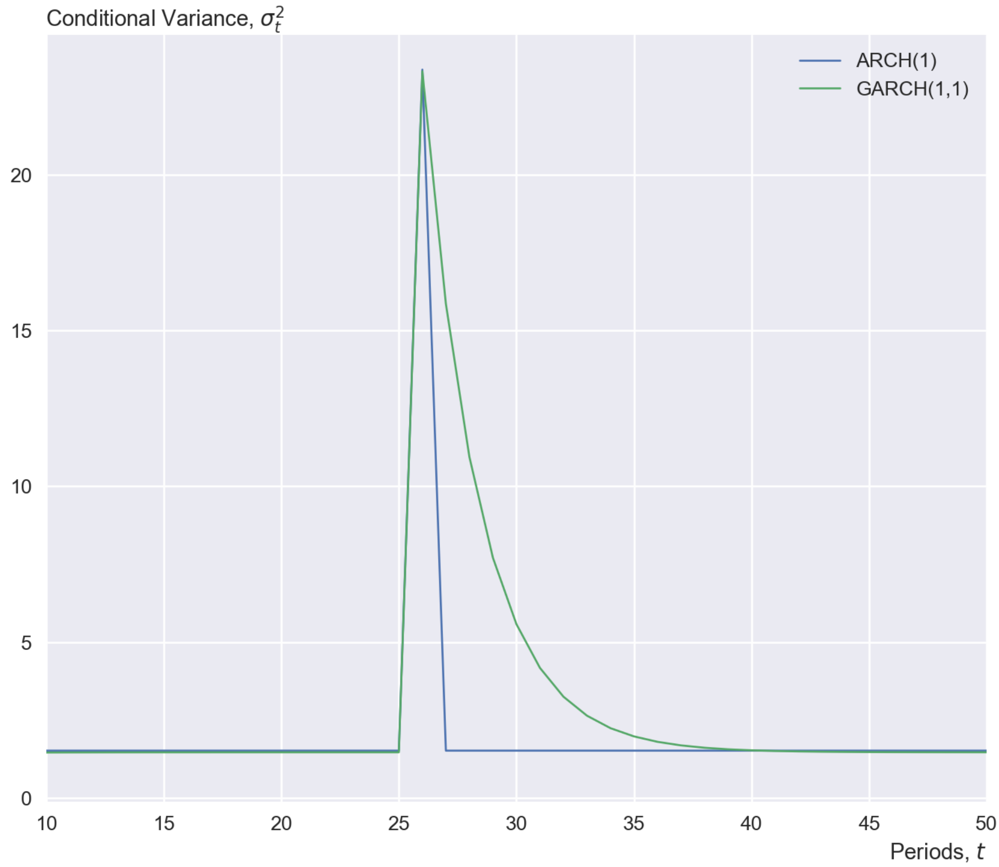
The ARCH model had one empirical weakness in that the conditional variances from the model converged back the unconditional variance after a large shock much quicker than empirical estimates indicate it should even for large lag lengths.

[Bollerslev, 1986] solved this problem with Generalized Autoregressive Conditional Heteroskedacity (GARCH) model which added persistence between the individual measures of the conditional variance across time by adding $\beta\sigma_{t-1}^2$ to the equation for the conditional variance in equation (8).

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \qquad \omega > 0, \alpha, \beta \geq 0 \tag{9}$$

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

Figure 3: Conditional Variance σ_t^2 response to a shock



Note: Parameters are chosen to give a similar unconditional variance for similar convergence.

From figure 3, we see that adding a lagged conditional variance term to the ARCH model, $\beta\sigma_{t-1}^2$, causes the conditional variance converge slower back to the unconditional variance. In the case of the ARCH(1), the conditional variance increase drastically as the shock hits and then decrease equally drastically the next period. In contrast, the GARCH(1,1) converge exponentially back to the unconditional variance, which, given weak stationarity, is given as

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

This slower converges is more inline with how the empirical variance behaves.

While the standard GARCH model does an admirable job of explaining the conditional variance of financial time series, econometricians have continuously worked to improved upon the work done by Engle and Bollerslev, like allowing for asymmetric effects of shocks which [Glosten et al., 1993] did with the GJR-GARCH 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 true and named the "leveraged effect".

Other approach exists like the Exponential GARCH (EGARCH) developed by [Nelson, 1991] which likewise captures the leverage effect with the conditional variance given as

$$\log\sigma_t^2 = \omega + \alpha(|z_{t-1}| - \mathbb{E}[|z_{t-1}|]) + \gamma z_{t-1} + \beta\log(\sigma_{t-1}^2) \qquad \omega > 0, \alpha, \beta, \kappa \geq 0 \tag{12}$$

For $\gamma < 0$ negative shocks will have a bigger impact on future volatility than positive shocks of the same magnitude. For these models to be weakly stationary, different conditions for the parameters apply, GJR-GARCH(1,1) $\alpha+\beta+0.5\kappa < 1$ with the conditions from the ARCH(1) and GARCH(1,1) contained with[5]. When the model is weakly stationary, the law of large numbers apply such that maximum likelihood estimates of the parameters will be consistent.

3.2.1 Univariate GARCH models in Portfolio theory

It should be clear by now that univariate GARCH models does far better at capturing key empirical features of financial time series than drawing IID from some distribution. An obvious question is, thus, where to apply these models?

The better fit to the empirical features should result in better predictions of the variance of the time series. Thus, instead of using the empirical estimates of the variance, $\sigma_{t+1}^2 = \hat{\sigma}^2$, we could use the GARCH model from equation (9) and set $\sigma_{t+1}^2 = \omega + \alpha\epsilon_t^2 + \beta\sigma_t^2$ as the variance of the assets and use these rather than $\hat{\sigma}^2$

Consider again the investor's problem with two asset from section 2.2.1. Now, the two asset are given by a univariate GARCH model given by equation (6), (7) and (9). The investor faces the problem in equation (2). The portfolio return and variance given are now given by

$$\mathbb{E}[B_t] = v_1\mu_1 + (1-v_1)\mu_2 \qquad \mathbb{V}[B_t] = v_1^2\sigma_{1,t+1}^2 + (1-v_1)^2\sigma_{2,t+1}^2 + 2v_1(1-v_1)\sigma_{1,t+1}\sigma_{2,t+1}\rho$$

where $\sigma_{1,t}^2$ and $\sigma_{2,t}^2$ are given by equation (9) or any other GARCH type model. Setting up the Lagrangian for the problem

$$\mathcal{L}(v_1) = v_1\mu_1 + (1-v_1)\mu_2 - \frac{\gamma}{2}\big(v_1^2\sigma_{1,t+1}^2 + (1-v_1)^2\sigma_{2,t+1}^2 + 2v_1(1-v_1)\sigma_{1,t+1}\sigma_{2,t+1}\rho\big)$$

Taking first order conditions with respect to the weight v_1

$$\frac{\partial\mathcal{L}}{\partial v_1} = \mu_1 - \mu_2 - \frac{\gamma}{2}\big(2v_1\sigma_{1,t+1}^2 - 2(1-v_1)\sigma_{2,t+1}^2 + (2-4v_1)\sigma_{1,t+1}\sigma_{2,t+1}\rho\big) = 0 \Leftrightarrow$$

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

Solving for v_1 by isolating v_1 on one side and factorizing

$$\mu_1 - \mu_2 + \gamma\sigma_{2,t+1}^2 - \gamma\sigma_{1,t+1}\sigma_{2,t+1}\rho = \gamma v_1\left(\sigma_{2,t+1}^2 + \sigma_{2,t+1}^2 - 2\sigma_{1,t+1}\sigma_{2,t+1}\rho\right) \Leftrightarrow$$

$$v_1^* = \frac{\mu_1 - \mu_2 + \gamma\sigma_{2,t+1}^2 - \gamma\sigma_{1,t+1}\sigma_{2,t+1}\rho}{\gamma\left(\sigma_{1,t+1}^2 + \sigma_{2,t+1}^2 - 2\sigma_{1,t+1}\sigma_{2,t+1}\rho\right)}$$

At first glance, the addition of the univariate GARCH model for asset 1 and 2 does not seem to change the result expect that the variances, $\sigma_{i,t+1}^2$, and by extension the standard deviations, $\sigma_{i,t+1}$, are now time varying and can be written as variables that are in the filtration, $\mathcal{F}_t$, by the GARCH equation in (9). Thus, the weight of asset 1, v_1, would now be time-varying if this problem was multi-period.

The univariate class of GARCH model is useful when modelling a single or a few assets. However, it is often more relevant to model many assets and more importantly their covariances. A natural development to the univariate GARCH models is to consider multivariate GARCH models or MGARCH.

3.3 Multivariate GARCH models

In the univariate case, we consider one assets and the volatility of this single assets. Now, consider p assets given by a $p \times 1$ vector, S_t, with a $p \times 1$ vector returns, r_t, given by a similar constant mean model now of vectors

$$r_t = \mu + \epsilon_t \tag{13}$$

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

In the univariate case, ensuring positive variance was simply to restrict a few parameters. However, for multivariate GARCH models or MGARCH, this is much more complicate. To ensure that the $p \times p$ covariance matrix, Ω_t, is indeed a covariance matrix it must positive definite. The question is thus parameterizing the model ensuring that Ω_t is positive definite for all t.

BEKK MGARCH

A mathematically simple yet empirically impractical approach is the BEKK GARCH by [Engle and Kroner, 1995] where the conditional covariance matrix resembles the univariate GARCH in form as a BEKK GARCH(1,1)

$$\epsilon_t = \Omega_t^{1/2} z_t \qquad z_t \sim IID.D(0, I_p) \tag{14}$$

$$\Omega_t(\theta) = \Omega + A\epsilon_{t-1}\epsilon_{t-1}' A' + B\Omega_{t-1}B' \tag{15}$$

with initial values taken as given and $t = 1, 2, ..., T$. Ω is positive definite and A and B are $p \times p$ dimensional matrices. θ is a vector of parameters where $\{\Omega, A, B\} \in \theta$ and D is some multidimensional distribution which could be a Multivariate Gaussian distribution for easy of computation, but other choices like a Multivariate Student's t-distribution or a Multivariate Generalized Error Distribution (GED) fit financial time series better. We have chosen to use a multivariate Student's t-distribution, the probability density function of which can be found in appendix A.1.

Note that the individual components of the matrix of parameters are next to impossible to get a meaningful interpretation of but the overall interpretation of A is completely analog to α in the univariate case and the same for B to β.
To better illustrate the workings of the model consider an example with $p = 2$

$$\Omega_t(\theta) = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \end{pmatrix} \begin{pmatrix} \epsilon_{1t-1} & \epsilon_{1t-1} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}$$
$$+ \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix}$$

A nice feature of the BEKK GARCH is that $\Omega_t(\theta)$ is positive definite for all t for any A and B and thus, the BEKK GARCH have an easy way around the paramterization problem. However, the number of parameters is $p(p+1)/2+2p^2$ meaning that the number of parameters to estimates explodes as the number of assets increase. For $p = 4$, it is 42 and for $p = 10$ it is 255, thus this model is only empirical practical when analyzing a small group of assets like $p < 10$. One way around this problem is to simplify the model in the Scalar BEKK(1,1) where A and B becomes scalars.

$$\Omega_t(\theta) = \Omega + \alpha\epsilon_{t-1}\epsilon'_{t-1} + \beta\Omega_{t-1} \qquad \alpha, \beta \geq 0 \tag{16}$$

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

$$\varrho((A \otimes A) + (B \otimes B)) < 1$$

where $\otimes$ is the tensor product and $\varrho(\cdot)$ is the spectral radius such that estimates of A and B can be estimated consistently by maximum likelihood. For the scalar-BEKK(1,1), the condition simplifies to $a + b < 1$.[6]

Similarly to univariate GARCH models, there exist multiple different Multivariate GARCH models, 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

Consider now a different variance process to the one presented above but an otherwise similar constant mean model for the returns. The DCC MGARCH use that the $p \times p$ covaraince

[6] [Engle, 1982], *Multivariate Simultaneous Generalized Arch*, proposition 2.7

matrix, Ω_t, can be decomposed into two variances matrices, Var_t, and a correlation matrix, Γ_t,

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

Where

$$\text{Var}_t = \text{diag}(\Omega_t) = \begin{pmatrix} \sigma^2_{1,t} & 0 & \cdots & 0 \\ 0 & \sigma^2_{2,t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{p,t} \end{pmatrix} \tag{18}$$

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

$$\sigma^2_{i,t} = \omega_i + \alpha_i \epsilon^2_{i,t-1} + \beta_i \sigma^2_{i,t-1} \quad \text{for} \quad i = 1, 2, ..., p \qquad \omega_i > 0, \alpha_i, \beta_i \geq 0 \tag{19}$$

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

$$\Gamma_t = \text{diag}(Q_t)^{-1} Q_t \text{diag}(Q_t)^{-1} = \begin{pmatrix} 1 & \rho_{12,t} & \cdots & \rho_{1p,t} \\ \rho_{21,t} & 1 & \cdots & \rho_{2p,t} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1,t} & \rho_{p2,t} & \cdots & 1 \end{pmatrix} \tag{20}$$

and the pseudo correlation, Q_t, 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 $\eta_t = \text{Var}_t^{-1} \epsilon_t \sim N(0, \Gamma_t)$ is a $p \times 1$ matrix of standardized disturbances of the correlations. This correlation model is weakly stationary when $(a + b) < 1$

The advantage of this is that it is possible to estimate this model in two stages. First, estimate the p univariate GARCH models for the p assets and second, estimate a multivariate scalar BEKK for the conditional correlation. The model has with a simple GARCH(1,1) and a scalar BEKK(1,1) has $3p + (p^2 - p)/2$ parameters. In comparison to the BEKK(1,1) for $p = 4$ the DCC MGARCH(1,1) has 6 parameters and 45 for $p = 10$. However, both models number of variables grows fast and thus to truly model a large number of assets, Eigenvalue MGARCH or λ-MGARCH models are preferable.

3.3.1 Multivariate GARCH models in Portfolio theory

Consider the investors problem for p assets presented in section 2.2.2. Now the p assets are given by a constant mean DCC MGARCH(1,1) model given by equation (13) and (17)-(21) with mean $\tilde{\mu}$. The investor minimize the risk given that the normalized weight sums to 1.

$$\min_{v_t} \{ \frac{1}{2} v'_t \Omega_{t+1} v_t \} \qquad \text{s.t.} \quad v'_t \mathbf{1} = 1$$

Note that the covariance matrix is now given as Ω_{t+1} which comes out of the DCC MGARCH model rather than the constant Σ. But note that Ω_{t+1} is $\mathcal{F}_t$ measurable as a result of the MGARCH model such that the investor with some accuracy can forecast Ω_{t+1} in period t, denoted $\Omega_{t+1|t}$. The Lagrangian is given as

$$\mathcal{L}(v_t) = \frac{1}{2}v_t'\Omega_{t+1|t}v_t - \lambda(v_t'\mathbf{1} - 1)$$

Taking the first order conditions and solving for v_t and λ as in section 2.2.2 yields the minimum variance portfolio after light algebra which can be found in appendix A.4.1

$$v_t = \frac{\Omega_{t+1|t}^{-1}\mathbf{1}}{\mathbf{1}'\Omega_{t+1|t}^{-1}\mathbf{1}} = v_t^{\text{MVP}} \tag{22}$$

The result is rather predictable in that $\Omega_{t+1|t}$ replaces the constant Σ. The result thus have the same interpretation. The main takeaway from this is that while it is easy to replace Σ with $\hat{\Sigma}$ and likewise for $\tilde{\mu}$, it is not necessarily the best option and other and more sophisticated options for $\tilde{\mu}$ and Σ exists. We have explored replacing Σ with the covariance matrix of a MGARCH model, $\Omega_{t+1|t}$ as they empirically fit the financial time series better. However, to truly benefit from the added complexity of GARCH models, we need to consider multi-period problem or dynamic trading strategies.

4 Dynamic Trading Strategies

Start with presenting the framework, where the investor can choose between two assets. Then extend the model and present the new solution. Compare the two and explain the intuition.

4.1 Dynamic programming by the Bellman Equation

Dynamic programming was by pioneered by [Bellman, 1966] to solve multi-period optimization problems. We have chosen to more closely follow the notation from [Gan and Lu, 2014] and [Gârleanu and Pedersen, 2013].

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

$$\max_{\{x_t\}_{t=0}^{\infty}} \mathbb{E}_0\left[\sum_{t=0}^{\infty}(1-\rho)^{t+1}f(x_t, s_t)\right] \quad \text{s.t.} \quad g(x_t, s_t) = 0 \tag{23}$$

Note that $\mathbb{E}_0[\cdot]$ is the conditional expectation given that filtration in period 0 i.e. $\mathbb{E}_0[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_0]$. The tools of dynamic programming that Bellman developed decomposes the multi-period problem into of two period models, "now" and "later". Recall that maximization and minimization are closely related such that a maximization problem can be turn into a minimization problem.[7], thus equation (23) is equivalent to

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

Consider an agent facing the minimization problem in (24). This could be to minimize the portfolio variance. Note we can write the sum as (i) and (ii), taking the first period, $t=0$, out of the sum.

$$\begin{aligned}
L_0 &= \min_{\{x_t\}_{t=0}^{\infty}} -\mathbb{E}_0\left[\sum_{t=0}^{\infty}(1-\rho)^{t+1}f(x_t,s_t)\right] - \lambda[g(x_t,s_t)] \\
&= \min_{\{x_t\}_{t=0}^{\infty}} -\mathbb{E}_0\left[\underbrace{(1-\rho)f(x_0,s_0)}_{(i)} + \underbrace{\sum_{t=1}^{\infty}(1-\rho)^{t+1}f(x_t,s_t)}_{(ii)}\right] - \lambda[g(x_t,s_t)] \\
&= \min_{\{x_t\}_{t=0}^{\infty}} -\mathbb{E}_0\left[(1-\rho)\left(f(x_0,s_0)+\mathbb{E}_1\left(\sum_{t=1}^{\infty}(1-\rho)^{t+1}f(x_t,s_t)\right)\right)\right] - \lambda[g(x_t,s_t)] \\
&= \min_{x_0} -\mathbb{E}_0\left[(1-\rho)\left(\underbrace{f(x_0,s_0)}_{(j)} + \underbrace{\min_{\{x_t\}_{t=1}^{\infty}} \mathbb{E}_1\left(\sum_{t=1}^{\infty}(1-\rho)^{t+1}f(x_t,s_t)\right)}_{(jj)}\right)\right] - \underbrace{\lambda[g(x_t,s_t)]}_{(jjj)}
\end{aligned}$$

Here we have the immediate reward, *(j)*, future reward, *(jj)*, under some constraint, *(jjj)*

$$= -\min_{x_t}\left[(1-\rho)f(x_0,s_0) + \min_{\{x_t\}_{t=1}^{\infty}} \mathbb{E}_0\left(\sum_{t=1}^{\infty}(1-\rho)^{t+1}f(x_t,s_t)\right)\right] - \lambda[g(x_t,s_t)]$$

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

$$V(s_0) = -\min_{x_0}\Big[(1-\rho)f(x_0,s_0)+\mathbb{E}_0[V(s_1)]\Big] - \lambda[g(x_t,s_t)]$$

Generalizing the value function to period t, we have

$$V(s_t) = -\min_{x_t}\left[\underbrace{(1-\rho)f(x_t,s_t)}_{(j)} + \underbrace{\mathbb{E}_t[V(s_{t+1})]}_{(ii)}\right] - \underbrace{\lambda[g(x_t,s_t)]}_{(iii)}$$

The value function captures the trade off between immediate reward, *(j)* future reward, *(jj)* under some constraint *(jjj)*. Given the value function, the optimal policies are the solutions to the optimization problem contained within the Bellman's equation, $V(s_t)$. The policies may be different real world objects of interest for instance, the optimal interest rate, the optimal

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

level of production for some firm or as for this paper, the optimal weight of wealth allocated into different assets. For our purposes, we will consider the following value function

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

This value function $V(v_{t-1})$ measures the value of entering period t with a portfolio of v_{t-1} assets

4.2 Many risky assets and no trading cost

Consider p assets given by a $p \times 1$ vector, S_t, with returns $r_t = (r_{1,t}, r_{2,t}, ..., r_{p,t})'$ which by the investor are assumed given by a multivariate Gaussian distribution $N_p(\mu, \Sigma)$.
An investor with mean-variance preferences given equation (1) seeks a dynamic trading strategy $\{v_t\}_{t=0}^{\infty}$ that solves the multi-period minimization problem presented in equation (23):

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

with $f(x_t, s_t) = \frac{1}{2} v_t \Sigma v_t'$. Multiplying the (-1) into minimization problem yields

$$\min_{\{v_t\}_{t=0}^{\infty}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} (1-\rho)^{t+1} \left(\frac{1}{2} v_t \Sigma v_t' \right) \right] \quad \text{s.t.} \quad v_t'\mathbf{1} = 1 \tag{25}$$

The problem can be solved via dynamic programming using the Bellman equation using the method in section 4.1. Consider an investor with preference as (1) and facing the minimization problem in (25), denoted L_t

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

Following the same argument as in section 4.1 we get

$$\begin{aligned}
&= \min_{v_0} \mathbb{E}_0 \left[(1-\rho)^1 \left[\frac{1}{2} v_0 \Sigma v_0' + \min_{\{v_t\}_{t=1}^{\infty}} \mathbb{E}_1 \left\{ \sum_{t=1}^{\infty} (1-\rho)^{t+1} \left(\frac{1}{2} v_t \Sigma v_t' \right) \right] \right\} \right] - \lambda(v_t'\mathbf{1} - 1) \\
&= \min_{v_0} \left[(1-\rho) \frac{1}{2} v_0 \Sigma v_0' + (1-\rho) \min_{\{v_t\}_{t=1}^{\infty}} \mathbb{E}_0 \left\{ \sum_{t=1}^{\infty} (1-\rho)^{t+1} \left(\frac{1}{2} v_t \Sigma v_t' \right) \right\} \right] - \lambda(v_t'\mathbf{1} - 1)
\end{aligned}$$

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

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

The value function (or Bellman equation) measures the value of the portfolio at time t with weight of v_{t-1} of the risky assets. Additionally, the value function captures the trade-off

between utility for this period which becomes clear as the static one period problem appears in *(i)* and future utility *(ii)* under the constraint the weights sum to 1, *(iii)*.
Solving for the optimal weight requires solving the first order conditions wrt. to the weights v_t. Now, the first order conditions are given by

$$\frac{\partial V(v_{t-1})}{\partial v_t} = (1-\rho)v_t\Sigma - \lambda\mathbf{1} = 0 \Leftrightarrow v_t = (1-\rho)^{-1}\Sigma^{-1}\lambda\mathbf{1}$$

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

$$\begin{aligned} 1 =& v_t'\mathbf{1} = \mathbf{1}'v_t = \mathbf{1}'(1-\rho)^{-1}\Sigma^{-1}\lambda\mathbf{1} = (1-\rho)^{-1}\mathbf{1}'\Sigma^{-1}\mathbf{1}\lambda \\ \lambda =& \frac{1}{(1-\rho)^{-1}\mathbf{1}'\Sigma^{-1}\mathbf{1}} \end{aligned}$$

Now, we insert the expression into the weights v_t

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

$$v_t = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \equiv v_t^{\mathrm{MVP}} \tag{26}$$

Note that given that variance, Σ, is assumed constant by the investor, the weight, v_t, are likewise constant. One could interpret this as an investor who rarely if ever updates his beliefs or information set and thus, would have no reason to update the weights of his portfolio. Thus, the optimal weight are identical to the static case in section 2.2.2 as the investor essentially solves infinitely many static allocation problems. Alternatively, one could consider that Σ is a estimated by its empirical counterpart, but as a rolling window estimation, such that the investor updates his weights whenever he updates his estimate for the covariance matrix, Σ.

Now consider the case where the investor have a different belief about the data generating process where the investor believes that the returns $r_t = (r_{1,t}, r_{2,t}, ..., r_{p,t})'$ are given by a constant mean DCC MGARCH(1,1) model given by equation (13) and (17)-(21). The investor now faces a slightly different minimization problem.

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

The solution to the problem is almost identical to the case with Gaussian returns as such the full problem can be found in appendix A.5.1. The value function of the problem is

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

where the optimal weight can be found from the first order conditions of the value function wrt. the weights, v_t and $\mathbb{E}_t[\Omega_{t+1}]$ can be given as the forecast value $\Omega_{t+1|t}$. After some light algebra, the optimal weights are given as

$$v_t = \frac{\Omega_{t+1|t}^{-1}\mathbf{1}}{\mathbf{1}'\Omega_{t+1|t}^{-1}\mathbf{1}} = v_t^{\mathrm{MVP}} \tag{28}$$

Note that given that covariance matrix, Ω_t is time-varying the weight, v_t, are likewise time-varying. Now the investor updates his beliefs about the future variances and covariance i.e. the risk and as such can make change to his portfolio and continuously update the optimal weights. Note that Ω_{t+1} can be written as variables that are $\mathcal{F}_t$-measurable i.e. the investor can with some accuracy forecast the covariance matrix within the near future or at the very least the investor believes he can forecast the covariance matrix one period into the future.

This result is not surprising as in the absences of trading cost, the investor can rebalance the portfolio each period to the new minimum variance portfolio at no cost. This is inline with [Gârleanu and Pedersen, 2013] and [Mei and Nogales, 2018] who both find that the aim portfolio i.e. the portfolio the investor aims to reach, is the Markowitz portfolio which corresponds to the minimum variance portfolio when only considering minimizing risk.

4.3 Many risky assets and trading costs

Consider p assets given by a $p \times 1$ vector, S_t, with returns with returns given by a constant mean DCC MGARCH(1,1) model given by equation (13) and (17)-(21). However, now the investor faces cost from trading and thus faces the following problem when seeking a dynamic investment strategy:

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

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

$$\Lambda_t = \Omega_t \gamma_D$$

meaning that trading costs are time-varying as Ω_t is time-varying and γ_D is the dealers risk-aversion. This might initially seem odd but to understand why it makes sense consider the dealer taking the opposite side of the trade that our investor makes. The dealer will hold this position for a period and then sell back to the market. In the period, the dealer hold the position, he holds risk equivalent to $\Delta v_t' \Omega_t \Delta v_t$. The trading cost can thus be interpreted as compensation for the dealers risk depending how risk averse the dealer is, γ_d. We solve this problem by using the method in section 4.1. Consider an investor with preferences given by equation (1), facing the maximization problem in (29), denoted L_t

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

Following the same argument as in section 4.1 we get

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

Generalizing to period t, we get that the value function, $V(v_t)$ is given by

$$
V(v_{t-1}) = \min_{v_t} \left[\frac{1}{2} (v_t - v_{t-1})' \Lambda_t (v_t - v_{t-1}) + (1-\rho) \left(\frac{1}{2} v_t' \mathbb{E}_t [\Omega_{t+1}] v_t + \mathbb{E}_t [V(v_t)] \right) \right] - \lambda_t (v_t' \mathbf{1} - 1) \tag{30}
$$

We note that $\mathbb{E}_t[\Omega_{t+1}]$ can be given as the forecast of Ω_{t+1}. Generally, Ω_{t+h} can be forecast with $\mathcal{F}_t$-measurable variables using equation (17)-(21) and recursive substitution which in this case is relevant for only $h = 1$. Let the forecast of Ω_{t+h} given $\mathcal{F}_t$ be denoted $\Omega_{t+h|t}$, similar so for $\mathrm{Var}_{t+h|t}$ and $\Gamma_{t+h|t}$. Recall that the conditional covariance matrix, Ω_{t+1}, can be decomposed into the conditional variance and conditional correlation matrices

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

Where Var_{t+1} given period t can be forecast as

$$
\mathrm{Var}_{t+1|t} = \mathrm{diag} \begin{pmatrix} \sigma^2_{1,t+1|t} \\ \sigma^2_{2,t+1|t} \\ \vdots \\ \sigma^2_{p,t+1|t} \end{pmatrix} = \mathrm{diag} \begin{pmatrix} \omega_1 + \alpha_1 \mathbb{E}_t[\epsilon^2_{1,t}] + \beta_1 \sigma^2_{1,t} \\ \omega_2 + \alpha_2 \mathbb{E}_t[\epsilon^2_{2,t}] + \beta_2 \sigma^2_{2,t} \\ \vdots \\ \omega_p + \alpha_p \mathbb{E}_t[\epsilon^2_{p,t}] + \beta_p \sigma^2_{p,t} \end{pmatrix}
$$

where $\mathbb{E}_t[\epsilon^2_{i,t}] = \epsilon^2_{i,t}$ as ϵ_t is known in period t.
Similarly for $\Gamma_{t+1|t} = \mathrm{diag}(Q_{t+1|t})^{-1} Q_{t+1|t} \mathrm{diag}(Q_{t+1|t})^{-1}$ which can be 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. As such Ω_{t+1} can be forecast with period t variables.

In contrast to the problems without trading costs in section 4.2, $\mathbb{E}_t[V(v_t)]$ also becomes relevant as trading costs add persistence between periods. Thus, we need to find an expression for $V(v_t)$. Following the method of [Gârleanu and Pedersen, 2013], we apply the 'guess and verify' method, which can divided into 6 steps:

1. Make a guess of 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 optimal policy (the weights)
4. Insert optimal policy (the weights) into the value function
5. Compare the new value function with the guessed one and verify it solves the problem
6. Solve for coefficients

We guess a solution for $V(v_t)$ which is given by

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

We evaluate the conditional expectation of the guessed value function, assuming that the forecast is the expectation

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

Inserting this into our value function in equation (30) yields the problem:

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

We redefine terms and see the investor minimizes the following quadratic problem:
$V(v_{t-1}) = (1-\rho)\big(\frac{1}{2}v_t' J_t v_t - v_t' j_t - d_t\big) - \lambda_t(v_t'\mathbf{1} - 1)$ with

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

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

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

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

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

Insert this back into the problem:

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

Recall the rewritten version of the quadratic problem $V(v_{t-1})$:

$$V(v_{t-1}) = (1-\rho)\big(\frac{1}{2}v_t' J_t v_t - v_t' j_t - d_t\big) - \lambda_t(v_t'\mathbf{1} - 1)$$

Insert the expression for v_t into $V(v_{t-1})$ and define $\bar{\lambda}_t = (1-\rho)^{-1}\lambda_t$: (Ignore λ_t for the time being)

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

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

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

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

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

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

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

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

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

now, we insert for j_t and d_t

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

Consider ($\mathbb{I}$)

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

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

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

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

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

This implies that the following restriction on the coefficient matrices must hold

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

Compare this to the guessed value function

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

We see that this is indeed a solution

A Appendix

A.1 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 Student's t random variable can be defined by

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

with the probability density function (pdf) given as

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

with ν degrees of freedom which determines the fatness of the tails and the number of moment 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 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

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

with ν degrees of freedom which determines the fatness of the tails and the number of moment which are finite. p is the number of dimensions. Note here that Σ is the covariance matrix of the multivariate normal distribution.

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 which determines the fatness of the tails and the number of moment 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 non-central normal distribution (See [Li and Nadarajah, 2020] for more).

A.3 Multi Assets Problem

A.3.1 MVP with return target

Insert the tendious algrebra from the multi asset problem

A.4 GARCH Asset Problems

A.4.1 Multi Asset Problem

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.5 Dymanic problems

A.5.1 Many GARCH Assets

Consider p assets given by a $p \times 1$ vector $S_t = (S_{1,t}, S_{2,t}, ..., S_{p,t})'$ with returns $r_t = (r_{1,t}, r_{2,t}, ..., r_{p,t})'$ given by a constant mean DCC MGARCH(1,1) model given by equation (13) and (17)-(21).

An investor with mean-variance preferences given equation (1) seeks as dynamic trading strategy $\{v_t\}_{t=0}^{\infty}$ which comes out of a multi-period maximization problem:

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

with $f(x_t, s_t) = \frac{1}{2}v_t\Omega_{t+1}v_t'$. We transform the maximization problem into a minimization problem

$$\min_{\{v_t\}_{t=0}^{\infty}} \mathbb{E}_0\left[\sum_{t=0}^{\infty}(1-\rho)^{t+1}\left(\frac{1}{2}v_t'\Omega_{t+1}v_t\right)\right] \quad \text{s.t.} \quad v_t'\mathbf{1}=1 \tag{31}$$

The problem can be solved via dynamic programming using the Bellman equation using the method in section 4.1. Consider an investor with preferences as (1) facing the minimization problem in (31)

$$L_0 = \min_{\{v_t\}_{t=0}^{\infty}} \mathbb{E}_0\left[\sum_{t=0}^{\infty}(1-\rho)^{t+1}\left(\frac{1}{2}v_t'\Omega_{t+1}v_t\right)\right] \quad \text{s.t.} \quad v_t'\mathbf{1}=1$$

Following the same argument as in section 4.1 we get

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

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

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

Solving for the optimal weight requires solving the first order conditions wrt. to the weights v_t. Now, the first order conditions are given by

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

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

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

Now, we insert the expression into the weights v_t

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

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

Note that given that covariance matrix, $\Omega_{t+1|t}$ is time-varying the weight, v_t, are likewise time-varying.