Advanced Empirical Finance - Topics in Data Science

Exam

June 7 2021

Exam part 1 - MA3

Exercise 1

We consider the portfolio maximization problem given by

$$\omega_{t+1}^{*} = \arg \max_{\omega_{t+1} \in \mathbb{R}^N, \iota' \omega_{t+1} = 1} \omega_{t+1}' \mu - \nu_t \left(\omega_{t+1}, \omega_{t^+}, \beta \right) - \frac{\gamma}{2} \omega_{t+1}' \Sigma \omega_{t+1}$$

where μ is the estimator of returns, Σ is the estimator of the covariance matrix and γ is the risk aversion parameter. The transaction costs are assumed to be quadratic in rebalancing and proportional to the stock illiquidity.

$$\nu_t \left(\omega_{t+1}, \omega_{t^+}, \beta \right) = \frac{\beta}{2} \left(\omega_{t+1} - \omega_{t^+} \right)' B \left(\omega_{t+1} - \omega_{t^+} \right)$$

The quadratic transaction costs function includes the Amihud measure as a measure of illiquidity of the assets expressed as the vector B. As Hautsch et. al. (2019) shows the inclusion of quadratic transaction costs can be interpreted as shrinkage of the covariance matrix towards a diagonal matrix. This improves portfolio allocation, as the shrinkage effects improves stability of the covariance estimates and reduce the frequency of rebalancing, implying a reduction in turnover costs. In addition the transaction cost are proportional to the illiquidity measure, hence we would expect transaction costs to be larger, the less liquid the given asset is.

We plug it into the maximization problem and rearrange terms:

$$\begin{aligned}
\omega_{t+1}^{*} &= \arg \max_{\omega_{t+1} \in \mathbb{R}^N, \iota' \omega_{t+1} = 1} \omega_{t+1}' \mu - \frac{\beta}{2} \left(\omega_{t+1} - \omega_{t^+} \right)' B \left(\omega_{t+1} - \omega_{t^+} \right) - \frac{\gamma}{2} \omega_{t+1}' \Sigma \omega_{t+1} \\
&= \arg \max_{\omega_{t+1} \in \mathbb{R}^N, \iota' \omega_{t+1} = 1} \omega_{t+1}' \mu - \frac{\beta}{2} \omega_{t+1}' B \omega_{t+1} + \beta \omega_{t+1} B \omega_{t^+} - \frac{\beta}{2} \omega_{t^+} B \omega_{t^+} - \frac{\gamma}{2} \omega_{t+1}' \Sigma \omega_{t+1} \\
&= \arg \max_{\omega_{t+1} \in \mathbb{R}^N, \iota' \omega_{t+1} = 1} \omega_{t+1}' \left(\mu - \beta B \omega_{t^+} \right) - \frac{\gamma}{2} \omega_{t+1}' \left(\Sigma + \frac{\beta}{\gamma} B \right) \omega_{t+1} \\
&= \arg \max_{\omega_{t+1} \in \mathbb{R}^N, \iota' \omega_{t+1} = 1} \omega_{t+1}' \mu_B^{*} - \frac{\gamma}{2} \omega_{t+1}' \Sigma_B^{*} \omega_{t+1}
\end{aligned}$$

where $\mu_B^{*} = \mu - \beta B \omega_{t^+}$ and $\Sigma_B^{*} = \Sigma + \frac{\beta}{\gamma} B$. From the rearrangement of terms it follows that the maximization problem with incorporated ex-ante transaction costs, boils down to the Markowitz mean-variance problem without transaction costs, if we adjust the expressions for μ and Σ. We find the optimal portfolio weights by solving the Lagrangian for the maximization problem:

$$\begin{aligned}
\mathcal{L}(\omega_{t+1}) &= \omega_{t+1}' \mu_B^{*} - \frac{\gamma}{2} \omega_{t+1}' \Sigma_B^{*} \omega_{t+1} - \lambda(\iota' \omega_{t+1} - 1) \\
\frac{\partial \mathcal{L}(\omega_{t+1})}{\partial \omega_{t+1}} &= \mu_B^{*} - \gamma \Sigma_B^{*} \omega_{t+1} - \lambda \iota = 0 \quad \Leftrightarrow \omega_{t+1} = \frac{1}{\gamma} \Sigma_B^{*-1} (\mu_B^{*} - \lambda \iota) \\
\frac{\partial \mathcal{L}(\omega_{t+1})}{\partial \lambda} &= \iota' \omega_{t+1} - 1 = 0 \qquad\qquad\quad \Leftrightarrow \iota' \omega_{t+1} = 1
\end{aligned}$$

Combining the FOC's yields the following expression for λ which can be reinserted into the first FOC and rearranged in order to find the optimal portfolio weights:

$$\begin{aligned}
1 &= \frac{1}{\gamma} (\iota' \Sigma_B^{*-1} \mu_B^{*} - \lambda \iota' \Sigma_B^{*-1} \iota) \\
\lambda &= \frac{1}{\iota' \Sigma_B^{*-1} \iota} (\iota' \Sigma_B^{*-1} \mu_B^{*} - \gamma) \\
\omega_{t+1} &= \frac{1}{\gamma} \Sigma_B^{*-1} (\mu_B^{*} - \left(\frac{1}{\iota' \Sigma_B^{*-1} \iota} (\iota' \Sigma_B^{*-1} \mu_B^{*} - \gamma) \iota \right) \\
\omega_{t+1}^{*} &= \frac{1}{\gamma} \left(\Sigma_B^{*-1} - \frac{1}{\iota' \Sigma_B^{*-1} \iota} \Sigma_B^{*-1} \iota \iota' \Sigma_B^{*-1} \right) \mu_B^{*} + \frac{1}{\iota' \Sigma_B^{*-1} \iota} \Sigma_B^{*-1} \iota
\end{aligned}$$

If comparing weights of a myopic investor disregarding transaction cost, the weight on an asset with a higher illiquidity measure (low liquidity of the given asset), will be higher for a investor who takes transaction costs into account, if we consider a price drop of the asset. This is due to the fact that is becomes more expensive to rebalance portfolio weights, the less liquid and in turn higher transaction cost are, for the given asset.

The function computing the optimal weights is found in the Rmarkdown file.

Exercise 2

As in Hautsch et al (2019), if we denote the initial wealth allocation as ω_0, we are able to study the long run effect, by examining the sequential rebalancing:

$$\omega_T = \sum_{i=0}^{T-1} \left(\frac{\beta}{\gamma} A(\Sigma_B^*) \right)^i \omega(\mu, \Sigma_B^*) + \left(\frac{\beta}{\gamma} A(\Sigma_B^*) \right)^T \omega_0$$

where $A(\Sigma_B^*) = \left(\Sigma_B^{*-1} - \frac{1}{\iota' \Sigma_B^{*-1} \iota} \Sigma_B^{*-1} \iota \iota' \Sigma_B^{*-1} \right)$. The equation shows that ω_T can be interpreted as the weighted average of the mean variance efficient portfolio weights and the initial allocation. Proposition 2 of Hautsch et al (2019) states that for infinitely large transaction costs the long run weights will be equal to the initial allocation. However proposition 3 states that there exist a range for β below a certain threshold β^* where the initial allocation can be ignored in the long run. Using $T \to \infty$ and $\beta < \beta^*$ the series of sequential rebalancing converges to a unique fix point given by the weights computed with no regard to the transaction costs as finally stated in Proposition 4 of hautsch et al (2019).

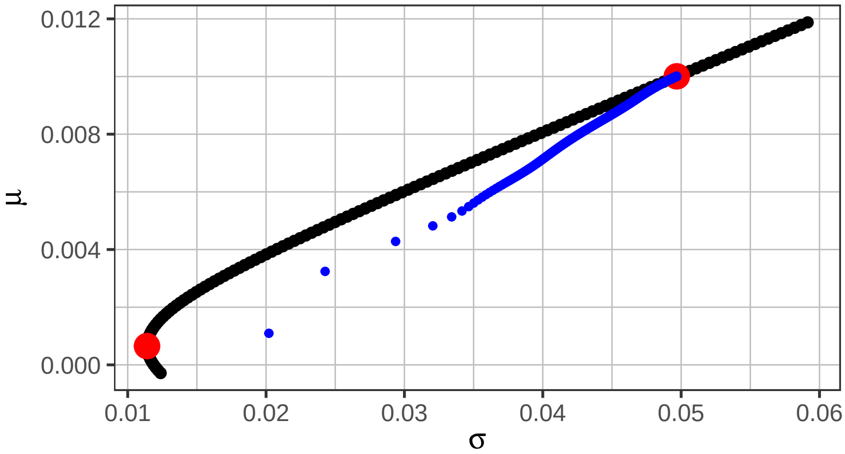
$$\omega_\infty = \left(I - \frac{\beta}{\gamma} \Sigma_B^* \right)^{-1} \omega(\mu, \Sigma_B^*) = \omega(\mu, \Sigma)$$

As stated in the article this shows that for large transaction costs above the threshold, the investor will not reallocate to the efficient portfolio. However if the transaction costs are moderate, we can obtain convergence towards the efficient portfolio of a setup where transaction costs are ignored.

In this exercise we show the convergence of the portfolio towards the efficient frontier as derived in proposition 4 when $T \to \infty$. The weights are computed using the function defined in exercise 1, and are computed using a loop over 10^4 values. μ is computed as the mean of each stock return using the full sample while Σ is computed with the Ledoit-Wolf shrinkage covariance estimator. We use the Ledoit-Wolf covariance estimator to estimate the covariance among the assets as proposed by Ledoit and Wolf (2003). As stated by Ledoit and Wolf (2003) the sample covariance matrix is unbiased but unfortunately often associated with a lot of estimation error when the number of observations is not significantly larger than the number of individual stocks. In our case we have 40 assets and 500 observations. However we would argue that estimation error may still be present and follow the advice of Ledoit and Wolf (2003) and avoid using the sample covariance matrix.

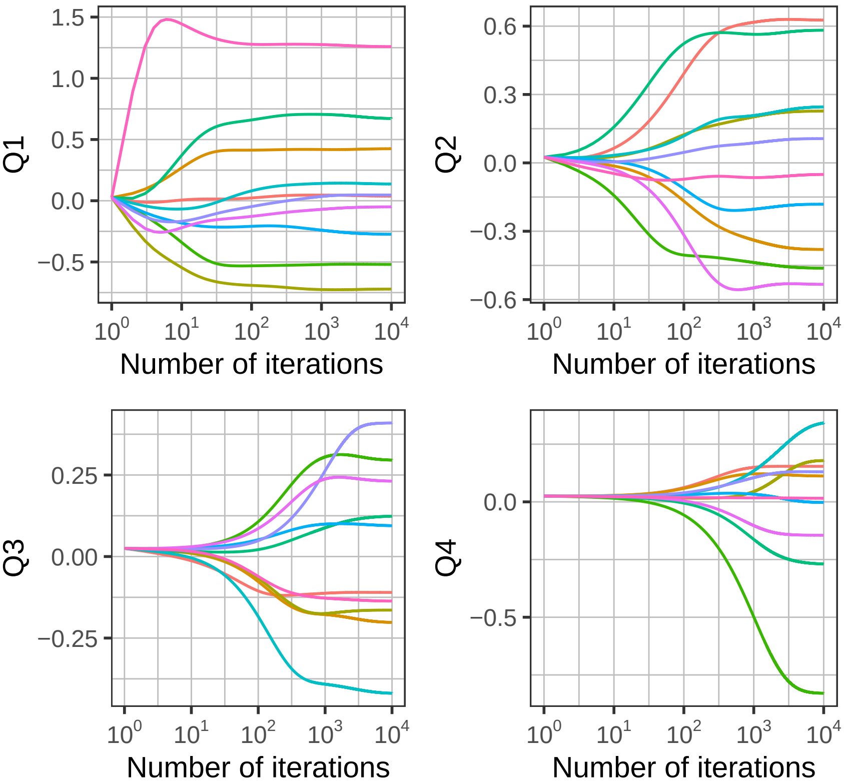
Below the convergence towards the efficient frontier is shown. The figure shows the convergence starting from the naive portfolio towards the theoretically efficient portfolio. The reallocation is notably larger in initial iterations, and the adjusting of portfolio weights decrease in speed as we loop through the iterations. We use $\beta = 1$ in this exercise.

Figure 1: Convergence towards efficient portfolio



In order to illustrate the weight dynamics towards the efficient portfolio we divide the stocks into quartiles based on the measure of illiquidity. This enables os to show the difference in convergence speed towards the efficient portfolio weights for the different groups of assets. The first quartile ($Q1$) are the most liquid assets, while the last quartile ($Q4$) are the assets with the highest illiquidity measure, hence the least liquid assets. The x-axis follows logarithmic scale for a better overview of the difference in convergence speeds.

Figure 2: Effect of illiquidity measure on weight dynamics



The plot shows a clear trend of decreasing speed in convergence towards the efficient weights as the illiquidity measure increase. The most liquid assets of the first quartile $Q1$ seem to reach the efficient weights almost immediately, while the less liquid the assets are, the slower the convergence is as seen from the latter quartiles.

Exercise 3

In this exercise we illustrate the effect of different values of the transaction cost parameter β in the optimization problem on the annualized out-of-sample Sharpe ratios. In the optimization problem we consider values of β between 0 and 100. However, when we evaluate the performance of the portfolio, this is done using a transaction cost parameter equal to the actual transaction cost. The true transaction cost parameter is assumed to be $\beta_{true} = 1$.

We consider two ways of estimating the variance-covariance matrix, $\hat{\Sigma}$, which is the sample covariance estimator and the shrinkage estimator proposed by Ledoit and Wolf (2003, 2004). However, we do not estimate the mean vector and simple set it to zero.

More specifically, we for each considered value of β do a backtest using a rolling window with a length of 250 days. We use the naive portfolio as the initial allocation and then each day re-estimate Σ and implement the optimal portfolio taking transaction cost into account. Since our data set consist of 500 observations for each asset and we set the rolling window length to 250, it follows that we for each β estimate a time series of optimal portfolios ω_t^β of length $500 - 250 = 250$. For each of the periods we calculate the realized gross portfolio return as:

$$r_t^\beta = r_t'\omega_t^\beta$$

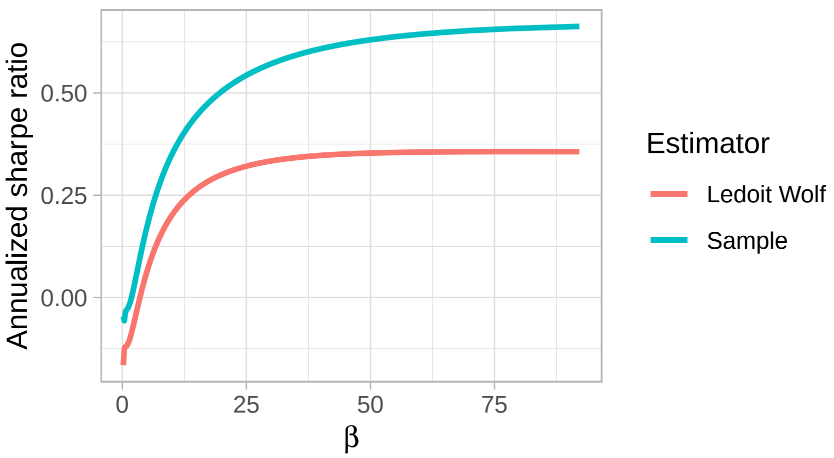
and the portfolio return net of transaction costs as, where we use the "actual" transactions cost parameter:

$$r_t^{\beta,nTC} = r_t^\beta - \beta_{true}||\omega_{t+1}^\beta - \omega_{t+}^\beta||_2$$

where the L_2-norm is calculated as $||\omega_{t+1}^\beta - \omega_{t+}^\beta||_2 = (\omega_{t+1} - \omega_{t+})'B(\omega_{t+1} - \omega_{t+})$ such that it takes the asset-specific transaction costs into account.

For each β we can then calculate the average net return and standard deviation, $\hat{\mu}^{\beta,pf}$, $\hat{\sigma}^{\beta,pf}$ based on the time series of realized net returns. The Sharpe-ratio is then calculated as the ratio between these. Results for both estimators of the the covariance matrix is reported below:

Figure 3: Annualized Sharpe ratios for different transaction costs parameter values



From the figure above we note, that the sample estimator of the covariance matrix clearly outperforms the Ledoit and Wolf shrinkage estimator, since it delivers higher out-of-sample Sharpe ratios for all β-values between 0 and 100.

On the one hand the transaction cost parameter improve the conditioning of the covariance matrix, but it also reduce the mean portfolio return. A change in the transaction cost parameter, β, therefore has an ambiguous effect of the Sharpe ratio. As evident from the figure, the annualized Sharpe ratio is increasing in the transaction cost parameter, β, for both estimators over the interval from 0 to 100. This means that positive effect from conditioning the covariance matrix dominates the negative effect from decreasing the mean. However, this overall positive effect on the Sharpe ratio seem to be largest for increases in small values of β, so the effect of further penalizing turnover vanishes as β goes to 100.

We assumed that the β-value reflecting the actual transaction costs is $\beta_{true} = 1$. When we use this value in the optimization problem it does not deliver the highest possible out-of-sample annualized Sharpe ratio. This means that it can be optimal to choose theoretically sub-optimal portfolios based on transaction cost parameter values that do not reflect the actual transaction costs. If we instead define our objective as maximizing the annualized out-of-sample Sharpe-ratio it is therefore tempting to view β as a parameter to be optimized, since our results indicate that it increases the Sharpe ratio if we choose a β-value which is higher than the true one.

Exercise 4

In the last exercise we perform a full-fledged backtesting strategy with proportional L_1 transactions costs. We compare the out-of-sample performance of a naive portfolio where the 40 assets are equally weighted; a portfolio which computes the theoretically optimal portfolio weights with optimal ex-ante adjustment for the L_1 transaction costs in the spirit of Hautsch et al. (2019) and a minimum variance portfolio with short sell constraint as suggested by Jagannathan and Ma (2003).

$$\nu_t\left(\omega_{t+1}, \omega_{t+}, \beta\right) = \beta \left\|\omega_{t+1} - \omega_{t+}\right\|_1 = \beta \sum_{i=1}^{N} \left|\omega_{i,t+1} - \omega_{i,t+}\right|$$

With L_1 transactions costs the cost is proportional to the sum of absolute re-balancing and thereby represent a more realistic proxy than the quadratic transaction costs we use in the exercises above. As is common in the literature we use the penalization term β of 50 bp.

To perform the portfolio backtesting strategy we start by defining functions that compute the optimal weights for each of the three portfolios. The naive portfolio places equal weights on each of the assets.

$$\omega = \frac{1}{N}\iota$$

In our cases with 40 assets each assets has a weight of 0.025.

As suggested by Hautsch et al. (2019) we compute the theoretical optimal portfolio weights with optimal ex-ante adjustment for the transaction costs.

$$\omega_{t+1}^{*} = \max_{\omega \in \mathbb{R}^N, \iota\omega = 1} \omega'\mu - \beta \left\|\omega_{t+1} - \omega_{t+}\right\|_1 - \frac{\gamma}{2}\omega'\Sigma\omega \text{ s.t. } \iota'\omega = 1$$

It should be noted that with L_1 transaction cost there does not exist a closed form solution, therefore we solve the optimization problem using a non-linear constraint optimizer. Jagannathan and Ma (2003) show that constructing a minimum variance portfolio with short-selling constraints is equivalent to shrinking the larger elements of the covariance matrix. When computing the the portfolio weights of the short-sell constrained minimum variance portfolio we face the following optimization problem:

$$\omega_{t+1}^{\text{mvp no s.}} = \min_{\omega \in \mathbb{R}^N} \omega'\Sigma\omega \text{ s.t. } \iota'\omega = 1 \text{ and } \omega_i \geq 0 \forall i = 1, \ldots N$$

As explained by Jagannathan and Ma (2003) introducing short-selling constraint may as other shrinking methods lead to a more precise estimate of true population covariance given that a large covariance is due to sampling error. However if the true population covariance is indeed large, short sell constraints and shrinkage will result in specification error.

We then use these functions to perform a out-of-sample backtest and evalute the performance of the three strategies. In our backtest we use a window-length of 250. The mean is computed as the sample mean of the returns, while the covariance is estimated using a Ledoit-Wolf shrinkage estimation.

To compare the performance of the three strategies we compute the annualized out-of-sample means, standard deviations, Sharpe ratios and turnover. The returns are the raw returns net of trading costs.

Table 1: Performance of portfolio strategies

strategy	Mean	SD	Sharpe	Turnover
Naive	0.3217	0.4236	0.7594	2.0390
MV (TC)	0.2695	0.4021	0.6702	0.0000
MV short-sell constrained	0.0768	0.2944	0.2607	4.3333

As shown in the table above the theoretical optimal portfolio that incorporate transaction costs ex-ante has the lowest turnover and second lowest standard deviation. The low turnover compared to the other portfolios is expected since this problem incorporates the transactions costs ex-ante. When the turnover is zero it might be an indication of the transaction cost parameter β being so large that reallocation is not profitable and the weights will be equal to the initial allocation as stated in proposition 2 by Hautsch et al (2019). On the other hand the the minimum variance portfolio with short-sell constraint had the lowest standard deviation but also the highest turnover and lowest annual net return with a annual return of just 7.7 pct. The lower return is partly due to the high turnover and the costs associated with re-balancing the portfolio, but might also be because we disregard the return when we aim to implement the minimum variance portfolio.

We further note that the naive portfolio has a higher mean and the highest Sharpe ratio. If we evaluate the strategies performance based on the Sharpe ratios an investor would be better of choosing the naive portfolio based on our analysis. This is in line with DeMiguel et al (2009) who states that the naive portfolio cannot be outperformed in high dimensions. This might be an indication that even though incorporating transaction costs ex ante or using short sell constraint can regularize the underlying covariance matrix it might not always be sufficient to outperform a simple naive portfolio.

Exam part 2

Exercise 1

We start by computing the daily realized volatility of SPY based on minute-level observations. In order to do so we start with the theoretically based assumption that asset prices behave like a semimartingale and more precisely assume the log asset prices, X_t, behave like a Geometric Brownian Motion

$$\log(S_t) = X_t = X_0 + \mu t + \sigma W_t$$

where S_t is the asset price at time t, μ and σ are constants, while W_t is a Brownian motion.

With n equidistant observations in each time interval and assuming $\Delta X_{t_{n,t+1}} = X_{t_{n,i+1}} - X_{t_{n,i}}, i = 0, \ldots, n-1$ are iid. with the distribution $N(\mu\Delta t_n, \sigma^2 \Delta t_n)$ the natural estimators are given by

$$\hat{\mu}_n = \frac{1}{n\Delta t_n} \sum_{i=0}^{n-1} \Delta X_{t_{n,i+1}} = (X_T - X_0)/T$$

$$\hat{\sigma}_n^2 = \frac{1}{(n-1)\,\Delta t_n} \sum_{i=0}^{n-1} \left(\Delta X_{t_{n,i+1}} - \overline{\Delta X}_{t_n}\right)^2$$

where $\overline{\Delta X}_{t_n} = \frac{1}{n} \sum_{i=0}^{n-1} \Delta X_{t_{n.i+1}}$.

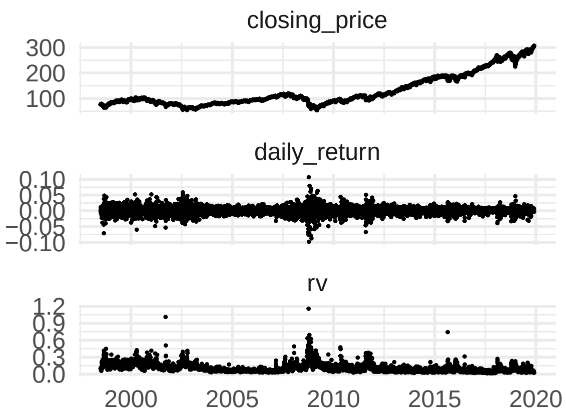
The estimator $\hat{\mu}_n$ is not consistent for a fixed time period but can be consistently estimated if $T \to \infty$. However the estimator $\hat{\sigma}_n^2$ is consistent for $n \to \infty$ and further we have that the non-centered version $\hat{\sigma}^2_{n,\text{nocenter}}$ is also consistent for $n \to \infty$. The realized variance can therefore be consistently estimated as the sum of the squared log returns and will in absence of jumps etc. converge to the integrated variance.

$$RV = \sum_{i=0}^{n} \Delta X_t^2$$

It is worth noting that our estimate of the daily realized variance and volatility could in theory be estimated more precisely if we used data with a higher frequency than 1 minute.

Figure 4: Returns and realized volatilities for SPY

(a) Daily return, closing price and return volatility



(b) Daily return distribution p.a.

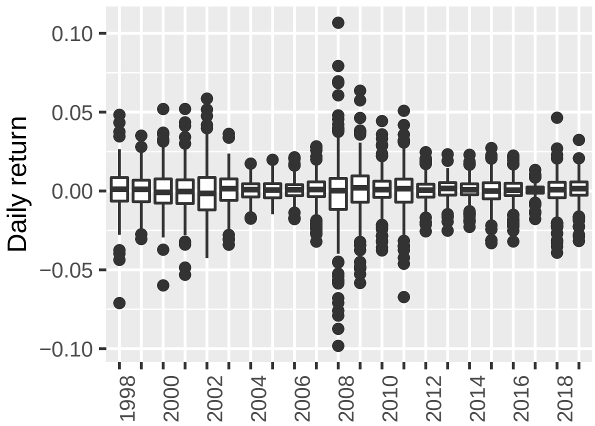


Figure 1 illustrates the closing price, close-to-close daily return and the realized volatilities for SPY. As shown there is a clear upward going trend in the closing price, indicating that the returns of SPY are on average positive. Further there is a clear tendency that the realized volatility increases during periods of financial distress such as 2000 and 2008 and is lower during financial booms. This illustrates the wellknown behavior of financial time series known as volatility clustering.

Table 2: Summary statistics of SPY

	rv	daily_return	closing_price
Mean	0.107	0.000	127.914
Variance	0.005	0.000	3854.139
SD	0.074	0.012	62.082
Min	0.015	-0.098	53.580
Max	1.156	0.107	307.010
Median	0.086	0.001	100.490

Financial markets are subject to microstructure frictions such as spreads and price discreteness. When we revert from our stylized world and introduce market microstructure frictions the prices we observe will be distorted and not efficient. Therefore, when estimating the realized volatility as the sum of the squared log returns the squared noise variance is included. Thus the notion that prices move in a continuous fashion and that the efficient price follows a Brownian motion can be deluded by the presence of microstructure friction.

As stated by Ait-Sahalia et al (2005) it may be better to sample at a lower frequency in order to reduce the noise in the estimate of the volatility due to microstructure frictions. It is worth noting that SPY which tracks the S&P 500 is an extremely liquid asset and therefore the degree of microstructure noise can be assumed to be relatively low. Assuming that it is within the group of assets with the lowest noise variance then the sample frequency of 1 minute should be optimal according to Ait-Sahalia et al (2005).

Exercise 2

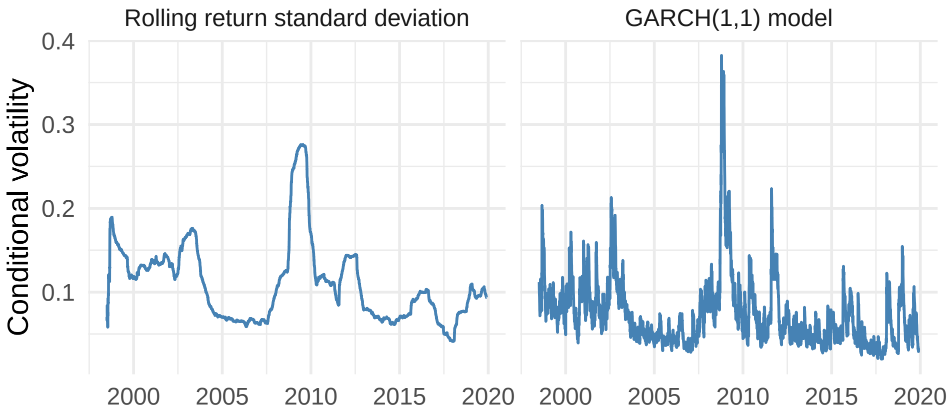
In this exercise we compute two measures of the conditional daily S&P 500 return volatility based on the daily close-to-close returns. The first measure is the rolling return standard deviation based on the last 250 trading days. The second measure is the conditional volatility based on a GARCH(1,1) model. The general idea behind volatility models such as a GARCH(1,1) is to use that the time series $\{r_t\}$ is serially uncorrelated, but dependent. Thus, the innovation of an asset return can be modelled as $a_t = r_t - \mu$, where μ is the sample mean. Further the innovation at time t is given as $a_t = \sigma_t \epsilon_t$, where ϵ_t is assumed to follow a standardized student t-distribution. The volatility equation within a GARCH(1,1) framework is then given by:

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

The parameters of the volatility equation are estimated using maximum-likelihood estimation. Given the parameters we can compute the fitted values of σ_t^2, which is the models prediction of the conditional variance.

The use of this GARCH(1,1) specification for the volatility equation can be justified on the basis of its conformity with a number a stylized facts for financial time series. Under the parameter restrictions $0 \leq \alpha_1, \beta_1 \leq 1, (\alpha_1 + \beta_1) < 1$, a large innovation or estimated volatility in the last period will increase the current volatility in line with the observed volatility clustering in return series. Next, it is noted in Tsay (2010) that under the parameter restriction $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$ the distribution of a GARCH(1,1) process is characterized by fatter tails than a normal distribution. This means that it is more likely to produce outliers in line with what seems to be the case for financial time series. A shortcoming of the GARCH(1,1) volatility equation is that the effects of positive and negative effect are symmetric, so it cannot account for the so-called leverage effect. Further it is also noted in Tsay (2010) that even though a GARCH(1,1) process has fatter tails than a normal distribution, the tail behavior is still too short compared to financial time series.

Figure 5: Conditional Daily S&P 500 return volatility



We note that the rolling return standard deviation in general predicts a higher and more persistent volatility than the GARCH(1,1) model, except from a few outliers. To discuss the implications of volatility estimates for portfolio optimization, we consider a risk-averse investor that allocates his or her portfolio between the market portfolio and a risk-free asset. An estimate of high future volatility implies that the investor should optimally allocate a smaller share of wealth in the market. We return to this issue in exercise 4. Further, volatility estimation is a major concern in terms risk management, since it is of interest for financial institutions to ensure that they have sufficient capital in the case of a large drop in asset prices. Whether a financial institution or investor use a GARCH(1,1) model or a rolling return standard deviation therefore has large implications, since the the rolling return standard deviation in general predicts a higher volatility.

Exercise 3

Empirically there has been found a link between risk premia and macroeconomic variables. This has given rise to the multifactor asset pricing models, where the general underlying idea is trying to model asset prices by observable variables that serve as a proxy for marginal utility of consumption. This is due to the fact that asset prices in theory are partly determined by the covariance between the stochastic discount factor (measured by marginal utility of consumption) and expected returns. Thereby the inclusion of macroeconomic predictor variables in a multifactor model can help explain the equity risk premia by analyzing co-movements of these predictor variables with risk premia.

However the causal direction between equity risk premium and fluctuations in the macroeconomic variables is not clear. The statement that macroeconomic variables are able to predict the equity risk premium lies on the assumption that the causality is unidirectional from macroeconomic volatility to equity risk premiums.

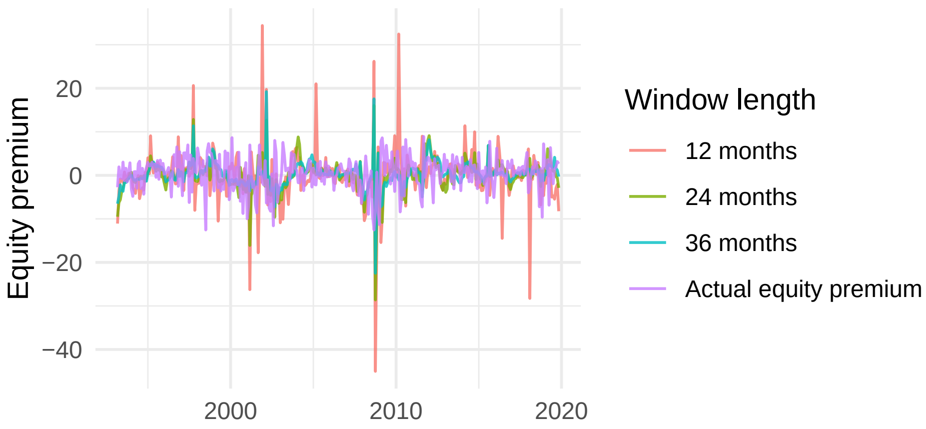
The efficient market hypothesis states that assets prices in general immediately reflects all available information and therefore no investor should be able to gain arbitrage from outperforming the market. The possible predictability of returns does not contradict the hypothesis of efficient markets as the variables are subject to risk as well as the returns. If the efficient market hypothesis holds, the only way to gain a higher return for an investor is by taking on more risk. This would imply a higher equity risk premium. If we then establish significant correlation between the macroeconomic predictors and the equity risk premium, it is not necessarily implying a potential arbitrage possibility.

The question on whether the notion of efficient markets holds is highly debated. For example the father of the multifactor model Eugene Fama argues that there is not a sufficient theoretical framework to thoroughly test whether the markets are in fact efficient, while others argue that due to the existence of market anomalies, financial markets cannot be efficient.

In order to compute the monthly one-step ahead equity premium forecasts we get rid of extreme outliers by winsorizing the observations of equity risk premium as it is common practice in the asset pricing literature. we apply rolling regressions on a sequence of window lengths spanning from one to three years. For each iteration we obtain the estimates based on the current window and predict the equity risk premium based on the parameter estimates.

When choosing the window length for the rolling estimations we have to be aware that the predictive performance can depend on the chosen window length. Given monthly observation a smaller window could be argued to be optimal as the span are large even for smaller windows. On the other hand to small a window makes estimates prone to larger estimation uncertainty and also prone to structural breaks in the data such as the dot-com-bubble or the financial crisis.

Figure 6: Predicted equity risk premium



The plot shows fluctuations of the realized and predicted equity risk premia for different window lengths. The predicted risk premium based on the narrower window produce larger outliers as visible from figure 3, while longer windows yield more persistence than the actual risk premium. We consider the window length of 24 to be optimal as it eludes large outliers but captures the movements of the actual risk premium fairly well. Trends in macroeconomic time series variables generally are known to be much more persistent than time series of stock returns. When including these as the only predictors in the model of the equity risk premium it makes sense to get more persistent and less volatile predictions of the returns if the windows are bigger.

We note from the regressions that the coeffecent of determination R^2 is generally very low for the rolling regressions, indicating a poor fit for the model. This is in line with the findings of Gu, Kelly and Xiu (2020), who find OLS to be performing very poorly in predicting assets returns, and being outperformed by machine learning methods such as penalized regressions, random forests and neural networks.

In addition Gu, Kelly, Xiu (2020) finds that when considering linear models using too many predictors, the efficiency of the linear regression deteriorates, and a vast set of predictors is only viable when using parameter penalization or dimension reduction. When boiling their OLS regression down to three benchmark macroeconomic predictors they obtain a better fit, while the R^2 still is below zero indicating a constant prediction still would outperform the linear regression model. However they argue that factor models can enhance predictability of returns through a larger coefficient of determination when applied to more sophisticated models such as random forests or neural networks.

Exercise 4

In order to derive the optimal share invested in the market, α_t^*, we rewrite the maximization problem by inserting the expectation and variance of the portfolio returns. Given that the risk-free rate is zero we know that $\mathbb{V}(r_f) = 0$ and $\mathbb{C}(r_{t+1}, r_f) = 0$. We use this to exclude terms and take the first order condition of the problem. Setting this equal to zero we are able to find the optimal α_t^* as a function of the parameters:

$$\alpha_t^* = \max_{\alpha} \alpha_t E_t(r_{t+1}) + (1-\alpha)r_f - \frac{\gamma}{2}\left(\alpha^2\mathbb{V}(r_{t+1}) + (1-\alpha_t)^2\mathbb{V}(r_f) + 2\alpha(1-\alpha)\mathbb{C}(r_{t+1}, r_f)\right)$$

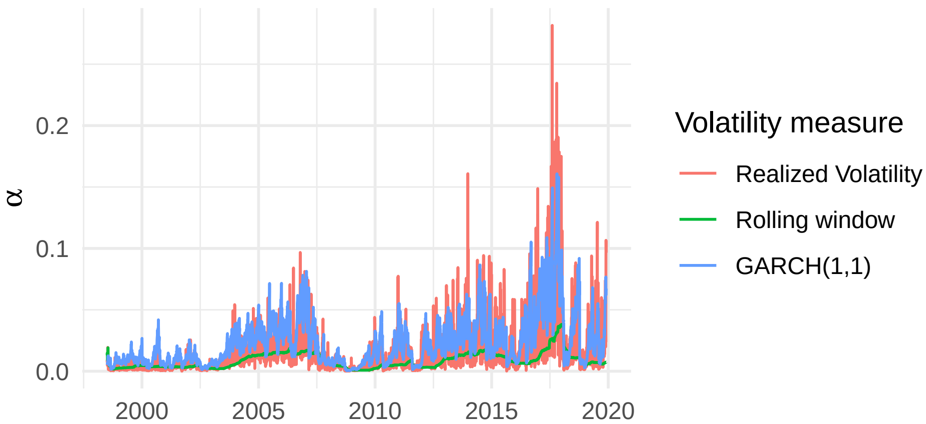
$$\alpha_t^* = \max_{\alpha} \alpha_t\mu_t + (1-\alpha_t)r_f - \frac{\gamma}{2}\left(\alpha_t^2\sigma_t^2\right)$$

$$\frac{\partial\alpha_t^*}{\partial\alpha_t} = \mu_t - r_f - \gamma\alpha_t\sigma_t^2 = 0$$

$$\alpha_t^* = \frac{\mu_t - r_f}{\gamma\sigma_t^2}$$

We can then determine the resulting allocation, α_t^*, each day given values of $\gamma, \sigma_t^2, \mu_t$, and r_f. Throughout our calculations we assume that $\gamma = 4$ and $r_f = 0$. However, we consider different estimates of both σ_t^2 and μ_t. First we set μ_t to the sample average of the annualized close-to-close returns from exercise 1. We consider as estimates of the variance both conditional measures from exercise 3 and the realized volatility from exercise 1. This yields three time series of optimal daily portfolio shares in the market.

Figure 7: Optimal share of wealth allocated to the market



In general an investor using the rolling return standard deviation as a measure of the variance should optimally allocate a smaller share of wealth into the market, when comparing to the other portfolio strategies. This is consistent with the conclusion from exercise 2 that the rolling return standard deviation predicts a higher volatility than the GARCH(1,1) model and the realized volatility. However, all three portfolio strategies implies that the investor should optimally implement a higher exposure to the market during periods where market volatility is low.

There are several ways of accounting for parameter uncertainty. However, only using the two first moments of return series, and assuming them to be true will be sub-optimal. If the investor is aware that there is estimation and model uncertainty it would be more reasonable to account for this ex-ante. More precisely the investors utility maximization problem should not just use the first two moments but rather the entire return distribution and thereby account for both higher moments and also the estimation uncertainty. Using a Bayesian approach the investor would deviate from the assumption that the parameters are fixed and instead consider them as random variables. Starting with a prior idea about their distribution the investor would then observe the data and update her beliefs summarizing them in a posterior distribution that characterises the distribution of the parameters. Due to the parameter uncertainty, the predicted return distribution will usually have fatter tails. Therefore if the investor adjusts for estimation uncertainty when solving their mean-variance problem the optimal allocation will result in a lower α.

Table 3: Annualized Sharpe ratios

	Mean	SD	Sharpe
GARCH, sample mean	0.001	0.003	0.374
GARCH, predicted returns	0.007	0.014	0.474
Rolling, sample mean	0.001	0.001	0.581
Rolling, predicted returns	0.004	0.009	0.413
RV, sample mean	0.004	0.001	2.511
RV, predicted returns	0.007	0.007	1.042

To compare the performance of the three portfolio strategies, we compute the annualized Sharpe ratios. The best performing portfolio is the one using realized volatility with a Sharpe ratio of 2.51. The implementation of this portfolio rests on the assumption that the volatility on the current day will be the same as the previous day. The Sharpe ratios for the portfolio strategies using the rolling return standard deviation and the GARCH(1,1) model are 0.58 and 0.37, respectively. We interpret the Sharpe ratio results as an indicator of how well the different volatility measures captures the actual volatility of the returns. Hence, the realized volatility measure outperforms the two conditional volatility measures from exercise 2.

Next we investigate how the portfolios using the three different volatility measures perform, if we instead set μ_t equal to the annualized equity premium predictions from exercise 3. This yields a higher Sharpe ratio of 0.47 for the GARCH(1,1) model, while we obtain lower Sharpe ratios for the other two portfolio strategies. Using the annualized predicted equity premium therefore has an ambiguous effect on the performance of the portfolio strategies. The fact that the Sharpe ratios decrease implies that the predicted equity premiums based on the linear multifactor model do not reflect the actual realized returns, which is in line with the fact that poor performing return predictions are inferior to constant predictions of the returns such as the sample mean.

If an investor does not account for transaction costs but are faced with these ex-post, portfolio strategies that may have been theoretically optimal before accounting for transaction costs will often involve too aggressive rebalancing. Due to the cost associated with the aggressive rebalancing these strategies will yield a lower net return compared to strategies that account for the transaction costs ex-ante as described in MA3. We would expect the models based on the realized volatility and GARCH specification to be associated with larger transaction costs due to the more volatile structure of their exposure to the market. This will decrease their relative performance compared to the strategy based on the rolling returns standard deviation.