Erasmus University Rotterdam - Erasmus School of Economics

Bayesian Extensions of the Black-Litterman Model

Master Thesis Quantitative Finance

Date final version: August 27, 2019

Abstract

To reduce the estimation error portfolio managers encounter when determining their portfolio allocation, this paper focuses on the application of the Black-Litterman model within a more extensive Bayesian framework. This framework allows to extend the existing Black-Litterman models to include prior specifications on the covariance matrix, hierarchical scaling parameter and non-normal data. Applying these models to a S&P 500 data set, the models using a Bayesian framework, taking into account additional parameter uncertainty, are in some challenging environments able to obtain a superior result compared to the benchmarks, making it a useful framework for portfolio managers. However, the models can not consistently outperform the less complex benchmark models.

Keywords: Black-Litterman model, portfolio optimization, Bayesian statistics

Student: J.F. Schepel

Student number: 483851

Supervisor: dr. A.M. Schnücker

Second assessor: dr. M.D. Zaharieva

Table of Contents

1 Introduction

Correctly estimated portfolios weights are essential for portfolio managers to achieve the highest possible utility from their investments. In order to arrive at the most favorable weights, the most commonly used setting is the Markowitz framework (Markowitz, 1952) for optimal portfolio weights, where weights are computed by taking into account both the mean and the variance of the underlying asset returns. Hence, a trade-off between the risk and return of the investment is established.

Although Markowitz developed a theoretically solid framework, it is prone to large estimation errors when used with sample moments for the mean and variance. Because it requires exact expected returns for all assets involved in the model, it does not allow for uncertainty. This results in extreme and volatile portfolio weights as a result of their sensitivity to the input, as indicated by Best and Grauer (1991), making them unreliable and costly to work with. It is for this reason that Michaud (1989) describes this estimation method as 'error maximization' in his research. Also Jobson and Korkie (1980) conclude that the out-of-sample results are poor when looking at the performance of the formed portfolios. In order to overcome this problem, the susceptibility to estimation errors due to the uncertainty in the input needs to be accounted for.

Much of the literature focuses on reducing the incorporated estimation error with the use of shrinkage techniques. With these techniques, it is possible to add more information to the estimation procedure and thereby potentially increase the accuracy of the estimation. This yields a better portfolio performance when the reduction in estimation error outweighs the increase in estimation bias. To achieve this, different approaches are developed to adjust the estimates of the parameters. For example, James and Stein (1992) introduce a biased estimator of the mean which dominates the sample mean in terms of mean squared error for any number of assets larger than two. Ledoit and Wolf (2003) and Kourtis, Dotsis, and Markellos (2012) do not focus on the mean but shrink the covariance matrix to reduce the estimation error present in the second moment of the asset returns. Golosnoy and Okhrin (2007), on the other hand, focus on shrinkage of the resulting portfolio weights and propose a multivariate shrinkage estimator. They find robust performances and significant gains. Furthermore, various Bayesian shrinkage strategies, benchmarked against frequentist approaches, are depicted in the paper by Frey and Pohlmeier (2016). Factor models and portfolio restrictions are applied as shrinkage methods as well. The former is implemented by Brandt, Santa-Clara, and Valkanov (2009), who model a common factor portfolio, and by Aguilar and West (2000), who use a Bayesian dynamic factor model in order to estimate the portfolio returns. The latter is widely employed, also as an addition to existing models, for example to account for a short selling constraint. A more extensive model is proposed by Yen (2015), where a reduction in risk is accomplished by penalizing the portfolio optimization.

As can be seen in the previous paragraph, some novel shrinkage techniques are available to more accurately estimate the optimal portfolio weights. In this paper, I focus on a shrinkage method called the Black-Litterman model (Black and Litterman, 1992). Introduced by Black and Litterman in 1992 while working for Goldman Sachs, the basis of this model is the blending of market equilibrium returns with investor's views on the asset returns as to obtain an estimation of the future asset returns. Where former research has a uniform distribution or the global minimum variance portfolio as initialization, the Black-Litterman model uses the equilibrium market portfolio. This portfolio is used as a starting point for determining the portfolio weights. The public market information is subsequently combined with the private information of the investor. This private information is depicted as a set of views, which can be relative or absolute, and is used as an extra layer of information from which asset returns can be estimated more accurately. As such, the views are shrunk towards the market equilibrium returns to find a balanced estimate. Furthermore, the views have their own level of uncertainty to account for misspecification of the input. For the combination of these two key elements, the model received most of its praise.

While the paper by Black and Litterman is successful for their introduction of a novel estimation technique, it lacks an in-depth review of the model and additional derivations. He and Litterman (1999) therefore provide further economic intuition and include a suitable example. Since then, a proliferation of Black-Litterman labeled models flooded the literature. By using point estimates and thereby neglecting the posterior distribution, Satchell and Scowcroft (2000) introduce a new branch of Black-Litterman models without Bayesian characteristics, coined the alternative reference model by Walters (2014). Because of the absence of Bayesian methods, which are the foundation of the Black-Litterman model, this model is not widely used in the literature and for this reason it also received critique by Michaud, Esch, and Michaud (2012). Meucci (2006) followed up on the Black-Litterman model by extending the framework with non-normal distributions for the market and investor's views. This is done by using a copula opinion pooling method. Looking at the Black-Litterman model from a frequentist perspective, Mankert and Seiler (2011) introduce derivations based on sampling theory. As a side product, new interpretations for the scaling parameter of the asset return variance and the uncertainty in the views are obtained. Another modeling approach is the augmented Black-Litterman model, introduced by Cheung (2013), which incorporates factor views to robustify the portfolio optimization. Lastly, Palczewski and Palczewski (2019) extend the literature of non-normal Black-Litterman models by using other risk measures than the moments of continuous distributions, where the paper concentrates on the expected shortfall as a measure of risk.

The mentioned literature shows the ability to outperform alternative portfolio construction models by making use of the intuitive blending of public and private information within the Black-Litterman model. In the meantime, however, more advanced estimation techniques have become readily available. While similar papers only discuss the theoretical background or implementation of one of the mentioned models, the goal of this paper is to focus on the implementation of a variety of extended Bayesian Black-Litterman models and comparing their performances.

I start with a simple market equilibrium model with no investor views, to determine easily obtainable shrunk portfolio weights, after which multiple existing Black-Litterman models are discussed. These models focus on accurately estimating the mean of the asset returns, while the variance parameter is mostly disregarded and assumed known. Although they form a solid basis, the existing models neglect a large part of the uncertainty incorporated in the estimation. Therefore, the first model I introduce considers the covariance matrix completely unknown, as it is an unobserved parameter and hence its uncertainty should be accounted for. As prior for the covariance matrix, I take the inverse-Wishart distribution. The second model corrects for the uncertainty of the scaling parameter for the variance in the classical Black-Litterman model. This element is set at a constant value in the base model, which lacks flexibility over time and does not allow for dispersion. Adjusting for this results in a hierarchical and more dynamic model. The third model focuses on another limitation of the mentioned models, which is the normality of the asset returns. While convenient for analytical computation, asset returns generally do not show conclusive normal behaviour. This model therefore accounts for non-normality by applying a Student's t-distribution to the data. Using these methods, the newly introduced models apply a more comprehensive Bayesian approach to the subject. Additionally, I provide a deeper understanding of the underlying variables and assumptions and utilize the Hamiltonian Monte Carlo sampling algorithm to obtain draws from complex posterior distributions. This sampler is theoretically more efficient than conventional samplers as it involves less randomness to explore the target distribution.

In order to derive results from the models discussed throughout the paper, a data set containing monthly S&P 500 stock prices is employed. The data set spans the time period January 2007 to December 2018, a period characterized by a financial crisis followed by predominantly positive asset returns. Examining the data, no conclusive normal behaviour is shown.

Applying the models to this data set, it is concluded that a more extensive Bayesian framework can improve the performance of the discussed models, but does not consistently outperform the benchmark models. Mainly the models where uncertainty associated with the covariance parameter is taken into account yield an outperformance over the alternative models. These models perform well in difficult scenarios, where the simpler models are not able to obtain an adequate performance. As such, applying a Bayesian framework is beneficial for portfolio managers as to improve their parameter estimations. However, as discussed in the sensitivity analysis, the performance of the models does depend on the initialization of the model and the data used. Which model to implement should therefore be determined while taking into account the current economical environment and the preferences of the investor. The framework set in this paper could help with such a decision.

In the following, I present and analyze the data at hand in Section 2 to provide a general overview of the information used for the models. Section 3 subsequently focuses on the theoretical background of the proposed Black-Litterman models and the necessary conditions. The benchmark models, sampling method and performance evaluation measures are introduced here as well. Results from the implementation of the models, an examination of the convergence diagnostics of the sampling process and a robustness analysis are given in Section 4, after which a conclusion and discussion follow in sections 5 and 6 respectively.

2 Data

The data set I use comprises monthly stock prices of companies listed on the S&P 500 index.[1] These companies are the 500 largest American companies measured by their market capitalization. The data set covers the time period January 2007 up to December 2018 and therefore starts just around the beginning of the credit crisis of 2007-2008. The number of monthly observations, T, equals 144. Only the stocks which are listed for the full time period are incorporated in the data set. Hence, a survivorship bias is expected. However, as investigated by Rohleder, Scholz, and Wilkens (2010), this bias is small for larger funds such as the one used in this paper. To account for stock splits, that is a change in the number of shares and therefore as a consequence a change in the stock price, I adjust the prices and the number of outstanding shares. In total, a number of $n = 451$ assets are included in the data set.

Table 1: Descriptive statistics of the asset returns

Mean return (%)	Mean 2007-2009 (%)	Mean 2010-2018 (%)	Min. (%)	Max. (%)	Average skewness	Average kurtosis	Median kurtosis
0.90	0.17	1.14	-84.50	259.66	0.58	7.21	3.12

Descriptive statistics for the asset returns of the S&P 500 data set, spanning the period 2007-01 to 2018-12. Mean 2007-2009 depicts the mean of the asset returns in the period of the credit crisis. Mean 2010-2018 is the average asset return after the concerning crisis. Min. and max. denote respectively the minimum and maximum returns.

In Table 1, some descriptive statistics for the data set are given. The average monthly return of the assets measured on the complete data set is 0.9%. As the data set starts amidst an economic crisis, the mean during and after the crisis are also given. Since the crisis was most noticeable in the years 2007-2009, I assume this period to be the crisis period. It can be seen that during the crisis, the average asset return was substantially lower than in the period after the crisis, although on average it was not negative over these two years. This result is supported by Figure 1, where the returns are plotted over time. In the crisis period, there is a considerable number of negative returns, while after 2009 the returns show a peak as the economy is expanding again. The minimum and maximum returns are extreme cases of -85% and 260%, as can also be seen in the histogram of the asset returns in Figure 2. For example, the minimum return is attributed to American International Group, a financial and insurance corporation which had to be bailed out in 2008 by the American government for $180 billion.

[1] Derived based on data from the Wharton Research Data Services 2019 Center for Research in Security Prices (CRSP), The University of Chicago Booth School of Business.

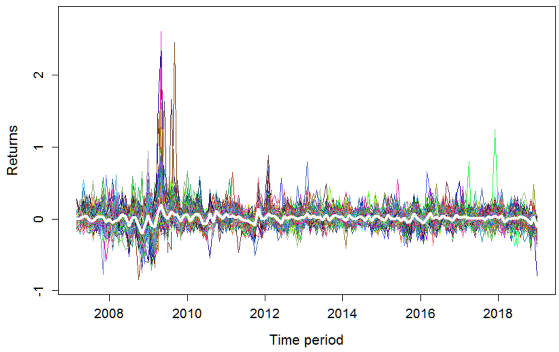

Figure 1: *Time series of the returns for each asset. The thick white line indicates the average asset return over the time frame, 2007-01 to 2018-12.*

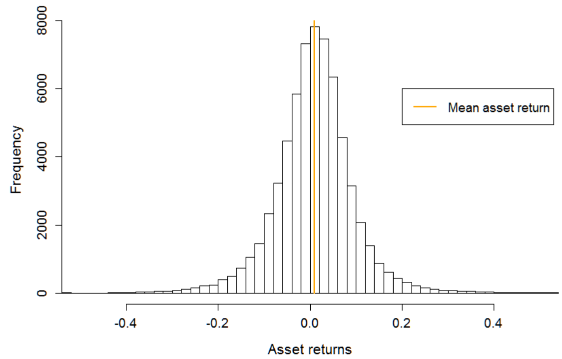

Figure 2: *Histogram of the asset returns in the time period 2007-01 to 2018-12. In orange, the average asset return is depicted.*

Regarding the distribution of the asset returns, some normal behaviour is shown; Figure 2 indicates a symmetric distribution and the value for the median kurtosis, 3.12, is close to the kurtosis of a normal distribution, which equals 3. The slightly higher kurtosis value indicates a leptokurtic distribution, which has more probability mass in its tails. I focus on the median kurtosis, as there are a few outliers right after the crisis period. During this period, some companies still have difficulties recovering, while other companies are flourishing as the economy is starting its expansion. This result can be seen in Figure 3, where the average cross-sectional kurtosis is plotted over time. However, when considering the quantile-quantile plot as shown in Figure 4, heavy-tailedness as exhibited by off diagonal values can be distinguished at the outer quantiles, as well as a right or positive skewness, which corresponds with the positive skewness value given in Table 1. For a normal distribution, only approximately diagonal elements are expected in this quantile-quantile plot. Hence, no conclusive evidence is found to support the assumption of normally distributed returns.

For the composition of the (multi-)factor models, data on the factors are needed. For example, the Fama-French three factor model (Fama and French, 1993) consists of the factors small minus big (SMB), which differentiates in the size of a certain company and highlights the outperformance of small companies compared to big companies, high minus low (HML), where companies with a high book-to-market ratio are expected to generate a higher return than companies with a low ratio, and market risk. Multiple data sets on these factor returns, not limited to the three factors mentioned before, are available via the web page of Kenneth French.[2] The data spans

[2]The data is obtained from `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html`.

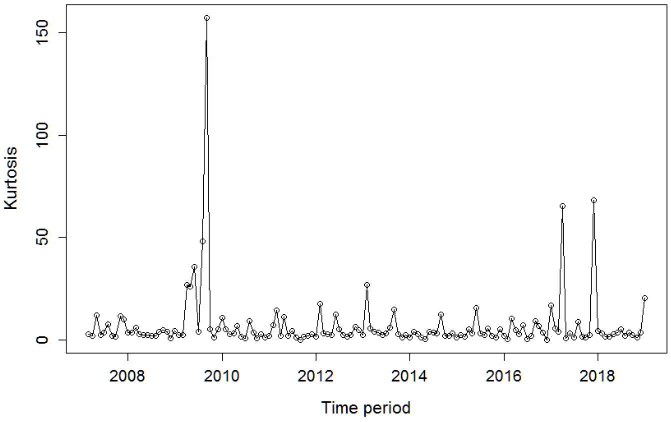

Figure 3: *The average kurtosis of the asset returns for each time point in the period 2007-01 to 2018-12.*

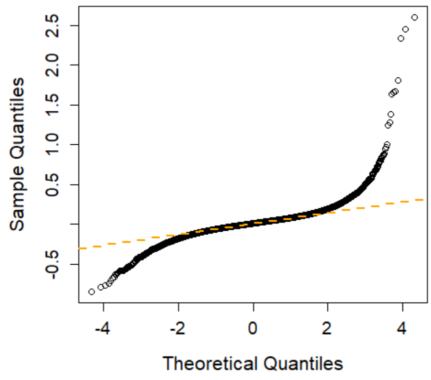

Figure 4: *Quantile-quantile plot for the asset returns in the period 2007-01 to 2018-12. The theoretical distribution is the normal distribution.*

the period July 1990 up to March 2019 and contains monthly returns. From this source, data on the market return and risk-free rate, defined as the one-month T-bill yield, are obtained as well. Figure 8 in Appendix B.1 shows the course of this risk-free rate over time, where it can be seen that the rate was at 0% for the larger part of the time period. Only recently, the risk-free rate increased again, as a consequence of the raising of the policy rate set by the American government.

For the estimation of the monthly asset returns, a rolling window approach with overlapping windows is employed. This implies that all estimates are based on the same fixed number of preceding data points, which also entails limited bias caused by historical extreme values as they drop out of the window at some point. I use a rolling window size of $h = 36$ month, thus equalling three years of information prior to each calculation. This window size, equalling 25% of the data set, allows for sufficient information to compute historical estimates while preserving an adequate proportion on which to backtest the models. The first h data points are used to generate an estimate for $h + 1$. With this estimate, the window shifts one month such that the window accommodates the data points 2 to $h + 1$ and this data is then used to compute an estimate for $h + 2$. Continuing with this method yields $T - h = 108$ estimates, all based on the h data points prior to the concerning estimate.

3 Methodology

In this section, the different asset selection models are derived and explained. First, the capital asset pricing model (CAPM) and its resulting market equilibrium returns are introduced, as these form an integral part of the proposed models. Subsequently, I describe a base model in Section 3.2, which does not incorporate investor views but solely focuses on the market equilibrium. Having defined a base model, three existing Black-Litterman models are discussed in Section 3.3, followed by my own extensions to the Black-Litterman model in Section 3.4. The Black-Litterman model aims to integrate both the market equilibrium returns of the base model and investor's view on asset returns to determine the portfolio weights. Each model has its own unique set of specifications, for example the number of parameters or the choice of the prior distributions. The characteristics of each model are highlighted in the respective sections, as well as their advantages and drawbacks. An overview of the models and their specifications is given in Section 3.5, Table 3. After defining the different models, the necessities needed to apply the mentioned models are introduced. Finally, the benchmark models and the performance evaluation measures are discussed.

In the next sections, the following variables are used for the n assets: a vector of asset returns $\boldsymbol{r} \in \mathbb{R}^{(n\times 1)}$, the mean of the asset returns $\boldsymbol{\mu} \in \mathbb{R}^{(n\times 1)}$, the assumed known covariance matrix of the asset returns $\dot{\boldsymbol{\Sigma}} \in \mathbb{R}^{(n\times n)}$ and its unknown counterpart $\boldsymbol{\Sigma} \in \mathbb{R}^{(n\times n)}$, and a vector containing portfolio weights, $\boldsymbol{w}_m \in \mathbb{R}^{(n\times 1)}$. Scalars are written as regular symbols (R, n), vectors are indicated by bold lowercase symbols ($\boldsymbol{a}$) and matrices are expressed as bold uppercase symbols ($\boldsymbol{A}$).

3.1 The CAPM and Market Equilibrium Returns

The CAPM, independently introduced by Treynor (1961), Sharpe (1964), Lintner (1965) and Mossin (1966), relies on the general equilibrium theory and is a factor model in which the only factor is the market return. It is derived under the following assumptions: i) every investor optimizes his mean-variance trade-off, ii) all investors have homogeneous expectations and iii) there is a risk-free rate (R_f) available against which investors can borrow and lend money. When these assumptions hold, the expected return of asset i is given by

$$\mathbb{E}(r_i) = R_f + \frac{\text{Cov}(r_i, r_m)}{\text{Var}(r_m)}(\mathbb{E}(r_m) - R_f) = R_f + \beta_i(\mathbb{E}(r_m) - R_f) \; , \tag{1}$$

where r_m denotes the market return, $\text{Cov}(r_i, r_m)$ the covariance between the returns of asset i and the market and $\text{Var}(r_m)$ the variance of the market return. Hence, the CAPM states that the excess return of an asset can be explained by its exposure to the non-diversifiable risk of the market.

Based on Equation 1, the market equilibrium returns can be derived using a technique called reverse optimization. Using the value weighted portfolio weights $\boldsymbol{w}_m$ for the assets, this approach yields a vector of market equilibrium returns given by $\boldsymbol{\pi} = \lambda \dot{\boldsymbol{\Sigma}} \boldsymbol{w}_m$. The full derivation can be found in Appendix A.1. In this formulation, λ is a constant defining the risk-return trade-off and can therefore be seen as the level of risk aversion of the investor. The value for the risk aversion parameter is greater than one for risk averse investors, which all investors are according to Markowitz. I assume its value to be equal to 2.5, which is a common choice and is also used in the paper by He and Litterman (1999).

3.2 The Market Equilibrium Model

The base asset allocation model I am using, which I call the market equilibrium model, is a simplification of the Black-Litterman model as it does not take into account the views of the investors. The asset returns in excess of the risk-free rate[3], $\boldsymbol{r}$, are assumed to be normally distributed with $\boldsymbol{r} \sim \mathcal{N}(\boldsymbol{\mu}, \dot{\boldsymbol{\Sigma}})$. The mean of the asset returns, $\boldsymbol{\mu}$, is assumed to be an unknown parameter, while $\dot{\boldsymbol{\Sigma}}$ is assumed to be a known covariance matrix. As $\boldsymbol{\mu}$ is unknown, a distribution is specified for this parameter. Following Black and Litterman (1992), the normal distribution centered around the market equilibrium return $\boldsymbol{\pi}$ is chosen. Once I have an estimate for the equilibrium market returns, the distribution of the mean asset returns is given by

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\pi}, \tau \dot{\boldsymbol{\Sigma}}) \ , \tag{2}$$

where $\tau \geq 0$ is an assumed to be known scaling parameter which defines the level of confidence in the equilibrium market returns. The closer the value for τ is to zero, the more confidence an investor has in the location of the equilibrium returns and thus the closer the asset returns are expected to be to the market equilibrium. Using this expression and by merging both normal distributions, the distribution of the asset returns is given by

$$\boldsymbol{r} \sim \mathcal{N}(\boldsymbol{\pi}, \dot{\boldsymbol{\Sigma}} + \tau \dot{\boldsymbol{\Sigma}}) \ . \tag{3}$$

It can be seen that the mean has not changed, but the variance is inflated since it accounts for both the uncertainty in the asset returns as well as for the uncertainty in the estimate of the mean asset return.

[3]For the remained of this paper, all mentioned returns are returns in excess of the risk-free rate unless stated otherwise.

3.3 Existing Black-Litterman Models

Having defined a model without incorporation of the views of the investor, I know discuss three models available in the current literature in which the two sources of information are merged. Of these models, the first is the original Black-Litterman model.

3.3.1 The Canonical Reference Model

The basis of the original Black-Litterman model is founded on three pillars: i) the semi-strong efficient market hypothesis, which states that only with superior (prior) private knowledge the market can be beaten, ii) the CAPM and iii) Bayes' theorem to blend the public and private information. As such, using the CAPM as a basis and adding private information in the form of investor's views, a model is composed which in theory is able to outperform the market. Following the distinction made by Walters (2014), I denote the original Black-Litterman model, as formulated by Black and Litterman, as the canonical reference model.

Two sources of information are available for the model: the market equilibrium returns and views on the asset returns. One of the key assumptions of the initial model is that both are normally distributed. The distribution of the expected asset returns $\boldsymbol{r}$ is given by $\boldsymbol{r} \sim \mathcal{N}(\boldsymbol{\mu}, \dot{\boldsymbol{\Sigma}})$, where similarly as before the covariance matrix $\dot{\boldsymbol{\Sigma}}$ is assumed to be known. Following the same logic as before, the distribution of the mean returns is given by $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\pi}, \tau\dot{\boldsymbol{\Sigma}})$ and the resulting distribution of the asset returns therefore reads $\boldsymbol{r} \sim \mathcal{N}(\boldsymbol{\pi}, \dot{\boldsymbol{\Sigma}} + \tau\dot{\boldsymbol{\Sigma}})$.

After exploring one side of the available information, Black and Litterman integrate the second source of information, namely the investor's views. These views can be seen as portfolios. A portfolio in this sense could consist of a single asset for which the investor sets a certain expected target for its return, which is appointed as an absolute view. A portfolio could also consist of multiple assets, where the investor defines a certain level of outperformance for (a group of) assets relative to other assets. This is a relative view, in which the views are expressed as a linear combination of assets. The k views are rows of the matrix $\boldsymbol{P} \in \mathbb{R}^{k \times n}$, where the concerning assets of each portfolio are depicted as values in $\boldsymbol{P}$. This matrix is also called the 'pick' matrix, as it locates the concerning assets for the views. For a more detailed example of the matrix $\boldsymbol{P}$, see Appendix A.2. Related to the pick matrix, there is the column vector $\boldsymbol{q} \in \mathbb{R}^{(k \times 1)}$, which reflects the expected returns of the views. The diagonal matrix $\boldsymbol{\Omega} \in \mathbb{R}^{(k \times k)}$ determines the level of uncertainty of a view, assuming mutually uncorrelated views, and is assumed known. The higher the value of the diagonal element ω_{ii}, the lower the confidence in a view. When $\omega_{ii} = 0$, there is full confidence in the corresponding view. Using this information, the system of views on the asset returns can be expressed in a more convenient form as

$$\boldsymbol{q} = \boldsymbol{P}\boldsymbol{\mu} + \boldsymbol{\eta} \,, \tag{4}$$

where $\boldsymbol{\eta}$ is the disturbance term of the expressed views related to the matrix $\boldsymbol{\Omega}$. In order to combine the equilibrium returns and the investor's views, Bayes' theorem is used. The views of the model, $\boldsymbol{q}$, are used as data such that its likelihood can be applied in the theorem. This yields

$$P(\ \boldsymbol{\mu} \mid \boldsymbol{q}\ ) = \frac{P(\ \boldsymbol{q} \mid \boldsymbol{\mu}\ )\ P(\boldsymbol{\mu})}{P(\boldsymbol{q})} \propto P(\ \boldsymbol{q} \mid \boldsymbol{\mu}\ )\ P(\boldsymbol{\mu}) \,. \tag{5}$$

The mean of the asset returns, centered around the market equilibrium return, is taken as the prior, $P(\boldsymbol{\mu})$. This prior, combined with the likelihood of the views of the investor conditional on the expected asset returns, $P(\ \boldsymbol{q} \mid \boldsymbol{\mu}\ )$, yields the mean return given the views of the investor, $P(\ \boldsymbol{\mu} \mid \boldsymbol{q}\ )$. The key Black-Litterman formula is now obtained and is given by

$$P(\ \boldsymbol{\mu} \mid \boldsymbol{q}\ ) = \mathcal{N}(\widehat{\boldsymbol{\mu}}_{BL}, \widehat{\boldsymbol{\Sigma}}_{BL}) \,, \tag{6}$$

where

$$\widehat{\boldsymbol{\mu}}_{BL} = [(\tau\dot{\boldsymbol{\Sigma}})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}]^{-1}[(\tau\dot{\boldsymbol{\Sigma}})^{-1}\boldsymbol{\pi} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{q}] \ \text{ and} \tag{7}$$

$$\widehat{\boldsymbol{\Sigma}}_{BL} = [(\tau\dot{\boldsymbol{\Sigma}})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}]^{-1} \,. \tag{8}$$

To arrive at this results, the derivation outlined in Appendix A.3, where conditional expectations are used, is followed. The result can also be obtained using a Bayesian approach, for which the derivation is given in Appendix A.4. Equation 6 thus yields a distribution of the expected returns where the views of the investors are taken into account as well. Using this distribution, the posterior asset returns can be derived, which subsequently can be used for the estimation of the portfolio weights. Zooming in on the mean of the distribution as shown in Equation 6, this parameter formulation can be rewritten as

$$\widehat{\boldsymbol{\mu}}_{BL} = \boldsymbol{\pi} + \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'[(\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}') + \boldsymbol{\Omega}]^{-1}[\boldsymbol{q} - \boldsymbol{P}\boldsymbol{\pi}] \,, \tag{9}$$

which is a result shown in Appendix A.3 as well. From this equation, it can be seen that when the investor is very confident in its views and thus for the diagonal elements of $\boldsymbol{\Omega}$ $\omega_{ii} \to 0\ \forall\ i$, it follows that $\widehat{\boldsymbol{\mu}}_{BL} \to \boldsymbol{P}^{-1}\boldsymbol{q}$. When the investor has almost no confidence in its own views ($\omega_{ii} \to \infty\ \forall\ i$), this changes to $\widehat{\boldsymbol{\mu}}_{BL} \to \boldsymbol{\pi}$. As the scale parameter of Equation 6 denotes the variance of the posterior mean estimate, a derivation for the posterior variance of the asset returns is still needed. This is given by

$$\widehat{\boldsymbol{\Sigma}}_{post.} = \dot{\boldsymbol{\Sigma}} + [(\tau\dot{\boldsymbol{\Sigma}})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}]^{-1} = \dot{\boldsymbol{\Sigma}} + \widehat{\boldsymbol{\Sigma}}_{BL} \,. \tag{10}$$

Using Equation 10, the conditional distribution of the asset returns is therefore written as

$$P(\ \boldsymbol{r} \mid \boldsymbol{q}) = \mathcal{N}(\widehat{\boldsymbol{\mu}}_{BL}, \widehat{\boldsymbol{\Sigma}}_{post.})\ . \tag{11}$$

3.3.2 The Alternative Reference Model

The alternative reference model introduces an additional assumption to the canonical reference model. The alternative reference model assumes the mean of asset returns to be deterministic, as opposed to the random behaviour in the previous model. As Walters (2014) describes the method, this would equate to omitting the parameter τ and hence introducing a point estimate for $\boldsymbol{\mu}$. The 'posterior' of the model is subsequently written as only the estimation of the mean asset return, thereby foregoing the updating of the posterior variance. With this assumption, the only parameter that can be varied is the parameter $\boldsymbol{\Omega}$. Rewriting Equation 9 to include this specification yields

$$\widehat{\boldsymbol{\mu}}_{ARM} = \boldsymbol{\pi} + \dot{\boldsymbol{\Sigma}}\boldsymbol{P}'[(\boldsymbol{P}\dot{\boldsymbol{\Sigma}}\boldsymbol{P}') + \boldsymbol{\Omega}]^{-1}[\boldsymbol{q} - \boldsymbol{P}\boldsymbol{\pi}]\ . \tag{12}$$

As the mean asset return is deterministic in this model, there is no need to update the variance of the asset returns, and the resulting distribution of the asset returns is given by

$$\boldsymbol{r} \sim \mathcal{N}\left(\widehat{\boldsymbol{\mu}}_{ARM}, \dot{\boldsymbol{\Sigma}}\right)\ . \tag{13}$$

3.3.3 The Extended Canonical Reference Model

In the canonical reference model, the parameter $\dot{\boldsymbol{\Sigma}}$ is assumed to be known. However, in general this parameter is unknown as it can not be directly observed or accurately measured. Hence, for this section, the approach by Roeder (2015) is followed, who still assume $\dot{\boldsymbol{\Sigma}}$ to be known but adds uncertainty around the covariance matrix when calculating the asset returns. He uses the result of the canonical reference model as input for the prior distribution when computing the asset returns. The views are therefore incorporated in this prior specification. As such, the model is an extension of the canonical reference model.

As prior for the covariance matrix, an inverse-Wishart distribution is chosen since it acts as the conjugate prior for the multivariate normal distribution. Such a distribution is denotes as $\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}, v)$, where $\boldsymbol{\Psi}$ is its scale and v the degrees of freedom. To ensure that $\boldsymbol{\Sigma}$ is a positive definite matrix, the degrees of freedom are required to be at least as large as the number of assets ($v \geq n$). These degrees of freedom of the inverse-Wishart distribution determine the level of confidence the investor has in the initialization of the parameter $\boldsymbol{\Psi}$. Using a normal distribution for the mean asset returns yields the following specification,

$$\boldsymbol{\Sigma}_{ECRM} \sim \mathcal{W}^{-1}\left(\widehat{\boldsymbol{\Sigma}}_{post.}, v\right) \quad \text{and} \quad \boldsymbol{\mu}_{ECRM}|\boldsymbol{\Sigma}_{ECRM} \sim \mathcal{N}\left(\widehat{\boldsymbol{\mu}}_{BL}, \frac{1}{l}\boldsymbol{\Sigma}_{ECRM}\right), \tag{14}$$

where $\widehat{\boldsymbol{\Sigma}}_{post.}$ follows from the canonical reference model and is therefore based on the known covariance matrix $\dot{\boldsymbol{\Sigma}}$. l represents a level of accuracy regarding $\boldsymbol{\Sigma}_{ECRM}$ and the subscript $_{ECRM}$ indicates the extended canonical reference model. Generally, l is defined as the number of observations used to compute the historical estimate of $\boldsymbol{\Sigma}$. As can be seen, the results from the canonical reference model are used here as information to specify the prior. This would theoretically yield a good starting point to approximate the posterior distribution. When a diffuse prior is preferred, one can use $\boldsymbol{\Sigma}_{ECRM} \sim \mathcal{W}^{-1}(\boldsymbol{I}_p, p)$, where $\boldsymbol{I}_p$ is the p-dimensional identity matrix. By definition, the joint distribution of these two parameters is given by a normal-inverse-Wishart distribution,

$$P(\boldsymbol{\mu}_{ECRM}, \boldsymbol{\Sigma}_{ECRM}) = \mathcal{NIW}\left(\widehat{\boldsymbol{\mu}}_{BL}, l, \widehat{\boldsymbol{\Sigma}}_{post.}, v\right). \tag{15}$$

Combining this joint distribution with the likelihood of the asset returns, $\boldsymbol{r} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, yields the posterior distribution,

$$P(\boldsymbol{\mu}_{ECRM}, \boldsymbol{\Sigma}_{ECRM} \mid \boldsymbol{r}) = \mathcal{NIW}\left(\widetilde{\boldsymbol{\mu}}_{ECRM}, \widetilde{l}, \widetilde{\boldsymbol{\Sigma}}_{ECRM}, \widetilde{v}\right). \tag{16}$$

The derivation of the complete posterior distribution as well as the formulas for the parameters $\widetilde{\boldsymbol{\mu}}_{ECRM}$, $\widetilde{l}$, $\widetilde{\boldsymbol{\Sigma}}_{ECRM}$ and $\widetilde{v}$ are given in Appendix A.5. Since the prior used is a conjugate prior, the posterior distribution can be solved analytically. From this normal-inverse-Wishart distribution, samples can be drawn and the means and covariances can be computed. As a last step, the posterior predictive distribution results from an additional Bayesian updating step and equals

$$\begin{aligned} P(\boldsymbol{r}_{T+1}|\boldsymbol{r}_1, \dots, \boldsymbol{r}_T)) &= \int_{\boldsymbol{\mu}}\int_{\boldsymbol{\Sigma}} P(\boldsymbol{r}_{T+1}|\boldsymbol{\mu}_{ECRM}, \boldsymbol{\Sigma}_{ECRM})\, P(\boldsymbol{\mu}_{ECRM}, \boldsymbol{\Sigma}_{ECRM}|\boldsymbol{r}_1, \dots, \boldsymbol{r}_T)\, d\boldsymbol{\Sigma}\, d\boldsymbol{\mu} \\ &= t_{\widetilde{v}-n+1}\left(\widetilde{\boldsymbol{\mu}}_{ECRM}, \widetilde{\boldsymbol{\Sigma}}_{ECRM}\, \frac{\widetilde{l}+1}{\widetilde{l}(\widetilde{v}-n+1)}\right). \end{aligned} \tag{17}$$

3.4 Introduction New Black-Litterman Models

The models depicted in the previous sections effectively blend the public and private information, but incorporate only a limited part of the uncertainty associated with the estimation of the asset returns. Hence, in the following, I introduce three extensions to the existing models as to capture the additional present dispersion of the data and further improve the estimation procedure.

3.4.1 The Black-Litterman Model With Unknown $\boldsymbol{\Sigma}$

In Section 3.3.3, I followed the approach by Roeder (2015), which extends the canonical reference model but where the uncertainty of the variance only became part of a second stage of the estimation procedure. While this accounts for some of the uncertainty, it still is not a full Bayesian approach and therefore the model is not able to capture the full extent of incorporated uncertainty. The known covariance matrix $\dot{\boldsymbol{\Sigma}}$ is still used to derive the parameters $\boldsymbol{\mu}_{BL}$ and $\widehat{\boldsymbol{\Sigma}}_{post.}$ resulting from the canonical reference model, which play an important role in the extended model. Hence, I introduce a new method, where the unknown covariance matrix is employed as an integral part of the Black-Litterman model. This can be seen as an empirically more relevant model, since it accounts for parameter uncertainty throughout the whole model. As such, this adjustment satisfies the ambition of modeling asset returns while incorporating investor's views in a truly Bayesian setting. I again apply an inverse-Wishart prior, given by $\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}\left(\dot{\boldsymbol{\Sigma}}, v\right)$, for the covariance matrix. For v, the degrees of freedom, I use $v = n$ as to accommodate a diffuse prior. This specification yields the following derivation of the posterior predictive distribution,

$$P(\boldsymbol{r}_{T+1}|\boldsymbol{q}_1, \ldots, \boldsymbol{q}_T) \propto t_{\widetilde{v}_{BLE}-n+1}\left(\widetilde{\boldsymbol{\mu}}_{BLE}, \widetilde{\boldsymbol{\Sigma}}_{BLE}\, \frac{\widetilde{l}_{BLE}+1}{\widetilde{l}_{BLE}(\widetilde{v}_{BLE}-n+1)}\right), \tag{18}$$

where the parameters of the model are indicated by the subscript $_{BLE}$. The full derivation of Equation 18 can be found in Appendix A.6. As I use conjugate priors for this case, the result can be derived analytically. The resulting posterior predictive distribution therefore again is a Student's t-distribution, with parameters defined as $\widetilde{v}_{BLE} = v + n$, $\widetilde{l}_{BLE} = (1/\tau) + n$, $\widetilde{\boldsymbol{\mu}}_{BLE} = \left(\frac{1}{\tau}\boldsymbol{\pi} + n\bar{\boldsymbol{q}}\right) / \widetilde{l}$ and $\widetilde{\boldsymbol{\Sigma}}_{BLE} = \boldsymbol{\Sigma} + \sum_{i=1}^{n}(\boldsymbol{q}_i - \bar{\boldsymbol{q}})(\boldsymbol{q}_i - \bar{\boldsymbol{q}})' + [(n/\tau)(\bar{\boldsymbol{q}} - \boldsymbol{\pi})(\bar{\boldsymbol{q}} - \boldsymbol{\pi})'] / \widetilde{l}$. As the predictive posterior distribution is conditional on the expected returns of the views, the distribution only yields information on the returns of assets for which a view has been given as input. As I work with views on all assets, this is not a problem. When working with an incomplete number of views, other estimation methods such as historical estimation can be used to derive expected returns for the remaining assets.

3.4.2 The Black-Litterman Model With Unknown τ

In the models defined above, the assumption is made that the scaling parameter τ is known, and thus subsequently the variance of the mean parameter is set proportional to the variance of the asset returns. This gives $p(\boldsymbol{\mu}) = \mathcal{N}(\boldsymbol{\pi}, \tau\dot{\boldsymbol{\Sigma}})$. Although this scaling parameter indicates a confidence level and thus a constant value could be assumed, it is difficult to set a specific value for such an abstract parameter. Working in a Bayesian framework, it is more fitting to assume some level of uncertainty associated with this specific parameter. Moreover, with a constant value, no variation is possible in different time periods. Therefore, my second model is based on the canonical reference model, but this time includes a distribution for the scaling parameter τ. I take a normal distribution truncated from below at zero as my prior, since it is unimodal and is less leptokurtic than for example the Student's t- or Laplace distribution, that is it has less mass in its tail. Combined with a relatively low value for the scaling parameter, this ensures a high concentration around the mean, but still allows for some deviation. In this manner, the initialization stays close to the value used in the models with a fixed parameter τ. Hence, for the distribution of τ, I use $\tau \sim \mathcal{N}(0.05, 0.02)$ with support $\tau \in [0, \infty)$. As this model has no analytically available posterior distribution, I use a sampling method to obtain draws for the parameters $\boldsymbol{\mu}$ and τ resulting from the model.

3.4.3 The Black-Litterman Model With Non-Normal Data

Up to this point, one of the main assumption is that the distribution of the asset returns is normal. Although normally distributed asset returns make computations easy to perform and are adapted at large in the financial industry, the distribution itself is questioned in the financial literature. It lacks the fat tailed behaviour which is commonly present in stock returns, as this accounts for the extreme price fluctuations that exist in asset return data. The rejection of the normal distribution is also supported by my preliminary data analysis in Section 2. Peiró (1994) concludes that the Student's t-distribution, which is leptokurtic and thus has more probability mass in its tails, best describes the stock returns. This conclusion is also drawn by Longin (2005), who states that the assumption of a Gaussian distribution for asset returns is rejected and that only the Student's t-distribution is not rejected for his data. As such, I introduce a model where the assumption of a normal distribution is relaxed and where a t-distribution for the data is adopted. In this model, as in the previous model, I assume the scaling parameter τ to be an unknown element of my model with $\tau \sim \mathcal{N}(0.05, 0.02)$. Furthermore, the same sampling method is used to obtain values for the posterior parameters.

3.5 Overview of the Models

To summarize the models discussed in Section 3.3 and 3.4, the main assumptions of each model are depicted in Table 3. In the subcaption of this table, abbreviations for the models are introduced as well.

Table 2: Overview of the models

Model	τ	Assumption $\mathbf{\Sigma}$	Distribution of the data	Sampling method
MEM	0.05	Known, $\dot{\mathbf{\Sigma}}$	Normal	Analytical distribution
CRM	0.05	Known, $\dot{\mathbf{\Sigma}}$	Normal	Analytical distribution
ARM	1	Known, $\dot{\mathbf{\Sigma}}$	Normal	Analytical distribution
ECRM	0.05	Partly known, partly $\mathcal{W}^{-1}(\widehat{\mathbf{\Sigma}}_{post.}, v)$	Normal	Analytical distribution
BLE	0.05	$\mathcal{W}^{-1}(\dot{\mathbf{\Sigma}}, v)$	Normal	Analytical distribution
CRM-τ	$\mathcal{N}(0.05, 0.02)$	Known, $\dot{\mathbf{\Sigma}}$	Normal	HMC
CRM-t	$\mathcal{N}(0.05, 0.02)$	Known, $\dot{\mathbf{\Sigma}}$	Student's t	HMC

Table 3: Overview of the discussed models and their specifications. The abbreviated model names correspond to respectively the Market Equilibrium Model (MEM, Section 3.2), Canonical Reference Model (CRM, Section 3.3.1), Alternative Reference Model (ARM, Section 3.3.2), Extended Canonical Reference Model (ECRM, Section 3.3.3), full Bayesian Black-Litterman model with unknown covariance matrix (BLE, Section 3.4.1), CRM model with unknown scaling parameter τ (CRM-τ, Section 3.4.2) and CRM model with Student's t-distributed data (CRM-t, Section 3.4.3). The abbreviation HMC denotes the sampling method Hamiltonian Monte Carlo.

3.6 The Known Parameters of the Black-Litterman Model

To be able to implement the models discussed in the previous sections, multiple parameters are needed as input. This set of parameters consists of the assumed to be known covariance matrix $\dot{\mathbf{\Sigma}}$, the vector with expected returns of the views $\boldsymbol{q}$, the covariance matrix indicating the confidence in the views $\mathbf{\Omega}$ and the scaling parameter τ. In this section, I discuss the methods available to obtain the concerning parameters and substantiate my choice for their initializations.

3.6.1 Factor Model to Estimate Covariance Matrix $\dot{\mathbf{\Sigma}}$

One of the key assumptions of the original Black-Litterman model is that the mean of the asset returns $\boldsymbol{\mu}$ is unknown, while the covariance matrix of the asset returns $\dot{\mathbf{\Sigma}}$ is presumed to have no uncertainty. As such, there is no need to focus on the dynamics of the covariance matrix. However, I still need a proper estimation which can be used in the asset selection models as this matrix is not observed. Since the dimensions of the covariance matrices that I am estimating are a multiple of my number of observations, the covariance matrix obtained by historical data is not of full rank and therefore not invertible. A possible approach to overcome this problem is by using the shrinkage estimator as introduced by Ledoit and Wolf (2003). This shrinkage estimator is defined as a linear combination of the shrinkage target, obtained by employing the CAPM, and the sample covariance matrix. Although this is an appropriate method to obtain an invertible covariance matrix, the usage of the single factor CAPM makes it less accurate. In order to obtain a more stable and accurate estimate, I use a multi-factor model. A factor model tries to capture the risk and return characteristics of an asset by decomposing it into a number of common factors. For such a factor model, the asset returns as well as the factor returns are required. The former are known and the latter can be estimated by splitting the per factor categorized assets into quantiles and using the difference in return between the first and last quantile as the factor return. Such a multi-factor model with f factors can be written as the multivariate regression

$$\boldsymbol{R} = \boldsymbol{A} + \boldsymbol{B}\boldsymbol{R}_F + \boldsymbol{\Xi} \ , \tag{19}$$

where the components are defined as the asset return matrix $\boldsymbol{R}$, intercept $\boldsymbol{A} \in \mathbb{R}^{n \times T}$, factor loadings $\boldsymbol{B} \in \mathbb{R}^{n \times f}$, factor returns $\boldsymbol{R}_F \in \mathbb{R}^{f \times T}$ and idiosyncratic returns $\boldsymbol{\Xi} \in \mathbb{R}^{n \times T}$. From this factor model, a structured covariance matrix for the asset returns is obtained, which is given by $\text{Var}(\boldsymbol{r}) = \boldsymbol{B}\boldsymbol{\Sigma}_F\boldsymbol{B}' + \boldsymbol{\Sigma}_{\Xi}$, since the factor and idiosyncratic returns are orthogonal by definition. In order to lower the number of parameters to estimate and at the same time reduce the estimation error, I use an additional layer of shrinkage and set $\boldsymbol{\Sigma}_{\Xi}$ equal to a diagonal matrix, assuming uncorrelated residuals.

A factor model frequently applied by academics and used by portfolio managers as basis to construct their own factor model is the Fama-French five factor model (Fama and French, 2015). This factor model is an extension of the three factor model also introduced by Fama and French. Therefore, I use this five factor model in order to compose a shrunk estimate of the covariance matrix for the asset returns. The factors applied in this model are the market sensitivity, Small Minus Big (SMB), High Minus Low (HML), Robust Minus Weak (RMW) and Conservative Minus Aggressive (CMA). These factors capture the difference in performance between companies with opposite characteristics and are composed by subtracting the return of the highest quantile from the return of the lowest quantile, or vice versa depending on the composition of the factor. As such, for the SMB factor, the return of the quantile composed of companies with a small market capitalization is subtracted by their opposite quantile, composed of companies with a large market capitalization. On average, this yields a positive return and thus outperformance of small companies compared to larger companies. The other factors are structured in a similar way and account for respectively the book-to-market ratio (HML), operating profitability (RMW) and investment rate (CMA) of the companies.

3.6.2 Determining the View Parameter $\boldsymbol{q}$

The views for the Black-Litterman model comprise of $\boldsymbol{q}$, the expected return of the view portfolio, $\boldsymbol{P}$, the portfolio pick matrix, and $\boldsymbol{\Omega}$, the uncertainty with regards to the views. Each asset for which there is a view available is located by the pick matrix $\boldsymbol{P}$ and subsequently combined with the mean return of the assets $\boldsymbol{\mu}$ to form the vector $\boldsymbol{P\mu}$. The pick matrix $\boldsymbol{P}$ is important, as it ensures the possibility of stating views on single assets or a combination of assets. These views can thus either be absolute, as in the first case, or relative, as in the second case. Additionally, the weighting scheme for relative views can be adjusted to meet the preference of the investor. For example, weighting can be based on equality or on the market capitalization of an asset. These two different methods of weighting are exemplified in Appendix A.2.

Regarding determining the expected returns of the views, $\boldsymbol{q}$, various methods can be used, including but not limited to historical returns, analyst recommendations, information not accessible to the general public or asset pricing models. Historical returns can be used with different time horizons, depending on the preference of the investor. Analyst recommendations can be employed to incorporate expert opinions in the model. The number of experts stating buy, hold or sell recommendations is then used both to determine the expected percentage of outperformance, as well as the reliability of the advice. Stock price targets are also related to this method, as they indicate the expected under- or outperformance of certain assets in the view of an expert. Furthermore, some investors may have access to private company information. This information can comprise of, for example, management meetings or information on the liquidity of an asset when the investor is already holding a large proportion of it. The main focus of this paper, however, is on the implementation of asset pricing models in order to arrive at an expectation for the return of the set of assets. Hence, I again use the output of the Fama-French five factor model to establish values for the returns of views according to the investor.

3.6.3 Providing a Method to Estimate the Parameter $\boldsymbol{\Omega}$

Using the same multi-factor model as used to obtain the view parameter $\boldsymbol{q}$, I determine the confidence in the view as denoted by the matrix $\boldsymbol{\Omega}$. When assuming independence of the different view portfolios, this matrix is diagonal. There are several ways to define the view estimation error matrix. The first method is to set it proportional to $\boldsymbol{\Sigma}$. He and Litterman apply this by defining $\boldsymbol{\Omega} = \tau \cdot \text{diag}(\boldsymbol{P\Sigma P'})$ and thus forcing the matrix to be diagonal. Meucci (2009) loosens this diagonality constraint by setting $\boldsymbol{\Omega} = \frac{1}{c}\boldsymbol{P\Sigma P'}$, where c in turn is set to $\frac{1}{\tau}$. In the paper by Mankert and Seiler (2011), the estimation error is seen as a confidence interval around the value of the views. Another option is proposed by Idzorek (2007), who tries to simplify the model. As such, the matrix comprises of $\boldsymbol{\Omega} = \alpha \boldsymbol{P\Sigma P'}$, where α is a scalar of confidence measured on the interval between the prior and conditional information. Lastly, when using a factor model in order to determine the expected return of the assets, the residuals can be used to determine a suitable value for $\boldsymbol{\Omega}$. This is applied in the paper by Beach and Orlov (2007), where they use EGARCH derived views. From the mentioned literature, it can be concluded that there is no agreed upon method to determine the parameter $\boldsymbol{\Omega}$. In this paper, I therefore choose for consistency and follow the approach by He and Litterman, who set the uncertainty proportional to the asset variance. In this manner, views on assets with a lot of dispersion have more uncertainty themselves, which is reasonable to assume. In my case, to ensure invertibility of the matrix $\boldsymbol{\Omega}$, I use the shrunk covariance matrix $\dot{\boldsymbol{\Sigma}}$ resulting from the factor model.

3.6.4 Setting the Scaling Parameter τ

Within the Black-Litterman framework, the confidence in the equilibrium return model is depicted by the scaling parameter τ, since it is incorporated in the dispersion around the equilibrium return $\boldsymbol{\pi}$ by $\tau\boldsymbol{\Sigma}$. In the alternative reference model, its value is set equal to one. Following this approach, however, forgoes the updating procedure of the posterior variance, as shown in the derivation of the alternative reference model. In this section, I therefore try to find a better founded value. Since the assigned value is affiliated with a degree of confidence, it is not possible to set a value that holds for every investor. The parameter remains a subjective element of the model. As such, Walters (2013) dedicated a whole paper on different methods to determine the value of τ. He suggests to estimate the parameter using either the standard error of the equilibrium returns, confidence intervals or a fraction of invested wealth. These methods remain somewhat arbitrary however, and in his example he ultimately uses a value of 0.05. He and Litterman (1999) assume the value to be 0.05 as well, corresponding to a confidence level of the CAPM as if it was estimated using 20 years of data. Furthermore, Lee (2000) describes in his book that the value of τ should be close to zero, since the uncertainty inherent to the mean should be far less than that of the returns ($\tau\boldsymbol{\Sigma} < \boldsymbol{\Sigma}$). Following this line of thought, Idzorek (2007) calibrates the parameter to have a value of 0.025. It can be seen that there is no common convention on the value of τ, although most authors agree that $\tau \ll 1$ must hold. As such, for most models discussed in this paper, I again follow the assumption of He and Litterman and define τ to be equal to 0.05. In addition, I define two Black-Litterman model with a full Bayesian interpretation of τ, where the parameter is not a point estimate but rather a distribution. This approach is used in the models defined in Section 3.4.2 and 3.4.3.

3.7 Portfolio Optimization

Having defined all necessities to predict the asset returns, the last step is to determine the weights of the assets in the portfolio. For this, I use the mean-variance framework as introduced by Markowitz (1952). This framework is known for maximizing a quadratic utility function, which equates to finding a balance between the expected return of an asset and its risk, defined as the variance of the asset returns. Investors prefer the former and dislike the latter. To compose the Markowitz portfolio, the vector of expected returns, $\boldsymbol{\mu}$, and the covariance matrix of these assets, denoted by $\boldsymbol{\Sigma}$, are needed. Defining $\boldsymbol{w} \in \mathbb{R}^{n \times 1}$ as a column vector of portfolio weights and r_p as a scalar depicting the return of the portfolio, the expected return of the portfolio is given by $\mathbb{E}(r_p) = \boldsymbol{w}'\boldsymbol{\mu}$ and the variance of the portfolio is given by $\text{Var}(r_p) = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}$. The Markowitz portfolio is the portfolio which maximizes the expected utility of an investor. Assuming a quadratic utility function, this results in maximizing the expected return of a portfolio whilst minimizing its variance. Hence, the unconstrained mean-variance optimization problem can be stated as

$$\max_{\boldsymbol{w}} \boldsymbol{w}'\boldsymbol{\mu} - \frac{\lambda}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \ , \tag{20}$$

with λ again the coefficient of the investor's risk aversion. To maximize this objective function, the first derivative with respect to $\boldsymbol{w}$ is taken and set equal to zero. This yields the optimal portfolio weights as follows,

$$\boldsymbol{\mu} - \lambda\boldsymbol{\Sigma}\boldsymbol{w} = 0 \iff \boldsymbol{w}^* = \frac{1}{\lambda}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \ . \tag{21}$$

Using this optimization procedure, it is possible to extract the optimal portfolio weights given the mean and variance resulting from the model. The sum of these portfolio weights, however, is not necessarily the same for each model. Hence, I add a restriction to the optimization procedure such that the sum of the portfolio weights must equal one: $\boldsymbol{w}'\boldsymbol{\iota} = 1$. Such a restriction is also called a budget restriction, as it determines the fraction of wealth that must be invested. With this restriction, the portfolio weights can still take on extreme values. For example, some weights can be assigned a large negative value, indicating a substantial short position in this asset. To compensate for this short position, there must also be a large long position for some of the assets, as the sum of the weights should still add up to one. Hence, I also include a restriction on the absolute weights of the assets in the portfolio in order to reduce such extreme weights. Considering a maximum asset weight of w_{max}, this yields $\boldsymbol{w} \leq \boldsymbol{\iota} w_{max}$. Furthermore, I consider a restriction on the short selling of assets, which is implemented using the constraint $\boldsymbol{w} \geq \mathbf{0}$, where $\mathbf{0}$ is a vector with zeros as input. This is a relevant constraint as well, as not all investment funds are allowed to take on short positions. Applying these restrictions yields the following two optimization problems,

$$(1) \quad \max_{\boldsymbol{w}} \boldsymbol{w}'\boldsymbol{\mu} - \frac{\lambda}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \text{ subject to } \boldsymbol{w}'\boldsymbol{\iota} = 1 \text{ and } \boldsymbol{w} \leq \boldsymbol{\iota} w_{max} \, , \tag{22}$$

$$(2) \quad \max_{\boldsymbol{w}} \boldsymbol{w}'\boldsymbol{\mu} - \frac{\lambda}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \text{ subject to } \boldsymbol{w}'\boldsymbol{\iota} = 1 \, , \, \boldsymbol{w} \leq \boldsymbol{\iota} w_{max} \text{ and } \boldsymbol{w} \geq \boldsymbol{0} \, . \tag{23}$$

The given optimization problems are solved using a numerical optimizer for quadratic programming problems. Moreover, the rebalancing frequency of the strategies is equal to once per month.

3.8 Sampling Process

For the more advanced models that I discussed, direct sampling from the distribution is not possible. Therefore, I use an alternative sampling method to obtain draws from the posterior distribution. A widely recognized sampling method is the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller, 1953; Hastings, 1970), which is also the basis for other sampling methods such as the Gibbs sampler (Geman and Geman, 1987). The approach that I choose is the Hamiltonian Monte Carlo method, which is introduced by Duane, Kennedy, Pendleton, and Roweth (1987) under the name Hybrid Monte Carlo in the field of quantum chromodynamics. This sampler, where the parameter state is represented by a particle, is a special case of the Metropolis-Hastings algorithm as well. The simulations are subsequently the paths of the particle when moving around the frictionless surface of the minus-log-posterior. These paths are transitions of the Markov chains. To determine which path to follow, the algorithm uses the gradient (the curvature) of the posterior distribution. Each simulation produces a proposal for the Metropolis-Hastings MCMC algorithm, and if accepted the next iteration will start at the location of the accepted proposal. This process is repeated for a predefined number of times. Hence, while moving along its paths, the particle explores the posterior distribution. Using these steps, draws from the approximated posterior distribution are obtained.

The advantage of the Hamiltonian Monte Carlo algorithm is that it is able to explore the state space faster than 'guess and check' Metropolis-Hastings algorithms such as the original Metropolis-Hastings algorithm. Especially for higher dimensional target distributions, it makes for quicker convergence. This is because guess and check algorithms tend to get stuck due to the large number of directions to explore, while only a fraction of these direction yields an accepted step.

The Hamiltonian Monte Carlo sampling algorithm has become popular among statisticians for its application in Bayesian statistical inference. A widely used implementation of the method can be found in the software package Stan (Carpenter et al., 2017), which implements the Hamiltonian Monte Carlo approach through the No-U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014). I use this software package, adjusted to work in R,[4] in order to obtain draws from the posterior distributions. The Hamiltonian Monte Carlo algorithm uses certain initial specifications to begin the sampling process, such as the number of iterations and step size. The initializations I use in the sampling procedures can be found in Appendix A.7. One thing to point out is that my computational resources are limited, which reflects on the settings of the algorithm. For example, the number of iterations is set at 100 when running a model using the whole data set, as this is roughly the maximum number for which the program is able to complete its estimations within reasonable time due to the complexity of the model. To investigate the accuracy when using such a small number of iterations, a subset of the data, for which efficient computation is possible, is used to perform a sensitivity analysis.

In order to check whether the formed Markov chains of the sampler have converged, the commonly used estimator $\widehat{R}$ is adopted. This estimator is defined as the ratio of the between and within dispersion among multiple chains and is given by

$$\widehat{R} = \sqrt{\frac{(df+3)\widehat{V}}{(df+1)W}}\,, \tag{24}$$

where df denotes the degrees of freedom, $\widehat{V}$ the pooled variance and W the mean variance within each chain. The derivation of the convergence diagnostic is given in Appendix A.8. The value of $\widehat{R}$ should not be substantially larger than 1, and it is argued by Vehtari, Gelman, Simpson, Carpenter, and Bürkner (2019) that at a value of $\widehat{R} < 1.01$, one can be fairly confident of convergence. Additional convergence measures available for the Hamiltonian Monte Carlo sampler are, among others, the number of divergent transitions and the tree depth, where the former is more severe than the latter. When there is a divergence noticeable in the Markov chain, this indicates that the sampler encountered a region in the target distribution with a high gradient which it cannot adequately explore. Such an occurrence is called a divergent transition. The information on the divergent transitions can be used to adjust the settings of the sampling process such that the sampler is better able to explore the target distribution, for example by increasing the number of iterations or by decreasing the step size per transition. Regarding the tree depth, this value determines the number of iteratively accumulated steps

[4] Stan Development Team. 2018. *RStan: the R interface to Stan.* R package version 2.17.3. `http://mc-stan.org`

per transition. To ensure that infinite loops of the algorithm are prevented, the tree depth has a predefined maximum. This maximum is reached, for example, in non-identified models, or when the model is so complex that the sampler reaches its maximum even though the model is identifiable. Hence, all convergence diagnostics measure to which extent the Hamiltonian Monte Carlo sampler is able to efficiently and accurately explore the target distribution.

3.9 Benchmark Models

In order to evaluate the performance of the different Black-Litterman models, I compare their results with a couple of (simpler) benchmark models. As the theory of the Black-Litterman model is based on the CAPM, the first benchmark used is the optimal strategy in such a setting. That is, I use a portfolio based on the market capitalization of each asset to form a value weighted market portfolio. In theory, as the weights increase and decrease with the growth and decline of each asset, the turnover of this portfolio is zero. However, as the number of available stocks of some assets also slightly changes in the data set over each period, the value weighted benchmark portfolio has a small turnover as well.

The second benchmark model is the equal weight portfolio. This asset selection model assigns a weight of $w_i = \frac{1}{N}$ to each asset, where N is the total number of assets. As there is a direct selection of the portfolio weights, there is no need to estimate the mean and the variance of the asset returns. Hence, the equal weight portfolio is not prone to estimation error, adequate diversification and a low turnover which results in less costs as compared to a model where more transactions are needed to maintain the strategy. It is shown by DeMiguel et al. (2007) that it is hard to consistently outperform this model in terms of the Sharpe ratio, and it is therefore frequently used to compare new portfolio strategies.

Lastly, since the multi-factor models as introduced by Fama and French (1993, 2015) receive much praise and are currently widely implemented, the Fama-French five factor model (Fama and French, 2015) is used as a benchmark as well. A general formulation of such a multi-factor model is given in Equation 19. The variance of this benchmark model is equal to the shrunk covariance matrix resulting from the Fama-French five factor model, $\dot{\mathbf{\Sigma}}$. Extracting the needed parameters from this set-up, the portfolio weights are calculated using these parameters as input for the Markowitz portfolio weights formula. In order to add to the comparability of this benchmark, the portfolio weights are restricted as to match the two scenarios introduced in Section 3.7.

3.10 Performance Evaluation

Having defined the various Black-Litterman and benchmark models, a measure is needed to compare the out-of-sample performance of the composed portfolios. For this, a rolling window is used in the computation of the asset weights. Denoting the different strategies by s and the rolling window size by h, the performance of the portfolios are determined using the following criteria:

(i) *Mean return:* Calculating the arithmetic average return over the out-of-sample period yields the mean return $\widehat{\mu}$ for each month. With this measure, first moments of the different portfolio strategies are compared. However, no adjustment in terms of risk is taken into account. The mean return is given by

$$\widehat{\mu}(s) = \frac{1}{T-h} \sum_{t=h+1}^{T} r_t(s) \ . \tag{25}$$

In this equation, $r_t(s)$ is the portfolio return of strategy or model s at time t.

(ii) *Volatility:* The volatility of each strategy, denoted by $\widehat{\sigma}$, quantifies the degree of dispersion of the portfolio returns. It is a measure of the risk incorporated in the returns, as a higher variation in the returns equates to less security of a stable outcome. The volatility, also called the standard deviation, is given by

$$\widehat{\sigma}(s) = \sqrt{\frac{1}{T-h-1} \sum_{t=h+1}^{T} [r_t(s) - \widehat{\mu}]^2} \ . \tag{26}$$

(iii) *Sharpe ratio:* By taking into account both the risk and return inherent to the portfolio returns, a more comprehensive view of the model performance is given. For this, the mean return and sample standard deviation are combined and used in the formula of the Sharpe ratio. This measure is widely used among portfolio managers, since it strikes an easily definable balance between risk and return. The ratio is given by

$$\widehat{\mathrm{SR}}(s) = \frac{\widehat{\mu}(s)}{\widehat{\sigma}(s)} \ . \tag{27}$$

(iv) *Compound annual growth rate:* Opposite to the arithmetic mean return $\widehat{\mu}$, the compound annual growth rate (CAGR) measures the yearly rate of return that is required to achieve the same result as the depicted strategy. The CAGR is given by

$$\widehat{CAGR}(s) = \sqrt[\frac{(T-h)}{12}]{(1+r_h) \cdot (1+r_{h+1}) \dots (1+r_T)} - 1 \ . \tag{28}$$

(v) *Expected shortfall:* As a measure for the risk embedded in the returns besides the standard deviation, the expected shortfall is used. The advantage of the expected shortfall when compared with the standard deviation, is that the expected shortfall takes into account the tail of a distribution to determine the risk incorporated in the model. The expected shortfall for the portfolio loss distribution X is given by

$$\widehat{\text{ES}}_\alpha(X) = \mathbb{E}(X|X \geq \text{VaR}_\alpha) = \frac{1}{1-\alpha}\int_\alpha^1 \text{VaR}_\gamma(X)d\gamma \ , \tag{29}$$

where $\text{VaR}_\alpha(X) = \min\{c : P(X \leq c) \geq \alpha\}$ denotes the value at risk. It thus equals the average expected loss when the loss is greater than the value at risk. For α, I use a value of 0.9. I use the expected shortfall instead of the value at risk since the latter is not a coherent risk measure. A risk measure is said to be coherent if it satisfies the axioms of subadditivity, monotonicity, positive homogeneity and translation invariance.

(vi) *Information ratio:* Many portfolio managers measure their performance relative to a certain benchmark they follow. Commonly used benchmarks are stock indices for equity returns and treasury yield indices for fixed income investors. As such, I add a performance evaluation criteria based on the relative performance compared to my asset universe, the S&P 500 index. The index data does not include dividend payments. The information ratio denotes the difference between the risk-adjusted return of the chosen strategy s and the benchmark. Its formula is given by

$$\widehat{\text{IR}}(s) = \frac{\mathbb{E}(r_p - r_b)}{\sqrt{\text{Var}(\boldsymbol{r}_p - \boldsymbol{r}_b)}} \ , \tag{30}$$

where r_b denotes the return of the benchmark. The difference in returns, as depicted in the numerator, is called the active return, and the standard deviation of the active return is known as the tracking error. As the information ratio compares the performance relative to a benchmark, this measure is only given for the Black-Litterman models and not for the benchmarks themselves.

(vii) *Turnover:* Since trading is costly due to the required transactions, it is also informative to get a grasp of the amount of trades that needs to be done in order to follow a given strategy. For this, the average turnover per rebalancing date of a portfolio can be computed by the formula

$$\widehat{\text{TO}}(s) = \frac{1}{T-h}\sum_{t=h+1}^{T}\sum_{j=1}^{n}(|w_{t+1,j}(s) - w_{t^+,j}(s)|) \ , \tag{31}$$

where $w_{t,j}(s)$ is the portfolio weight of asset j at time t, and $w_{t^+,j}(s)$ is the portfolio weight just before rebalancing at time $t+1$. As such, the turnover indicates which fraction of the portfolio is traded on average at a certain rebalancing date.

4 Results

Following the theoretical exploration of the different proposed models, it is time to look at the resulting portfolios and their performances. For the conciseness of the tables containing the results, the abbreviations as introduced in the model overview table, Table 3, are used.

4.1 Model Results With Budget and Short Selling Restrictions

The first results are obtained using both a budget restriction and a short selling restriction and can be found in Table 4. When focusing solely on the mean portfolio return, the restricted factor model stands out as it performs best with an average return of 3.2% per month. To obtain this result, more risk has been taken in terms of both volatility and expected shortfall. For example, observing the value for the expected shortfall, an average loss of 18.1% is expected in the worst 10% of monthly portfolio return cases. This increase in risk results in a sub-optimal value for the risk-adjusted return as measured by the Sharpe ratio.

Table 4: Model results with budget and short selling restrictions

Model	μ (%)	σ (%)	SR (%)	CAGR (%)	$ES_{0.9}$ (%)	TO (%)	IR (%)
MEM	0.9	3.5	24.6	10.0	-6.3	3.8	12.0
CRM	2.8	12.8	21.6	26.8	-18.1	66.1	16.3
ARM	3.2	13.4	23.6	31.9	-18.1	62.4	19.0
ECRM	2.6	11.9	22.3	25.5	-20.9	55.5	17.2
BLE	2.3	9.8	23.4	24.0	-14.9	59.2	18.0
CRM-τ	3.2	13.4	23.7	32.1	-18.1	62.3	19.1
CRM-t	1.9	6.6	28.1	21.6	-10.9	78.3	18.6
Market capitalization	0.9	3.6	23.9	9.9	-6.5	2.9	
Equal weight portfolio	1.1	3.9	28.6	13.2	-6.8	5.6	
Restricted factor model	3.2	13.5	23.8	32.3	-18.1	61.0	

Monthly performance evaluation results for the outlined models and their benchmarks. A rolling window size of $h = 36$ *is used in the computations and for the expected shortfall (ES),* α *is set at* 0.9. *The benchmark for the information ratio (IR) is the S&P 500 index. As the benchmark models are a benchmark themselves, no information ratio is provided. Furthermore, the portfolio weights are restricted by a budget and a short selling constraint. The Sharpe ratio, compounded annual growth rate and turnover are abbreviated by respectively SR, CAGR and TO.*

Because of its high diversification and therefore low risk profile, the highest Sharpe ratio is obtained by the equal weight portfolio. Together with the CRM-τ model, this model is the second best performing model measured by its ranking, which is shown in Table 5. Where the equal weight strategy excels in its high risk-adjusted return, that is less risk for the same return, the CRM-τ obtains a superior score when looking at the pure return measures μ and CAGR. Moreover, the CRM-τ model outperforms the other Black-Litterman models in terms of the information ratio. This indicates that the model is capable of most consistently outperforming the S&P 500 index, as its active return corrected for the taken risk is higher.

When investigating the turnover of the strategies, it can be seen that, for the portfolios in which views are involved, the simpler CRM and ARM models have a higher turnover value. For the CRM model for example, the turnover of 66.1% implies that every month, on average this percentage of the assets is bought and sold, which indicates a very active trading posture. This suggests that incorporating more uncertainty by means of a more extensive Bayesian framework yields more stable portfolio weights over time, as the turnover of the ECRM and BLE models for example is substantially lower. The exception to this is the CRM-t model, but since its convergence is doubtful, it is expected to result in a more unstable model. When trading costs are a concern for the investor, the MEM model performs better with a much lower turnover value of 3.8%. As its mean return parameter is equal to the market equilibrium returns with some incorporated dispersion, and therefore more consistently follows the variation in the market values of the companies, the lower turnover value is as expected. The closer a strategy stays to the market value weights, the lower the turnover value, as can also be seen in the market capitalization model. Such a strategy approaches a buy-and-hold strategy, as the assets are longer in the possession of the investor.

The worst performing models in this restricted setting are the CRM and ECRM models. Apparently, as the result of the CRM model is used in the prior specification of the ECRM model, the ECRM model is not able to improve the performance by using a Bayesian adjustment. In fact, as a result of the added uncertainty, the risk of the model in terms of expected shortfall is enlarged. Overall, the Black-Litterman models with the given budget and short selling restrictions are not able to consistently outperform the benchmark models when investigating the current set of performance evaluation measures.

Table 5: Ranking of the (benchmark) models with budget and short selling restrictions

Model	μ	SR	CAGR	$ES_{0.9}$	TO	Total score	Overall rank
MEM	9	3	9	1	2	24	**4**
CRM	4	10	4	6	9	33	**9**
ARM	3	7	3	6	8	27	**5**
ECRM	5	9	5	10	4	33	**9**
BLE	6	8	6	5	5	30	**7**
CRM-τ	2	6	2	6	7	23	**2**
CRM-t	7	2	7	4	10	30	**7**
Market capitalization	10	4	10	2	1	27	**5**
Equal weight portfolio	8	1	8	3	3	23	**2**
Restricted factor model	1	5	1	6	6	19	**1**

This table indicates the ranking of the portfolios measured on the performance evaluation criteria. The lower the value of the score, the better the performance of the model compared to the other models. The overall rank is determined using the sum of the scores. A lower rank indicates a relatively better performing model. When the total score of multiple models is equal, both obtain an equal rank.

The portfolio returns resulting from the discussed investment strategies can also be regressed on the five Fama-French factors, as discussed in Section 3.6.1, to investigate whether the outperformance of the models can be explained by a tilt to one of the given factors. The results of such a regression can be found in Table 11 in Appendix C.1. It can be seen that especially the market factor has a large impact on the returns. This factor is highly significant for all models, and thus a tilt towards this factor results. Additionally, a slight tilt towards the value factor can be distinguished, which indicates that the outperformance is partly due to an overweight of smaller companies.

4.2 Model Results With Budget, Short Selling and Weight Restrictions

The following scenario has tighter restrictions regarding the composition of the portfolio weights. Not only are the portfolio weights restricted by a budget and short selling constraint, the maximum value has been limited as well, to 1% of the total portfolio. As such, highly overweighted assets in the portfolios are prevented, forcing the portfolios to be better diversified. The performance measure results for this scenario can be found in Table 6 and the ranking of the models can be found in Table 7.

Table 6: Model results with budget, short selling and weight restrictions

Model	μ (%)	σ (%)	SR (%)	CAGR (%)	$ES_{0.9}$ (%)	TO (%)	IR (%)
MEM	0.9	3.4	25.5	10.3	-6.2	5.8	13.7
CRM	1.1	3.6	29.6	12.7	-6.1	25.9	14.7
ARM	1.1	3.6	29.9	12.9	-6.1	26.0	15.4
ECRM	1.3	4.1	32.5	16.2	-6.9	30.4	31.0
BLE	1.1	5.1	22.5	12.9	-9.6	27.0	14.0
CRM-τ	1.1	5.1	29.9	12.9	-6.1	26.1	15.4
CRM-t	1.2	3.7	32.1	14.5	-6.4	48.6	24.8
Market capitalization	0.9	3.6	23.9	9.9	-6.5	2.9	
Equal weight portfolio	1.1	3.9	28.6	13.2	-6.8	5.6	
Restricted factor model	1.1	3.6	29.8	12.9	-6.1	26.1	

Monthly performance evaluation results for the outlined models. A rolling window size of $h = 36$ is used in the computations and for the expected shortfall (ES), α is set at 0.9. The benchmark for the information ratio (IR) is the S&P 500 index. As the benchmark models are a benchmark themselves, no information ratio is provided. Furthermore, the portfolio weights are restricted by a budget and short selling constraint and the maximum asset weight in the portfolio is set at 1% of the total portfolio. The Sharpe ratio, compounded annual growth rate and turnover are abbreviated by respectively SR, CAGR and TO.

As the degree of freedom in choosing the portfolio weights is lower in this scenario, the values of the performance measures show more similarities. The measures indicating risk, that is the volatility and expected shortfall, have lower values compared to the scenario discussed before, as the dispersion in the weights and therefore portfolio returns is lower. Furthermore, because of the limited trading possibilities, the turnover values decrease, which reduces costs for investors. Again, the CRM-t model forms an exception with a much higher turnover value, but the same instability rationale explains this behaviour. The turnover of the MEM model increases as well. Apparently, due to the added weight restriction, the model is forced to increase the frequency or size of its transactions in order to maintain the chosen strategy.

The best performing model, however, is the ARM model. Measured along the performance measures, it does not outperform any other model within a certain area, but has an above average performance over all measures. When looking at the CRM-τ model, the performance is very similar, albeit the CRM-τ model just underperforms when compared with the ARM model. In the previous scenario, this was the other way around. Adding uncertainty to the mean of the asset returns hence decreases the performance when tighter restrictions are used as this results in less freedom in the choice of asset weights.

Table 7: Ranking of the (benchmark) models with budget, short selling and weight restrictions

Model	μ	SR	CAGR	$ES_{0.9}$	TO	Total score	Overall rank
MEM	9	8	9	5	3	34	**8**
CRM	8	6	8	1	4	27	**6**
ARM	5	3	5	2	5	20	**1**
ECRM	1	1	1	9	9	21	**2**
BLE	3	10	4	10	8	35	**9**
CRM-τ	6	4	6	2	6	24	**4**
CRM-t	2	2	2	6	10	22	**3**
Market capitalization	10	9	10	7	1	37	**10**
Equal weight portfolio	4	7	3	8	2	24	**4**
Restricted factor model	7	5	7	2	7	28	**7**

This table indicates the ranking of the portfolios measured on the performance evaluation criteria. The lower the value of the score, the better the performance of the model compared to the other models. The overall rank is determined using the sum of the scores. A lower rank indicates a relatively better performing model. When the total score of multiple models is equal, both obtain an equal rank.

Despite the additional restriction, the ECRM model performs generally good as well compared to the other models in terms of its mean return, Sharpe ratio and specifically its information ratio. The model is able to maintain a relatively high mean return statistic whilst benefiting from a lower volatility, which results in higher ratios. The model thus excels in its active portfolio return. Hence, due to the stricter weight optimization limitations, the model profits from a lower variance, increasing its risk-adjusted return compared to the previous scenario. There is, however, also a relation with the turnover rate, which is second highest only to the CRM-t model. In order to maintain its superior portfolio return, the active trading characteristic must be maintained as well. Furthermore, it is interesting to see that there is a considerable improvement of the ECRM model over the simpler CRM model. The added complexity and incorporation of the additional uncertainty benefit the model in this scenario.

Lastly, regressing the portfolio returns on the five factors, it can again be seen in Table 12, Appendix C.1, that there is a tilt towards the market factor. However, in contrast to the previous results, the models now suffer from a tilt away from the size factor. Due to the added weight restriction, the models are forced to be more diversified and are thus less able to collect on this factor.

4.3 Convergence Diagnostics

For the CRM-τ and CRM-t models, I use the Hamiltonian Monte Carlo sampler to obtain draws from the posterior distribution of parameters. As it is an MCMC method, the following diagnostics are declared in Section 3.8 in order to identify whether the corresponding Markov chains have converged: the number of divergent transitions, the $\widehat{R}$ statistic and the tree depth. Whenever the Markov chains show no convergence, the draws obtained from the chains are deemed not reliable.

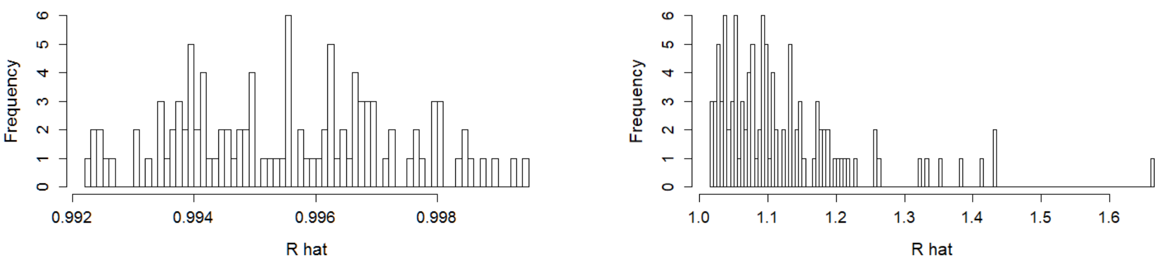

Figure 5: *Histogram of the average* $\widehat{R}$ *values for the CRM-*τ *(left) and CRM-t (right) model. The time period spans 2010-01 to 2018-12. 100 iterations per chain and four chains are used in the sampling process.*

Starting with the divergent transitions, for both the CRM-τ and CRM-t model there were no divergent transitions detected. This indicates that the sampler has no problem exploring the target distribution. However, continuing with the $\widehat{R}$ diagnostic, it can be seen in Figure 5 that for the CRM-t model on the right, the average $\widehat{R}$ values are quite high. As stated in Section 3.8, the desired value for the diagnostic does not exceed 1.01. With average values reaching a maximum higher than 1.6 and a vast majority of the values exceeding the threshold value, it is clear that there is no sufficient confidence in convergence of the parameter values for the CRM-t model. For the CRM-τ model, all values appear to remain within the bound of the threshold and therefore the confidence in the convergence of the CRM-τ model is, on the contrary, adequate. The same pattern can be recognized from Figure 6, where the average tree depth of the models is presented. Especially for the time period from 2014 onward, the CRM-τ model is able to efficiently explore the target distribution and therefore has low average tree depth values. As a rolling window is used, this period defined by the years 2014-2018 is not biased anymore by the aftermath of the financial crisis, yielding a less complex model to explore. For the CRM-t model, however, the complexity seems to increase as the tree depth has an incline during this period. This may be due to the fact that this period is predominantly a period of economic expansion, which can be seen in Figure 9 in Appendix B.2, with less extreme and opposite asset returns. For this, the adopted Student's t-distribution is less suited as it has heavier tails.

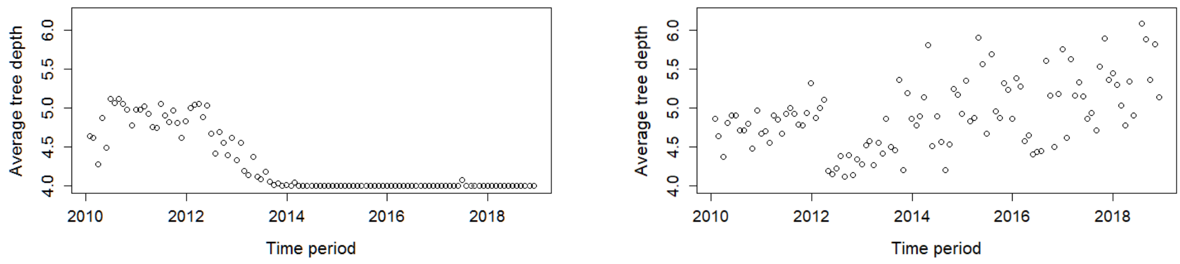

Figure 6: *Plot of the average tree depth for the CRM-τ (left) and CRM-t (right) model. The time period spans 2010-01 to 2018-12. 100 iterations per chain and four chains are used in the sampling process.*

When investigating the given convergence diagnostic values, the target distribution of the CRM-τ model seems to be explored sufficiently by the Hamiltonian Monte Carlo sampler, with no divergent transitions and low values for both the average $\widehat{R}$ statistic and the tree depth. The CRM-t model, on the other hand, seems to be more complex and the 100 iterations used probably are not enough to explore the posterior distribution adequately.

4.4 Robustness Analysis

As shown in the previous section, the convergence of the CRM-t model is doubtful. By means of a robustness analysis, the level of confidence in the models could be defined more accurately. Since an increase in iterations for the whole data set is not possible with my current computational resources, I perform a sensitivity analysis on a subset of the data as to examine the robustness of the (sampling) models in various settings. The subset consists of 30 assets, chosen to eliminate bias created by the size of the companies and the sectors in which they operate. For this, the Standard Industrial Classification (SIC) code for each company is retrieved such that an equal distribution over the industries is achieved. Furthermore, from each industry, two asset belonging to the 40% largest of the group as measured by their market capitalization are chosen, and vice versa for two smaller companies. On this subset, a sensitivity analysis is performed and discussed in the following sections.

4.4.1 The Value for τ

When using the CRM-τ and CRM-t models with the subset of assets, the first element that seems to decrease the efficiency of the sampling algorithm to obtain posterior draws is the mean of the scaling parameter τ. This value was previously set at 0.05 such that the distribution for τ was given by $\mathcal{N}(0.05, 0.02)$. As can be seen in Figure 7, the number of divergent transitions for the

CRM-t model decrease with an increase in this mean value for τ. This indicates that, to improve the accuracy of the CRM-t model for a low number of iterations, a higher value for τ should be used. Therefore, in the following sections, where the subset is employed, $\tau \sim \mathcal{N}(0.1, 0.02)$ is used as this increases the efficiency of the subset. In turn, this implies that the algorithm is able to perform better when there is less confidence in the market equilibrium, since a higher value for τ denotes a larger dispersion around the market equilibrium returns.

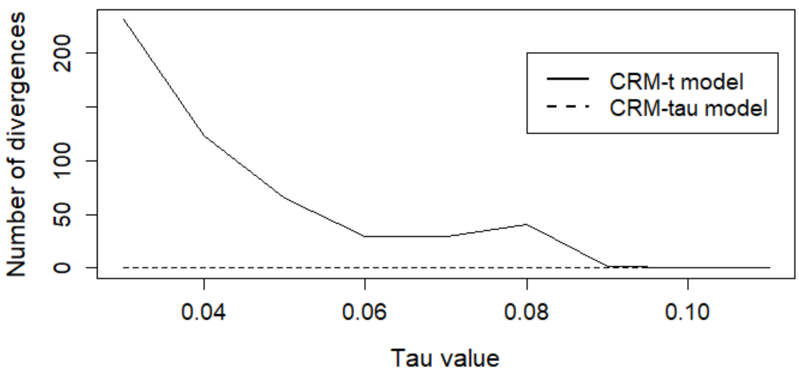

Figure 7: *Shown in this graph are the number of divergent transitions per value of the parameter τ for the CRM-τ (dashed) and CRM-t (solid) model. The number of iterations is set at 100.*

For comparison, the rankings of the models with $\tau = 0.1$ when using the complete data set are given in Appendix C.2, Tables 13 and 14. It can be seen that for the first scenario, the restricted factor model is still the best performing model. The performance of the CRM-τ model slightly decreases, but overall the ranking stays largely the same. For the second scenario, the performance is comparable to the situation with $\tau = 0.05$ as well, although the CRM-t model depletes most of its performance. However, as it was already concluded that the outcome of this model is less reliable as measured by the convergence diagnostics, not too much emphasis is laid on the model. Moreover, the average performance in the second scenario slightly decreases, which indicates that the higher confidence in the equilibrium returns, as illustrated by a lower value for τ and as suggested by the literature, seems valid. When looking at the performance of the previously best performing model, the ARM model, it can be seen that it does not benefit from the change in τ as it does not incorporate this value, and thus its performance reduces compared to the other models.

4.4.2 Iterations

As a next step, also the number of iterations should be able to increase the accuracy of the model. An increase in the number of iterations does not only increase the number of drawn samples which can be used to determine the parameter values, but it also increases the number of warm-up iterations used to set the optimal sampling conditions. The results for varying iteration values for the CRM-τ and CRM-t model in the first scenario can be found in Table 8. In this table, it can be seen that the CRM-τ model performs very consistently as no difference can be distinguished from the results when using 100 or 10,000 iterations. This yields the expectation that the model is able to supply reliable results with iterations as low as 100, which was also concluded in section 4.3. The CRM-t model, on the other hand, shows much more dispersion between the resulting values for various iteration settings. It seems to converge, as the values with 1,000 and 10,000 iterations are much closer than when using 100 iterations, and as such a larger number of iterations is recommended for this model. Moreover, as the number of iterations increases, the turnover decreases while maintaining or increasing the overall performance, which indicates that the model becomes more stable and employable as a result of such an increase. The same behaviour is seen in the setting in which there is an additional weight restriction, for which the results can be found in Table 15 in Appendix C.3. This table furthermore shows that, due to the added restriction, the values of the measures are more similar per set of iterations.

Table 8: Model results for the CRM-τ and CRM-t model with a varying number of iterations

Model	Iterations	μ (%)	σ (%)	SR (%)	CAGR (%)	$ES_{0.9}$ (%)	TO (%)	IR (%)
	100	-0.2	11.8	-2.0	-10.4	-20.9	47.2	-9.9
CRM-τ	1,000	-0.2	11.8	-2.0	-10.4	-20.9	47.2	-9.9
	10,000	-0.2	11.8	-2.0	-10.4	-20.9	47.2	-9.9
	100	0.3	6.2	4.2	0.8	-10.9	139.2	-13.9
CRM-t	1,000	0.6	5.8	10.3	5.2	-10.7	68.9	-5.9
	10,000	0.6	6.0	10.2	5.3	-11.8	34.9	-5.2

Performance evaluation measure values for the CRM-τ and CRM-t models when using different numbers of iterations. For τ, $\mathcal{N}(0.1, 0.02)$ is used. Moreover, a rolling window size of $h = 36$ is used in the computations and for the expected shortfall (ES), α is set at 0.9. The benchmark for the information ratio (IR) is the S&P 500 index. Furthermore, the portfolio weights are restricted by a budget constraint and a short selling constraint. The Sharpe ratio, compounded annual growth rate and turnover are abbreviated by respectively SR, CAGR and TO.

4.4.3 Time Periods

To further test the robustness of the model results, I take into account different time periods. The standard time period spanning 2007 to 2018 is compared with two periods with a shorter duration, spanning 2007 to 2015 and 2010 to 2018. In Appendix B.2, Figure 9, the total market capitalization over the whole data set is shown. From this graph, it can be seen that the periods beginning in 2007 start just before the financial crisis takes effect, while the period beginning in 2010 starts after this tumultuous period and experiences a predominantly growing market size. When investigating how the models perform during these periods, Table 9 shows that especially the more complex models, being the BLE and the CRM-t models, perform well in more turbulent periods, while the simpler models, such as the MEM and market capitalization benchmark model, outperform in periods of consistent growth. Moreover, the ECRM model is able to improve the performance of the CRM model, unless the market growth is more stable, as in the last period. When comparing the three different time periods, the equal weight benchmark portfolio is in no period the top ranking model, but is able to maintain above average performance within all periods and thus ranks as the best overall model.

Table 9: Ranking of the (benchmark) models in various time periods

Model	Ranking period			Sum of ranks	Overall rank
	2007-2015	2007-2018	2010-2018		
MEM	8	3	4	15	**3**
CRM	10	9	1	20	**8**
ARM	6	7	5	18	**6**
ECRM	4	1	8	13	**2**
BLE	1	5	10	16	**4**
CRM-τ	7	7	7	21	**9**
CRM-t	3	6	9	18	**6**
Market capitalization	9	4	3	16	**4**
Equal weight portfolio	2	2	2	6	**1**
Restricted factor model	5	10	6	21	**9**

Ranking of the different models when using data from a specific time period. The periods span respectively 2007-01 to 2015-12, 2007-01 to 2018-12 and 2010-01 to 2018-12. A lower ranking indicates a relatively better performing model. The models are restricted by a budget and short selling constraint. For the computation of the models, $\tau = 0.1$ *and 1,000 iterations are used. The ranking criteria are the same as before.*

For the environment with a supplemental weight restriction, especially the performance of the CRM-τ and the restricted factor model increases as they are better able to benefit from the tighter restrictions in turbulent periods. The results can be found in Table 16 in Appendix C.3. Furthermore, due to the tighter restrictions, the overall risk of the models decreases and thus the advantage of the ECRM and BLE models, that is incorporating more uncertainty, fades as can be seen by their decreases performances. The equal weight portfolio strategy loses its edge over the competing strategies as well, as the other models are by construction more diversified due to the weight restriction.

4.4.4 Rolling Window Sizes

In addition to the period from which the data is obtained, as discussed in the previous section, the amount of data to be utilized in the calculations can be modified as well. In general, an increase in data is expected to increase the accuracy of the models. Again, the best overall performing model, as can be seen in Table 10, is the equal weight portfolio. As the amount of data has no effect on this strategy, it is able to achieve a consistent performance independent of the rolling window size.

When comparing the ECRM and the BLE model, where the BLE model has a more extensive Bayesian implementation, it can be seen that the BLE model more efficiently uses the increase in data as it is able to outperform the ECRM model when using larger window sizes. Apparently, with more data available, it is beneficial to account for additional uncertainty using a Bayesian framework. This conclusion is also strengthened by the results of the CRM-t model. As for the CRM and CRM-τ model, these models abide well when there is less data available, which is one of the properties of a Bayesian approach.

Furthermore, historical market equilibrium returns seem to be a good indication of future asset returns, as the MEM and market capitalization models, where this element plays an important role, also perform notably better with an increase in available data. Investigating the results when applying the additional portfolio weights restriction, as presented in Table 17 in Appendix C.3, the results show to be similar, although the CRM and ARM models seem to have switched places in the ranking. As was also concluded for the whole data set, the ARM model performs better in the presence of a weight restriction.

Table 10: Ranking of the (benchmark) models using various rolling window sizes

Model	Ranking per rolling window size			Sum of ranks	Overall rank
	$h = 12$	$h = 36$	$h = 60$		
MEM	6	3	3	12	**3**
CRM	1	9	7	17	**5**
ARM	6	7	9	22	**9**
ECRM	4	1	6	11	**2**
BLE	10	5	2	17	**5**
CRM-τ	3	7	10	20	**7**
CRM-t	9	6	5	20	**7**
Market capitalization	6	4	4	14	**4**
Equal weight portfolio	2	2	1	5	**1**
Restricted factor model	5	10	8	23	**10**

Ranking of the different models when using different rolling window sizes (h). A lower ranking indicates a relatively better performing model. The models are restricted by a budget and short selling constraint. For the computation of the models, $\tau = 0.1$ *and 1,000 iterations are used. The ranking criteria are the same as before.*

5 Conclusion

Throughout this paper, multiple Black-Litterman models have been discussed in which public information of asset returns is combined with the private information of investors. These models yield a competent foundation for portfolio managers to be used as asset selection methods. Besides interlinking various sources of data, the models allow for a more flexible specification of the model initialization in order to accommodate the preferences of the investor. For example, the choice of the views, the different prior specifications and the hierarchy of the model are all adaptable elements. As such, when specifying a model, the implementer should carefully determine the input parameters, such as the views and the confidence in the model, to optimize its performance.

For the specifications given in this paper, the Black-Litterman models are only in some instances able to convincingly outperform the benchmark models, but their performance is dependent on the data and the specified parameters per model. When compared to the S&P 500 benchmark however, which is used in the information ratio, all models are able to obtain a superior performance. This performance is, in turn, dependent on the set restriction which the portfolio weights should satisfy. The stricter these conditions, the lower the degrees of freedom that are left for the models to benefit from their flexibility. This lower flexibility translates predominantly in a lower exposure to the risk incorporated in the models, as measures by their volatility and expected shortfall. Some models, such as the ECRM model, seem to benefit from these tighter restriction, as it dampens its losses. Depending on the chosen restriction scenario, the restricted factor and ECRM models perform best. For the models with a more extensive Bayesian approach, namely the ECRM, BLE and CRM-τ models, their performance varies, but overall they are able to perform well in the different scenarios investigated in the sensitivity analysis. However, the diversified equal weight portfolio remains a fierce contender for these models. As such, mainly in certain specific environments, portfolio construction benefits from the Bayesian frameworks. Hence, while currently frequently neglected by practitioners, it is beneficial to incorporate such a framework in the asset allocation process, as it reduces the estimation error present in the estimation process and demonstrates to be able to outperform competing strategies.

However, the best performing models also demonstrate to have their downsides. As shown by the robustness analysis, the more complex Bayesian CRM-t model incorporates the extended Bayesian framework as well, but requires extensive computational resources to yield acceptable convergence and is thus more demanding to implement. The performance of this model therefore still needs additional research. Moreover, the outperformance of the leading models almost always comes at a cost, namely an increase in turnover and consequently a more costly trading strategy due to their activity. It is therefore up to the investor to choose the strategy fitting his preferences best.

The goal of this paper was to introduce and compare various Black-Litterman models and to provide an adequate understanding of the operation of the models. Hence, a framework has been established within which to implement the (extended) Black-Litterman models which can be adapted to the preferences of the investor. By carefully composing the models and taking into account the uncertainties and available sources of information, superior performing models can result. As such, the Black-Litterman model is a proper shrinkage method to reduce the estimation error present in the input parameters for the Markowitz portfolio. The performance could, however, be improved by using more accurate private information and further substantiating the initialization of the model parameters. Applying this in an extensive Bayesian framework, the Black-Litterman model has the potential to increase its performance relative to alternative asset allocation models.

6 Discussion

The Black-Litterman models discussed in this paper are dependent on various sources of information, assumptions and model choices and are therefore submissive to bias. First of all, the data set at hand consists of roughly 450 assets and 12 years of monthly data. However, the research could benefit from a more extensive data set. A longer time period would be able to capture more fluctuations present in the economy, where the current data set only starts at the beginning of the financial crisis of 2007-2008 and experiences a period of economic expansion for the largest part of its remainder. This gives a limited view of the economic cycle. As shown in the sensitivity analysis, different models have varying performances during segments of this cycle. Also, the data sets suffers by construction from survivorship bias, which should be corrected for to obtain more accurate and realistic results. Furthermore, the inclusion of a larger or contrasting asset universe could yield a more robust analysis of the model, as it would assess the applicability of the model in alternative scenarios. An extended set of assets could for example have a better geographical spread or include additional asset classes such as fixed income or commodities. Moreover, when implemented, the set of restrictions should meet the mandate of the investor. In this paper, an absolute restriction is applied, but a relative restriction proportional to the benchmark portfolio could also be considered. This might consequently yield a lower turnover value and therefore reduce transaction costs.

Secondly, some of the assumptions made to derive the results might benefit from further research. For example, although the scaling parameter τ is used in a Bayesian setting in the hierarchical models, its value is assumed fixed at $\tau = 0.05$ for the remaining models. This value has been substantiated by conventions in the literature, but may not be suitable for all models. The same holds for the parameter of risk aversion, λ. A value of 2.5 has been used as is common in the literature, but this may not be appropriate for this specific data set. As the data set focuses on American companies, these parameters should reflect the level of confidence and risk aversion of American investors, which may differ from global levels. Furthermore, due to computational limitations, the number of iterations in the Hamiltonian Monte Carlo sampling process has been set to 100 per chain. While the CRM-τ model was able to yield sufficient convergence as measured by the convergence diagnostics, this can not be concluded for the CRM-t model. A higher number of iterations should, as discussed, increase the accuracy for this model, but during the research such a setting was not computationally feasible. For further confidence in its convergence, a smaller step size, higher proposal acceptance probability or more Markov chains should give more comfort regarding the results. This requires, however, more computing power.

Thirdly, there still are some directions for further research that have not been touched upon in the paper. My first suggestion is implementing an other base model to be used in the Black-Litterman model. I followed the original model in the sense that I used the CAPM as the foundation for the market equilibrium returns. This assumption could be substituted by another asset pricing model such as the Intertemporal CAPM (Merton, 1973), which is an extension of the original CAPM. This would also imply a different approach in the derivation of the equilibrium returns $\boldsymbol{\pi}$, as the reverse optimization technique may not be applicable. Secondly, when looking at the prior specifications of the model, only the normal inverse-Wishart distribution has been used as prior for the covariance matrix. An alternative to the model could be investigated with a different prior. As a different, non-conjugate prior would imply that the parameters should be obtained by sampling, I suggest using an LKJ prior (Lewandowski, Kurowicka, and Joe, 2009) for the correlation matrix as it is better suited for sampling. Using such a prior does, however, requires many computations and is therefore less convenient for models with many assets. When incorporating a limited number of assets, for example when only comparing different classes of assets, this prior specification could improve the robustness of the model. Lastly, the influence of the views on the performance of the models is an interesting additional research direction. The views used in this paper are derived from a factor model, but portfolio managers may be able to specify better founded views using their experience and their own current models. Investigating the effect of these adjusted views would improve the understanding of the model and enables to more accurately map the applicability of the framework.

7 References

Aguilar, O. & West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics 18*(3), 338–357.

Beach, S. L. & Orlov, A. G. (2007). An application of the Black–Litterman model with EGARCH-M-derived views for international portfolio management. *Financial Markets and Portfolio Management 21*(2), 147–166.

Best, M. J. & Grauer, R. R. (1991). Sensitivity analysis for mean-variance portfolio problems. *Management Science 37*(8), 980–989.

Black, F. & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal 48*(5), 28–43.

Brandt, M. W., Santa-Clara, P., & Valkanov, R. (2009). Parametric portfolio policies: exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies 22*(9), 3411–3447.

Brooks, S. P. & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics 7*(4), 434–455.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: a probabilistic programming language. *Journal of Statistical Software, Articles 76*(1), 1–32.

Cheung, W. (2013). The augmented Black–Litterman model: a ranking-free approach to factor-based portfolio construction and beyond. *Quantitative Finance 13*(2), 301–316.

DeMiguel, V., Garlappi, L., & Uppal, R. (2007). Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? *The Review of Financial Studies 22*(5), 1915–1953.

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid monte carlo. *Physics Letters B 195*(2), 216–222.

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

Fama, E. F. & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Frey, C. & Pohlmeier, W. (2016). Bayesian shrinkage of portfolio weights. *Available at SSRN 2730475*.

Geman, S. & Geman, D. (1987). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In *Readings in Computer Vision*, 564–584. Elsevier.

Golosnoy, V. & Okhrin, Y. (2007). Multivariate shrinkage for optimal portfolio weights. *The European Journal of Finance 13*(5), 441–458.

Hamilton, J. D. (1994). State-space models. *Handbook of Econometrics 4*, 3039–3080.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*(1), 97–109.

He, G. & Litterman, R. (1999). The intuition behind Black–Litterman model portfolios. *Available at SSRN 334304*.

Henderson, H. V. & Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM Review 23*(1), 53–60.

Hoffman, M. D. & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research 15*(1), 1593–1623.

Idzorek, T. (2007). A step-by-step guide to the Black–Litterman model: incorporating user-specified confidence levels. *Forecasting Expected Returns in the Financial Markets*, 17–38.

James, W. & Stein, C. (1992). Estimation with quadratic loss. *Breakthroughs in Statistics*, 443–460.

Jobson, J. D. & Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association 75*(371), 544–554.

Kourtis, A., Dotsis, G., & Markellos, R. N. (2012). Parameter uncertainty in portfolio selection: shrinking the inverse covariance matrix. *Journal of Banking & Finance 36*(9), 2522–2531.

Ledoit, O. & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance 10*(5), 603–621.

Lee, W. (2000). *Theory and methodology of tactical asset allocation*, Volume 65. John Wiley & Sons.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis 100*(9), 1989–2001.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The Journal of Finance 20*(4), 587–615.

Longin, F. (2005). The choice of the distribution of asset returns: how extreme value theory can help? *Journal of Banking & Finance 29*(4), 1017–1035.

Mankert, C. & Seiler, M. (2011). Mathematical derivations and practical implications for the use of the Black–Litterman model. *Journal of Real Estate Portfolio Management 17*(2), 139–159.

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

Merton, R.C. (1973). An intertemporal capital asset pricing model. *Econometrica 41*(5), 867–887.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics 21*(6), 1087–1092.

Meucci, A. (2006). Beyond Black–Litterman in practice: a five-step recipe to input views on non-normal markets. *Available at SSRN 872577*.

Meucci, A. (2009). Enhancing the Black–Litterman and related approaches: views and stress-test on risk factors. *Journal of Asset Management 10*(2), 89–96.

Michaud, R. O. (1989). The Markowitz optimization enigma: is 'optimized' optimal? *Financial Analysts Journal 45*(1), 31–42.

Michaud, R. O., Esch, D., & Michaud, R. (2012). Deconstructing Black–Litterman: how to get the portfolio you already knew you wanted. *Available at SSRN 2641893*.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, 768–783.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective.* MIT Press.

Palczewski, A. & Palczewski, J. (2019). Black–Litterman model for continuous distributions. *European Journal of Operational Research 273*(2), 708–720.

Peiró, A. (1994). The distribution of stock returns: international evidence. *Applied Financial Economics 4*(6), 431–439.

Roeder, D. E. (2015). *Dealing with data: an empirical analysis of Bayesian Black-Litterman model extensions*. Ph. D. thesis, Duke University Durham.

Rohleder, M., Scholz, H., & Wilkens, M. (2010, 09). Survivorship bias and mutual fund performance: relevance, significance, and methodical differences*. *Review of Finance 15*(2), 441–474.

Satchell, S. & Scowcroft, A. (2000). A demystification of the Black–Litterman model: managing quantitative and traditional portfolio construction. *Journal of Asset Management 1*(2), 138–150.

Sharpe, W. F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *The Journal of Finance 19*(3), 425–442.

Treynor, J. L. (1961). Market value, time, and risk. *Unpublished manuscript, dated 08/08/1961*.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.C. (2019). Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. *arXiv preprint arXiv:1903.08008*.

Walters, J. (2013). The factor tau in the Black–Litterman model. *Available at SSRN 1701467*.

Walters, J. (2014). The Black–Litterman model in detail. *Available at SSRN 1314585*.

Yen, Y. (2015). Sparse weighted-norm minimum variance portfolios. *Review of Finance 20*(3), 1259–1287.

Appendix A Additional Derivations

A.1 Derivation Equilibrium Market Returns

Assuming that the CAPM holds and following the approach by Satchell and Scowcroft (2000), the equilibrium market returns can be derived as follows,

$$
\begin{aligned}
\mathbb{E}(r_i) - R_f &= \left(\frac{\text{Cov}(r_i, r_m)}{\text{Var}(r_m)}\right)(\mathbb{E}(r_m) - R_f) \\
\boldsymbol{\pi} &= \left(\frac{\text{Cov}(\boldsymbol{r}, r_m)}{\sigma_m^2}\right)\mu_m \\
\boldsymbol{\pi} &= \left(\frac{\text{Cov}(\boldsymbol{r}, \boldsymbol{r}'\boldsymbol{w}_m)}{\sigma_m^2}\right)\mu_m \\
\boldsymbol{\pi} &= \left(\frac{\text{Cov}(\boldsymbol{r}, \boldsymbol{r})\boldsymbol{w}_m}{\sigma_m^2}\right)\mu_m \\
\boldsymbol{\pi} &= \frac{\mu_m}{\sigma_m^2}\boldsymbol{\Sigma}\boldsymbol{w}_m \\
\boldsymbol{\pi} &= \lambda\boldsymbol{\Sigma}\boldsymbol{w}_m \ . \qquad (32)
\end{aligned}
$$

A.2 Explanation Pick Matrix $\boldsymbol{P}$

The pick matrix $\boldsymbol{P}$ contains absolute and/or relative views. For the purpose of interpretability, I give an example of the construction of this matrix, where I use five assets and three views. This implies I have a 3×5 matrix $\boldsymbol{P}$ and a 3×1 matrix $\boldsymbol{q}$. The first example view is the absolute view that asset two have an expected return of 3%. The second and relative view is that assets three and four outperform asset one by 0.4%. The third view is the expectation that asset one and five have a return of 1.2%, and this view is weighted by the market capitalization of respectively €100 and €300 for illustrative purposes. Given these views, $\boldsymbol{P}$ and $\boldsymbol{q}$ are given by

$$
\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0.5 & 0.5 & 0 \\ 0.25 & 0 & 0 & 0 & 0.75 \end{pmatrix}, \ \boldsymbol{q} = \begin{pmatrix} 3\% \\ 0.4\% \\ 1.2\% \end{pmatrix}. \qquad (33)
$$

It can be seen that for absolute views the rows sum to one, but for relative views this summation equals zero. Bullish views furthermore consist of positive elements, whereas bearish views receive negative values.

A.3 Derivation Black-Litterman Master Formula, Conditional Approach

The derivation of the posterior distribution coined as the Black-Litterman formula is similar to the updating step of the Kalman filter (Hamilton, 1994). Both use conditional distributions to obtain their result. To derive the conditional distribution of the expected asset returns given the investor's views, I have the normal distribution of the mean of the asset returns, given by $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\pi}, \tau\dot{\boldsymbol{\Sigma}})$, and the normal distribution of the views of the investor at my disposal. By the rules of adding and multiplying normal distributions, the distribution of the views is as follows,

$$\begin{aligned}\boldsymbol{q} &\sim \boldsymbol{P}\mathcal{N}(\boldsymbol{\pi}\ ,\ \tau\dot{\boldsymbol{\Sigma}}) - \mathcal{N}(\boldsymbol{0}\ ,\ \boldsymbol{\Omega}) \\ &\sim \mathcal{N}(\boldsymbol{P\pi}\ ,\ \tau\boldsymbol{P}\dot{\boldsymbol{\Sigma}}\boldsymbol{P}' + \boldsymbol{\Omega})\ .\end{aligned} \tag{34}$$

Furthermore, in order to obtain the joint distribution, I need to derive the covariance between $\boldsymbol{q}$ and $\boldsymbol{\mu}$, which is given by

$$\mathrm{Cov}(\boldsymbol{q}, \boldsymbol{\mu}) = \mathrm{Cov}(\boldsymbol{P\mu} - \boldsymbol{\eta}, \boldsymbol{\mu}) = \mathrm{Cov}(\boldsymbol{P\mu}, \boldsymbol{\mu}) = \boldsymbol{P}\mathrm{Var}(\boldsymbol{\mu}) = \boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\ . \tag{35}$$

Because $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ are orthogonal by definition, the $\boldsymbol{\eta}$-term can be omitted in Equation 35. Having the proper definitions, $\boldsymbol{q}$ and $\boldsymbol{\mu}$ can be written as a joint normal distribution given by

$$\begin{pmatrix} \boldsymbol{q} \\ \boldsymbol{\mu} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{P\pi} \\ \boldsymbol{\pi} \end{pmatrix}, \begin{pmatrix} \tau\boldsymbol{P}\dot{\boldsymbol{\Sigma}}\boldsymbol{P}' + \boldsymbol{\Omega} & \boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}} \\ \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}' & \tau\dot{\boldsymbol{\Sigma}} \end{pmatrix}\right)\ . \tag{36}$$

Using this joint normal distribution, it states that $\boldsymbol{\mu}|\boldsymbol{q} \sim \mathcal{N}(\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\pi}})$ with

$$\widehat{\boldsymbol{\pi}} = \boldsymbol{\pi} + \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'(\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}' + \boldsymbol{\Omega})^{-1}(\boldsymbol{q} - \boldsymbol{P\pi}) \tag{37}$$

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\pi}} = \tau\dot{\boldsymbol{\Sigma}} - \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'(\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}' + \boldsymbol{\Omega})^{-1}\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\ . \tag{38}$$

This can be proven by firstly defining $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ for $\boldsymbol{\mu}$ and $\boldsymbol{q}$ and $\boldsymbol{M}$ as the covariance matrix of the joint distribution. Secondly, define a third variable, $\boldsymbol{z}_3 := \boldsymbol{z}_1 + \boldsymbol{A}\boldsymbol{z}_2$, with $\boldsymbol{A} = -\boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}$. It can be shown that $\boldsymbol{z}_3$ and $\boldsymbol{z}_2$ are uncorrelated since

$$\mathbb{E}(\boldsymbol{z}_3\boldsymbol{z}_2') = \mathbb{E}(\boldsymbol{z}_1\boldsymbol{z}_2 + A\boldsymbol{z}_2\boldsymbol{z}_2') = \boldsymbol{M}_{12} - \boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{22} = 0\ . \tag{39}$$

Hence, since $\boldsymbol{z}_3$ and $\boldsymbol{z}_2$ are uncorrelated, and using the fact that any two uncorrelated components of a multivariate normal distribution are independent as well, the independence of $\boldsymbol{z}_3$ and $\boldsymbol{z}_2$ is obtained. Furthermore, the expectation of $\boldsymbol{z}_3$ equals $\mathbb{E}(\boldsymbol{z}_3) = \boldsymbol{\mu}_1 + \boldsymbol{A}\boldsymbol{\mu}_2 = \boldsymbol{\pi} + \boldsymbol{A}\ \boldsymbol{P\pi}$. Therefore, the conditional expectation of $\boldsymbol{z}_1$ given $\boldsymbol{z}_2$ is given by

$$\begin{aligned}\mathbb{E}(\boldsymbol{z}_1|\boldsymbol{z}_2) &= \mathbb{E}(\boldsymbol{z}_3 - A\boldsymbol{z}_2|\boldsymbol{z}_2) \\ &= \mathbb{E}(\boldsymbol{z}_3|\boldsymbol{z}_2) - \mathbb{E}(\boldsymbol{A}\boldsymbol{z}_2|\boldsymbol{z}_2) \\ &= \mathbb{E}(\boldsymbol{z}_3) - \boldsymbol{A}\boldsymbol{z}_2 \\ &= \boldsymbol{\pi} + \boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}(\boldsymbol{z}_2 - \boldsymbol{P\pi}) \\ &= \boldsymbol{\pi} + \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'(\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}' + \boldsymbol{\Omega})^{-1}(\boldsymbol{q} - \boldsymbol{P\pi})\ .\end{aligned} \tag{40}$$

This proves the first moment, $\widehat{\boldsymbol{\pi}}$. Using the Woodbury Matrix Identity, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\pi}}$ can also be rewritten to match the form given in Equation 6. The Woodbury Matrix Identity is given by

$$(\boldsymbol{A}+\boldsymbol{UCV})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1}+\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\boldsymbol{VA}^{-1}\ , \tag{41}$$

where $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{U} \in \mathbb{R}^{n\times k}$, $\boldsymbol{C} \in \mathbb{R}^{k\times k}$ and $\boldsymbol{V} \in \mathbb{R}^{k\times n}$. Hence, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\pi}}$ can be written as $((\tau\dot{\boldsymbol{\Sigma}})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P})^{-1}$. Furthermore, multiplying Equation 41 with $\boldsymbol{U}$ and using the 'push-through' identity Henderson and Searle (1981) stated as $(\boldsymbol{A}+\boldsymbol{UCV})^{-1}\boldsymbol{UC} = \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}+\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}$, I obtain

$$(\boldsymbol{A}+\boldsymbol{UCV})^{-1}\boldsymbol{U} = [(\boldsymbol{A}+\boldsymbol{UCV})^{-1}\boldsymbol{UC}]\boldsymbol{C}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}+\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\boldsymbol{C}^{-1}\ . \tag{42}$$

Using these two identities, $\widehat{\boldsymbol{\pi}}$ can also be reformulated as to match its counterpart in Equation 6 by using the following steps,

$$\begin{aligned}
\widehat{\boldsymbol{\pi}} &= \boldsymbol{\pi} + \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'(\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'+\boldsymbol{\Omega})^{-1}(\boldsymbol{q}-\boldsymbol{P\pi}) \\
&= \boldsymbol{\pi} - \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'(\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'+\boldsymbol{\Omega})^{-1}\boldsymbol{P\pi} + \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'(\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'+\boldsymbol{\Omega})^{-1}\boldsymbol{q} \\
&= [\tau\dot{\boldsymbol{\Sigma}} - \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'(\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'+\boldsymbol{\Omega})^{-1}\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}](\tau\dot{\boldsymbol{\Sigma}})^{-1}\boldsymbol{\pi} + \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'(\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'+\boldsymbol{\Omega})^{-1}\boldsymbol{q} \\
&= [(\tau\dot{\boldsymbol{\Sigma}})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}]^{-1}(\tau\dot{\boldsymbol{\Sigma}})^{-1}\boldsymbol{\pi} + \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'(\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'+\boldsymbol{\Omega})^{-1}\boldsymbol{q} \qquad (43) \\
&= [(\tau\dot{\boldsymbol{\Sigma}})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}]^{-1}(\tau\dot{\boldsymbol{\Sigma}})^{-1}\boldsymbol{\pi} + \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'(\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'+\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}\boldsymbol{q} \\
&= [(\tau\dot{\boldsymbol{\Sigma}})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}]^{-1}(\tau\dot{\boldsymbol{\Sigma}})^{-1}\boldsymbol{\pi} + [(\tau\dot{\boldsymbol{\Sigma}})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}]^{-1}\boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{q} \qquad (44) \\
&= [(\tau\dot{\boldsymbol{\Sigma}})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}]^{-1}[(\tau\dot{\boldsymbol{\Sigma}})^{-1}\boldsymbol{\pi} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{q}]\ , \qquad (45)
\end{aligned}$$

where the Woodbury Matrix Identity is used in Equation 43 and the push-through identity in Equation 44. For the second moment, an equivalent approach is used as for the first moment and yields

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{z}_1|\boldsymbol{z}_2) &= \mathrm{Var}(\boldsymbol{z}_3 - \boldsymbol{A}\boldsymbol{z}_2|\boldsymbol{z}_2) \\
&= \mathrm{Var}(\boldsymbol{z}_3|\boldsymbol{z}_2) + \mathrm{Var}(\boldsymbol{A}\boldsymbol{z}_2|\boldsymbol{z}_2) - \boldsymbol{A}\mathrm{Cov}(\boldsymbol{z}_3, -\boldsymbol{z}_2) - \mathrm{Cov}(\boldsymbol{z}_3, -\boldsymbol{z}_2)\boldsymbol{A}' \\
&= \mathrm{Var}(\boldsymbol{z}_3|\boldsymbol{z}_2) \\
&= \mathrm{Var}(\boldsymbol{z}_3) \\
&= \mathrm{Var}(\boldsymbol{z}_1 + \boldsymbol{A}\boldsymbol{z}_2) \\
&= \mathrm{Var}(\boldsymbol{z}_1) + \boldsymbol{A}\mathrm{Var}(\boldsymbol{z}_2)\boldsymbol{A}' + \boldsymbol{A}\mathrm{Cov}(\boldsymbol{z}_1, \boldsymbol{z}_2) + \mathrm{Cov}(\boldsymbol{z}_1, \boldsymbol{z}_2)\boldsymbol{A}' \\
&= \boldsymbol{M}_{11} + \boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{22}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{21} - 2\boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{21} \\
&= \boldsymbol{M}_{11} + \boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{21} - 2\boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{21} \\
&= \boldsymbol{M}_{11} - \boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{21} \\
&= \tau\dot{\boldsymbol{\Sigma}} - \tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'(\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\boldsymbol{P}'+\boldsymbol{\Omega})^{-1}\boldsymbol{P}\tau\dot{\boldsymbol{\Sigma}}\ , \qquad (46)
\end{aligned}$$

which proves the variance part, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\pi}}$. I have now fully derived the Black-Litterman formula given in Equation 6, as well as the intermediate result shown in Equation 9.

A.4 Derivation Black-Litterman Master Formula, Bayesian Approach

In Appendix A.3, I derived the master Black-Litterman formula using conditional distributions. It is also possible to derive this formula using a Bayesian framework. For this, the following information is used:

- Prior: $P(\boldsymbol{\mu}|\boldsymbol{\Sigma}) = \mathcal{N}(\boldsymbol{\pi}, \tau\boldsymbol{\Sigma})$;
- Likelihood: $P(\boldsymbol{q}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}(\boldsymbol{P}\boldsymbol{\mu}, \boldsymbol{\Omega})$;
- Bayes' rule as given in Equation 5;
- Assumption that $\boldsymbol{\Sigma}$ is known.

Using the corresponding probability density functions and writing them out fully yields

$$\begin{aligned}
P(\boldsymbol{\mu}|\boldsymbol{q}, \boldsymbol{\Sigma}) &\propto P(\boldsymbol{\mu}|\boldsymbol{\Sigma})P(\boldsymbol{q}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \frac{1}{\sqrt{(2\pi)^n}}|\tau\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\pi})'\tau\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\pi})\Big\} \\
&\qquad \cdot \frac{1}{\sqrt{(2\pi)^n}}|\boldsymbol{\Omega}|^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}(\boldsymbol{q}-\boldsymbol{P}\boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\boldsymbol{q}-\boldsymbol{P}\boldsymbol{\mu})\Big\} \\
&\propto \exp\Big\{ -\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\pi})'\tau\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\pi}) - \frac{1}{2}(\boldsymbol{q}-\boldsymbol{P}\boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\boldsymbol{q}-\boldsymbol{P}\boldsymbol{\mu})\Big\} \\
&\propto \exp\Big\{ -\frac{1}{2}[\boldsymbol{\mu}'\tau\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}] + \boldsymbol{\mu}'\tau\boldsymbol{\Sigma}^{-1}\boldsymbol{\pi} - \frac{1}{2}[\boldsymbol{\mu}'(\boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P})\boldsymbol{\mu}] + \boldsymbol{\mu}'(\boldsymbol{P}'\boldsymbol{\Omega}^{-1})\boldsymbol{q}\Big\} \\
&= \exp\Big\{ -\frac{1}{2}\boldsymbol{\mu}'[\tau\boldsymbol{\Sigma}^{-1} + (\boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P})]\boldsymbol{\mu} + \boldsymbol{\mu}'[\tau\boldsymbol{\Sigma}^{-1}\boldsymbol{\pi} + (\boldsymbol{P}'\boldsymbol{\Omega}^{-1})\boldsymbol{q}]\Big\}\ ,
\end{aligned} \tag{47}$$

which can be recognized as the kernel of a normal distribution with the same characteristics as in Equation 6. Continuing the derivation to find the posterior predictive distribution of the asset returns yields

$$\begin{aligned}
P(\boldsymbol{r}_{T+1}|\boldsymbol{q}, \boldsymbol{\Sigma}) &= \int P(\boldsymbol{r}_{T+1}, \boldsymbol{\mu}|\boldsymbol{q}, \boldsymbol{\Sigma})d\boldsymbol{\mu} \\
&= \int P(\boldsymbol{r}_{T+1}|\boldsymbol{\mu}, \boldsymbol{q}, \boldsymbol{\Sigma})P(\boldsymbol{\mu}|\boldsymbol{q}, \boldsymbol{\Sigma})d\boldsymbol{\mu} \\
&= \int \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})\, \mathcal{N}(\widehat{\boldsymbol{\mu}}_{BL}, \widehat{\boldsymbol{\Sigma}}_{BL})d\boldsymbol{\mu} \\
&= \mathcal{N}(\widehat{\boldsymbol{\mu}}_{BL}, \widehat{\boldsymbol{\Sigma}}_{\text{post.}})\ .
\end{aligned} \tag{48}$$

A.5 Derivation Normal-Inverse-Wishart Posterior Distribution

The posterior distribution of the canonical reference model with the unknown covariance matrix $\boldsymbol{\Sigma}$ from Section 3.3.3 is derived as follows,

$$
\begin{aligned}
P(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \boldsymbol{r}) &\propto P(\boldsymbol{r} | \boldsymbol{\mu}, \boldsymbol{\Sigma})\ P(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&\propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}n} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{r}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{r}_i - \boldsymbol{\mu}) \right\} \\
&\quad \cdot |\boldsymbol{\Sigma}|^{-\frac{v_0+p+1}{2}} \exp\left\{ -\frac{1}{2} \mathrm{tr}[\boldsymbol{\Psi} \boldsymbol{\Sigma}^{-1}] - \frac{l}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) \right\} \\
&= |\boldsymbol{\Sigma}|^{-\frac{v_0+p+n+1}{2}} \exp\Bigg\{ -\frac{1}{2} \mathrm{tr}\left[\left(\boldsymbol{\Psi} + \sum_{i=1}^{n} (\boldsymbol{r}_i - \bar{\boldsymbol{r}})(\boldsymbol{r}_i - \bar{\boldsymbol{r}})' \right) \boldsymbol{\Sigma}^{-1} \right] \\
&\quad - \frac{n}{2} (\bar{\boldsymbol{r}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{r}} - \boldsymbol{\mu}) - \frac{l}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) \Bigg\} \\
&= |\boldsymbol{\Sigma}|^{-\frac{v_0+p+n+1}{2}} \exp\Bigg\{ -\frac{1}{2} \mathrm{tr}\left[\left(\boldsymbol{\Psi} + \sum_{i=1}^{n} (\boldsymbol{r}_i - \bar{\boldsymbol{r}})(\boldsymbol{r}_i - \bar{\boldsymbol{r}})' \right) \boldsymbol{\Sigma}^{-1} \right] \qquad (49) \\
&\quad - \frac{\widetilde{l}}{2} (\boldsymbol{\mu} - \widetilde{\boldsymbol{\mu}}_0)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \widetilde{\boldsymbol{\mu}}_0) - \frac{ln}{\widetilde{l}} (\bar{\boldsymbol{r}} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{r}} - \boldsymbol{\mu}_0) \Bigg\} \\
&= |\boldsymbol{\Sigma}|^{-\frac{v_0+p+n+1}{2}} \exp\Bigg\{ -\frac{\widetilde{l}}{2} (\boldsymbol{\mu} - \widetilde{\boldsymbol{\mu}}_0)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \widetilde{\boldsymbol{\mu}}_0) \\
&\quad - \frac{1}{2} \mathrm{tr}\left[\left(\boldsymbol{\Psi} + \sum_{i=1}^{n} (\boldsymbol{r}_i - \bar{\boldsymbol{r}})(\boldsymbol{r}_i - \bar{\boldsymbol{r}})' + \frac{ln}{\widetilde{l}} (\bar{\boldsymbol{r}} - \boldsymbol{\mu}_0)(\bar{\boldsymbol{r}} - \boldsymbol{\mu}_0)' \right) \boldsymbol{\Sigma}^{-1} \right] \Bigg\} \\
&= |\boldsymbol{\Sigma}|^{-\frac{\widetilde{v}+p+1}{2}} \exp\left\{ -\frac{\widetilde{l}}{2} (\boldsymbol{\mu} - \widetilde{\boldsymbol{\mu}}_0)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \widetilde{\boldsymbol{\mu}}_0) - \frac{1}{2} \mathrm{tr}\left[\widetilde{\boldsymbol{\Psi}} \boldsymbol{\Sigma}^{-1} \right] \right\}, \qquad (50)
\end{aligned}
$$

where

$$\widetilde{v} = v + n \ , \qquad (51)$$

$$\widetilde{l} = l + n \ , \qquad (52)$$

$$\widetilde{\boldsymbol{\mu}} = (l \boldsymbol{\mu}_0 + n \bar{\boldsymbol{r}}) \ / \ \widetilde{l} \ , \qquad (53)$$

$$\widetilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + \sum_{i=1}^{n} (\boldsymbol{r}_i - \bar{\boldsymbol{r}})(\boldsymbol{r}_i - \bar{\boldsymbol{r}})' + [(ln)(\bar{\boldsymbol{r}} - \boldsymbol{\mu}_0)(\bar{\boldsymbol{r}} - \boldsymbol{\mu}_0)'] \ / \ \widetilde{l} \ . \qquad (54)$$

As discussed in the regarding section, $\boldsymbol{\mu}_0$ is defined as $\widehat{\boldsymbol{\mu}}_{BL}$ and $\boldsymbol{\Psi}$ is defined as $\widehat{\boldsymbol{\Sigma}}_{post.}$, both resulting from the canonical reference model. Furthermore, I used the complete the squares operation in step 49 as in Murphy (2012), page 143. Looking at the result, a normal-inverse-Wishart distribution can be described by the parameters $\widetilde{\boldsymbol{\mu}}$, $\widetilde{l}$, $\widetilde{\boldsymbol{\Psi}}$ and $\widetilde{v}$.

A.6 Derivation Posterior Predictive Distribution With Unknown $\boldsymbol{\Sigma}$

To arrive at the posterior predictive distribution as given in Section 3.4.1, I follow the steps described below,

$$
\begin{aligned}
P(\boldsymbol{r}_{T+1}|\boldsymbol{q}_1,\dots,\boldsymbol{q}_T) &= \int_{\boldsymbol{\mu}}\int_{\boldsymbol{\Sigma}} P(\boldsymbol{r}_{T+1},\boldsymbol{\mu},\boldsymbol{\Sigma}|\boldsymbol{q})\; d\boldsymbol{\Sigma}\; d\boldsymbol{\mu} \\
&= \int_{\boldsymbol{\mu}}\int_{\boldsymbol{\Sigma}} P(\boldsymbol{r}_{T+1}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{q})\; P(\boldsymbol{\mu},\boldsymbol{\Sigma}|\boldsymbol{q})\; d\boldsymbol{\Sigma}\; d\boldsymbol{\mu} \\
&\propto \int_{\boldsymbol{\mu}}\int_{\boldsymbol{\Sigma}} P(\boldsymbol{r}_{T+1}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{q})\; P(\boldsymbol{\mu},\boldsymbol{\Sigma})\; P(\boldsymbol{q}|\boldsymbol{\mu},\boldsymbol{\Sigma})\; d\boldsymbol{\Sigma}\; d\boldsymbol{\mu} \\
&= \int_{\boldsymbol{\mu}}\int_{\boldsymbol{\Sigma}} P(\boldsymbol{r}_{T+1}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{q})\; P(\boldsymbol{\mu}|\boldsymbol{\Sigma})\; P(\boldsymbol{\Sigma})\; P(\boldsymbol{q}|\boldsymbol{\mu},\boldsymbol{\Sigma})\; d\boldsymbol{\Sigma}\; d\boldsymbol{\mu} \\
&= \int_{\boldsymbol{\mu}}\int_{\boldsymbol{\Sigma}} \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})\; \mathcal{N}(\boldsymbol{\pi},\tau\dot{\boldsymbol{\Sigma}})\; \mathcal{W}^{-1}(\dot{\boldsymbol{\Sigma}},v)\; \mathcal{N}(\boldsymbol{P}\boldsymbol{\mu},\boldsymbol{\Omega})\; d\boldsymbol{\Sigma}\; d\boldsymbol{\mu} \\
&= \int_{\boldsymbol{\mu}}\int_{\boldsymbol{\Sigma}} \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})\; \mathcal{NIW}(\widetilde{\boldsymbol{\mu}},\widetilde{l},\widetilde{\boldsymbol{\Sigma}},\widetilde{v})\; d\boldsymbol{\Sigma}\; d\boldsymbol{\mu} \\
&= \int_{\boldsymbol{\mu}}\int_{\boldsymbol{\Sigma}} \mathcal{NIW}(\widetilde{\boldsymbol{\mu}}_{BLE},\widetilde{l}_{BLE},\widetilde{\boldsymbol{\Sigma}}_{BLE},\widetilde{v}_{BLE})\; d\boldsymbol{\Sigma}\; d\boldsymbol{\mu} \\
&= t_{\widetilde{v}_{BLE}-n+1}\left(\widetilde{\boldsymbol{\mu}}_{BLE},\widetilde{\boldsymbol{\Sigma}}_{BLE}\,\frac{\widetilde{l}_{BLE}+1}{\widetilde{l}_{BLE}(\widetilde{v}_{BLE}-n+1)}\right)\;.
\end{aligned} \tag{55}
$$

The formulations of the resulting parameters are given in the respective section and equal

$$\widetilde{v}_{BLE} = v + n\;, \tag{56}$$

$$\widetilde{l}_{BLE} = \left(\frac{1}{\tau}\right) + n\;, \tag{57}$$

$$\widetilde{\boldsymbol{\mu}}_{BLE} = \left(\frac{1}{\tau}\boldsymbol{\pi} + n\bar{\boldsymbol{q}}\right)\;/\;\widetilde{l}\;, \tag{58}$$

$$\widetilde{\boldsymbol{\Sigma}}_{BLE} = \boldsymbol{\Sigma} + \sum_{i=1}^{n}(\boldsymbol{q}_i-\bar{\boldsymbol{q}})(\boldsymbol{q}_i-\bar{\boldsymbol{q}})' + [(n/\tau)(\bar{\boldsymbol{q}}-\boldsymbol{\pi})(\bar{\boldsymbol{q}}-\boldsymbol{\pi})']\;/\;\widetilde{l}\;. \tag{59}$$

A.7 Hamiltonian Monte Carlo Algorithm

To sample from the models for which there is no analytical distribution readily available, I use the Hamiltonian Monte Carlo sampler as discussed in Section 3.8. To initialize the sampler, some specification need to be defined. For my sampling processes, I use the following settings:

- *Number of Markov chains:* When using more Markov chains, the convergence of the sampler can be determined more accurately. It is therefore discouraged to use only one Markov chain in the sampling process. However, more chains also increase the computational load. I follow the recommended setting given by the Stan development team to use four Markov chains.

- *Iterations:* As my computational resources are limited, I am only able to efficiently run roughly 100 iterations per chain when executing calculations using the whole data set. For smaller data sets, more iterations can be applied.

- *Warm-up iterations:* Of the sets of iterations, half are intended as warm-up iterations, also called the burn-in period, which are not used to form the Markov chain from which the samples are drawn. During this warm-up period, the program calibrates the starting specifications as to improve the efficiency and accuracy of the algorithm.

- *Starting values:* The starting values are determined randomly for each Markov chain seperately.

- *Step size:* At each step, the negative log probability function and the gradient of the target distribution are computed. The step size, or discretization time interval, defines how meticulous the algorithm is in its exploration, as a smaller step size ensures more frequent computations. Therefore, by taking smaller steps, target distributions with a complex geometry can be modeled more accurately. The Stan program is able to determine the optimal step size to match the acceptance probability of the algorithm and thus adaptively set its value.

- *Acceptance probability:* The standard Metropolis-Hastings target average proposal acceptance probability for adaptation in the Stan program is set at 0.8. I use this value as well, as a larger value increases the required computing power. Increasing the acceptance probability is, however, an effective method to reduce the number of divergent transitions, as it implicitly decreases the step size and therefore increases the robustness of the inference algorithm.

A.8 Convergence Diagnostic $\widehat{R}$

The convergence diagnostic used for the Hamiltonian Monte Carlo sampler, defined as $\widehat{R}$, consists of the within-chain variance and the between-chain variance. Assume M chains, i iterations per chain and *df* degrees of freedom. Furthermore, define the mean of the variance within each chain to be equal to $W = (1/M)\sum_{m=1}^{M} \widehat{\sigma}_m^2$, where $\widehat{\sigma}_m^2$ denotes the variance of chain m, and the between chain variance to equal B/n, where $B = (i/(M-1))\sum_{m=1}^{M}\left(\widehat{\theta}_m - \widehat{\theta}\right)^2$ with $\widehat{\theta}_m$ the posterior sample mean of parameter θ for chain m and $\widehat{\theta}$ the overall sample posterior mean. Following the steps outlined by Brooks and Gelman (1998) yields for the pooled variance

$$\widehat{V} = \frac{i-1}{i}W + \frac{M+1}{Mi}B\,. \tag{60}$$

Estimating the degrees of freedom *df* by the method of moments yields $df = \left(2\widehat{V}^2\right) / \operatorname{Var}(\widehat{V})$. The convergence diagnostic is now given by

$$\widehat{R} = \sqrt{\frac{(df+3)\widehat{V}}{(df+1)W}}\,. \tag{61}$$

Appendix B Additional Figures

B.1 Data Analysis

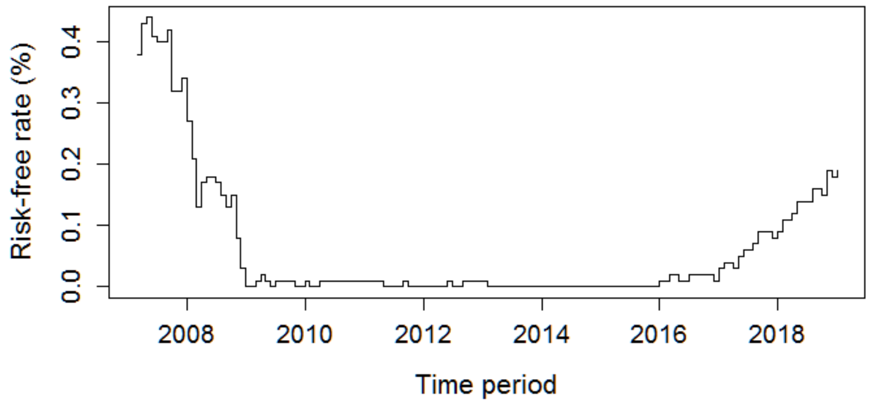

Figure 8: *Stepwise plot of the risk-free rate, defined as the one-month T-bill yield. The time period is 2007-01 to 2018-12.*

B.2 Convergence Diagnosis

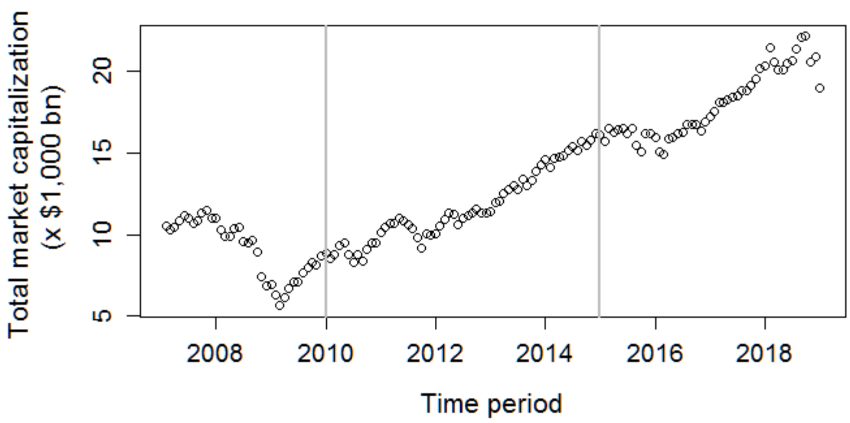

Figure 9: *Plot of the total market capitalizations (in USD) for the time period 2007-01 to 2018-12. The grey lines indicate the respective beginning or end of the smaller time periods 2010-01 to 2018-01 and 2007-01 to 2015-01.*

Appendix C Additional Tables

C.1 Regression Portfolio Results on Fama-French Factors

Table 11: Regression of portfolio returns of the five Fama-French factors, first scenario

Model	Intercept	Market	Size	Value	Profitability	Investment	R^2
Average factor return		0.0073	-0.0002	-0.0013	0.0027	0.0003	
MEM	0.00	0.89***	-0.09	-0.15	0.15	0.35**	91.1%
ARM	0.03*	1.15**	0.69	-2.32*	-3.48	2.32	23.5%
CRM	0.02*	1.03**	1.02	-2.17*	-3.04	2.23	21.6%
ECRM	0.02	1.51***	0.97	-0.92	-0.69	-2.72*	39.5%
BLE	0.01	1.45***	1.42*	0.60	-0.48	0.26	43.2%
CRM-t	0.01*	0.91***	0.21	-0.13	-0.51	0.62	31.8%
CRM-τ	0.03*	1.15**	0.69*	-2.31*	-3.49	2.31	23.6%

*This table shows the regression results of the portfolio returns regressed on the five Fama-French factors accounting for the market and the size, value, profitability and investment rate of the assets. The confidence levels are indicated by superscripts: ***: 0.1%, **: 1%, *: 5%. The applied restrictions for the formation of the portfolios are a full investment and short selling restriction.*

Table 12: Regression of portfolio returns of the five Fama-French factors, second scenario

Model	Intercept	Market	Size	Value	Profitability	Investment	R^2
Average factor return		0.0073	-0.0002	-0.0013	0.0027	0.0003	
MEM	0.00	0.89***	-0.09	-0.15	0.15	0.35**	91.1%
ARM	0.00	0.85***	0.08	-0.41*	0.12	0.82***	75.2%
CRM	0.00	0.85***	0.07	-0.41*	0.14	0.81***	75.1%
ECRM	0.01***	0.97***	0.20	-0.65***	-0.18	-0.02	86.2%
BLE	0.00	1.18***	0.61***	0.06	-0.23	0.12	86.0%
CRM-t	0.01*	0.89***	0.13	-0.08	0.13	0.28	77.4%
CRM-τ	0.00	0.85***	0.08	-0.41*	0.12	0.82***	75.2%

*This table shows the regression results of the portfolio returns regressed on the five Fama-French factors accounting for the market and the size, value, profitability and investment rate of the assets. The confidence levels are indicated by superscripts: ***: 0.1%, **: 1%, *: 5%. The applied restrictions for the formation of the portfolios are a full investment, short selling and weight (at most 1%) restriction.*

C.2 Ranking Results Complete Data Set When Using $\tau = 0.1$

Table 13: Ranking of the (benchmark) models with budget and short selling restrictions for $\tau = 0.1$

Model	μ	SR	CAGR	$ES_{0.9}$	TO	Total score	Overall rank
MEM	9	3	9	1	2	24	**3**
CRM	4	10	4	6	9	33	**9**
ARM	3	8	3	6	8	28	**6**
ECRM	5	9	5	10	4	33	**9**
BLE	6	4	6	5	5	26	**5**
CRM-τ	2	7	2	6	7	24	**3**
CRM-t	7	2	7	4	10	30	**8**
Market capitalization	10	5	10	2	1	28	**6**
Equal weight portfolio	8	1	8	3	3	23	**2**
Restricted factor model	1	6	1	6	6	20	**1**

This table indicates the ranking of the portfolios measured on the performance evaluation criteria. A lower rank indicates a relatively better performing model. When the total score of multiple models is equal, both obtain an equal rank. For these results, $\tau = 0.1$ is used and the applied restrictions are a full investment and short selling restriction.

Table 14: Ranking of the (benchmark) models with budget, short selling and weight restrictions for $\tau = 0.1$

Model	μ	SR	CAGR	$ES_{0.9}$	TO	Total score	Overall rank
MEM	9	8	9	1	3	30	**7**
CRM	8	5	8	2	4	27	**5**
ARM	7	4	7	4	5	27	**5**
ECRM	1	1	1	9	9	21	**1**
BLE	2	9	2	10	8	31	**8**
CRM-τ	6	3	6	5	6	26	**4**
CRM-t	4	6	4	7	10	31	**8**
Market capitalization	10	10	10	6	1	37	**10**
Equal weight portfolio	3	7	3	8	2	23	**3**
Restricted factor model	5	2	5	3	7	22	**2**

This table indicates the ranking of the portfolios measured on the performance evaluation criteria. A lower rank indicates a relatively better performing model. When the total score of multiple models is equal, both obtain an equal rank. For these results, $\tau = 0.1$ is used and the applied restrictions are a full investment, short selling and weight (at most 1%) restriction.

C.3 Additional Robustness Results With Weight Restrictions

Table 15: Model results for the CRM-τ and CRM-t model with a varying number of iterations

Model	Iterations	μ (%)	σ (%)	SR (%)	CAGR (%)	$ES_{0.9}$ (%)	TO (%)	IR (%)
	100	0.3	5.2	5.7	2.0	-9.2	33.4	-14.7
CRM-τ	1,000	0.3	5.2	5.7	2.0	-9.2	33.3	-14.7
	10,000	0.3	5.2	5.8	2.0	-9.2	33.3	-14.6
	100	0.6	5.7	10.6	5.5	-10.6	117.3	-6.6
CRM-t	1,000	0.5	5.6	9.8	4.8	-10.4	56.1	-7.9
	10,000	0.6	5.7	11.0	5.7	-10.9	32.1	-5.5

Performance evaluation measure values for the CRM-τ and CRM-t models when using different numbers of iterations. For τ, $\mathcal{N}(0.1, 0.02)$ is used. Moreover, a rolling window size of $h = 36$ is used in the computations and for the expected shortfall (ES), α is set at 0.9. The benchmark for the information ratio (IR) is the S&P 500 index. Furthermore, the portfolio weights are restricted by a budget and short selling constraint and the maximum asset weight in the portfolio is set at 18% of the total portfolio. The Sharpe ratio, compounded annual growth rate and turnover are abbreviated by respectively SR, CAGR and TO.

Table 16: Ranking of the (benchmark) models in various time periods

Model	Ranking period			Sum of ranks	Overall rank
	2007-2015	2007-2018	2010-2018		
MEM	8	2	7	17	**5**
CRM	8	9	2	19	**9**
ARM	6	7	3	16	**4**
ECRM	4	3	10	17	**5**
BLE	4	5	9	18	**7**
CRM-τ	2	6	4	12	**1**
CRM-t	2	8	8	18	**7**
Market capitalization	10	4	6	20	**10**
Equal weight portfolio	7	1	5	13	**3**
Restricted factor model	1	10	1	12	**1**

Ranking of the different models when using data from a specific time period. The periods span respectively 2007-01 to 2015-12, 2007-01 to 2018-12 and 2010-01 to 2018-12. A lower ranking indicates a relatively better performing model. The models are restricted by a budget, short selling and weight constraint where the weight constraint limits the asset weights to 18% of the total portfolio. The ranking criteria are the same as before. For the computation of the models, $\tau = 0.1$ and 1,000 iterations are used.

Table 17: Ranking of the (benchmark) models using various rolling window sizes

Model	Ranking per rolling window size			Sum of ranks	Overall rank
	$h = 12$	$h = 36$	$h = 60$		
MEM	5	2	2	9	**2**
CRM	10	9	8	27	**9**
ARM	2	7	5	14	**5**
ECRM	7	3	3	13	**3**
BLE	3	5	10	18	**7**
CRM-τ	4	6	7	17	**6**
CRM-t	8	8	6	22	**8**
Market capitalization	6	4	3	13	**3**
Equal weight portfolio	1	1	1	3	**1**
Restricted factor model	8	10	9	27	**9**

Ranking of the different models when using different rolling window sizes (h). A lower ranking indicates a relatively better performing model. The models are restricted by a budget and short selling constraint and the maximum asset weight in the portfolio is set at 18% of the total portfolio. For the computation of the models, $\tau = 0.1$ *and 1,000 iterations are used. The ranking criteria are the same as before.*