

Constructing Robust Portfolios, the Role of Parameter Uncertainty in Dynamic Optimal Portfolio Allocations

Erasmus University Rotterdam - Erasmus School of Economics

Master Thesis Quantitative Finance

Supervisor: A. Naghi, Second Assessor: X. Xiao

Van Pelt, T.G.

409137

June 30, 2020

Abstract

This paper investigates the ability to improve traditional portfolio optimisation rules in practice. Specifically, I examine the effect of parameter uncertainty on Markowitz portfolio performance and quantify the corresponding losses. Frequentist methods, in the form of direct parameter shrinkage and assigned portfolio weight shrinkage are employed to suppress effects of estimation errors. A Bayesian approach of the traditional Markowitz portfolio is used to account for estimation risks implicitly and novel Bayesian portfolio combinations are defined in search of an optimal investment rule. Moreover, the assumption of normal returns is relaxed by considering Markov Switching Gaussian Mixture models. I demonstrate the mean-variance modifications to be effective in improving out-of-sample performance and show ability to beat the equally weighted benchmark for empirical samples of the 48 Industry portfolios and Fama-French 100 portfolios. Lastly, I validate the competence of producing robust portfolios over periods of changing economic times when parameter uncertainty is considered in the Markowitz model.

Key words: Portfolio Optimisation, Parameter Uncertainty, Shrinkage, Bayesian Statistics

JEL Classifications: G11, C38, C11, C53

Contents

1	Introduction	4
1.1	Problem Definition	4
1.2	Literature Review	5
1.3	Contributions and Findings	7
2	Methodology	8
2.1	Characterizing Estimation Error	10
2.1.1	Traditional Portfolio Allocation Methods	10
2.1.2	Quantifying Parameter Uncertainty	12
2.2	Parameter Shrinkage Techniques	15
2.2.1	Direct Shrinkage of Parameters	15
2.2.2	The Black-Litterman Model with Investor Views	17
2.2.3	Factor Models	19
2.3	Portfolio Weight Shrinkage	19
2.3.1	Target Portfolio Shrinkage	20
2.4	Accounting for Parameter Uncertainty	23
2.4.1	Diffuse Prior	24
2.4.2	Informative Priors	25
2.4.3	Hierarchical Bayes Models	27
2.5	Robustification of Models	28
2.5.1	Accounting for Non-Normal Returns	28
2.5.2	Robustification via Bayesian Portfolio Combinations	30
3	Out-of-Sample Performance Evaluation	31
3.1	Initialisation and Base-Case Investor	31
3.2	Evaluation Criteria and Performance Measures	32
4	Data	33
4.1	Descriptive Statistics	33
4.2	In-Sample Results for Factor Models	35
5	Results	36
5.1	Direct Shrinkage of Parameters	36
5.2	Portfolio Weight Shrinkage	38
5.3	Accounting for Parameter Uncertainty	39
5.4	Combination Strategies	41
5.5	Measuring Robustness	45
6	Conclusion	46
7	Discussion and Further Research	48

Appendices	54
A Expected Utility Loss for Varying Risk Aversion	54
B Derivation Alternative Three-Fund Rule, Markowitz and Equally Weighted Portfolio	55
C Derivation Alternative Three-Fund Rule, Equally Weighted and Global Minimum Variance Portfolio	57
D Bayesian Inference for the Diffuse Prior	58
E Bayesian Inference for the Conjugate Prior	60
F Bayesian Inference for the Hierarchical Bayes Gibbs Sampler	63
G Specifications of the Markov Switching Gaussian Mixture Model	65
H Models Out-of-Sample Performance for Fama-French 100 Portfolios	69
I Robustness Analysis of the Winner Markowitz Modifications for Fama-French 100 Portfolios	71

1 Introduction

1.1 Problem Definition

One of the great hurdles in portfolio management regards the performance of theoretically optimal portfolios. Since the introduction of the mean-variance portfolio in Markowitz' seminal paper (Markowitz, 1952), countless sophisticated techniques for portfolio formation have been derived, yet no method seems to consistently be able to produce positive returns. In particular it turns out to be especially difficult to significantly outperform the equally weighted portfolio in terms of out-of-sample Sharpe ratios (DeMiguel et al., 2009). Experienced investors have derived numerous theoretically grounded portfolio constructing methods but they all seem to fall apart in practice. Throughout this investigation, I will try to bring textbook and real-life economics closer to one another in attempt to improve the empirical applicability of theoretical asset management techniques.

Traditional portfolio optimisation (as derived in Markowitz (1952)) that targets risk and return levels proved a major advancement in asset allocation. Its intuitive and widely applicable approach of maximising returns given a level of risk made the model a popular choice for investors and explains why it is considered one of the building blocks of modern portfolio theory. The models' dynamic applicability, and ability to integrate investor (mean-variance) constraints are reasons for its common use in practice today, and the reason for which investigating improvements of such a model remain of relevance.

Though useful, the Markowitz portfolio has been shown to have several drawbacks that have proved it inadequate for practical implementations. Simple suboptimal techniques such as the equally weighted and value weighted portfolios tend to perform better out-of-sample (Jobson and Korkie, 1989) and positive risk-adjusted returns are not made consistently (DeMiguel et al., 2009). The reason for this is that mean-variance portfolios have the tendency to assign extreme weights to certain assets. Moreover, these portfolio allocations fluctuate significantly over time, making the method increasingly unreliable for asset management firms to use in practice. The cost of financing these irregularities is simply too large. It is therefore desirable to develop robust portfolios that can stabilise these weights over time.

The main reason for the large fluctuations of Markowitz weights, is the model sensitivity to its input parameters, the expected returns and the covariance of returns. It has been shown that small changes of the input parameters lead to significant changes in asset weights, making the portfolio extremely volatile (Hurley and Brimberg, 2015). M. Best and R.Grauer (1991) demonstrate this volatility by reporting differences when subtly increasing the returns of an arbitrary asset in their investment space. An increase of 2% in returns of an arbitrary asset can lead to reallocation of more than half the securities from the starting equally weighted strategy. However, the reallocation has negligible effects on the risk and return of the resulting portfolio. The traditional optimisation technique takes *exact* (sample) inputs for the expected returns and covariance, and as these tend to deviate from the (unknown) true parameters, the assigned weights can be very different from the true underlying optimal allocations. Producing robust portfolios in practice therefore turns out problematic, especially considering the fact that moment sample estimates have the tendency to be rather noisy (Chopra and Ziemba, 1993). Moreover, Michaud (1989) denotes the mean-variance strategy as error maximising. The intuition stems from the fact that the model tends to assign larger weights to the assets with higher returns, that are most negatively correlated and have the smallest variances. These correspond to cases of greatest estimation error, causing the optimal solution to simultaneously be maximised in mean-variance as well as in errors. The fundamental problem of the Markowitz portfolio hence lies in the parameter uncertainty involved in parameter estimation. Throughout I will investigate the possibilities to suppress this effect.

1.2 Literature Review

Since Markowitz (1952), mean-variance portfolios have always been at the centre of portfolio management research. A vast amount of literature has been devoted to testing its applicability and uses, of which many, such as Frankfurter et al. (1971), Jobson and Korkie (1989) and Putnam (1991), identify the significance of estimation risk. Empirical demonstrations of this problem are portrayed in Levy and Sarnat (1970), Chopra and Ziemba (1993) and Chopra et al. (1993). These conclude that parameter estimations errors in combination with model sensitivity to input parameters makes the Markowitz allocation incompetent for practical uses. Previous works have posed a number of solutions for these problems. I provide an overview of the literature, along with the shortcomings that ought to be accounted for and further researched.

Shrinkage techniques have been popular in literature to try and mitigate the impact of estimation error. This is particularly due to its intuitive approach that proves easily applicable for large dimensions. I distinguish two main frequentist methods that have been used to reduce the effect of parameter uncertainty. The first approach targets the input parameters of the Markowitz portfolio. Jorion (1986) and Ledoit and Wolf (2004), shrink the sample estimates by combining them with more stable targets. Adding some bias to the sample estimates could result beneficial as long as the reduction in estimation error outweighs the addition of bias. Jorion (1986) shrinks the vector of means to a more certain shrinkage target, a vector of ones. Ledoit and Wolf (2004), apply a similar trick to the covariance parameter. DeMiguel et al. (2013) show that these new estimators can replace the sample moments for improved performance. Though this approach results in more ‘certain’ estimates of moments, it does not solve the issue regarding the sensitivity of the Markowitz portfolio to its input parameters. In an attempt to reduce this effect, the second shrinkage approach is applied on portfolio weights. This way extreme weight allocations on certain assets are reduced. Noteworthy research include Kan and Zhou (2007) and DeMiguel et al. (2015), who shrink mean-variance weights of each asset towards target portfolios to stabilise portfolio weights. Pollak (2011), and Frahm and Memmel (2010) carry out analogous investigations. A drawback of this method could be that optimal weights are shifted away from the theoretical solutions which can lead to worse portfolio performance. Investigating the trade off between more stable weights and portfolio returns is of interest. Furthermore, the shrinkage intensities for these portfolio combinations are also based on sample estimates that are subject to parameter uncertainty. The empirical applicability of these frequentist approaches is therefore yet to be confirmed.

More pragmatic avenues such as explicitly implementing weight restrictions have also been explored to lower weight allocations. Jagannathan and Ma (2002) consider a no short selling constraint, which is shown to be a form of indirect shrinkage of the expected returns. Alternative specifications are found in Brodie et al. (2009) and Li (2015), who apply norm-constrained portfolios. Furthermore, Brandt (2009) provides combinations of portfolio constraints and factor models. These papers show that weight constraints prove successful in targeting the volatility of weight allocations but come at cost of weight flexibility, restricting the feasible region of the mean-variance portfolio. A different, less restrictive approach is taken by Hautsch and Voigt (2019). Instead of targeting weights directly, portfolio turnovers are penalised as a means to improve portfolio allocations. A combination of the mentioned portfolio constraints can be found in the procedure of Han (2019). Han reports improved performance of a portfolio model that penalizes the deviation from a defined reference portfolio. Though successful in providing stable portfolios, these methods are more a means to performance improvement than a solution to the underlying Markowitz portfolio deficiencies. Instead, I will focus on improving the model application by accounting for its weaknesses and specifically targeting the fundamental problem of parameter uncertainty.

For that reason, it would be worthwhile to investigate the problem from a Bayesian perspective. With Bayesian methods, uncertainty can be accounted for by taking (adequate) distributions for the input parameters rather than frequentist point estimates. S.J. Brown (1976) and Klein and Bawa (1976) conduct such an analysis using a diffuse or flat prior. Frost et al. (1986) and Frey (2016) rely on conjugate prior specifications to account for estimation errors for the Markowitz and Global Minimum Variance Portfolio respectively. Other prior specifications are derived from economic objectives in Jun Tu and Guofu Zhou (2010) or from asset pricing models in Pastor and Stambaugh (2000) and Pastor (2000). The large body of literature devoted to Bayesian analysis highlights some of its attractive features. Prior choice specification allows for a great degree of flexibility as well as (economic) intuition. Furthermore Bayesian methods can be used for cases of very high dimensions making it suited to financial data sets.

The bulk of literature dedicated to modelling posterior predictive excess returns via Bayesian statistics assumes normal asset returns. However, countless papers have reported that asset returns do, in fact, not follow a Gaussian distribution. As pointed out by papers as Fama (1965) and Ané and Geman (2000), stylised facts of excess returns show returns to behave more leptokurtic than normal, with properties as skewness, fatter tails and extreme values. Some works that have proposed other, more fitting distributions include Robert C. Blattberg and Nicholas J. Gonedes (1974), who consider fatter tails of returns by Student's t-distributions, and Harvey et al. (2010), who use a multivariate skew normal distribution to account for the skewness of returns. Gaussian Mixtures are examined in Buckley et al. (2008) and Qian (2009). These prove very useful as a mixture of enough normal distributions can replicate almost any other distribution. In Gaussian mixture models, the latent variables are random draws defined by a multinomial distribution (Qian, 2011). These can be generalised however, by allowing the latent regimes to follow a Markov process. As derived in, Baum and Petrie (1966) and Baum et al. (1970), the resulting model is defined as a Hidden Markov Model (HMM), or more commonly referred to as Markov Switching (MS) model. Qian (2011) provides a Bayesian application of the MS model for portfolio selection.

Over the years, researchers have also deviated from the original implementation of the Markowitz portfolio. Reiterations of the original mean-variance problem have been derived to account for different investor preferences. Gârleanu and Pedersen (2013) and DeMiguel et al. (2015), for instance, define a longer investment horizon with the inclusion of transaction costs in the mean-variance specification. Other alternative models as the Global Minimum Variance Portfolio and the Black-Litterman model (Black and Litterman, 1992) are two of the more popular substitutes for the Markowitz portfolio. Both these have shown to provide more robust portfolios and shall be included as benchmarks. The Global Minimum Variance Portfolio disregards expected returns, thus reducing parameter uncertainty, while the Black-Litterman relies on the combination of a market equilibrium and investor views, to solve the problem of extreme weights.

The relevance of investigating portfolio optimisation lies in its abundant usage by (institutional) investors in practice. Traditional methods employed by financial institutions are based on frequentist theory which omit embedded uncertainty and prove sub-optimal. In this research more fitting Bayesian methods are explored that can prove valuable for practitioners. Furthermore, by postulating a novel framework of Bayesian model combinations that is more robust to changing economic times, a step is taken in the right direction towards finding the holy grail among investment strategies within applied finance. Combining optimal frequentist methods with Bayesian estimates extends previous efforts in literature that mitigate effects of estimation error and proves a relevant addition to both empirical and theoretical portfolio management.

1.3 Contributions and Findings

Throughout this research, comments are provided on the ability to suppress parameter uncertainty in dynamic portfolio optimisation. A range of methods are employed to mitigate the effect of estimation error on the Markowitz portfolio, and the competence of these methods in providing robust portfolios is evaluated on empirical data sets. The contributions of this paper are fivefold.

To address sensitivity of the Markowitz portfolio to parameter estimates, frequentist parameter shrinkage methods are implemented. An alternative specification of the Black-Litterman model is employed providing a new shrinkage interpretation of the model. More specifically, it is shown that a mean-variance portfolio based on Black-Litterman parameter estimates is comparable to a direct parameter shrinkage method where sample estimates are shrunk based on investor views. The reported out-of-sample measures show the model adaptation to provide minimal improvements over traditional Markowitz performance. However, ‘better’ investor views can potentially enhance performance. With more accurate (forecast) views, the Black-Litterman model would shrink parameters closer to true values to boost mean-variance performance.

The second contribution lies in portfolio weight shrinkage. These methods combine the Markowitz portfolio with other target portfolios that are less susceptible to estimation risk. Consequently, the magnitude and volatility of weight allocations are targeted. I propose other combinations towards strategies that are subject to less parameter uncertainty. New optimal combinations of Markowitz with the naive equally weighted portfolio and the Global Minimum Variance Portfolio with equally weighted are derived in attempt to improve over traditional Markowitz performance. Though successful in diminishing portfolio volatility, the proposed methods are not able to outperform the empirical performance of the optimal three-fund rule of Kan and Zhou (2007), which combines Markowitz and the Global Minimum Variance Portfolio.

The third contribution targets the shortcomings of frequentist portfolio optimisation techniques. Frequentist solutions rely on sample estimates which are subject to parameter uncertainty. This paper proposes to use Bayesian portfolios that are able to account for the embedded parameter uncertainty. A number of Bayesian specifications are regarded, of which the hierarchical prior of Greyserman et al. (2006) and the Markov Switching Gaussian Mixture model of Qian (2011) specifications prove most successful in capturing stylised facts of returns. By adopting a Bayesian perspective for ‘optimal’ frequentist methods this paper is novel in its approach to enhance portfolio optimisation techniques. In particular, defining the Kan and Zhou (2007) three-fund rule in a Bayesian manner such that estimation error is better accounted for, and return characteristics are better captured, extends the works of the previous authors. The out-of-sample performance measures improve substantially and the tough equally weighted benchmark portfolio is beat. The Markov Switching Gaussian Mixture extended three-fund rule reports an out-of-sample Sharpe ratio of 1.442 and 2.077, compared to 0.243 and 0.089 of the traditional Markowitz model, and 0.395 and 0.321 of the equally weighted portfolio, for the 48 Industry portfolios and Fama-French 100 portfolios data sets. Additionally, the model proves a lot more robust to changing economic times, remarkably outperforming the benchmarks during the global financial crisis. Over the years 2007 to 2009, the three-fund rule extension reports out-of-sample Sharpe ratios of 0.712 and 1.645, as opposed to -0.123 and 0.553 for Markowitz, and 0.134 and 0.033, for the equally weighted portfolio over the 48 Industry portfolios and Fama-French 100 portfolios data sets, respectively.

Moreover, the Bayesian implementation also allows for additional contributions. In this paper an alternative approach to construct portfolio combinations is proposed. By applying Bayesian statistics, strategy weights can be defined as distributions. The resulting weight distributions of different portfolios are then ‘merged’ to construct novel Bayesian portfolio combinations.

The technique is presented for a combination of the three pillars of portfolio optimisation, Markowitz, Global Minimum Variance Portfolio and Black-Litterman. Different combination schemes of the three models are explored based on their respective expected utilities. Equally weighted, maximum utility and utility weighted combinations are presented. The equally weighted combination of portfolios that is based on the Bayesian Markov Switching Gaussian Mixture model reports the highest out-of-sample performance in terms of Sharpe ratios over the whole sample, equal to 1.615 and 2.338 for the 48 Industry Portfolio and Fama-French 100 portfolios data sets, respectively. As for the Bayesian extension of the three-fund, remarkably high performance is reported, deeming the Bayesian combination extensions the most effective Markowitz modifications to construct robust portfolios. Furthermore, the proposed Bayesian combination framework opens up new avenues in the search of a universal portfolio optimisation model.

Finally, by considering detailed evaluation criteria and conducting a comprehensive robustness analysis over different time periods, further understanding of the performance of existing portfolio is enabled.

The rest of this paper is structured as follows. Section 2 reports traditional portfolio optimisation models and characterises the impact of estimation errors. Both frequentist and Bayesian techniques are presented to suppress the effects of parameter uncertainty. Section 3 and section 4 then provide the out-of-sample performance measures and the regarded empirical data sets. In section 5, the out-of-sample performances of the modified Markowitz models are discussed, and concluding remarks are made in Section 6. Evaluations and avenues for further research are provided in section 7.

2 Methodology

In attempt to improve the out-of-sample performance and practical applicability of the traditional portfolio allocation methods, extensions and alternative procedures are pursued that are able to overcome the commonly known shortcomings. I commence by presenting the traditional multiperiod investment strategies: the Mean-Variance or Markowitz portfolio (MW), the Global Minimum Variance Portfolio (GMVP) and the Black-Litterman model (BL). These will be taken as starting points of the investigation. The role of parameter uncertainty in mean-variance optimisation is then made evident in the form of a loss function and used as motivation for the search of methods to tackle it. Underlying reasons for the malperformance of the traditional portfolio allocations are defined and methods to address these causes are presented. Section 2.2.1 regards the quality of the model input parameters, section 2.3 focuses on shrinking the allocated weights and section 2.4 attempts to account for estimation error through Bayesian approaches. The Bayesian model is extended to account for non-normality of asset returns in section 2.5.1, and portfolio combinations for more robust performance are discussed in section 2.5.2. An overview of the considered models is provided in Table 1.

Table 1: Overview of Models

No.	Model	Abbreviation	Section
1 -	Traditional optimisation Models		
	Markowitz portfolio	MW	2.1.1
	Global Minimum Variance portfolio	GMVP	2.1.1
	Black-Litterman	BL	2.1.1
2 -	Direct Shrinkage of Parameters		
	Markowitz with shrunk estimate for μ	MW μ_{sh}	2.2.1
	Markowitz with shrunk estimate for Σ	MW Σ_{sh}	2.2.1
	Markowitz with shrunk estimates for μ and Σ	MW μ_{sh}, Σ_{sh}	2.2.1
	Alternative Black-Litterman specification, μ and Σ shrunk by views	MW-BL	2.2.1
	Markowitz with μ and Σ estimated by Fama-French 3-factor model	MW 3FF	2.2.3
	Markowitz with μ and Σ estimated by Principal Component Analysis	MW PCA	2.2.3
3 -	Portfolio Weight Shrinkage		
	Two fund rule, tangency portfolio and risk-free asset	2f	2.3.1
	Three fund rule, Markowitz, Global Minimum Variance and Risk-free asset	3f	2.3.1
	Three fund rule, Markowitz, Equally weighted and Risk-free asset	3f EW	2.3.1
	Three fund rule, Equally weighted, Global Minimum Variance and Risk-free asset	3f EW x GMVP	2.3.1
4 -	Accounting for Parameter Uncertainty		
	Bayesian Markowitz with diffuse prior	MW diff	2.4.1
	Bayesian Markowitz with conjugate prior - Normal, Inverse Wishart	MW conj	2.4.2
	Bayesian Markowitz with conjugate hyperparameter prior - Jorion	MW Jor	2.4.2
	Bayesian Markowitz with hierarchical prior	MW hier	2.4.3
	Bayesian Global Minimum Variance with conjugate prior - Normal Inverse Wishart	GMVP conj	2.4.2
	Bayesian Global Minimum Variance with hierarchical prior	GMVP hier	2.4.3
	Bayesian Black-Litterman with conjugate prior - Normal Inverse Wishart	BL conj	2.4.2
	Bayesian Black-Litterman with hierarchical prior	BL hier	2.4.3
5 -	Robustification of Models		
	Bayesian Markowitz Markov Switching Gaussian Mixture	MW MSGM	2.5.1
	Bayesian Global Minimum Variance Markov Switching Gaussian Mixture	GMVP MSGM	2.5.1
	Bayesian Black-Litterman Markov Switching Gaussian Mixture	BL MSGM	2.5.1
	Bayesian three-fund rule, Markowitz, Global Minimum Variance and Risk-free asset with hierarchical prior	3f hier	2.4.3
	MSGM Bayesian three-fund rule, Markowitz, Global Minimum Variance and Risk-free asset	3f MSGM	2.5.1
	Hierarchical Bayesian equally weighted combination of Markowitz, Global Minimum Variance and Black-Litterman based on utility	U EWC hier	2.5.2
	Hierarchical Bayesian max utility weighted combination of Markowitz, Global Minimum Variance and Black-Litterman	U MWC hier	2.5.2
	Hierarchical Bayesian weighted combination of Markowitz, Global Minimum Variance and Black-Litterman based on utility	U WC hier	2.5.2
	MSGM Bayesian equally weighted combination of Markowitz, Global Minimum Variance and Black-Litterman based on utility	U EWC MSGM	2.5.2
	MSGM Bayesian max utility weighted combination of Markowitz, Global Minimum Variance and Black-Litterman	U MWC MSGM	2.5.2
	MSGM Bayesian weighted combination of Markowitz, Global Minimum Variance and Black-Litterman based on utility	U WC MSGM	2.5.2
6 -	Benchmark Model		
	Equally Weighted, naive 1/N portfolio	EW	2.1.1

Note: This table presents an overview of the considered Markowitz modification strategies and alternatives portfolio rules that are investigated throughout this paper. Abbreviations that are used to refer to each strategy are given along with the corresponding sections in which they are explained. μ denotes the expected excess return parameter and Σ is the covariance parameter of excess returns.

2.1 Characterizing Estimation Error

In this section the traditional portfolios allocation methods are presented. The Markowitz mean-variance model (Markowitz (1952)), the Global Minimum Variance, and the Black-Litterman model (without investor views) (Idzorek (2007)) are considered starting points of the investigation. Following common practice, these efficient portfolios are constructed with sample moments. The effects of estimation error on MW performance are quantified in the case where sample estimates are used by the investor. This is used to highlight the impact of parameter uncertainty and used to motivate the search for methods to suppress its effects.

2.1.1 Traditional Portfolio Allocation Methods

As in DeMiguel et al. (2009), a daily rolling window approach is adopted for a dynamic implementation of the investment strategies. Sample moments calculated over the rolling window are used to determine the fraction of wealth to be held in each of the ($i = 1, \dots, N$) assets. To be consistent with previous literature, the methodology section assumes a sample of T observations to calculate the weight allocations and returns of a strategy at time t . The process is moved over one day, to $t + 1$, and is repeated. This is done over the whole out-of-sample set such that the portfolio strategies are rebalanced every day. A rolling window is opted for as opposed to an expanding window due to its ability to better respond to structural breaks (Bessler et al., 2017).

The Markowitz Portfolio

The MW portfolio relies on investors seeking to maximise their returns given a level of risk. The optimisation problem can be expressed in terms of Utility, $U(w_t)$, given asset weights, w_t , at time period, t . The MW mean-variance problem is known to be

$$\max_{w_t} U(w_t) = w_t' \mu - \frac{\gamma}{2} w_t' \Sigma w_t. \quad (1)$$

Here $\mu \in \mathbb{R}^{N \times 1}$ is a vector of the assets' expected returns¹, and $\Sigma \in \mathbb{R}^{N \times N}$ is a matrix for the covariance of returns. Both are constructed from the information available in that window. Scalar γ reflects the investor's risk aversion. The optimisation problem can be solved to obtain a vector of the weights that the MW trading rule assigns to each asset at time t ,

$$w_t^M = \frac{1}{\gamma} \Sigma^{-1} \mu. \quad (2)$$

Original MW construction relied on normally distributed returns. However, more recent studies such as Landsman and Nešlehová (2008) and Bessler et al. (2017) have shown mean-variance preferences to be reasonable as long as the returns are symmetric². The lack of a required specification for the assets returns is one of the key reasons for which the MW approach has been so popular in practice. As the true moments are unknown, sample moments using T observations are taken as input parameters³ to approximate the portfolio.

¹All expected returns are taken in excess of the risk free.

²"It is sufficient that returns are elliptically symmetrically distributed so that all investor preference are equivalent to mean-variance preferences" (Bessler et al., 2017).

³ $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T r_t$ and $\hat{\Sigma} = \frac{1}{N} \sum_{t=1}^T (r_t - \hat{\mu})(r_t - \hat{\mu})'$.

The Global Minimum Variance Portfolio

The GMVP only takes into account the variance or risk of the portfolio. The strategy is less subject to estimation error as it is not composed of the noisy estimates of the expected returns parameter. Chopra and Ziemba (1993) show the estimation errors for the mean to be about 10 times larger than that of the covariance. The risk minimization problem including budget constraint, and minimum variance solution w^{GMV} are given by,

$$\min_{w_t} w_t' \Sigma w_t, \quad \text{s.t.} \quad w_t' \mathbf{1} = 1, \quad (3)$$

$$w_t^{GMV} = \frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}' \Sigma^{-1} \mathbf{1}}. \quad (4)$$

Sample estimates are used to approximate the GMVP. The model is included in the analysis as it can provide further insights in to the amount of uncertainty that each individual parameter adds to the MW portfolio. Furthermore, the motivation behind the GMVP is similar to that of MW, focusing on risk minimisation, and hence making it a straightforward alternative.

The Black-Litterman Model

The BL model makes use of a combination of the discussed mean-variance structure and the Capital Asset Pricing Model (CAPM) of Sharpe (1964). The market equilibrium portfolio is used as ‘neutral’ basis, and is then tweaked according to investor views and expectations. Lee (2000) states that the BL model is able to “largely mitigate” the problem of estimation error-maximization making the strategy a reasonable alternative to the mean-variance portfolio trading rule. Though it has an unfair advantage over the traditional mean-variance implementations by making use of investor views, I include the BL in the analysis as a benchmark and for completeness. Taken together with MW and GMVP, the three most prominent strategies of portfolio optimisation are considered. Furthermore, in section 2.2.1, the BL model is shown to be equivalent to a form of parameter shrinkage for the MW portfolio. For the base-case, the canonical BL model as described in Walters (2011) is considered.

The Black-Litterman without views

The most simple implementation of the Black-Litterman model disregards investor views. It is often opted for in practice as investor views are subjective and are thus difficult to estimate. Without investor views, BL takes the same position as the market equilibrium as derived by the CAPM. The factor model reads,

$$E(r_i) = r_f + \beta_i(E(r_m) - r_f), \quad (5)$$

where the expected returns of asset i are denoted by $E(r_i)$, the risk-free rate by r_f and the expected excess market return by $E[r_m]$. The sensitivity of the expected excess asset returns to the expected excess market returns, β_i , can also be expressed as $\frac{\text{Cov}(r_i, r_m)}{\text{Var}(r_m)}$.

To obtain the market equilibrium returns one uses a reverse optimisation method (Idzorek, 2007). Starting from the optimal weight allocation for the market weights, one can back out, $r_\pi \in \mathbb{R}^{N \times 1}$, the implied excess equilibrium returns,

$$r_\pi = \gamma \Sigma w_{mkt}. \quad (6)$$

Again γ is the risk aversion coefficient, $\Sigma \in \mathbb{R}^{N \times N}$ is the covariance of returns, and $w_{mkt} \in \mathbb{R}^{N \times 1}$ is the market capitalization weight of the assets.

To visualise how the BL model without investor views produces more robust portfolios than the MW model, assume that the distribution of expected returns r_{t+1} is normal, $r_{t+1} \sim \mathcal{N}(\mu, \Sigma)$. For the base-case take Σ to be the sample estimate. The unknown mean return can then be defined as a random variable, $\mu \sim \mathcal{N}(r_\pi, \tau\Sigma)$, where r_π , the implied excess returns, and $\tau\Sigma$ are estimates of the mean and a scaled variance of the unknown mean returns. Parameter τ determines the investors confidence in r_π . By taking the distribution μ in combination with that of the expected excess returns, one can show that $r_{t+1} \sim \mathcal{N}(r_\pi, \Sigma + \tau\Sigma)$. Note how the variance of returns is inflated by a scaled variance term. Next to taking on a different expected excess return mean, the BL model is subject to greater variance than the MW portfolio. This results in less extreme weight allocations and therefore ‘safer’ and more stable portfolios.

This research focuses primarily on improving out-of-sample performance of the MW portfolio. This section however, presents all three pillars of portfolio optimisation for reference and completeness. Definitions of the GMVP and BL are used in later sections to mitigate effects of estimation errors for MW and thus required introduction. Moreover, the inclusion of the alternative optimisation strategies provides additional adequate benchmarks for the empirical out-of-sample performance of MW. Lastly, in section 2.5.2, I investigate whether the three traditional models can be combined effectively to specify a (more) universal portfolio optimisation model.

2.1.2 Quantifying Parameter Uncertainty

To characterise the effect of parameter uncertainty on the MW portfolio I make use of the investors utility as done in Kan and Zhou (2007). The mean-variance preferences are depicted in equation 1. Substituting in the true optimal MW weight w_t^M of (2) gives the corresponding optimal level of investor utility,

$$U(w_t^M) = \frac{1}{2\gamma} \mu' \Sigma^{-1} \mu = \frac{\theta^2}{2\gamma}. \quad (7)$$

Here $\theta^2 = \mu' \Sigma^{-1} \mu$ is the squared Sharpe ratio of the ex ante tangency portfolio of the risky assets. Identically, the sample weight equivalent can be obtained by substituting in sample estimates. From a frequentist perspective, and under the normality assumption of returns, $\hat{\mu}$ and $\hat{\Sigma}$ are independent of one another and follow a normal and Wishart distribution respectively for N assets and T time periods,

$$\hat{\mu} \sim \mathcal{N}(\mu, \Sigma/T), \quad (8)$$

$$\hat{\Sigma} \sim W_N(T-1, \Sigma)/T. \quad (9)$$

Robb J. Muirhead (1982) show that the expected value of the inverse of a Wishart defined covariance is $E[\hat{\Sigma}^{-1}] = \frac{T}{T-N-2} \Sigma^{-1}$. The resulting expected value of portfolio weights is then $E[\hat{w}_t^M] = \frac{T}{T-N-2} w_t^M$ when $T > N + 2$. As the factor by which the inverse covariance is multiplied is greater than one, it can be concluded that $|w_t^M| < |\hat{w}_t^M|$. Therefore, a sample plug-in MW portfolio tends to take larger positions in risky assets than a MW portfolio with true parameters.

It is possible to construct risk functions to quantify the estimation risk involved when using the sample estimates. As in

Kan and Zhou (2007), I provide the risk function in the case of both input parameters unknown, as well as for assumed known mean and covariance to dissect the effect of each parameter individually.

In the first scenario, the parameter uncertainty associated with μ is considered. Assume Σ to be known, such that, $\hat{w}_t^M = \frac{1}{\gamma}\Sigma^{-1}\hat{\mu}$. Using that $\hat{\mu}'\Sigma^{-1}\hat{\mu} \sim \chi_N^2(T\mu'\Sigma^{-1}\mu)/T$,

$$\begin{aligned} E[\tilde{U}(\hat{w}_t^M)|\Sigma] &= E[\hat{w}_t^M]'\mu - \frac{\gamma}{2}E[\hat{w}_t^M'\Sigma\hat{w}_t^M] = \frac{1}{\gamma}\mu'\Sigma^{-1}\mu - \frac{1}{2\gamma}E[\hat{\mu}'\Sigma^{-1}\hat{\mu}], \\ &= \frac{1}{\gamma}\mu'\Sigma^{-1}\mu - \frac{1}{2\gamma}\left(\frac{N + T\mu'\Sigma^{-1}\mu}{T}\right), \\ &= \frac{\theta^2}{2\gamma} - \frac{N}{2\gamma T}. \end{aligned} \quad (10)$$

The resulting risk function denoted by $\rho(\cdot)$ is

$$\begin{aligned} \rho(w_t^M, \hat{w}_t^M|\Sigma) &= U(w_t^M) - E[\tilde{U}(\hat{w}_t^M)|\Sigma], \\ &= \frac{N}{2\gamma T}. \end{aligned} \quad (11)$$

Implying that the investor certainty equivalent loss is expected to be $\frac{N}{2\gamma T}$. Furthermore, as expected, the loss decreases when sample size T increases, this makes sense as then samples estimates get closer to the true values. Also note how an increase in the number of assets N , which would be reflected by more elements in μ , increases the expected loss. By considering more assets, there is more estimation error involved. Clearly, for large portfolios the impact of parameter uncertainty is substantial, even when Σ is known.

Next, the loss function for known μ and unknown Σ is derived. Literature such as Chopra and Ziemba (1993), state the estimation error to be greater for the expected returns than for the variance. I check whether the impact on the optimal portfolio weight is consistent with their finding. Take for this scenario the optimal weights $\hat{w}_t^M = \frac{1}{\gamma}\hat{\Sigma}^{-1}\mu$, where Σ is estimated. It turns out useful to define the variable $W = \Sigma^{-\frac{1}{2}}\hat{\Sigma}\Sigma^{-\frac{1}{2}} \sim W_N(T-1, I_N)/T$ for simplification of the derivation.

Haff (1979) show the inverse moments of this variable to be

$$E[W^{-1}] = \left(\frac{T}{T-N-2}I_N\right), \quad (12)$$

$$E[W^{-2}] = \left[\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)}\right]I_N, \quad (13)$$

where $T > N + 4$. The expected out-of-sample performance can now be derived,

$$\begin{aligned} E[\tilde{U}(\hat{w}_t^M)|\mu] &= \frac{1}{\gamma}E[\mu'\hat{\Sigma}^{-1}\mu] - \frac{1}{2\gamma}E[\mu'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\mu], \\ &= \frac{1}{\gamma}E[\mu'\Sigma^{-\frac{1}{2}}W^{-1}\Sigma^{-\frac{1}{2}}\mu] - \frac{1}{2\gamma}E[\mu'\Sigma^{-\frac{1}{2}}W^{-2}\Sigma^{-\frac{1}{2}}\mu], \\ &= k_1\frac{\theta^2}{2\gamma}, \end{aligned} \quad (14)$$

where,

$$k_1 = \left(\frac{T}{T-N-2}\right)\left[2 - \frac{T(T-2)}{(T-N-1)(T-N-4)}\right]. \quad (15)$$

The loss function for the MW portfolio given known μ is

$$\rho(w_t^M, \hat{w}_t^M) = (1 - k_1) \frac{\theta^2}{2\gamma}, \quad (16)$$

such that $1 - k_1$ is the percentage loss of the expected out-of-sample performance due to the estimation error of $\hat{\Sigma}$. As in the case where the expected returns were unknown, this loss function is also increasing in N and decreasing in T . The relation also implies that $k_1 < 1$, and that k_1 can become negative for N large relative to T . That would translate to an investor avoiding investments in risky assets.

In the case of unknown μ and known Σ , the expected out-of-sample performance is affected by a fixed quantity $\frac{N}{2\gamma T}$. On the other hand, when Σ is unknown and μ is known, the out-of-sample performance is affected respective to the magnitude of the true parameters, namely a constant proportional amount dependent on θ^2 . The impact of the individual parameters on the expected losses differ. By calculating expected losses for different θ values, Kan and Zhou (2007) point out that in cases where N/T is large, one can expect estimation error of Σ to be sizeable compared to μ . This finding stresses the importance of considering the uncertainty of Σ as well as μ for MW portfolio construction.

Lastly, repeat the same analysis, but now for both parameters μ and Σ unknown. Both sample estimates, $\hat{\mu}$ and $\hat{\Sigma}$, are used to determine the MW weights. Using the properties of the inverse of moments of the Wishart distribution the expected loss function can be derived,

$$\begin{aligned} E[\tilde{U}(\hat{w}_t^M)] &= \frac{1}{\gamma} E[\hat{\mu}' \hat{\Sigma}^{-1} \mu] - \frac{1}{2\gamma} E[\hat{\mu}' \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \hat{\mu}], \\ &= \frac{1}{\gamma} E[\hat{\mu}' \hat{\Sigma}^{-\frac{1}{2}} W^{-1} \hat{\Sigma}^{-\frac{1}{2}} \mu] - \frac{1}{2\gamma} E[\hat{\mu}' \hat{\Sigma}^{-\frac{1}{2}} W^{-2} \hat{\Sigma}^{-\frac{1}{2}} \hat{\mu}], \\ &= k_1 \frac{\theta^2}{2\gamma} - \frac{NT(T-2)}{2\gamma(T-N-1)(T-N-2)(T-N-4)}, \end{aligned} \quad (17)$$

assuming that $T > N + 4$. The resulting loss function for out-of-sample performance is given by

$$\rho(w_t^M, \hat{w}_t^M) = (1 - k_1) \frac{\theta^2}{2\gamma} + \frac{NT(T-2)}{2\gamma(T-N-1)(T-N-2)(T-N-4)}. \quad (18)$$

In line with the first two scenarios, the loss is characterised by N, T, γ and θ^2 . The loss is greater for larger N or θ^2 , and smaller for larger T and γ . I highlight the fact that the second term in (18) is always greater than the loss function for known Σ , implying that the losses associated with uncertainty of the parameters are not additive.

This section quantified the impact of estimation error on the MW portfolio when sample estimates are used. The presented analytical solution provides intuition for the incurred losses in terms of out-of-sample performance. Especially in cases where N is large relative to T , utility loss is significant. Furthermore, I highlight the misconception of expected asset return estimation errors being much greater than that of variance estimation. In cases of large N/T , the incurred utility losses for unknown Σ become larger, such that the cost of estimating Σ can be larger than the cost of estimating μ (Kan and Zhou, 2007). This motivates the search for shrinkage methods for both parameters. Also note how the loss function depends on specification of risk aversion parameter. In Appendix A I provide the expected utility loss for varying values of γ for the mean-variance portfolios. As expected, smaller values of γ lead to larger expected utility losses. As the derivation of the risk aversion parameter is somewhat subjective and falls outside the scope of this research, I assume a fixed value for γ for the considered

models⁴. Having established that effects of parameter uncertainty on MW performance are significant, the next sections focus on suppressing the impact of estimation errors.

2.2 Parameter Shrinkage Techniques

I distinguish two frequentist approaches to suppress the effects of parameter risk: direct parameter shrinkage and portfolio weight shrinkage. Parameter shrinkage aims to reduce estimation error to improve the quality of input parameters, whilst portfolio weight shrinkage targets the sensitivity of the MW model to input parameters.

2.2.1 Direct Shrinkage of Parameters

In this section, the expected returns and covariance are shrunk towards parameters that are more ‘certain’ or less subject to parameter uncertainty. Such combinations allow for more accurate parameter estimates when the added bias is outweighed by the reduction in parameter uncertainty. Implementation is rather straightforward, the modified MW portfolio is derived by the plug-in approach where shrinkage estimates replace the standard sample estimates.

Shrinkage techniques define the parameter of interest as a convex combination of the sample parameter and a target parameter. Mathematically this takes the form of,

$$X_{sh} = (1 - \alpha_x)\hat{X} + \alpha_x X_{tg}, \quad (19)$$

where X_{tg} represents a target parameter and α_x defines the shrinkage intensity. X_{sh} is the resulting shrunk estimate of the parameter. First, consider shrinking the sample estimate of the expected excess returns. Jorion (1986) take a scaled vector of ones, denoted by $\iota \in \mathbb{R}^{N \times 1}$ for shrinking the sample mean. Accordingly, define $\mu_{tg} = v\iota$, with v as scale parameter. The shrinkage estimator is given by

$$\mu_{sh} = (1 - \alpha_\mu)\hat{\mu} + \alpha_\mu \mu_{tg}. \quad (20)$$

Where v is chosen to minimise the bias of the target, $v_\mu = \min_v \|\nu\iota - \mu\|_2^2$ (DeMiguel et al., 2013). Taking the derivative and setting it equal to zero results in an estimate of $v_\mu = \frac{1}{N} \sum_{i=1}^N \mu_i = \bar{\mu}$. The resulting scale parameter is the estimate for the grand mean, the cross-sectional average over the expected excess returns of all assets.

A similar procedure is followed for deriving shrinkage intensity α_μ . The expected quadratic loss of the shrinkage estimator is minimised,

$$\min_{\alpha_\mu} E[\|\mu_{sh} - \mu\|_2^2] \quad \text{where} \quad \|X_i\|_2^2 = \sum_{i=1}^N X_i^2. \quad (21)$$

The quadratic minimisation proves beneficial as it can be applied without making assumptions regarding the distribution of

⁴I use a as small as possible γ value such that the investor is still risk-averse, $\gamma = 1$.

returns (DeMiguel et al., 2013). The resulting shrinkage intensity is⁵,

$$\alpha_\mu = \frac{E(\|\hat{\mu} - \mu\|_2^2)}{E(\|\hat{\mu} - \mu\|_2^2) + \|v_{\mu t} - \mu\|_2^2} = \frac{(N/T)\bar{\sigma}^2}{(N/T)\bar{\sigma}^2 + \|v_{\mu t} - \mu\|_2^2}. \quad (22)$$

With $\bar{\sigma}^2 = \text{trace}(\Sigma)/N$, the sum over the diagonal elements of Σ . The ‘less uncertain’ estimate of the expected excess returns will replace the sample parameter in the construction of portfolios. The shrinkage estimator can be shrunk to any arbitrary known value and this will result in reduction of the uncertainty of sample estimates. However, the magnitude of reduction will be larger if the shrinking target is closer to the true value. A target equal to the grand mean therefore seems a suitable candidate. The Markowitz model with shrunk expected returns parameter will be referred to as MW μ_{sh} throughout.

Next, shrinkage of the covariance matrix is carried out. A convex combination of the sample parameter and a scaled shrinkage target is opted for as in Ledoit and Wolf (2004),

$$\Sigma_{sh} = (1 - \alpha_\Sigma)\hat{\Sigma} + \alpha_\Sigma \cdot c_1 \Sigma_{tg}. \quad (23)$$

Where the identity matrix, I , is taken as shrinkage target. The identity matrix seems an adequate candidate due to its constancy and simplicity. Furthermore, as it requires no estimation, its corresponding estimation error is zero. Scalar c_1 is determined such that it minimises the added bias, $c_1 = \min_{c_1} \|c_1 I - \sigma\|_F^2$. Taking the derivative and setting it to zero results in $(1/N)\sum_{i=1}^N \sigma_i^2 = \sigma^2$, the variance-equivalent of the grand mean.

For an analytical expression of the shrinkage intensity α_Σ , the expected quadratic loss is minimised with respect to α_Σ ,

$$\min_{\alpha_\Sigma} E[\|\Sigma_{sh} - \Sigma\|_F^2] \quad \text{with} \quad \|X\|_F^2 = \text{trace}(X'X). \quad (24)$$

The minimisation problem can be solved and rewritten into the following expression for the shrinkage intensity,

$$\alpha_\Sigma = \frac{E(\|\hat{\Sigma} - \Sigma\|_F^2)}{E(\|\hat{\Sigma} - \Sigma\|_F^2) + \|\nu_\Sigma I - \Sigma\|_F^2}. \quad (25)$$

Observe that α_Σ is the relative expected loss of sample covariance against the expected losses of the identity *and* sample covariance matrices (DeMiguel et al., 2013). An analytical form (and corresponding) derivation of the shrinkage estimator can be found in DeMiguel et al. (2013). Importantly, given the assumption of IID normal returns and that $T > N + 4$, the paper provides an expression for $E(\|\hat{\Sigma} - \Sigma\|_F^2)$. I provide the expression for completeness and ease of reproducibility,

$$E(\|\hat{\Sigma} - \Sigma\|_F^2) = \frac{N}{T-1} \left(\frac{\text{trace}(\Sigma^2)}{N} + N(\bar{\sigma}^2)^2 \right), \quad (26)$$

where $\bar{\sigma}^2 = \text{trace}(\Sigma)/N$.

Again, the shrunk parameters shall be used as new plug-in parameters for MW allocations. Throughout, the Markowitz portfolio with shrunk covariance parameter will be referred to as MW Σ_{sh} , and the Markowitz portfolio with shrinkage estimates for both the expected returns and covariance will be referred to as MW μ_{sh}, Σ_{sh} . The out-of-sample performance of these frequentist MW modifications will provide insights on the ability to mitigate the impact of parameter uncertainty through improvements of the quality of inputs.

⁵The complete derivation of the shrinkage intensity can be found in Appendix A.1 of DeMiguel et al. (2013).

2.2.2 The Black-Litterman Model with Investor Views

The basic BL model introduced in section 2.1.1 is extended by a second, ‘subjective’ source of information which comes in the form of investor views. The views can be incorporated in the portfolio decision, and are taken in combination with market equilibrium returns to derive the model. The formulation of Idzorek (2007) is presented in this section. The BL model will be regarded as benchmark, and is also used in constructing combination strategies of the three traditional portfolio optimisation strategies in section 2.5. Additionally, by specifying a different equilibrium portfolio for traditional BL, I establish an alternative shrinkage interpretation of the model. In particular, I show that BL can be interpreted as a MW portfolio with shrunk parameter estimates when the mean-variance portfolio is taken as market equilibrium.

First, consider the implementation of investor views. There are contrasting methods to obtain these, of which the main schools of thought rely on historical estimates or incorporation of analyst opinions. Though very relevant, the derivation of investor views does not fall under the scope of this investigation. Therefore, the works of Schepel (2019) are followed in order to take the investor views as given. Investor views are derived by means of a factor model, such that there a view, Z , is assumed for each asset, N . More specifically, the Fama-French 3-factor model is used to determine these investor views. Define a column vector $Q \in \mathbb{R}^{Z \times 1}$, to represent the views on returns, and a matrix $\Omega \in \mathbb{R}^{Z \times Z}$ to adopt the uncertainties associated to the views. Q and Ω are set to the factor model mean and covariance estimates, respectively⁶. For simplicity Ω is taken to be diagonal such that investor views are independent and uncorrelated. The diagonal elements of Ω , be denoted by ω_i represent the confidence in investor views. The higher ω_i , the lower the confidence in investors views on asset i . Following Idzorek (2007), a more practical form of the system of views can be adopted. The views can be summarised by,

$$Q = P\mu + \eta, \quad (27)$$

where disturbances η are defined by the confidence matrix Ω . Bayes theorem can then be applied to combine the excess return distributions with investor views,

$$P(\mu|Q) = \frac{P(Q|\mu) \cdot P(\mu)}{P(Q)} \propto P(Q|\mu) \cdot P(\mu). \quad (28)$$

By multiplying the expected return prior, $P(\mu)$, by the likelihood of investor views given the expected returns, $P(Q|\mu)$, the distribution of the expected returns given investor views is obtained⁷. The resulting solution, given a scale parameter τ and implied equilibrium vector r_π is,

$$P(\mu|Q) \sim \mathcal{N}(\hat{\mu}_{BL}, \hat{\Sigma}_{BL}), \quad (29)$$

with parameters,

$$\hat{\mu}_{BL} = [(\tau\Sigma)^{-1} + P'\Omega^{-1}P]^{-1}[(\tau\Sigma)^{-1}r_\pi + P'\Omega^{-1}Q], \quad (30)$$

$$\hat{\Sigma}_{BL} = [(\tau\Sigma)^{-1} + P'\Omega^{-1}P]^{-1}. \quad (31)$$

Observe that scalar τ determines whether the assigned weights are allocated more towards the implied returns, r_π , or more towards investor views, $P^{-1}Q$. Furthermore, it is important to highlight the effects that investor confidence levels has on the

⁶Section 2.2.3 elaborates on the definition and calculation of estimates of Fama-French 3-factor model.

⁷The complete derivation of the Black-Litterman model can be found in Idzorek (2007).

BL estimate of the mean. Rewriting the equation for $\hat{\mu}_{BL}$ can make the effect more apparent,

$$\hat{\mu}_{BL} = r_\pi + \tau \Sigma P' [(P \tau \Sigma P') + \Omega]^{-1} [Q - P r_\pi]. \quad (32)$$

Recall how confidence in the investors views is reflected by the diagonal elements, ω_i , of the Ω matrix. The more certain the investor is about his views, the closer ω_i gets to zero. Consequently, $\hat{\mu}_{BL} \rightarrow P^{-1}Q$, the investor views. On the other hand, when the investor is less certain about his views, and $\omega_i \rightarrow \infty$, $\hat{\mu}_{BL} \rightarrow r_\pi$, the implied equilibrium returns. This is rather intuitive, as when the investor is 100 % confident of their views then BL should invest in their views only, and in the case that the investor is not confident at all, BL should hold the market equilibrium portfolio.

BL is included in the analysis due to its popular use by firms in practice. Comparison of BL to MW can help assess the reliability of the presented MW modifications. Furthermore, later sections assess the ability to construct a universally applicable portfolio by combining the three traditional portfolio optimisation methods. In addition the BL is shown to mimic a form of parameter shrinkage for the MW portfolio when specified differently.

Shrinkage Interpretation of Black-Litterman

When taking an alternative equilibrium portfolio for the BL model, one can show its equivalence to a MW model with shrunk parameters. Recall the definition of the implied equilibrium excess returns of (6). By taking the mean-variance portfolio of equation 2, as w_{mkt} , I show that the specification of BL is equivalent to that of parameter shrinkage. First, define the implied excess equilibrium return as,

$$r_\pi = \gamma \Sigma \times \left(\frac{1}{\gamma} \Sigma^{-1} \hat{\mu} \right) = \hat{\mu}. \quad (33)$$

where the equilibrium (MW) portfolio is defined under similar assumptions as the BL model in Idzorek (2007) - with known Σ . Substituting this into equation 32, the resulting BL estimate for μ becomes,

$$\begin{aligned} \hat{\mu}_{BL} &= \hat{\mu} + \tau \Sigma P' [(P \tau \Sigma P') + \Omega]^{-1} [Q - P \hat{\mu}], \\ &= \hat{\mu} (1 - \tau \Sigma P' [(P \tau \Sigma P') + \Omega]^{-1} P) + \tau \Sigma P' [(P \tau \Sigma P') + \Omega]^{-1} Q, \\ &= (1 - \alpha_\mu) \hat{\mu} + \alpha_\mu \mu_{tg}. \end{aligned} \quad (34)$$

Coinciding with the shrinkage definition of equation 19, with a shrinkage intensity of $\alpha_\mu = \tau \Sigma P' [(P \tau \Sigma P') + \Omega]^{-1} P$ and shrinkage target of $\mu_{tg} = P^{-1}Q$. The intensity of shrinkage is thus defined by investor confidence. If investor is more certain about his views, the $\hat{\mu}_{BL}$ estimate tends to the target $P^{-1}Q$, and if less certain, the estimate tends to $\hat{\mu}$.

Taking a normal distribution for future excess returns, the BL model can be summarised by,

$$r_{t+1} \sim \mathcal{N}(\hat{\mu}_{BL}, \Sigma + \hat{\Sigma}_{BL}), \quad (35)$$

where Σ is taken as its sample estimate $\hat{\Sigma}$. Under mean-variance preferences the Markowitz portfolio with a shrunk mean of $\hat{\mu}_{BL}$ and inflated variance of $\hat{\Sigma} + \hat{\Sigma}_{BL}$ remains. Given correct investor views, this should provide more conservative and accurate predictions. The model is denoted as MW-BL throughout and will be considered as an alternative parameter shrinking method in the investigation.

2.2.3 Factor Models

Another approach used in literature for the estimation of excess return moments relies on factor models. These prove useful when data sets are large. Summarizing the variations in the data in fewer factors makes the data more manageable and comprehensible. In order to suppress estimation error, describing asset returns in fewer factors, $J < N$, that capture the bulk of the variation may prove beneficial. By omitting non-representative/outlier stock returns, parameter uncertainty can be reduced when estimating μ and Σ . I follow the methodology of De Nard et al. (2018) to estimate the parameters.

I employ dynamic factor models, such that a rolling window is used to estimate coefficients. For every asset i , at time t , I assume the excess returns $r_{i,t}$ to follow a factor model,

$$r_{i,t} = \beta_i' f_t + u_{i,t}, \quad (36)$$

where excess returns are taken in excess of the risk-free rate, hence the missing intercept. Furthermore, $\beta_i := (\beta_{i,1}, \dots, \beta_{i,j})'$ and $E[u_{i,t}|f_t] = 0$, where $f_t = (f_{1,t}, \dots, f_{j,t})$ are the factors, β_i are the factor loadings, and $u_{i,t}$ is the error term for asset i at time period t .

Under these assumptions, the (time-varying) parameters that are to be used as replacements for the sample estimates are defined as,

$$\hat{\mu}_{r,t} = E[\beta_i' f_t], \quad (37)$$

$$\hat{\Sigma}_{r,t} = B' \Sigma_{f,t} B + \Sigma_{u,t}. \quad (38)$$

Here B is the $J \times N$ loading matrix with vectors β_i along its columns, $\Sigma_{f,t}$ is the covariance of factors and $\Sigma_{u,t}$ is the covariance of residuals.

Two prominent factor models used in finance are considered. First, the Fama-French 3-factor model of Fama and French (1993). The model makes use of the factors: firm size, book-to-market values and excess return on the market. The market return is in excess of the risk-free rate, the SMB - *small minus big* - factor differentiates the (out)performance of small firms against large firms, and HML - *high minus low* - factor is derived from the fact that firms with higher book-to-market ratios are expected to attain higher returns than lower ratio firms. This extension of the Capital Asset Pricing Model is often used as a benchmark in financial practices, hence the reason for its use in this research. Second, I opt for Principal Component Analysis (PCA) as a factor model. This data reduction technique can be used to shrink the amount of dimensions/assets to a number of factors that explain the bulk of variations in the assets space. A dynamic factor modelling approach is adopted such that for each rolling window the principal components explain at least 95% of the total asset return variance. The resulting factor model estimates for the expected returns and covariance will be used in the MW portfolio. The MW model based on parameter estimates of the 3-factor model is referred to as MW 3FF, and the model based on PCA is referred to as MW PCA.

2.3 Portfolio Weight Shrinkage

This section tackles the sensitivity of the MW portfolio to input parameters, and the portfolios tendency to assign extreme weights to assets. By shifting MW weight allocations to weights of safer portfolios, the model becomes more stable and robust to input parameters. The optimal two-fund and three-fund rules of Kan and Zhou (2007) are reviewed and used as inspiration

for weight shrinkage towards other, theoretically more stable portfolios.

2.3.1 Target Portfolio Shrinkage

MW portfolios are known to allocate very extreme weights to assets, making the model infeasible for practical purposes. By shrinking these weights to those of a more stable target portfolio, the overall weight allocations can be reduced. Moreover, shifting towards a safer strategy can mitigate sensitivity to input parameters, as these become less influential in the final weight assignments. As the resulting weight allocations mitigate the effects of MW deficiencies, the modifications can prove beneficial in terms of out-of-sample performance and empirical applicability. I first consider shrinkage towards the risk-free asset and shrinkage towards the GMVP, denoted by the two-fund and three-fund rules of Kan and Zhou (2007), respectively.

The Optimal Two-Fund Rule

The two-fund (2f) rule stems from two fund separation theorem of Wenzelburger (2008). The theorem postulates that mean-variance investors allocate their wealth over combinations of the riskless asset and the tangency portfolio. Accordingly, the assigned asset weights become a scaled form of the Markowitz trading rule,

$$\hat{w}_t^{2f} = \frac{c}{\gamma} \hat{\Sigma}^{-1} \hat{\mu}. \quad (39)$$

The shrinkage scalar c determines the amount invested in the tangency portfolio θ^2 (and the risk-free). It is set to maximise expected out-of-sample performance of the two-fund rule,

$$E[\tilde{U}(c\hat{\Sigma}^{-1}\hat{\mu}/\gamma)] = \frac{c\theta^2}{\gamma} \left(\frac{T}{T-N-2} \right) - \frac{c^2}{2\gamma} \left(\theta^2 + \frac{N}{T} \right) \left[\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \right], \quad (40)$$

where $T > N + 4$ is assumed. Maximising this equation with respect to c , results in the optimal shrinkage intensity for the tangency portfolio⁸,

$$c^* = \underbrace{\left[\frac{(T-N-1)(T-N-4)}{T(T-2)} \right]}_{c_\Sigma} \underbrace{\left(\frac{\theta^2}{\theta^2 + \frac{N}{T}} \right)}_{c_\mu}. \quad (41)$$

The term can be distinguished into two components. If Σ is known, there is no embedded uncertainty related to Σ , and only the second term, c_μ , remains to characterize shrinking intensity c^* . On the other hand if μ is known, then only the first term, c_Σ , characterises shrinking intensity c^* . This means that we can define c_Σ to be associated with the error in Σ and the c_μ to be associated with errors in μ . Under the optimal choice of $\hat{w}_t^{2f*} = c^* \hat{\Sigma}^{-1} \hat{\mu} / \gamma$, the expected out-of-sample performance is higher than that of the sample MW. This can be derived by comparing the expected utilities associated with both strategies. The utility function of \hat{w}_t^{2f*} is,

$$E[\tilde{U}(\hat{w}_t^{2f*})] = \frac{\theta^2}{2\gamma} \left[\frac{(T-N-1)(T-N-4)}{(T-2)(T-N-2)} \right] \left(\frac{\theta^2}{\theta^2 + \frac{N}{T}} \right), \quad (42)$$

which can be observed to be larger than the expected utility of the sample MW in equation 17 (Kan and Zhou, 2007). In theory, the two-fund rule should thus outperform the sample Markowitz. In practice however, optimal c^* is proves incalculable

⁸The complete derivation of the optimal shrinkage intensity can be found in Kan and Zhou (2007).

as the true Sharpe ratio of the tangency portfolio is unknown. This issue is solved by using the sample estimates to estimating c^* . The resulting approximated optimal two-fund rule is,

$$\hat{w}_t^{2f*} = \frac{1}{\gamma} \hat{c}^* \hat{\Sigma}^{-1} \hat{\mu}, \quad \hat{c}^* = c_{\Sigma} \left(\frac{\hat{\theta}^2}{\hat{\theta}^2 + \frac{N}{T}} \right). \quad (43)$$

As $\hat{c}^* < 1$, the two-fund rule invests less in risky assets than the MW portfolio. Intuitively this follows from the fact that the sample MW riskier than the true MW portfolio making it less attractive. Accordingly, the weights should be scaled downwards.

Given that the two-fund rule allocates assets over the risk free asset and the sample tangency portfolio, the trading strategy is still subject to parameter risk. To diversify this risk, trading towards another, safer portfolio may prove valuable. As long as the estimation errors of the considered portfolios are not perfectly correlated, the total amount of risk can be reduced by such a combination. This gives rise to the three-fund rule of Kan and Zhou (2007).

The Optimal Three-Fund Rule with Global Minimum Variance Portfolio

This trading rule invests in three different funds, the riskless asset, the sample tangency and the GMVP. The GMVP is chosen as target portfolio to shrink to as it is per definition the portfolio with minimum risk. Furthermore, as estimation of the GMVP only relies on parameter Σ , the models estimation errors are lower. Trading towards GMVP thus reduces the portfolio risk (diversification), as well as the total amount of parameter uncertainty for the portfolio. Define the weight of the three-fund strategy (3f) as a linear combination of the mentioned funds,

$$\hat{w}_t^{3f} = \frac{1}{\gamma} (c \hat{\Sigma}^{-1} \hat{\mu} + d \hat{\Sigma}^{-1} \iota_N). \quad (44)$$

Where shrinkage intensities c and d are found to maximise expected out-of-sample performance,

$$\begin{aligned} E[\tilde{U}(\hat{w}_t^{3f})] &= E[\hat{w}_t^{3f}]' \mu - \frac{\gamma}{2} E[\hat{w}_t^{3f} \Sigma \hat{w}_t^{3f}], \\ &= \left(\frac{T}{T-N-2} \right) \frac{1}{2\gamma} \left[2(c\mu' \Sigma^{-1} \mu + d\mu' \Sigma^{-1} \iota_N), \right. \\ &\quad \left. - \frac{T(T-2)}{(T-N-1)(T-N-4)} \times \left(\left(\mu' \Sigma^{-1} \mu + \frac{N}{T} \right) c^2 + 2(\mu' \Sigma^{-1} \iota_N) cd + (\iota_N' \Sigma^{-1} \iota_N) d^2 \right) \right]. \end{aligned} \quad (45)$$

It is assumed that $T > N + 4$. Taking the first derivative with respect to parameters c, d and setting to zero, Kan and Zhou (2007) find the optimal shrinkage rates to be,

$$c^{**} = c_{\Sigma} \left(\frac{\psi^2}{\psi^2 + \frac{N}{T}} \right), \quad d^{**} = c_{\Sigma} \left(\frac{\frac{N}{T}}{\psi^2 + \frac{N}{T}} \right) \mu_g. \quad (46)$$

Here μ_g is the expected excess return of the GMVP, and ψ^2 is the squared slope of the asymptote to the ex ante MV frontier. Mathematically it reads,

$$\psi^2 = \mu' \Sigma^{-1} \mu - \frac{(\mu' \Sigma^{-1} \iota_N)^2}{\iota_N' \Sigma^{-1} \iota_N} = (\mu - \mu_g \iota_N)' \Sigma^{-1} (\mu - \mu_g \iota_N). \quad (47)$$

Shrinkage intensity c^{**} determines the degree of shrinkage for the MW portfolio and shrinkage intensity d^{**} determines the degree of shrinkage for the GMVP. Both are estimated by their sample counterparts.

The Optimal Three-Fund Rule with Equally Weighted Portfolio

Seeking further diversification of the MW portfolio estimation risks, Tu and Zhou (2011) investigate the use of incorporating the equally weighted (EW) portfolio as extra investment fund. As the portfolio tends to outperform more sophisticated methods in practice, and it is not subject to any parameter uncertainty, the $1/N$ portfolio proves to be an adequate extra investment fund alongside the MW portfolio and the risk-free asset. As opposed to Tu and Zhou (2011), I do not constrain the shrinkage intensities to sum to one. The restriction is not needed for derivation of the optimal intensities and omitting it allows for increased flexibility (and feasible region of the optimisation) in portfolio combinations. Following the intuition of Kan and Zhou (2007), I maximise expected utility to define the three-fund rule with equally weighted portfolio (3f EW),

$$\hat{w}_t^{3fEW} = c_2 \frac{1}{\gamma} \hat{\Sigma}^{-1} \hat{\mu} + d_2 \frac{1}{N} \iota_N. \quad (48)$$

Where the equally weighted portfolio is defined by $w_t^{EW} = \frac{\iota_N}{N}$, investing equal amounts over all assets. One can calculate the corresponding out-of-sample performance for this investment strategy and differentiate with respect to c_2 and d_2 to find the optimal shrinkage intensities. I derive these in Appendix B. The resulting optimal shrinkage intensities are,

$$c_2^* = \frac{\theta^2 \iota_N' \Sigma \iota_N - (\iota_N' \mu)^2}{\frac{T(T-2)(N/T+\theta^2)\iota_N' \Sigma \iota_N}{(T-N-1)(T-N-4)} - \frac{T(\iota_N' \mu)^2}{T-N-2}}, \quad (49)$$

$$d_2^* = \frac{N \iota_N' \mu}{\gamma \iota_N' \Sigma \iota_N} \left(1 - \frac{(T-N-1)(T-N-2)(T-N-4)(\theta^2 - (\iota_N' \mu)^2)}{T^2(T-2)(N/T+\theta^2)\iota_N' \Sigma \iota_N - ((T-N-1)(T-N-2)(T-N-4))(\iota_N' \mu)^2} \right), \quad (50)$$

corresponding to maximum expected out-of-sample utilities. Shrinkage intensity c_2 determines the degree of shrinkage for the MW portfolio and shrinkage intensity d_2 determines the degree of shrinkage for the EW. Both are estimated by their sample counterparts.

The Optimal Three-Fund Rule for Global Minimum Variance and Equally Weighted Portfolios

An even more certain portfolio rule could be found in the combination of the EW portfolio and the GMVP. I include this rule for completeness of the analysis. Define the three-fund rule combining the equally weighted and GMVP portfolios (3f EWxGMVP) rule as,

$$\hat{w}_t^{3fEG} = c_3 \frac{1}{\gamma} \hat{\Sigma}^{-1} \iota_N + d_3 \frac{1}{N} \iota_N. \quad (51)$$

The derivations of optimal shrinkage intensities are reported in Appendix C. The expected out-of-sample utility maximising shrinkage intensities are,

$$c_3^* = \frac{\iota_N' \Sigma \iota_N \cdot \iota_N' \Sigma^{-1} \mu - \iota_N' \mu}{\frac{T(T-2)(\iota_N' \Sigma \iota_N)(\iota_N' \Sigma^{-1} \iota_N)}{(T-N-1)(T-N-4)} - \frac{NT}{T-N-2}}, \quad (52)$$

$$d_3^* = \frac{1}{\gamma \iota_N' \Sigma \iota_N} \left[\iota_N' \mu - \left(\frac{NT}{T-N-2} \right) \cdot \frac{\iota_N' \Sigma \iota_N \cdot \iota_N' \Sigma^{-1} \mu - \iota_N' \mu}{\frac{T(T-2)(\iota_N' \Sigma \iota_N)(\iota_N' \Sigma^{-1} \iota_N)}{(T-N-1)(T-N-4)} - \frac{NT}{T-N-2}} \right]. \quad (53)$$

Where the shrinkage intensity c_3 determines the degree of shrinkage for the GMVP and d_3 determines the degree of shrinkage for the EW. Again, both these are estimated by their sample counterparts.

The resulting portfolio combinations can end up investing more in to risky assets than the traditional MW model. By

construction, the GMVP relies on a budget constraint such that all wealth is invested in risky assets. Similarly, the EW portfolio invests all wealth over all assets. In section 5.2 I evaluate whether the increased investments in risky positions by methods that are less prone to estimation risk is beneficial for out-of-sample performance.

I also acknowledge an alternative method to boost MW performance through weight targeting. Works as Frost and Savarino (1988) and Jagannathan and Ma (2002) opt for the application of explicit weight restrictions on the MW portfolio. These studies show this approach to be rather effective. However, I argue that these methods *enforce* robustness and do not solve the underlying deficiencies of mean-variance portfolios. Input parameters remain the same, and with it effects of estimation error are still present. Furthermore the restrictions need to be subjectively determined and greatly impact the flexibility of allocations. As this paper investigates avenues to improve traditional performance by means of targeting parameter uncertainty, I deem weight constraint methods to fall outside the scope of this paper.

2.4 Accounting for Parameter Uncertainty

A more sophisticated method to account for parameter uncertainty can be found in the form of Bayesian analysis. By specifying input parameters as distributions, uncertainty can be implicitly accounted for in the construction of portfolio weights. This section investigates the ability to boost out-of-sample performance of the aforementioned strategies when estimation error is taken into consideration. I specify a Bayesian framework for portfolio models and provide shrinkage interpretations for portfolio weights under different priors.

Following the intuition behind the traditional MW model, I aim to maximise the investors mean-variance utility. As Zellner and Chetty (1965), the predictive distribution of asset returns are used to define an optimal Bayesian portfolio. Let Φ_t be all data available up to t and $U(w)$ the utility for holding strategy w at $t + 1$. The predictive distribution is denoted by $p(r_{t+1}|\Phi_t)$. The optimal Bayesian portfolio can then be formulated by,

$$\hat{w}_{t+1}^{Bayes} = \underset{w}{\operatorname{argmax}} \int_{r_{t+1}} U(w)p(r_{t+1}|\Phi_t)dr_{t+1}. \quad (54)$$

The predictive function can be rewritten as a combination of the data likelihood of the returns and posterior of the parameters. By integrating out the uncertainty in the parameters μ and Σ the predictive posterior distribution is shown to be,

$$p(r_{t+1}|\Phi_t) = \int_{\mu} \int_{\Sigma} p(r_{t+1}|\mu, \Sigma, \Phi_t)p(\mu, \Sigma|\Phi_t)d\Sigma d\mu, \quad (55)$$

such that estimation error is accounted for. The joint posterior density for (μ, Σ) is $p(\mu, \Sigma|\Phi_t)$, and $p(r_{t+1}|\mu, \Sigma, \Phi_t)$ is the likelihood of excess returns given μ , Σ and the data available up to time t . Following common practice in Bayesian analysis of returns, an IID multivariate normal distribution is chosen for the likelihood of the excess returns. Let r_t denote a vector of returns for the N assets. The likelihood of the N-dimensional multivariate normal population over sample $t = 1, \dots, T$ is given by,

$$\begin{aligned} p(r|\mu, \Sigma) &= \prod_{t=1}^T p(r_t|\Phi_t), \\ &\propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \sum_{t=1}^T (r_t - \mu)' \Sigma^{-1} (r_t - \mu)\right). \end{aligned} \quad (56)$$

This assumption allows for convenient analytical computation of the Bayesian parameter estimates. The resulting unconditional posterior means are used as replacements for the sample estimates of μ and Σ . This means that variants of the traditional portfolio optimisation models and proposed MW modifications can be defined that take parameter uncertainty into consideration implicitly. I investigate whether frequentist methods that are subject to estimation risk benefit from the Bayesian specifications. In particular the traditional allocations, MW, GMVP and BL are extended in order to assess their exposure to - and ability to account for - parameter uncertainty. A Bayesian extension of the optimal three-fund rule is also considered. In the next sections I conduct the Bayesian analysis for different prior specifications that have been motivated by literature.

2.4.1 Diffuse Prior

First, consider an uninformative prior. The diffuse specification makes no assumptions regarding parameter values, such that vague information or knowledge is provided regarding parameter estimation risks. The flat prior assumes that all values of μ and Σ are equally likely in order to avoid any subjective decisions regarding uncertainties. In this case, the data likelihood mostly determines the Bayesian parameter estimates. Define the (Jeffrey's) diffuse prior for μ and Σ as conventional,

$$p(\mu, \Sigma) \propto |\Sigma|^{-\frac{N+1}{2}}. \quad (57)$$

By applying Bayes rule (prior \times likelihood), the resulting predictive density, $p(r_{t+1}|\tilde{\mu}, \tilde{\Sigma}, \Phi_T)$, is shown to be students t distributed with $\tilde{\mu} = \hat{\mu}$, equal to the sample mean, and $\tilde{\Sigma} = \frac{T+1}{T-N}\hat{\Sigma}$, equal to the sample covariance multiplied by a scalar. The derivations are provided in Appendix D. Substituting the (diffuse) posterior parameters into the MW framework of (2) results in the optimal Bayesian portfolio weight. The resulting diffuse MW portfolio (MW diff) is defined as,

$$\hat{w}_t^{diff} = \frac{1}{\gamma} \left(\frac{T-N-2}{T+1} \right) \hat{\Sigma}^{-1} \hat{\mu}. \quad (58)$$

The Bayesian optimal portfolio under diffuse specification is very similar to that of the frequentist sample plug-in \hat{w}_t^M . In fact, in expectation,

$$E[\hat{w}_t^{diff}] = \frac{T}{T+1} \hat{w}_t^M. \quad (59)$$

The uninformative Bayesian prior reduces the weight allocations from the sample MW by a slight factor of $T/(T+1)$. This shows the MW diff strategy to be slightly more conservative than its sample counterpart as a result of taking parameter uncertainty into consideration. The degree of conservation (shrinkage) is determined by the relative size between T and N . A relatively larger N increases estimation errors due to relatively fewer observations and a greater number of assets. Consequently, the investor becomes more conservative and reduces the amount invested in the sample MW portfolio. As in section 2.1.2, I provide an analytical expression for the expected out-of-sample utility of \hat{w}_t^{diff} to compare the expected performance of the traditional and Bayesian method. The expected utility for MW diff is,

$$E[\tilde{U}(\hat{w}_t^{diff})] = k_4 \frac{\theta^2}{2\gamma} - \frac{NT(T-2)(T-N-2)}{2\gamma(T+1)^2(T-N-1)(T-N-4)}, \quad (60)$$

where it is assumed that $T > N + 4$, and

$$k_4 = \left(\frac{T}{T+1} \right) \left[2 - \frac{T(T-2)(T-N-2)}{(T+1)(T-N-1)(T-N-4)} \right]. \quad (61)$$

A comparison of expected out-of-sample performance expressions of the sample MW (17), the two-fund rule (40), three-fund rule (46) and MW diff equation (60) allows for judgements regarding which methods' weight estimates lie theoretically closer to the true MW portfolio and thus which should perform better. These are derived in Kan and Zhou (2007), for conciseness I summarise their findings: the MW diff theoretically outperforms the sample plug-in approaches, but is outperformed by the two- and three- fund rules. I argue however, that the two- and three- fund rules are still subject to estimation risk due to the frequentist approach and portfolios that are used. It is therefore of interest to investigate whether the findings of Kan and Zhou (2007) hold empirically.

2.4.2 Informative Priors

In the works of Klein and Bawa (1976) and Kan and Zhou (2007) the sub-optimal performance of the diffuse Bayesian method is pointed out. The uninformative priors do not add enough information for an improvement of portfolio performance over frequentist approaches. Furthermore, a non-informative prior may take on extreme (unrealistic) values that make no economic sense, affecting the reliability of the posterior. This motivates the search for a theoretically grounded, and useful informative prior. In this section informative specifications are investigated.

Conjugate Prior

A conjugate prior defines the same type of distribution for the posterior. Consequently, the posterior distribution is of a known class and can be derived analytically. As a result, informative priors have been common in Bayesian decision making. Following standard practice for stock returns, the normal and inverted Wishart distributions are taken as informative priors for μ and Σ , respectively. Mathematically,

$$\mu | \Sigma \sim \mathcal{N}(\mu_0, \frac{1}{\kappa_0} \Sigma), \quad (62)$$

$$\Sigma \sim \text{Inv-Wishart}_{\nu_0}(\Sigma_0), \quad (63)$$

with μ_0 as the prior mean, κ_0 a hyperparameter reflecting the prior precision on μ_0 , Σ_0 the prior variance and ν_0 a prior precision parameter for Σ_0 . As mentioned, the resulting posterior distribution of μ and Σ are of the same type as the conjugate priors. S.J. Brown (1976) show that the posterior mean of μ is a weighted average of the prior μ_0 and the sample means $\hat{\mu}$,

$$\tilde{\mu}^{Conj} = \frac{\kappa_0}{T + \kappa_0} \mu_0 + \frac{T}{T + \kappa_0} \hat{\mu}. \quad (64)$$

Similarly, the posterior mean of Σ is a weighted average of the prior variance, sample variance, and deviations of the prior mean to the sample means $(\mu_0 - \hat{\mu})$,

$$\tilde{\Sigma}^{Conj} = \frac{T+1}{T(\nu_0 + N - 1)} \left(\Sigma_0 + T\hat{\Sigma} + \frac{T\kappa_0}{T + \kappa_0} (\mu_0 - \hat{\mu})(\mu_0 - \hat{\mu})' \right). \quad (65)$$

The derivations are shown in Appendix E. The hyperparameters are defined as in Frost et al. (1986). The grand, or cross-sectional mean is taken for the prior mean, and the shrunk estimate of the covariance matrix of Ledoit and Wolf (2004) as defined in section 2.2.1 is used for Σ_0 . Furthermore, κ_0 is taken equal to T and ν_0 is taken equal to N , such that they are diffuse (Avramov et al., 2009).

The conjugate prior specification results in a similar form of shrinkage as the frequentist parameter shrinkage methods of section 2.2.1. $\hat{\mu}$ is shrunk to the average of the prior, and $\hat{\Sigma}$ is shrunk to prior variance and deviations of prior to sample mean. Differences with the frequentist methods are found in shrinkage targets and shrinkage intensities which are determined by the hyperparameters. Recall that the frequentist direct parameter shrinkage methods aim to minimise expected quadratic loss of the shrinkage estimator with respect to the parameter, whereas the Bayesian parameter shrinkage is characterised by a priori defined distributions that resemble the uncertainty in parameters. As a result, the Bayesian implementation is expected to better account for estimation errors than the frequentist model. Moreover, the shrinkage targets for $\hat{\Sigma}$ in the Bayesian method, the prior variance and the deviations of prior to sample mean, seem more adequate than the identity matrix used for direct parameter shrinkage. One would also expect the conjugate prior specifications to provide better parameter estimates than the diffuse case. As more information regarding the location, scale and shape of μ and Σ are incorporated, the estimates are expected to be more accurate when the informative priors are appropriate. The resulting conjugate MW portfolio, referred to as MW conj throughout, should account for estimation errors better and outperform the diffuse specification.

Jorion Hyperparameter Prior

Another commonly used prior for asset returns is that of Jorion (1986). A prior for the mean excess returns is defined based on the frequentist works that shrunk the expected returns parameter, μ , in section 2.2.1. Accordingly, the prior for μ is taken to be a (conjugate) normal distribution. The Jorion specification is similar to that the conjugate case but does not specify a prior on parameter Σ . This stems from the idea that estimation errors associated with μ are significantly greater.

Given prior mean η and precision parameter λ , the prior for μ is defined as

$$p(\mu|\Sigma, \lambda) \propto |\Sigma|^{-1} \exp\left[-\frac{1}{2}(\mu - \eta)'(\lambda\Sigma^{-1})((\mu - \eta))\right]. \quad (66)$$

Observe that the Jorion hyperparameter prior is a generalization of the diffuse case (for which $\lambda = 0$). The richer Jorion model is therefore expected to perform better. The derivation of the predictive density given the informative prior of (66) is similar to that of the conjugate case and can be found in Jorion (1986). The resulting predictive distribution is shown to be conditionally multivariate normal with mean,

$$E[r_{t+1}] = (1 - \tilde{a})\hat{\mu} + \tilde{a}'\hat{\mu}_0, \quad (67)$$

where

$$\tilde{a} = \frac{\lambda}{T + \lambda} \quad \text{and} \quad \hat{\mu}_0 = \frac{\iota'\hat{\Sigma}^{-1}\hat{\mu}}{\iota'\hat{\Sigma}^{-1}\iota}, \quad (68)$$

and covariance

$$V(r_{t+1}) = \hat{\Sigma} \left(1 + \frac{1}{T + \lambda}\right) + \frac{\lambda}{(T(T + 1 + \lambda))} \frac{\iota\iota'}{\iota'\hat{\Sigma}^{-1}\iota}. \quad (69)$$

The Jorion specification differs from the conjugate method by empirically estimating hyperparameter λ . For that, the approach of Jorion (1986) is adopted. A density is adopted for λ , denote that by $p(\lambda|\mu, \eta, \Sigma)$, and it is taken to be a gamma distribution with a mean of $(N+2)/d$. Constant d is defined as $d = (\mu - \iota\eta)'(\Sigma^{-1})(\mu - \iota\eta)$. The distribution is estimated using sample variants of the moments, and is then used to determine the optimal portfolio weights. Zellner and Chetty (1965) show the optimal trading strategy w to be approximated by,

$$\hat{w}_t^{Jor} = \frac{N+2}{(N+2) + (\hat{\mu} - \iota\hat{\mu}_0)'T\hat{\Sigma}^{-1}(\hat{\mu} - \iota\hat{\mu}_0)}. \quad (70)$$

Again the resulting parameter estimates can be given a shrinkage interpretation. Here $\hat{\mu}$ is shrunk towards $\hat{\mu}_0$ and $\hat{\Sigma}$ is shrunk towards $\frac{\iota\iota'}{\iota'\Sigma^{-1}\iota}$. Due to the Bayesian specification Jorion is expected to outperform the frequentist case as well. Notice that the sample mean is shrunk towards the average return of the GMVP portfolio. This differs from the conjugate specification where the mean was shrunk to the subjectively defined prior mean. It makes sense to shrink towards GMVP estimates as per definition this portfolio targets minimising the portfolio risk. Shrinking expected returns towards its average can therefore result in a safer and more stable weight allocation estimate than shrinking towards the grand mean. The MW model based on the Jorion hyperparameter is referred to as MW Jor throughout.

2.4.3 Hierarchical Bayes Models

In light of providing stable portfolio strategies that are robust to estimation error, it makes sense to consider all relevant parameter uncertainty. In this section the fixed hyperparameters of the conjugate specification are relaxed in attempt to minimise the subjectivity involved in prior definitions. Greyserman et al. (2006) provide a concise overview of the empirical implementations of Bayesian portfolio strategies, among which the hierarchical Bayes method. Following their works, I define a hyperparameter $\tilde{\mu}_0$ for μ as uniformly distributed, and a scale hyperparameter δ_0^2 for Σ as inverse gamma distributed. Adopting distributions for the hyperparameters allows for priors on return moments, μ and Σ , but avoids any subjective decisions regarding the location and precision of these parameters. Instead investors can express confidence in their beliefs/prior-specifications by changing the parameters of the hyperpriors. Assuming an uninformed investor, that has no confidence in their beliefs, the hyperpriors are taken as diffuse. The complete data model is formulated as,

$$\begin{aligned} p(r|\mu, \Sigma) &\sim \mathcal{N}(\mu, \Sigma), \\ p(\mu|\tilde{\mu}_0, \Sigma) &\sim \mathcal{N}(\tilde{\mu}_0, \Sigma/\kappa_0), \\ \Sigma &\sim \text{Inv-Wishart}_{\nu_0}(\delta_0^{-2}P_0^{-1}), \\ \tilde{\mu}_0 &\sim \text{Uniform}(-\infty, \infty), \\ \delta_0^2 &\sim \text{Inv-Gamma}(\epsilon_1, \epsilon_2). \end{aligned} \quad (71)$$

Here the values of ϵ_1, ϵ_2 are taken such that the hyperprior is diffuse, and are set to 0.0001. Similarly, the uniform distribution for $\tilde{\mu}_0$ is over the set of all real numbers. The degrees of freedom parameter $\nu_0 \geq N$ so that it is greater or equal to the number of assets and $\kappa_0 = 0.1N$, equalling ten percent of the sample size, in conform with previous literature (Greyserman

et al., 2006). The correlation matrix P_0 is,

$$P_0 = \begin{bmatrix} 1 & \rho_0 & \dots & \rho_0 \\ \rho_0 & 1 & \dots & \rho_0 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_0 & \rho_0 & \dots & 1 \end{bmatrix}, \quad (72)$$

where I assume a weak positive correlation of $\rho_0 = 0.5$ between assets. This is justified as I consider assets of the same industry in 48IndP and same value sorts in 100FF data sets.

The inclusion of the hyperparameter priors complicates the distribution of the predictive returns making it of unknown type. I therefore rely on Markov Chain Monte Carlo (MCMC) sampling to derive the posterior distributions of μ , Σ and r_{T+1} for the hierarchical Bayes model. A Gibbs sampler seems appropriate. For tractability the sampler and conditional distributions used for sampling are presented in Appendix F. A Markov Chain is constructed of which the limiting distribution approximates the posterior distribution. An iterative simulation procedure is used to sample from this Markov chain and once the chain has reached convergence, the subsequent draws are considered draws of the joint posterior distribution which can be used to compute posterior means, variances and marginal densities. The MW model based on the hierarchical Bayes specification is referred to as MW hier throughout.

The presented Bayesian estimation techniques are applicable to all three traditional mean-variance portfolio strategies. Just as in the frequentist case, one can use the posterior mean estimates for input parameters of the different models. Denote GMVP conj, and BL conj, as the conjugate specifications of the models, and GMVP hier, and BL hier, as the hierarchical versions. As the optimal three-fund rule suffers from estimation risk, a hierarchical version of the model is also considered. This extends previous efforts of optimal frequentist methods. The model is referred to as 3f hier, throughout. Investigation of performance improvements for frequentist models when Bayesian analysis is conducted is done in sections 5.3 and 5.4.

2.5 Robustification of Models

The MW modifications described in previous sections give rise to extensions that can further improve out-of-sample performance for mean-variance portfolios. The methods that target parameter uncertainty can be further developed to tackle the stability of MW weight allocations and the corresponding performance volatility over changing economic times. In this section a Bayesian method that accounts for the non-normality of returns is presented, and portfolio combinations of MW, GMVP and BL are introduced as means to improve portfolio robustness.

2.5.1 Accounting for Non-Normal Returns

There is much evidence in the literature suggesting that excess returns do not follow a normal distribution. Stylised facts as derived by papers as Ané and Geman (2000) and Fama (1965) have shown that returns tend to exhibit skewness and higher kurtosis than normal distributions, such that fatter tails and more extreme values are observed. In this paper I take the approach of Qian (2011) for a more robust Bayesian method that can take into account the non-normality of the data. In

particular, two modifications to the earlier suggested Bayesian models are considered. First, a Markov Switching regime is considered to account for common shocks on all assets. This allows for a more reactive model and is sensible to use when considering a sample that includes the global financial crisis. Second, a Gaussian mixture form of returns is implemented for the idiosyncratic risks on individual assets. The mixture of normals are better able to capture the fatter tails of returns. The combination of these two models is the Markov Switching Gaussian Mixture (MSGM) model for multivariate asset returns. The Markov process allows for relaxation of IID latent variables of Gaussian Mixture models. A Bayesian MCMC approach is opted for that enables modelling a number of regimes or states at relatively low computation costs. Furthermore, the Bayesian setting accounts for parameter uncertainty similar to section 2.4. In a sense, this model is a generalization of the Bayesian models of section 2.4, where only *one* normal distribution for the returns was considered.

First I introduce the parameters. As before, take asset returns as $r_t = (r_{1,t}, \dots, r_{N,t})'$ with $t = 1, \dots, T$, but now let these be driven by a latent Markov process with S regimes at time t . Denote the regimes by $\tau_t \in \{1, \dots, S\}$. Also define the collection of returns over time as $r_1^T = \{r_t\}_{t=1}^T$. Lastly, define a starting distribution for every regime S as $\pi = (\pi_1, \dots, \pi_S)'$ and a transition matrix,

$$Q = \begin{pmatrix} Q_1 \\ \vdots \\ Q_S \end{pmatrix} = \begin{pmatrix} Q_{1,1} & \dots & Q_{1,S} \\ \vdots & & \\ Q_{S,1} & \dots & Q_{S,S} \end{pmatrix}. \quad (73)$$

Conditional on the latent regime τ_t , r_t follows a Gaussian mixture with K latent states. The latent states $\lambda_t \in \{1, \dots, K\}$ follow a multinomial distribution with $\eta_S = (\eta_{S,1}, \dots, \eta_{S,K})'$ as probabilities for the corresponding regime. By taking the latent regime, τ_t , and latent state, λ_t , as known, the distribution of r_t can be defined. Take $\phi(\cdot)$ as the density of multivariate normal and $I(\cdot)$ as an indicator function for its input. The distribution of returns is,

$$P(r_t | \tau_t, \lambda_t) = \prod_{s=1}^S \prod_{k=1}^K [\phi(r_t; \mu_{s,k}, \Sigma_{s,k})]^{(I(\tau_t=s)) \cdot (I(\lambda_t=k))}. \quad (74)$$

Note how the model allows for different means and covariances of excess returns dependent on the regime and state at time t . This is key in improving model robustness for changing economic conditions. Next specify the priors of model parameters. These are taken to be conjugate as before, but per regime and state s, k ,

$$\begin{aligned} \mu_{s,k} &\sim \mathcal{N}(b_{s,k}, V_{s,k}), \\ \Sigma_{s,k}^{-1} &\sim \text{Wishart}_{\nu_{s,k}}(\Omega_{s,k}), \\ Q_s &\sim \text{Dirichlet}(a_{s,1}, \dots, a_{s,S}), \\ \pi &\sim \text{Dirichlet}(c_1, \dots, c_S), \\ \eta_s &\sim \text{Dirichlet}(f_{s,1}, \dots, f_{s,K}), \end{aligned} \quad (75)$$

where $s = 1, \dots, S$ and $k = 1, \dots, K$. The Dirichlet distribution is a conjugate distribution for categorical distributions, making it suitable for the choice of latent regimes, states and Markov probabilities. The full posterior conditionals for the parameters of interest and the corresponding sampler as used by Qian (2011) are reported in Appendix G.

In this investigation, I take three regimes for the Markov switching specification and three states to construct the Gaussian

mixture in order to generalize the Bayesian model based on normal returns. The three latent Markov regimes can capture common shocks to assets for periods of extremely high volatility (crisis), periods of ‘usual’ levels of volatility and periods of higher stability. Furthermore, the three latent Gaussian states should be able to capture idiosyncratic shocks better by better fitting the leptokurtic distribution of returns. With these extra modifications, the MSGM model is expected to provide a more robust strategy that can perform better over time.

The estimated predictive posterior distribution of excess asset returns is used to implement MW, GMVP and BL strategies that account for parameter uncertainty and non-normal returns. Define these as MW MSGM, GMVP MSGM and BL MSGM, respectively. This is possible as the traditional strategies do not depend on the normality assumptions in their derivation, and can therefore accommodate other return distributions. Moreover, as the frequentist three-fund rule also suffers from estimation risk, a Bayesian variant of the model is also explored to determine whether performance of Kan and Zhou (2007) frequentist methods can be augmented through Bayesian analysis. This model will be referred to as 3f MSGM throughout.

2.5.2 Robustification via Bayesian Portfolio Combinations

The previous section demonstrated the use of Bayesian analysis in accounting for embedded estimation risks. The ‘safer’/more accurate moments based on the posterior means of μ and Σ were used to derive portfolio weights that account for parameter uncertainty. In this section I prove the use of carrying on the parameters uncertainty into weight allocations, so that more uncertainty information than just the first moments are used to derive portfolio allocations. I use these weight distributions to construct Bayesian portfolio combinations, extending the frequentist portfolio combination rules of section 2.3 to a Bayesian setting that includes the BL.

First, I point out the relevance of carrying on uncertainty into the portfolio weights. Even though the same prior distributions are used for deriving portfolio allocations, the resulting distribution of the weights are not of the same class. Recall how the GMVP portfolio does not make use of the parameter μ and thus different distributions are expected⁹. Similarly, BL significantly manipulates the uncertainty from its input parameters when taking into account investor views. Consequently, the uncertainties/distributions for the weights of each strategy will be different. This needs to be considered when constructing robust portfolio combinations.

The method is illustrated by means of the MCMC output for the hierarchical specifications. For each of the three traditional allocations, MW, GMVP and BL have an approximated ‘distribution’ for their respective input parameters made up of the Markov Chain draws. More specifically, the MW and BL portfolios have a set of M simulations for the posterior mean and corresponding M simulations of the posterior covariance. Similarly, GMVP had M simulations for the covariance. These posterior simulations can be used to construct a distribution for the portfolio weights given the specifications of section 2.1.1. A portfolio combination strategy can be derived by taking these approximated weight distributions together.

It is desirable to find a combination between MW, GMVP and BL as the three models are at the forefront of portfolio optimisation, and a combination of these more popular methods would seem a step in the right direction towards finding a universally applicable trading rule. Finding a suitable combination of these weight allocations remains unexplored. Given

⁹Bodnar et al. (2017) show the distribution of the GMVP weights to be t-distributed for conjugate and diffuse priors. The MW weights would thus be a t-distribution multiplied by the normal posterior for μ .

the approximated weight distributions of each strategy, I propose 3 combination methods based on the certainty equivalence (CEQ) of the portfolio weight draws: the equally weighted (U EWC), the maximum CEQ (U MWC), and weighted CEQ (U WC). The combinations aim to suitably mix the simulated weights of each strategy. As investors aim to maximise utility, CEQ seems an adequate measure to define the model mixtures. For the equally weighted, U EWC, one can simply take the simulated weights of MW, GMVP and BL together. The aggregated weight simulations (of $3M$ draws) then mimic a distribution for an equal combination of the three strategies. Taking the posterior mean of this sample of weight draws provides the equally weighted combination strategy. For the maximum CEQ combination, U MWC, I rely on strategy utility of the draw at $m = 1, \dots, M$. The model with highest mean-variance utility is adopted as draw of the approximated weight distribution. The posterior mean of this sample of weight draws (of M draws) provides the U MWC allocation. Finally, I consider a combination weighted by draw utilities. This requires careful definition as utilities can be negative in this setting. For fair comparison, I include a combination over negative utility strategies that, as for aforementioned strategies, also makes use of an investor that invests wealth at every time period. In the case that all strategies report a negative utility for draw m , I choose the least negative strategy for that simulation. This would lead to highest attainable utility if the investor must invest. In the case of only one positive utility I choose that strategy. For the case of 2 or more positive utilities I choose a weight of the strategy weighted over the total sum of all positive strategy utilities. The posterior is taken over the collected weight draws to define U WC.

The specifications of the combination strategies allow for weight shifting towards the traditional optimisation portfolio that is most attractive to the investor at that time period. Assuming that utility provides a good indication for the optimal strategy as done in classical portfolio optimisation, one would expect combinations to provide more robust portfolios. Next to diversification, the inclusion of the Bayesian BL should also counter the usual high weights of the MW and GMVP allocations. The combination strategies based on utility are implemented by the hierarchical Bayes and MSGM models. U EWC hier, U MWC hier and U WC hier represent the portfolios derived off of the hierarchical specification and U EWC MSGM, U MWC MSGM and U WC MSGM are the portfolios derived off of the MSGM sampler.

3 Out-of-Sample Performance Evaluation

3.1 Initialisation and Base-Case Investor

For tractability an overview of the investor characteristics is provided. Similar specifications to that of DeMiguel et al. (2015) are chosen for compatibility. The investor takes on a dynamic portfolio allocation approach by using a rolling window of size $h = 500$ past observations to determine his portfolio at time t . More specifically, the data available from $t - h$ up to t to construct portfolios. The investor risk aversion parameter is taken as $\gamma = 1$. In literature, there seems to be no agreed upon value for the parameter for risk aversion but values tend to lie between 1 and 10 for mean-variance investors (Ang, 2014). In section 2.1 I show that the estimation error effects are greater for smaller risk aversion. To ensure that the effects of estimation error are apparent, it thus makes sense to take γ as small as possible such that the investor is still risk-averse. Accordingly, conclusions on model abilities to mitigate impacts of parameter uncertainty become clearer. Assuming a risk aversion parameter of $\gamma = 1$ is therefore reasonable.

For reproducibility, some supplementary definitions regarding the BL model are also in order. In particular, I provide specifications for the parameter of expected views q , pick matrix P , the confidence in views Ω and the scaling parameter τ .

As the derivation of investor views are not central to this investigation, similar definitions of the views as in Schepel (2019) are opted for. The pick matrix P is taken to be an Identity matrix such that there is an absolute view for every asset, $Z = N$. The expected views are set equal to the estimated returns $\hat{\mu}$ based on the Fama-French 3-factor model of section 2.2.3. For the confidence in views, He and Litterman (2005) define a diagonal Ω matrix such that views of different assets are independent. Following their works, I set $\Omega = \tau \cdot \text{diag}(P\Sigma P')$, where Σ is taken to be the Fama-French 3-factor model covariance estimate. Schepel (2019) argues that “there is no common convention on the value of τ , although most authors agree that $t \ll 1$ ”. Accordingly, τ is taken equal to $1/h$.

3.2 Evaluation Criteria and Performance Measures

In order to provide an extensive overview of the impact of parameter uncertainty and abilities to suppress it, a number of evaluation criteria are regarded. The performance measures are chosen such that they are in accordance with previous literature and enable evaluations regarding stability and robustness of portfolios. The chosen measures are commonly used in portfolio management and allow for clear conclusions to be drawn regarding model out-of-sample performance. Define T as the total amount of observations and h as the sub-sample of observations in the moving window.

Out-of-Sample Mean Returns for Strategy j

The out-of-sample mean returns are the arithmetic mean of the reported out-of-sample returns for a strategy j ,

$$\hat{\mu}_j = \frac{1}{T-h} \sum_{t=h+1}^T r_{t,j}, \quad (76)$$

where $r_{t,j}$ is the portfolio return of strategy j , at time t and is defined as the weight allocation of time t multiplied by the returns in the next period $t+1$: $r_{t,j} = (w_t^j)'r_{t+1}$.

Out-of-Sample Standard Deviation of Returns for Strategy j

The out-of-sample standard deviation of returns quantifies the degree of dispersion of the out-of-sample returns for every strategy j ,

$$\hat{\sigma}_j = \sqrt{\frac{1}{T-h-1} \sum_{t=h+1}^T [r_{t,j} - \hat{\mu}_j]^2}. \quad (77)$$

The standard deviation is considered a measure for the risk of the portfolio. A higher standard deviation indicates more volatile portfolio returns.

Annualised Out-of-Sample Sharpe Ratio of Strategy j

The annualised out-of-sample Sharpe ratio is a measure for the risk-adjusted returns over the out-of-sample period. Define it as,

$$SR_j = \frac{\hat{\mu}_j}{\hat{\sigma}_j} \times \sqrt{252}. \quad (78)$$

As it accounts for both risk and return of the portfolio, it allows for a more complete evaluation of strategy performance. The Sharpe Ratio is annualised by multiplying by square root of the average number of trading days in a year, 252. To test

whether the out-of-sample SRs of two strategies are significantly different I rely on methodology of Ledoit and Wolf (2008) (Remark 2.1). A stationary bootstrap of Politis and Romano (1994) with $B = 1000$ bootstrap samples and block size $b = 1$ is adopted to compute the bootstrap p-values¹⁰.

Certainty Equivalent for Strategy j

The certainty equivalence (CEQ) is defined as the risk-free rate that a mean-variance investor would be willing to receive such that he is indifferent between the risk-free rate and the rule. It can also be interpreted as a measure of investor utility. CEQ is given by,

$$CEQ_j = \hat{\mu}_j - \frac{\gamma}{2} \hat{\sigma}_j. \quad (79)$$

Turnover for Strategy j

The turnover provides a measure for the amount of assets that are re-balanced at each trading period. It is used as a measure of portfolio stability, as more stable portfolios are expected to have lower turnovers. Mathematically the turnover is defined as,

$$\text{turnover}_j = \frac{1}{T - N} \sum_{t=1}^{T-h} \sum_{l=1}^N (|\hat{w}_{j,l,t+1} - \hat{w}_{j,l,t}|). \quad (80)$$

Here $\hat{w}_{j,l,t}$ is the portfolio weight for strategy j , in asset l , at time period t . $\hat{w}_{j,l,t+}$ is the portfolio weight before rebalancing at time period $t + 1$ and $\hat{w}_{j,l,t+1}$ is portfolio weight at time period $t + 1$ after rebalancing.

Mean Weight Invested in Risky Assets for Strategy j

The mean weight invested in risky assets (MWR) is used to gauge the amount of wealth that is invested in risky assets. Define it as,

$$MWR_k = \frac{1}{T - N} \sum_{t=1}^{T-h} \sum_{l=1}^N \hat{w}_{j,l,t}. \quad (81)$$

Observing how much wealth the investor allocates to risky assets can aid in determining robustness of portfolios.

4 Data

4.1 Descriptive Statistics

To be consistent with previous literature, I will be using the same data as DeMiguel et al. (2015). As this research focuses primarily on portfolio performance in practice, their empirical data sets are considered. These contain a large number of assets such that parameter uncertainty is substantial. I focus on the 48 Industry portfolios (48IndP) and the Fama-French 100 portfolios (100FF) formed on size and book-to-market ratio. Daily frequency trading data from July 7, 2004, until September 19, 2012 is used to construct daily returns¹¹. The analysis is performed on daily returns to ensure feasibility of parameter estimation, in particular for the covariance matrix Σ that has high dimensions in both data sets. I note that due to high

¹⁰Matlab code available on the personal website of Ledoit: http://www.ledoit.net/jef2008_abstract.htm.

¹¹This data is available on the Kenneth French Website, https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

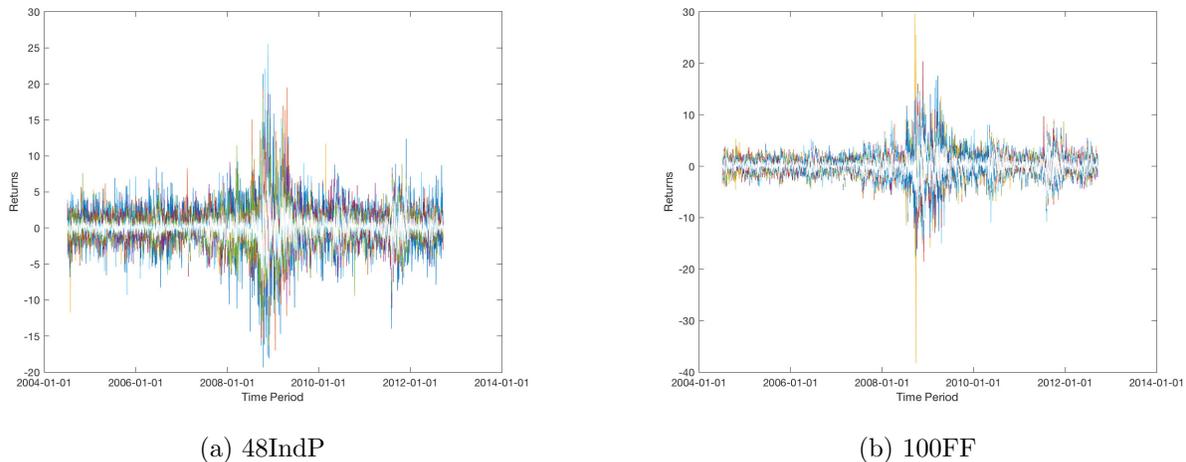
turnovers, trading at daily frequency may not be feasible for the institutional investor, but the higher frequency data can help capture further performance insights of the considered portfolio strategies. For instance, more detailed conclusions regarding model performance stability over time can be drawn. To further investigate this matter, the data set is chosen to include the financial crisis of 2007-2008. Performance of the considered strategies during this time can provide insights on whether the modifications enhance model robustness. The analysis is carried out on returns of assets/portfolios as these are more likely to suffer from parameter uncertainty than the more diversified indices. Furthermore, it is of interest to apply the methodology to data sets with high-dimensions, or many assets such that the allocations are implementable in practice.

Table 2: Descriptive Statistics

	Mean Return (%)	Mean (%) 2004-2006	Mean (%) 2007-2009	Mean (%) 2010-2012	Min (%)	Max (%)	Average Skewness	Average Kurtosis	Median Kurtosis
48IndP	0.042	0.057	0.014	0.058	-19.340	25.560	0.036	10.100	9.477
100FF	-0.202	-0.137	-0.496	0.059	-99.990	29.580	-0.128	9.299	8.045

Note: This table presents descriptive statistics of daily returns of all assets in the 48 Industry portfolios (48IndP) and the Fama-French 100 portfolios (100FF) data sets over a time frame of July 7, 2004, until September 19, 2012. The Mean return is reported over the whole sample as well as over subperiods of the data sets: pre-crisis (2004-2006), crisis (2007-2009) and post-crisis (2010-2012). Min and Max denote the minimum and maximum returns, respectively.

Figure 1: Time Series Plot of Daily Asset Returns

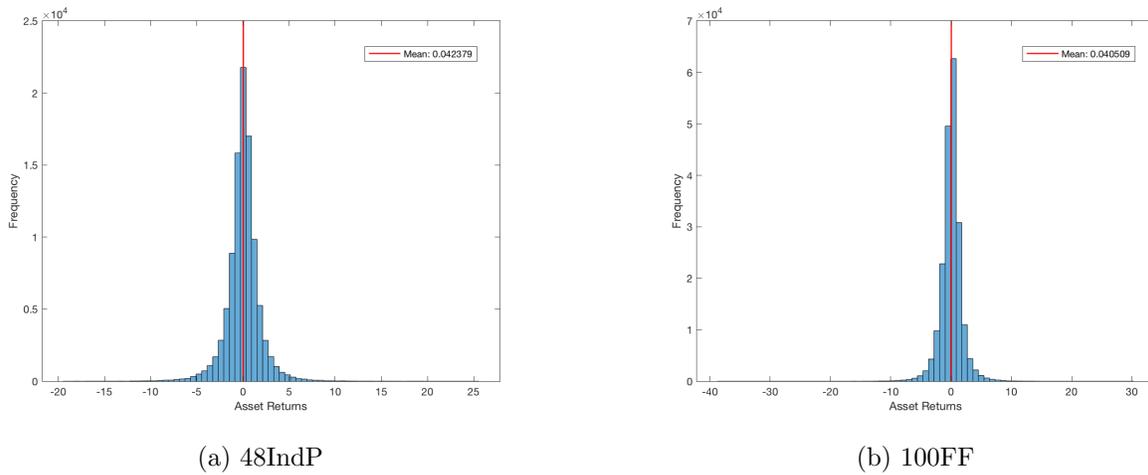


Note: This figure shows the time series of daily returns for all assets in the 48 Industry portfolios (48IndP) and the Fama-French 100 portfolios (100FF) data sets, respectively. The average daily asset return is reported by a white line over the data time frame, July 7, 2004, until September 19, 2012.

Table 2 reports descriptive statistics for the considered data sets. Over the whole sample period a mean daily return of 0.042 % and -0.202 % is reported for the 48IndP and 100FF data samples, respectively. In line with the scope of this research, the means for subperiods before, during and after the global financial crisis are reported. A robust portfolio should withstand a stable performance over the whole set. The period of 2007-2009 is taken as Crisis period where the effects of the global financial crisis were most prominent in accordance with the National Bureau of Economic Research¹². The mean returns,

¹²<https://www.nber.org/cycles.html>.

Figure 2: Histogram of Daily Asset Returns



Note: This figure shows histograms of daily returns for all assets in the 48 Industry portfolios (48IndP) and the Fama-French 100 portfolios (100FF) data sets, respectively. The average (mean) daily asset return is reported. The time frame is from July 7, 2004, until September 19, 2012.

0.014% and -0.496%, are substantially lower in this period than the pre- and post- crisis periods, for assets in 48IndP and 100FF, respectively. Moreover, as is expected and can be seen from Figure 1, the crisis years are characterised by a period of increased volatility. The maximum and minimum returns values for both samples are found within this time frame. Figure 2 displays a histograms of all daily returns of the data sets. Although the distributions resemble that of a normal distribution, the high mean and median kurtosis values, 10.10 and 9.48 for 48IndP returns, and 9.30 and 8.05 for 100FF returns, indicate the possibility of the daily returns following a distribution with fatter tails. This remains a point of attention throughout the investigation.

4.2 In-Sample Results for Factor Models

In this section the fits for the Fama-French 3-factor, and the 95% explained variance PCA models of section 2.2.3 are investigated. A multivariate regression of the daily excess returns on factors is carried out for each rolling window sample of 500 observations over the time frame July 7, 2004, until September 19, 2012. The average R^2 is taken as a measure for the factor model fit. Average R^2 for pre-, during and post-crisis subperiods are reported for both 48IndP and 100FF data sets.

The Fama-French 3-factor model has a relatively low $R^2 = 0.643$ for the 48IndP data set over the full sample. The worsened fit over the pre-crisis period, equal to 0.499 seems to be the main reason for this. Over other subsets better fits are reported. For the 100FF data set, the model fit seems better, $R^2 = 0.868$, but also suffers from a lesser pre-crisis period. The rather low fits for both data sets indicate that the FF3 parameter approximations may not be able to produce significant reductions in parameter uncertainty as it arguably omits part of the variation in the data. For completeness and due to its popularity in literature, the FF3 is kept in the analysis.

The 95% of total variance explained PCA model performs a lot better reporting R^2 of 0.971 and 0.955, for the 48IndP and 100FF data sets, respectively. Furthermore, the fit seems stable over time as shown by all subperiods reporting similar R^2 .

Table 3: In-sample Performance of Factor Models

	Full Sample \bar{R}^2	Pre Crisis \bar{R}^2 (2004-2006)	Crisis \bar{R}^2 (2007-2009)	Post Crisis \bar{R}^2 (2010-2012)
FF3 48IndP	0.643	0.499	0.732	0.771
PCA 48IndP	0.971	0.976	0.968	0.965
FF3 100FF	0.868	0.836	0.882	0.925
PCA 100FF	0.955	0.954	0.956	0.952

Note: This table presents the in-sample fits of the Fama-French 3-factor (FF3) model and the 95% of total variance explained Principal Component Analysis (PCA) for the 48 Industry portfolios (48IndP) and the Fama-French 100 portfolios (100FF) data sets. The model fits are presented in the form of average R^2 for a rolling window with 500 observations over the time period July 7, 2004, until September 19, 2012. Subperiods defined by pre-, during- and post- crisis are reported as well.

This suggests that the parameters approximations derived from this model could provide improved and more robust estimates for μ and Σ than the sample variants.

5 Results

Presented in the following subsections are the out-of-sample performances of the MW modification models on empirical data. First, the mitigation effects of direct parameter shrinkage are evaluated, followed by the portfolio weight shrinkage and Bayesian methods. Subsequently, the out-of-sample performances for the defined combination strategies are reported. The best performing strategies that prove most applicable in practice are chosen as 'winner strategies' and are used to make concluding remarks regarding the effectiveness of parameter uncertainty suppression to produce robust portfolios in section 5.5. The discussion focuses on the out-of-sample results for the 48IndP data set for conciseness and clarity of the report. The 48IndP set proves representative for the performance of the portfolio allocation methods as similar conclusions are drawn for the methods applied on 100FF data. The model out-of-sample performances for 100FF are reported in Appendix H. The 100FF data set is used to further assess robustness of portfolio strategies in section 5.5.

5.1 Direct Shrinkage of Parameters

In accordance with the well-known drawbacks of the sample MW portfolio, Table 4 reports a relatively high turnover for the mean-variance strategy, 11.236. Compared to the turnover of the EW, 4.588, which is by definition the most stable portfolio, I point out the high volatility of portfolio allocations when sample estimates are used for MW. By also reporting a significantly lower out-of-sample Sharpe ratio than the EW, 0.243 against 0.395, the incompetence of using the MW portfolio in practice is reinforced. Note also that on average MW invests only about 10% of total wealth in risky assets. As a risk-aversion parameter of 1 is taken, the low weights for risky assets indicate high covariance and/or low expected excess return estimates. I evaluate whether direct shrinking of the sample estimates allows for improved performance.

The input parameter shrinkage methods for μ of Jorion (1986) and Σ of Ledoit and Wolf (2004) provide minimal improvement. MW risk-adjusted returns do not improve significantly when shrinkage estimators for μ and/or Σ are used as plug-ins, and no significant differences are made in CEQ or turnovers. The poor performance implies that reducing estimation error through parameter shrinkage is unable to offset the increase in parameter bias. As this rather direct method does not address the issue of the model sensitivity to inputs, the impact of parameter uncertainty is not suppressed sufficiently. For

Table 4: Out-of-Sample Performance of Direct Parameter Shrinkage Models on 48 Industry Portfolios

	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
MW	0.007	0.429	0.243	-0.049	11.236	0.104
MW μ_{sh}	0.006	0.403	0.226	-0.041	10.449	0.109
MW Σ_{sh}	0.006	0.408	0.247	-0.041	10.374	0.102
MW μ_{sh}, Σ_{sh}	0.006	0.384	0.231	-0.034	9.649	0.107
MW 3FF	0.006	0.436	0.208	-0.051	11.489	0.109
MW PCA	0.014	0.531	0.419	-0.087	10.059	0.170
GMVP	0.028	0.813	0.546	-0.302	16.715	1.000
BL	0.006	0.436	0.208	-0.051	11.338	0.104
MW-BL	0.006	0.436	0.208	-0.051	11.469	0.109
EW	0.040	1.609	0.395	-1.000	4.558	0.980

Note: This table presents the performance measures for the parameter shrinking methods of section 2.2.1. The presented models regard the Markowitz portfolio (MW) with improved input parameters. MW μ_{sh} uses a shrunk estimate of expected returns, MW Σ_{sh} uses a shrunk estimate for covariance and MW μ_{sh}, Σ_{sh} uses both shrunk estimates. MW 3FF and MW PCA use input parameters based on estimates of the Fama-French 3-factor and PCA factor models, respectively. Furthermore, alternative portfolio optimisation models are considered. Provided are the Global Minimum Variance Portfolio (GMVP), Black-Litterman (BL), and a shrinkage variant of the Black-Litterman (MW-BL). The equally weighted (EW) portfolio is also included as benchmark. Evaluation criteria are calculated for each strategy on daily returns of the 48 Industry portfolios data set over time frame June 27, 2006, until September 19, 2012. The reported performance measures for each strategy are: mean return, standard deviation, Sharpe ratio (SR), certainty equivalence (CEQ), turnover and Mean Weight invested in Risky assets (MWR). P-values for difference in Sharpe ratios with respect to Markowitz (MW) are derived from methods of Ledoit and Wolf (2008). *, **, *** represent 10%, 5% and 1% significance levels, respectively.

the factor models' approximation of input parameters, more promising results are reported. The MW PCA model retrieves a high SR of 0.419 such that principal components that explain 95% of the variance prove able to summarize the bulk of return information effectively. By disregarding possible outliers, better estimates for input parameters of MW are provided. MW PCA is considered a 'winner' strategy and further investigated in section 5.5. The 3FF on the other hand performs worse than sample MW. Considering the low R^2 that Fama-French 3-factor model reported for in-sample results in section 4.2, this is not surprising.

The GMVP minimises portfolio risk by construction, this can be seen from its relatively low standard deviation of returns 0.813, which are approximately half that of the EW strategy. As a result it obtains the highest Sharpe Ratio, 0.546. Note also that, by construction, the GMVP invests all wealth into risky assets. Considering that it invests approximately ten times more wealth than sample MW, and reports a turnover that is only slightly higher than the sample MW, 16.715 compared to 11.23, shows the GMVP weight allocations to be more stable. This can be explained by the model disregarding noisy estimates for the expected returns parameter μ , and highlights the significance of estimation error in optimal portfolio derivation.

The sample BL model has similar performance to sample MW. With a low SR of 0.208 I note that the model is not able to outperform its substitutes, MW and GMVP. The alternative specification, MW-BL, does not provide for much improvement over MW either. Reporting similar evaluation criteria to the traditional BL model, I conclude that investor views of the 3FF model do not shrink MW parameter estimates effectively. Other investor views could however potentially allow for better shrinkage as the 3FF model proved a poor fit for returns, see section 4.2. Furthermore, I point out the similarity in performance between the MW 3FF and MW-BL model. This is an intriguing finding regarding the different implementations of the 3FF model. Exploring the dynamics of how these two models relate theoretically can be considered in further research.

On a final note, in conform with DeMiguel et al. (2009), I stress the difficulty in beating the equally weighted strategy. The EW outperforms the classical MW portfolio implementation regardless of the considered shrunk input parameters. The naive strategy proves hard to beat and thus serves as an adequate benchmark. This section confirms the underlying issues of using the MW portfolio in practice. It is beat by the EW due to instability, and sensitivity to input parameters. Furthermore, direct parameter shrinkage offers little improvements over the traditional MW implementation. I conclude by posing that only the MW PCA and GMVP prove feasible alternatives and proceed to investigate effectiveness of other solutions.

5.2 Portfolio Weight Shrinkage

Table 5: Out-of-Sample Performance of Weight Shrinkage Models on 48 Industry Portfolios

	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
2f	0.002	0.170	0.190	-0.016	3.363	-0.027
3f	0.018	0.429	0.682*	-0.065	9.935	0.498
3f EW	0.003	0.184	0.273	-0.003	4.327	0.062
3f EW x GMVP	0.000	0.025	0.153	0.001	0.479	0.028
EW	0.040	1.609	0.395	-1.000	4.558	0.980

Note: This table presents the performance measures for the weight shrinking methods explained in Section 2.3. 2f and 3f are the two- and three- fund rules of Kan and Zhou (2007). 3f EW is an optimal three-fund combination between the Markowitz portfolio and the equally weighted, and 3f EW x GMVP represents the optimal three-fund combination between the equally weighted and Global Minimum Variance portfolios. The equally weighted (EW) portfolio is also included as benchmark. The evaluation criteria are calculated for each strategy on daily returns of the 48 Industry portfolios data set over time frame June 27, 2006, until September 19, 2012. The reported performance measures for each strategy are: mean return, standard deviation, Sharpe ratio (SR), certainty equivalence (CEQ), turnover and Mean Weight invested in Risky assets (MWR). P-values for difference in Sharpe ratios with respect to Markowitz (MW) are derived from methods of Ledoit and Wolf (2008). *, **, *** represent 10%, 5% and 1% significance levels, respectively.

Kan and Zhou (2007) demonstrate the use of shrinking the derived MW portfolio weights towards alternative target portfolios. In their paper, the two-fund portfolio, built on the premises of being able to decompose the MW portfolio into the tangency portfolio and the risk free asset, and the three-fund rule, where the MW portfolio is shrunk towards the GMVP are tested on simulated normal and t-distributed returns. Reporting promising results, I apply their methods to empirical data of the 48IndP and the 100FF data set. Furthermore the ability of further suppressing parameter uncertainty, by constructing other model combinations that are less subject to estimation errors, is investigated. The findings are reported in Table 5.

The 2f rule shrinks the MW weights towards zero as can be seen from the lower MWR value. The lower amount of wealth invested in risky assets is paired with lower returns and risk. The resulting SR is lower than the sample MW benchmark, deeming the 2f rule less suited for practical purposes. Reducing the amount (optimally) invested in risky assets does not lead to improved MW performance. The 3f rule however, proves a lot more applicable in practice. With a much greater SR than MW, 0.682 against 0.243, and a lower turnover, 9.935 against 11.236, the method is rather successful in robustifying the MW model. Note also that this strategy invests more into risky assets than the MW by combination with the GMVP. On average 50% of wealth is invested in risky assets, allowing for much higher returns. This motivates the search for other suitable target portfolios.

I examine shrinking towards the EW and a combination of the GMVP and EW portfolios. Combination with the naive portfolio is attractive as this portfolio does not suffer from estimation risk. This is backed by the substantially lower standard deviation in returns reported for the 3f EW strategy, 0.184. However, as the EW does not make use of risk return information,

its allocation proves suboptimal as shown by lower returns of 0.003. The overall risk-adjusted return provides a slight improvement over the sample MW, 0.273 against 0.243. The 3f EW x GMVP is less successful in improving out-of-sample performance. The combination of GMVP and EW reports a lower risk, with an out-of-sample standard deviation equal to 0.025, but also much lower returns. The low SR of 0.153, shows that this combination is less attractive for investors seeking large returns, but may be of interest for a more conservative and risk-averse investor. Causes of this could be found in matters of oversimplification when considering GMVP and EW. This way risk-return patterns are not exploited sufficiently, resulting in sub optimal performance.

The 3f rule with GMVP is considered a ‘winner strategy’ or candidate improvement for traditional MW and will be further investigated in section 5.5. I conclude that including the EW in portfolio combinations that are subject to estimation errors can aid in reducing portfolio risk, but not necessarily performance.

5.3 Accounting for Parameter Uncertainty

Table 6: Out-of-Sample Performance of Bayesian Models on 48 Industry Portfolios

	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
MW diff	0.006	0.385	0.243	-0.037	10.093	0.094
MW conj	0.002	0.043	0.572**	0.002	1.048	0.026
MW Jor	0.004	0.189	0.306*	-0.004	4.614	0.094
MW hier	0.007	0.395	0.279**	-0.038	9.910	0.100
MW MSGM	0.029	0.345	1.317**	-0.029	17.063	0.104
GMVP conj	0.029	0.815	0.562	-0.302	16.678	1.000
GMVP hier	0.099	2.568	0.615	-5.647	43.668	3.466
GMVP MSGM	0.054	0.775	1.112	-0.226	18.607	1.000
BL conj	0.007	0.346	0.338	-0.027	8.189	0.115
BL hier	0.006	0.436	0.205	-0.051	11.499	0.109
BL MSGM	0.033	0.392	1.346**	-0.055	17.004	0.132
EW	0.040	1.609	0.395	-1.000	4.558	0.980

Note: This table presents the performance measures for the Bayesian models explained in section 2.4. Different priors are adopted for the optimal portfolios, Markowitz (MW), Global Minimum Variance Portfolio (GMVP) and the Black-Litterman (BL). The prior specifications are denoted by diffuse (diff), conjugate (conj), Jorion (Jor) and hierarchical (hier). Furthermore, MSGM refers to the Markov Switching Gaussian Mixture specification of the model. The equally weighted (EW) portfolio is included as a benchmark. The evaluation criteria are calculated for each strategy on daily returns of the 48 Industry portfolios data set over time frame June 27, 2006, until September 19, 2012. The reported performance measures for each strategy are: mean return, standard deviation, Sharpe ratio (SR), certainty equivalence (CEQ), turnover and Mean Weight invested in Risky assets (MWR). P-values for difference in Sharpe ratios with respect to Markowitz (MW) are derived from methods of Ledoit and Wolf (2008). *, **, *** represent 10%, 5% and 1% significance levels, respectively.

The ability of improving MW performance by accounting for parameter uncertainty can be deduced off of the results of Table 6. As pointed out theoretically in section 2.4.1, the diffuse prior MW is unable to beat the classical MW out of sample. The non-informative specification of the prior, inherent to a more conservative investor who has little to no knowledge about the model parameters, reports similar performance and out-of-sample SR as the sample MW model, 0.243. The conjugate specification, which takes on a normal distribution for the expected excess returns μ , and Inverse-Wishart distribution for the covariance of excess returns Σ , shows the use of including extra parameter information. Significant improvements over the traditional implementation are made. The conjugate specification accounts for parameter uncertainty by shrinking the sample mean towards the grand mean, and the sample covariance towards the prior covariance estimate (of Ledoit and Wolf (2004)) and deviations of the prior to sample mean. The improved estimators allow for a much safer strategy than allocations

based solely on sample counterparts. The much lower standard deviation of returns contributes to a substantially higher SR of 0.572.

Though accounting for a great part of the parameter uncertainty, the conjugate portfolio still suffers from uncertainty and subjectivity in definitions of the hyperparameters. Defining extra information on these hyperparameters in the form of distributions (hyperpriors) can further reduce exposure to parameter risk. The Jorion hyperparameter model specifies a hyperparameter for the mean excess returns such that the sample estimate for μ is shrunk to the grand mean of GMVP portfolios. Given the superior out-of-sample performance of GMVP to MW, as shown in section 5.1, it makes sense that shrinking the sample estimate of the μ to the average of the GMVP portfolios can lead to enhanced performance. No prior is taken for Σ . The MW Jor improves performance over sample MW reporting higher risk-adjusted returns of 0.306, but is outperformed by the conjugate specification. This can be considered evidence for the presence of significant estimation risk for parameter Σ . Specifying priors for both parameters such that uncertainties of both parameters are considered therefore proves desirable. Due to its abundant use in portfolio management, MW Jor is included as benchmark for the 'winner' strategies in section 5.5.

The hierarchical specification extends the conjugate specification by including priors for the hyperparameters. It takes on an inverse-gamma and uniform distribution for the hyperparameters to account for the estimation errors involved at a deeper level. The MW SR is improved from 0.243 to 0.395 for the hierarchical specification. Noteworthy is that the hierarchical set up is outperformed by the less informative conjugate model. The reason for this could be that the chosen hyperparameter values for the conjugate model, the grand mean for μ and the covariance approximation of Ledoit and Wolf (2004) for Σ , lie very close to the true values. The diffuse hyperparameter priors, that imply little confidence in the defined priors, can take on hyperparameter values that lie further away from the true values worsening estimation. Consequently, the user specified values work better. However, as the hierarchical prior eliminates user subjectivity in hyperparameter determination, the hierarchical approach is deemed more applicable in practice and is thus considered as (the better) MW modification candidate in section 5.5.

Finally, consider the MSGM model that accounts for parameter uncertainty *and* non-normal returns. With a very high mean return of 0.026, and risk-adjusted return of 1.165, the method accounting for different regimes and fatter tails proves the clear winner for the mean-variance investor. The model is able to capture changes in volatility levels as well as shifts in the mean expected excess returns over time. This added flexibility, especially over a sample period involving a time of crisis, proves particularly effective in boosting MW performance. The increased returns however seem to come at cost of a higher turnover and possibly stability. Robustness analysis is therefore in order. I conclude that the MW portfolio benefits significantly from implementation of a (hierarchical) Bayesian structure and from considering non-normal returns. The Robustness of these models are investigated further in section 5.5. Next, some comments are in order referring to the alternative strategies. First consider the GMVP. Clearly the investment strategy also benefits from the Bayesian specifications. Both the conjugate and the hierarchical specifications outperform the sample GMVP portfolio with SRs of, 0.562 and 0.612 respectively, compared to sample SR of 0.546. The hierarchical GMVP is slightly safer with lower turnover and standard deviation of returns. The magnitude of the SR improvement from sample MW to MW hier compared to the improvement from sample GMVP to GMVP hier, provides evidence for the magnitude of estimation risk involved in estimating μ . The out-of-sample performance improvements are much greater for MW than for GMVP. In line with conclusions from previous literature such as Chopra et al. (1993), this shows μ to indeed face sizeable uncertainty and hence reinforces the reason for which it is referred to as a

'noisy' parameter.

Furthermore, the risks for the GMVP portfolios are significantly higher than the MW models. This can be noted from the much higher standard deviations of returns. The reason for this lies within the definition of the GMVP, which invests all wealth into risky portfolios. Consequently, returns are higher but the volatility of these returns is higher too. The constructed Bayesian weight combinations evaluated in next sections aim to find a suitable balance/combination between these two optimal portfolios. For the BL alternative, similar results to the MW are reported. The inclusion of views however does not lead to substantial improvements over the traditional MW implementation. The BL obtains highest SR when incorporating underlying regimes and states for the returns, BL MSGM reports high risk-adjusted returns of 1.346. Based on the magnitude of performance improvements of the traditional portfolios when taking into account estimation error, judgements can be made regarding the allocations' sensitivity to parameter uncertainty. It seems BL and MW face similar impacts of estimation risk, both which are greater than the effects on the GMVP. When the embedded parameter risk is accounted for however, the performance of the three traditional models seems to converge.

I conclude this section by emphasising the use of specifying portfolio models in a Bayesian format as well as the added value of adopting a non-normal distribution for the asset returns. Including extra knowledge about parameter locations can prove very useful in improving the out of sample performance of portfolio strategies. The next sections investigate the ability to provide robust portfolios strategies.

5.4 Combination Strategies

Table 7: Out-of-Sample Performance of Combination Models on 48 Industry Portfolios

	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
U EWC hier	0.038	0.939	0.635*	-0.676	17.490	1.229
U MWC hier	0.006	0.435	0.211	-0.051	11.462	0.109
U WC hier	0.006	0.435	0.211	-0.051	11.462	0.109
U EWC MSGM	0.039	0.381	1.615***	-0.038	13.487	0.412
U MWC MSGM	0.023	0.307	1.192	-0.020	14.673	0.085
U WC MSGM	0.023	0.307	1.192	-0.020	14.673	0.085
3f hier	0.019	0.427	0.690*	-0.067	9.167	0.511
3f MSGM	0.036	0.391	1.442**	-0.040	11.258	0.467
EW	0.040	1.609	0.395	-1.000	4.558	0.980

Note: This table presents the performance measures for the Bayesian portfolio combination methods explained in Section 2.5.2. The combinations are constructed using Bayesian implementations of Markowitz, Global Minimum Variance Portfolio and Black-Litterman. The Bayesian implementations are based on a hierarchical prior specification (hier) and a Markov Switching Gaussian Mixture (MSGM) model. The hierarchical combinations include the equally weighted (U EWC hier), maximum utility (U MWC hier) and utility weighted (U WC hier) combinations. The MSGM model combinations include the equally weighted (U EWC MSGM), maximum utility (U MWC MSGM) and utility weighted (U WC MSGM) combinations. Additionally, Bayesian implementations of the optimal three-fund rule of Kan and Zhou (2007) are also presented for hierarchical (3f hier) and MSGM specifications (3f MSGM). The equally weighted (EW) portfolio is included as a benchmark. Evaluation criteria are calculated for each strategy on daily returns of the 48 Industry portfolios data set over time frame June 27, 2006, until September 19, 2012. The reported performance measures for each strategy are: mean return, standard deviation, Sharpe ratio (SR), certainty equivalence (CEQ), turnover and Mean Weight invested in Risky assets (MWR). P-values for difference in Sharpe ratios with respect to Markowitz (MW) are derived from methods of Ledoit and Wolf (2008). *, **, *** represent 10%, 5% and 1% significance levels, respectively.

Table 7 shows improvement in out-of-sample performance for the Kan and Zhou (2007) three-fund rule when parameter risk is implicitly accounted for. This demonstrates the gains of using Bayesian portfolios in combination with frequentist optimisation techniques. By using parameter estimates that regard estimation error, derived from the hierarchical Bayes and the MSGM models, the 3f rule reports lower return volatility, and higher risk-adjusted returns. The MSGM model is best able to capture stylised facts of daily returns, and thus it is no surprise that its implementation results in the higher SR gain. The 3f rule improves from a SR of 0.682 to 1.442 for 3f MSGM. As the Bayesian MW model, MW MSGM, is also outperformed in mean and risk-adjusted returns, I conclude the worth of including the portfolio GMVP in the allocation. Both 3f hier and 3f MSGM are considered 'winner' strategies and viable modifications of the MW portfolio. The performance robustness is tested in section 5.5.

In the search of a combination of the three pillars of portfolio management, the MW and GMVP models are also merged with the BL. These merges are based on distributions of the portfolio weights. I argue that Bayesian combinations based on weight distributions should perform better than plug-in of the Bayesian estimates into frequentist model combinations. In this manner, manipulations to parameter distributions that are carried out when constructing portfolio weights are taken into account. The equally weighted combination of the Bayesian portfolios reports the highest out-of-sample performance. The hierarchical Bayesian combination of the three traditional portfolio rules, U EWC hier, seems able to produce very high mean returns. However, it also reports very high risk, 0.939. The combination based on switching regimes and fatter tails, U EWC MSGM, proves to be a safer option. The strategy has a low volatility of 0.381 and SR of 1.615, reporting even better performance measures than 3f MSGM. The fact that the (sub-optimal) equally weighted combination performs so well, supports the feasibility of the proposed combination framework based on weight *distributions*. However, as turnovers are relatively high for these strategies, their robustness is to be evaluated in the next ('winner') section.

The search for a more 'optimal' Bayesian combination proved less useful. The highest utility choosing method, and the weighted combination method based on draw utilities, report the same SRs for both hierarchical and MSGM implementations, 0.211 and 1.192, respectively. This indicates that for most draws the three traditional portfolio methods have retrieved negative utilities, such that the maximum utility and weighted max utility strategies coincide. The proposed weighted combination based on utility therefore seems inadequate. Other weighting schemes or measures for designing a combination method may prove better.

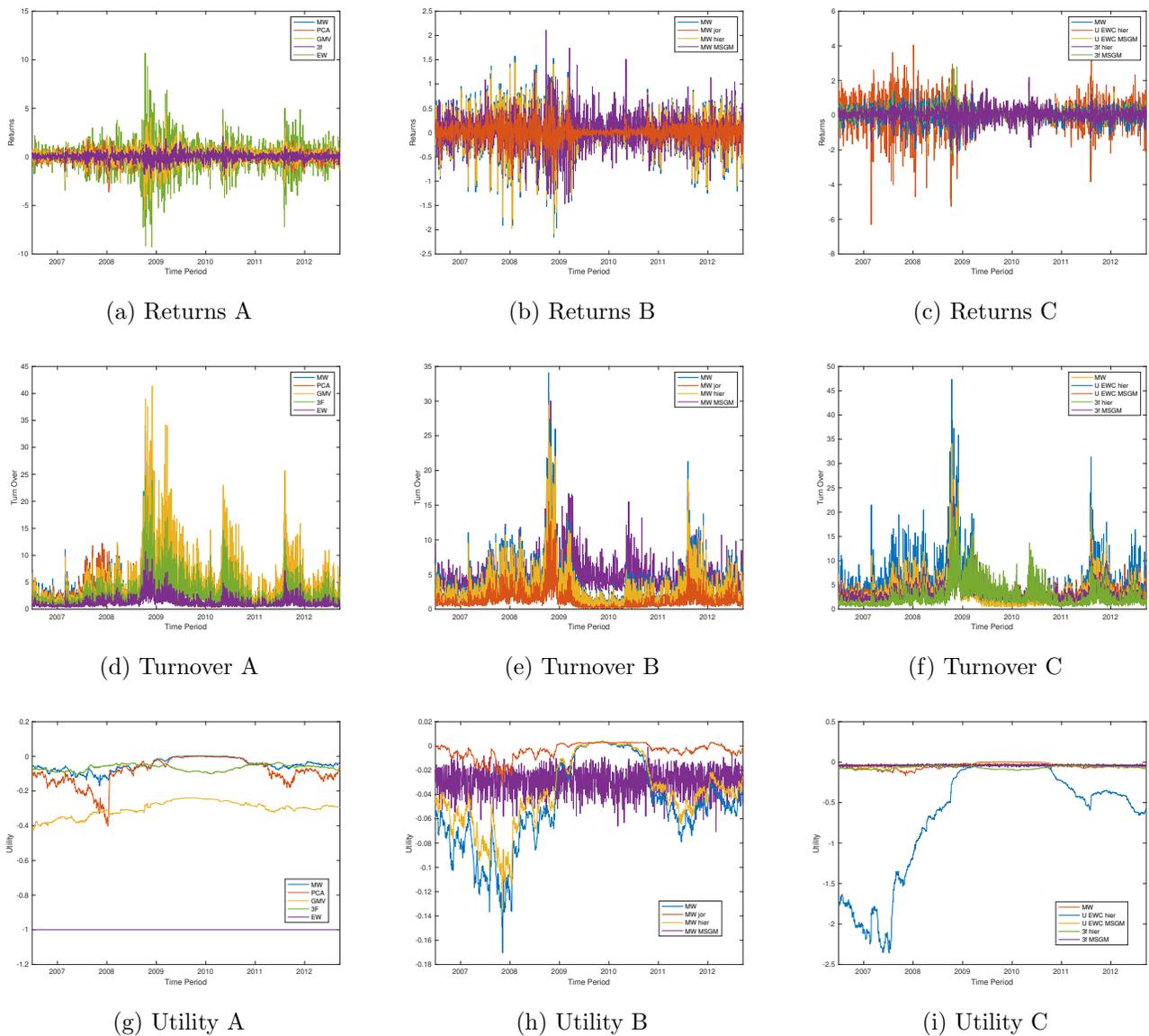
This section highlights the main contributions of this paper. Using Bayesian statistics for the formation of portfolio combinations proves greatly beneficial. The Bayesian three-fund rule extension is shown to be a valuable modification to the works of Kan and Zhou (2007). Especially the MSGM implementation allows for increased mean returns while simultaneously mitigating the portfolio risk. Furthermore, the proposed Bayesian combination framework proves competent and feasible. Out of all the considered models, the equally weighted Bayesian combination has proved superior in terms of out-of-sample Sharpe ratios. This shows the value and potential of combinations based on weight distributions. More sophisticated combination schemes can potentially lead to further performance improvements. Additionally, based on the evaluation criteria reported for combinations of the three pillars of portfolio optimisation, a strategy that considers MW, GMVP and BL, seems a good starting point for a universally applicable trading rule.

Table 8: Out-of-Sample Performance of Winner Markowitz Modifications for subperiods of 48 Industry Portfolios

Pre Crisis	Strategy	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
	MW	-0.003	0.329	-0.145	-0.07	2.331	0.130
	GMVP	0.108	0.420	4.079***	-0.387	2.083	1.000
	MW PCA	0.005	0.356	0.207	-0.126	2.067	0.256
	EW	0.095	0.706	2.131	-1.000	0.782	0.980
	3f	0.044	0.218	3.206***	-0.078	1.331	0.478
	MW Jor	0.004	0.132	0.456	-0.004	0.948	0.117
	MW hier	0.007	0.298	0.356	-0.053	2.051	0.121
	MW MSGM	0.053	0.270	3.139	-0.029	3.981	0.103
	U EWC hier	0.212	0.864	3.903	-1.859	4.235	2.131
	U EWC MSGM	0.065	0.276	3.713***	-0.048	3.175	0.435
	3f hier	0.044	0.212	3.283***	-0.082	1.169	0.491
	3f MSGM	0.054	0.251	3.421***	-0.040	2.389	0.470
Crisis	Strategy	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
	MW	-0.004	0.524	-0.123	-0.057	3.872	0.094
	GMVP	-0.019	0.961	-0.315	-0.302	5.800	1.000
	MW PCA	0.008	0.648	0.184	-0.099	3.451	0.126
	EW	0.016	1.920	0.134	-1.000	1.574	0.980
	3f	-0.003	0.494	-0.089	-0.066	3.501	0.490
	MW Jor	-0.002	0.234	-0.124	-0.006	1.643	0.085
	MW hier	-0.002	0.484	-0.049	-0.044	3.424	0.090
	MW MSGM	0.014	0.402	0.561	-0.029	5.432	0.104
	U EWC hier	-0.021	1.123	-0.297	-0.812	6.155	1.317
	U EWC MSGM	0.024	0.437	0.884**	-0.037	4.308	0.407
	3f hier	-0.003	0.490	-0.090	-0.067	3.236	0.502
	3f MSGM	0.020	0.454	0.712*	-0.039	3.701	0.466
Post Crisis	Strategy	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
	MW	0.020	0.314	0.993	-0.036	2.709	0.111
	GMVP	0.065	0.677	1.522	-0.285	4.254	1.000
	MW PCA	0.022	0.400	0.885	-0.067	2.445	0.203
	EW	0.059	1.334	0.700	-1.000	1.127	0.980
	3f	0.037	0.378	1.553***	-0.062	2.452	0.512
	MW Jor	0.009	0.137	1.103***	-0.001	1.055	0.099
	MW hier	0.018	0.289	1.015**	-0.028	2.381	0.108
	MW MSGM	0.040	0.284	2.242***	-0.029	4.529	0.103
	U EWC hier	0.068	0.693	1.568*	-0.302	3.961	0.960
	U EWC MSGM	0.049	0.327	2.396	-0.038	3.558	0.413
	3f hier	0.037	0.379	1.565***	-0.064	2.268	0.526
	3f MSGM	0.049	0.332	2.322	-0.040	2.904	0.469

Note: This table presents the performance measures for the chosen winner MW portfolio modifications over subperiods of the 48 Industry portfolios data set. The presented frequentist models are Markowitz (MW), Global Minimum Variance Portfolio (GMVP), Markowitz based on Principal Component Analysis estimates (MW PCA) and the optimal three-fund rule (3f). The Markowitz portfolio and three-fund rule are also considered in a Bayesian setting. Markowitz using a Jorion (MW Jor) prior and both using hierarchical (MW hier, 3f hier) and Markov Switching Gaussian Mixture (MW MSGM, 3f MSGM) prior specifications. Equally weighted Bayesian portfolio combinations of Markowitz, Global Minimum Variance Portfolio and Black-Litterman are also considered for hierarchical (U EWC hier) and MSGM specifications (U EWC MSGM). The equally weighted (EW) portfolio is also included as benchmark. The out-of-sample period spans from June 27, 2006, until September 19, 2012. The time period is split up into pre-crisis (2006), crisis (2007-2009) and post-crisis (2010-2012) periods. The reported performance measures for each strategy are: mean return, standard deviation, Sharpe ratio (SR), certainty equivalence (CEQ), turnover and Mean Weight invested in Risky assets (MWR). P-values for difference in Sharpe ratios with respect to the EW are derived from methods of Ledoit and Wolf (2008). *, **, *** represent 10%, 5% and 1% significance levels respectively.

Figure 3: Performance Measures for Winner Markowitz Modifications over Time for 48 Industry portfolios



Note: This figure shows the evaluation criteria for the winner Markowitz modification models over time. It presents the out-of-sample returns, turnover and corresponding utility for the period June 27, 2006 until September 19, 2012 on returns of assets from the 48 Industry portfolios (48IndP) data set. Figures A provide performance of the frequentist Markowitz (MW) alternatives: Global Minimum Variance (GMV), equally weighted (EW) and three-fund rule (3f). Figures B provide the performance for Bayesian MW implementations: Jorion prior (Jor), hierarchical MW (MW hier) and Markov Switching Gaussian Mixture MW (MSGM MW). Figures C report the proposed Bayesian combination methods: Equally weighted combination of hierarchical MW, GMVP and Black-Litterman (U EWC hier), equally weighted combination of MSGM MW, GMVP and Black-Litterman (U EWC MSGM), hierarchical three-fund rule (3f hier) and MSGM three-fund (3f MSGM).

5.5 Measuring Robustness

Table 8 provides results for the evaluation criteria split over different sub periods of the out-of-sample set. I look at the performance of the most competent MW modifications of previous sections over pre financial crisis (2006), financial crisis (2007-2009) and post financial crisis (2010-2012) periods. The considered strategies are GMVP, MW PCA, 3f, MW Jor, MW hier, MW MSGM, 3f hier, 3f MSGM, U EWC hier and U EWC MSGM. The traditional sample MW and EW portfolios are also included as benchmarks. In order to provide judgment on the ability of beating the EW portfolio out of sample, the SR significance is taken with respect to the naive portfolio in this section.

Notable is the worth of including the GMVP portfolio in portfolio allocations. Combinations with GMVP improve MW performances over all considered subperiods. Table 8 reports superior Sharpe ratios for the three-fund rule over traditional MW implementations for both frequentist and Bayesian methods. Observe that in all three subperiods the MW model is outperformed by the 3f rule, and MW hier and MW MSGM are outperformed by 3f hier and 3f MSGM respectively. Including GMVP in portfolio strategy increases investments in minimum-risk assets (MWR) allowing for an enhanced performance. This effect is greatest for pre and post crisis periods where the GMVP portfolio obtains higher returns. The substantially deteriorated performance of GMVP in the crisis period, which reports worse SRs than MW over that time span, still improves performance when strategies are taken together, but to a lesser degree. I also consider Bayesian combinations of MW, GMVP and BL models. Similar effects as for the MW, GMVP mixtures are observed but result in riskier strategies. By aggregating BL, U EWC hier and U EWC MSGM, shift their portfolio weights away from the optimal risk-return allocations. Additionally, as no optimal fractions are assigned to the weight distributions of MW, GMVP and BL, a riskier portfolio results. This can be seen from the higher standard deviation values of U EWC hier than of the 3f hier over the three subperiods.

The Bayesian implementations of MW portfolio allow for more subtle improvements in out-of-sample performance over time. MW Jor and MW hier only improve slightly in terms of SR for the three periods. The strategies that account for parameter uncertainty implicitly show to take on safer portfolios than the MW. Mean returns and standard deviation are lower, leading to slight improvements of the risk-adjusted returns. Though these are not as large as the SRs for the GMVP combinations, I do stress the substantially lower volatility for Jorion and hierarchical MW portfolios, in both returns and weight allocations. This can be seen in Figure 3. Sub Figure 3b shows that the more stable returns and Sub Figure 3e, shows that the lowest turnovers are observed for MW Jor and MW hier. Given that the Bayesian MW models report lowest standard deviations of returns and turnovers, such a strategy may be more suitable for an investor that is more risk-averse. The MWR is lowest for these strategies too. Considering associated transaction costs, the lower investments and turnovers of the Bayesian modifications may therefore prove most applicable in practice. Further relevant research could incorporate transaction costs to enable a better verdict of empirical application.

Similarly, the MW allocation that uses estimates from PCA factor model seems to provide robust performance over time. The strategy reports positive SRs over all three time periods, but is not able to beat the EW portfolio significantly. Table 8 also proves the superior competence of the MSGM estimates over sample estimates of μ and Σ . MSGM implements a Markov Switching regime for common shocks on assets and a Gaussian mixture form for returns for idiosyncratic risks. These modifications that enable better capturing of the true return nature help MW over the subperiods. The added flexibility leads to significant improved performance over time, and most notably in the Crisis period, where SR improves from -0.123 for sample MW, to 0.561 for MW MSGM. MSGM shows to be the most reactive Bayesian implementation method over time,

boosting performance of MW and the combination methods, U EWC MSGM and 3f MSGM.

Next, consider the research goal of being able to beat the equally weighted strategy out of sample. Table 8 reinforces the difficulty in beating the naive portfolio. Especially in periods of crisis the EW shows to be one of the few allocation methods with positive returns. By considering different regimes, the MSGM models are able to outperform the EW strategy. Even in the crisis period, MW MSGM, U EWC MSGM and 3f MSGM retrieve higher out-of-sample Sharpe ratios. Furthermore, the significance levels of Table 8, which report the level of significant difference of portfolio Sharpe Ratios with respect to the EW portfolio, show ability to outperform the EW significantly. Moreover, Figure 3 also shows ability to attain higher utilities for MW modifications than EW, implying that the MW modifications are much more attractive than the EW to the mean-variance investor over time. Figure 3 is used to address whether the proposed MW modifications lead to more robust portfolios. All methods except the hierarchical Bayesian combination U EWC hier, show ability to produce more stable portfolio returns over time. Moreover, turnovers are also noticeably lower. I therefore conclude that the regarded methods are able to address weaknesses of mean-variance portfolios and that there is clear evidence for robustifying MW portfolios by tackling parameter uncertainty. Additionally, ability to beat the equally weighted strategy significantly in terms of out-of-sample performance is also established. Identical analysis is also carried out for Fama-French 100 portfolios, of which similar conclusions can be drawn. For 100FF the 3f MSGM and U EWC MSGM methods too report the best performance over changing economic conditions and show ability to significantly beat the EW over time. Given the fact that MW modifications also prove successful on other data sets, robustness of the proposed methods is supported.

6 Conclusion

This paper provides a wide range of methods to suppress the effect of parameter uncertainty in portfolio optimisation. Frequentist and Bayesian approaches are reviewed and extended in attempt to enhance the out-of-sample performance and empirical applicability of the Markowitz portfolio.

For the frequentist shrinkage of parameters, sample estimates of return moments, which are used as inputs for the Markowitz portfolio, are shrunk to more certain values. Existing shrinkage methods for the mean, shrunk to the grand mean (Jorion, 1986), and covariance, shrunk to the Identity matrix (Ledoit and Wolf, 2004), are reviewed. In addition, this paper establishes equivalence of the Black-Litterman model to such a form of parameter shrinkage. By altering the original model definition, Black-Litterman can be interpreted as shrinkage of both the mean and covariance of returns, with shrinkage intensities based on- and towards- investor views. For both 48 Industry portfolios and Fama-French 100 portfolios, directly shrinking the sample inputs of the Markowitz portfolio does not lead to significant performance improvements. The reduction in parameter uncertainty is not able to outweigh the added bias for the defined shrinkage intensities. Furthermore, the sensitivity of Markowitz to the input parameters is not acknowledged. Accordingly, parameter quality improvements alone prove unable to diminish the volatility in weight allocations sufficiently.

The sensitivity of Markowitz to input parameters is addressed by the second frequentist method. Portfolio weights are shrunk towards other portfolios that have more stable weights and are less susceptible to estimation error. Aside from the two-fund rule and three-fund rule of Kan and Zhou (2007), novel combinations of the Markowitz to the equally weighted portfolio and, Global Minimum Variance Portfolio to equally weighted portfolios are provided. The correspondent optimal shrinkage intensities are derived such that the expected utility loss is minimised. The new model combinations prove able to

reduce portfolio volatility and turnover, but the three-fund rule of Kan and Zhou (2007) remains most effective in tackling input sensitivity and maintaining the optimality of portfolio weights. It reports the highest out-of-sample Sharpe ratios of the considered frequentist portfolio approaches and proves most attractive to the frequentist mean-variance investor.

Furthermore, Bayesian specifications of the Markowitz portfolio prove effective in stabilising portfolios that are subject to parameter uncertainty. Different prior specifications motivated by previous literature are used to best capture estimation error in the decision problem. Diffuse, conjugate, Jorion and hierarchical prior definitions are considered. Of these, The hierarchical prior model proves to be the most conservative and can therefore be regarded as most applicable in practice. Moreover, by additionally accounting for non-normality of returns with a Markov Switching Gaussian Mixture model, which specifies level shifts in the mean and accounts for idiosyncratic risks of assets, performance of Markowitz is further improved. Higher out-of-sample Sharpe ratios are reported over subperiods of the data indicating improved robustness for the Markowitz modification.

The main contributions lie in Bayesian specifications of frequentist optimal portfolios. By considering Bayesian estimates to construct portfolios that can account for embedded parameter uncertainty, shortcomings of the frequentist methods are addressed. In this manner, the three-fund rule of Kan and Zhou (2007) is modified to better account for estimation errors. The hierarchical prior and Markov Switching Gaussian Mixture model specifications prove successful in capturing styles facts of returns and boosting model performance. The modified three-fund rule reports very high out-of-sample performance and proves robust to changing economic times. As the model is able to beat the tough equally weighted benchmark over different subperiods, including the global financial crisis, the Bayesian three-fund rule is considered a viable extension to the traditional Markowitz portfolio.

The Bayesian implementations also allow for novel portfolio combinations. By considering portfolio weights as distributions, combinations of traditional portfolio optimisation methods can be constructed. The technique is demonstrated for a combination of the three ‘pillars’ of portfolio optimisation; the Markowitz-, the Global Minimum Variance- and Black-Litterman- portfolio. A number of combination schemes are proposed, of which the equally weighted strategy provides highest performance in terms of out-of-sample Sharpe ratios over the whole sample. In particular the combination that accounts for different return regimes and fatter tails proves effective, reporting similar robust performance to the three-fund rule extension. The empirical applicability of the proposed Bayesian combination framework opens up new avenues in the search of a universal portfolio optimisation model.

Methods that tackle parameter uncertainty prove effective in mitigating the weaknesses of the traditional Markowitz portfolio. The proposed adaptations produce more more robust and stable portfolio weights that perform better over time. For the 48 Industry portfolios and the Fama-French 100 portfolios, ability is reported to outperform the $1/N$ portfolio in terms of risk-adjusted returns over subperiods that include the global financial crisis. The Markowitz modifications presented in this research allow for a ‘modern’ take on Markowitz portfolio optimisation that is more appropriate in practice.

7 Discussion and Further Research

The Markowitz modification methods prove viable extensions to the traditional Markowitz portfolio. Improvements are reported in out-of-sample performance, weight stability and performance robustness. Additionally, the tough $1/N$ benchmark can be beaten significantly in terms of out-of-sample Sharpe ratios over changing economic times. These findings propose numerous possibilities for further research. This section evaluates the presented analysis and provides ideas for future investigations.

First, some comments regarding the used data sets are in order. The models were imposed on the 48IndP and the 100FF data sets over the time period of July 7, 2004, until September 19, 2012, in conform with literature of DeMiguel et al. (2015). The 48IndP constitutes of 48 Industry portfolios and the 100FF contains 100 portfolios sorted on size and book-to-markets. Both these sets are formed by portfolios, implying that part of the estimation risk has already been diversified. Investigating whether the discussed methods perform better when applied directly on stocks, of the S&P 500 for example, is of interest. Applying the proposed methods on data that suffers more from estimation risk can prove more effective. Furthermore, it would be interesting to regard less frequent data such as monthly returns. Not only is that more realistic for the institutional investor, but the associated estimation errors are greater as well. Correspondingly, the gains made by the proposed Markowitz adaptations should become greater.

For illustrative purposes this investigation took a simplifying assumption regarding the risk aversion of the investor. Parameter γ is taken fixed and equal to one. Graphs in Appendix A show that for smaller investor risk aversion expected utility losses resulting from parameter uncertainty are greater. Consequently, γ was chosen equal to 1 so that risk aversion is at its smallest but the investor remains risk-averse. In previous literature values between 1 and 10 tend to be chosen. Increasing the risk aversion parameter may diminish the differences in MW and proposed solution performances. Such a sensitivity analysis is in order. The same argument holds for alternative window sizes.

Another limitation of the analysis regards the hierarchical and MSGM models. Due to the computational cost, MCMC samples of 2000 draws were run for the estimation of the parameters. Improved performance over traditional implementations is reported. However, it remains unclear whether the Markov chains have reached a point of convergence, and whether convergence is to a local or global optimal. Initializing MCMC samples at different points may provide better convergence to global optimums, but requires sufficient computational power. Additionally, the MSGM model may benefit from a two-level prior Dirichlet distribution specification. This way, the ‘upper’ Dirichlet defines parameters of the ‘lower’ distribution, which defines the distribution of the states. This way, information on which states are more likely to occur can be implemented.

Moreover, the feasibility of the proposed methods in practice should be further investigated. Throughout the investigation, the proposed methods provide useful (theoretical) insights regarding the impact of estimation error on portfolio optimisation and how this effect can be mitigated. The implementation of Bayesian statistics and the novel combination methods prove able to improve Markowitz portfolio performance and are of relevance for all investors. However, given the high weight allocations, turnovers and rebalancing frequency, the empirical applicability of the MW modifications for institutional investors remains arguable. Especially considering the fact that transaction costs ought to be taken into account. Validating empirical applicability when transaction costs are considered is necessary. Further research could investigate ways to make the presented models more suitable for real life traders. This involves tweaking the Markowitz portfolio definition by redefining the investor problem/utility. For instance, one could investigate the problem for a multiperiod investment horizon as done in DeMiguel

et al. (2015). Additionally, the investor utility function can be made to account for transaction costs and possible uncertainty aversion. These extensions match the interests of true investors better. The Markowitz model can also be derived such that a weight constraint or turnover constraint is implemented. Not only does this complement the idea of investments of a finite amount, but works as Jagannathan and Ma (2002) have shown Markowitz models to benefit from such constraints as well. The ‘new’ Markowitz model defined from the modified investor preferences will be more suited for real-life trading but will still suffer from parameter uncertainty. Therefore, the model can still benefit from the proposed estimation error mitigation effects provided in this paper.

Throughout, the use of specifying the portfolio problem in a Bayesian manner was emphasised. It is therefore of interest to research alternative informative priors. The Bayesian methods used in this paper use diffuse, conjugate and hierarchical priors, reporting better accountability of uncertainty for the more detailed models. Having concluded that the conjugate specification already provided useful additional information, other methods for further/better calibrating model (hyper)parameters are in order. For instance, Pastor (2000) and Pastor and Stambaugh (2000) provide a way of incorporating asset pricing models as priors which could result effective in this setting also. Motivating which priors to use remains a subjective discussion however, and thus new definitions require careful motivation. Other avenues for further research can be found in the implementation the Bayesian portfolio combination framework. The technique enables portfolios to be merged based on their weight distributions. In this paper combinations of the traditional portfolio optimisation portfolios, Markowitz, Global Minimum Variance and Black-Litterman, were considered. These models are at the forefront of portfolio optimisation and therefore such a combination seems to be a good starting point for a universally applicable portfolio. The proposed combination-schemes were based on draw utilities and proved less effective than the equally weighted combination strategy. It would be worthwhile to investigate alternative (optimal) combination-weights of models. Additionally, using other measures (such as past performance) or portfolios to construct combinations may prove more useful.

References

- Ané, T. and Geman, H. (2000). Order Flow, Transaction Clock, and Normality of Asset Returns. *The Journal of Finance*, 55(5):2259–2284.
- Ang, A. (2014). *Asset management : a systematic approach to factor investing*. Oxford University Press.
- Avramov, D., Zhou, G., and Louis, S. (2009). Bayesian Portfolio Analysis. Technical report, Annual Review of Financial Economics.
- Bauder, D., Bodnar, T., Parolya, N., and Schmid, W. (2020). Bayesian mean–variance analysis: optimal portfolio selection under parameter uncertainty. *Quantitative Finance*, pages 1–22.
- Baum, L. E. and Petrie, T. (1966). Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *Annals of Mathematical Statistics*, 37(6):1554–1563.
- Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *Annals of Mathematical Statistics*, 41(1):164–171.
- Bessler, W., Opfer, H., and Wolff, D. (2017). Multi-asset portfolio optimization and out-of-sample performance: an evaluation of Black–Litterman, mean-variance, and naïve diversification approaches. *The European Journal of Finance*, 23(1):1–30.
- Black, F. and Litterman, R. (1992). Global Portfolio Optimization. *Financial Analysts Journal*, 48(5):28–43.
- Bodnar, T., Mazur, S., and Okhrin, Y. (2017). Bayesian estimation of the global minimum variance portfolio. *European Journal of Operational Research*, 256(1):292–307.
- Brandt, M. W., S.-C. P. . V. R. (2009). Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns. *The Review of Financial Studies*, 22(9):3411–3447.
- Brodie, J., Daubechies, I., De Mol, C., Giannone, D., and Loris, I. (2009). Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences of the United States of America*, 106(30):12267–12272.
- Buckley, I., Saunders, D., and Seco, L. (2008). Portfolio optimization when asset returns have the Gaussian mixture distribution. *European Journal of Operational Research*, 185(3):1434–1461.
- Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, 75(1):79–97.
- Chopra, V. K., Hensel, C. R., and Turner, A. L. (1993). Massaging Mean-Variance Inputs: Returns from Alternative Global Investment Strategies in the 1980s. *Management Science*, 39(7):845–855.
- Chopra, V. K. and Ziemba, W. T. (1993). The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. *The Journal of Portfolio Management*, 19(2):6–11.
- De Nard, G., Ledoit, O., and Wolf, M. (2018). Factor Models for Portfolio Selection in Large Dimensions: The Good, the Better and the Ugly. *SSRN Electronic Journal*.
- DeMiguel, V., Alberto, M.-U., Nogales, F. J., DeMiguel, V., Alberto, M.-U., and Nogales, F. J. (2015). Parameter Uncertainty in Multiperiod Portfolio Optimization with Transaction Costs. *Journal of Financial and Quantitative Analysis*, 50(6):1443–1471.
- DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953.
- DeMiguel, V., Martin-Utrera, A., and Nogales, F. J. (2013). Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, 37(8):3018–3034.

- Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, 38(1):34.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.
- Frahm, G. and Memmel, C. (2010). Dominating estimators for minimum-variance portfolios. *Journal of Econometrics*, 159(2):289–302.
- Frankfurter, G. M., Phillips, H. E., and Seagle, J. P. (1971). Portfolio Selection: The Effects of Uncertain Means, Variances, and Covariances. *The Journal of Financial and Quantitative Analysis*, 6(5):1251.
- Frey, C. (2016). Bayesian Shrinkage of Portfolio Weights. *SSRN Electronic Journal*.
- Frost, P. A. and Savarino, J. E. (1988). For better performance. *The Journal of Portfolio Management*, 15(1):29–34.
- Frost, P. A., Savarino, J. E., Frost, P. A., and Savarino, J. E. (1986). An Empirical Bayes Approach to Efficient Portfolio Selection. *Journal of Financial and Quantitative Analysis*, 21(3):293–305.
- Gârleanu, N. and Pedersen, L. (2013). Dynamic Trading with Predictable Returns and Transaction Costs. *The Journal of Finance*, 68(6):2309–2340.
- Geweke and John (2007). Interpretation and inference in mixture models: Simple MCMC works. *Computational Statistics & Data Analysis*, 51(7):3529–3550.
- Greyserman, A., Jones, D. H., and Strawderman, W. E. (2006). Portfolio selection using hierarchical Bayesian analysis and MCMC methods. *Journal of Banking and Finance*, 30(2):669–678.
- Haff, L. R. (1979). Estimation of the Inverse Covariance Matrix: Random Mixtures of the Inverse Wishart Matrix and the Identity. *The Annals of Statistics*, 7(6):1264–1276.
- Han, C. (2019). How much should portfolios shrink? *Financial Management*, page fima.12282.
- Harvey, C. R., Liechty, J. C., Liechty, M. W., and Peter, M. (2010). Portfolio selection with higher moments. *Quantitative Finance*, 10(5):469–485.
- Hautsch, N. and Voigt, S. (2019). Large-scale portfolio allocation under transaction costs and model uncertainty. *Journal of Econometrics*, 212(1):221–240.
- He, G. and Litterman, R. (2005). The Intuition Behind Black-Litterman Model Portfolios. *SSRN Electronic Journal*.
- Hurley, W. J. and Brimberg, J. (2015). A note on the sensitivity of the strategic asset allocation problem. *Operations Research Perspectives*, 2:133–136.
- Idzorek, T. M. (2007). *A step-by-step guide to the Black-Litterman model: Incorporating user-specified confidence levels*. Elsevier Ltd.
- Jagannathan, R. and Ma, T. (2002). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. Technical report, National Bureau of Economic Research, Cambridge, MA.
- Jobson, J. D. and Korkie, B. (1989). A Performance Interpretation of Multivariate Tests of Asset Set Intersection, Spanning, and Mean-Variance Efficiency. *JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS*, 24(2).
- Jorion, P. (1986). Bayes-Stein Estimation for Portfolio Analysis. *The Journal of Financial and Quantitative Analysis*, 21(3):279.
- Jun Tu and Guofu Zhou (2010). Incorporating Economic Objectives into Bayesian Priors: Portfolio Choice under Parameter Uncertainty. *The Journal of Financial and Quantitative Analysis*, 45(4):959–986.
- Kan, R. and Zhou, G. (2007). Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656.

- Klein, R. W. and Bawa, V. S. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3(3):215–231.
- Landsman, Z. and Nešlehová, J. (2008). Stein’s Lemma for elliptical random vectors. *Journal of Multivariate Analysis*, 99:912–927.
- Ledoit, O. and Wolf, M. (2004). Honey, I Shrunk the Sample Covariance Matrix. *Journal of Portfolio Management*, 30(4):110–119.
- Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, pages 850–859.
- Lee, W. (2000). *Advanced Theory and Methodology of Tactical Asset Allocation*. New York: John Wiley & Sons.
- Levy, H. and Sarnat, M. (1970). International diversification of investment portfolios. *The American Economic Review*, 60(4):668–675.
- Li, J. (2015). Sparse and Stable Portfolio Selection With Parameter Uncertainty. *Journal of Business and Economic Statistics*, 33(3):381–392.
- M. Best and R. Grauer (1991). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. *Review of Financial Studies*, 4(2):315–342.
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1):77–91.
- Michaud, R. O. (1989). The Markowitz Optimization Enigma: is ‘Optimized’ Optimal? *Financial Analyst Journal*.
- Murphy, K. P. (2007). Conjugate Bayesian analysis of the Gaussian distribution. Technical report, University of British Columbia.
- Pastor, L. (2000). Portfolio Selection and Asset Pricing Models. *The Journal of Finance*, 55(1):179–223.
- Pastor, L. and Stambaugh, R. F. (2000). Comparing asset pricing models: An investment perspective. *Journal of Financial Economics*, 56(3):335–381.
- Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.
- Pollak, I. (2011). Weight shrinkage for portfolio optimization. In *2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2011*, pages 37–40.
- Putnam, Bluford H; Quintana, J. M. (1991). Mean-Variance Optimal Portfolio Models and the Inappropriateness of the Assumption of a Time-Stable Variance-Covariance Matrix. *Review of Financial Economics*, 1(1).
- Qian, H. (2009). Bayesian Portfolio Selection with Gaussian Mixture Returns. Technical report, MPRA Paper 32688, University Library of Munich, Germany.
- Qian, H. (2011). Bayesian Portfolio Selection in a Markov Switching Gaussian Mixture Model. Technical report, MPRA Paper 35561, University Library of Munich, Germany.
- Robb J. Muirhead (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, Inc., Hoboken, NJ, USA.
- Robert C. Blattberg and Nicholas J. Gonedes (1974). A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices. *The Journal of Business*, 47(2):244–280.
- Schepel, J. (2019). *Bayesian Extensions of the Black-Litterman Model*. (MSc Thesis). Available from Erasmus University Thesis Repository, Erasmus University Rotterdam.
- Sharpe, W. F. (1964). Capital Asset Prices: A theory of Market Equilibrium Under Conditions of Risk. *The Journal of Finance*, 19(3):425–442.

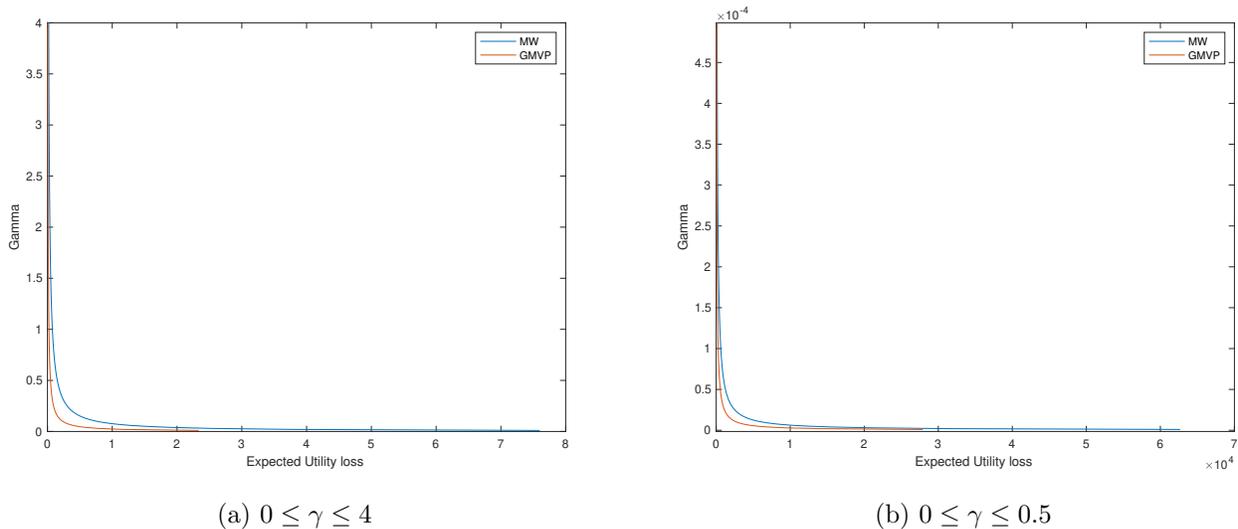
- S.J. Brown (1976). *Optimal portfolio choice under uncertainty*. PhD thesis, University of Chicago, Ph.D. Diss.
- Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1):204–215.
- Walters, J. (2011). The Black-Litterman Model in Detail. *SSRN Electronic Journal*.
- Wenzelburger, J. (2008). A Note on the Two-fund Separation Theorem A Note on the Two-fund Separation Theorem *. Technical report, Munich Personal RePEc Archive.
- Wolf, M. (2018). *Frequentist and Bayesian inference in multi-period portfolio analysis*. (Unpublished MSc Thesis) Erasmus School of Economics, Erasmus University Rotterdam.
- Zellner, A. and Chetty, V. K. (1965). Prediction and Decision Problems in Regression Models from the Bayesian Point of View. *Journal of the American Statistical Association*, 60(310):608.

Appendices

A Expected Utility Loss for Varying Risk Aversion

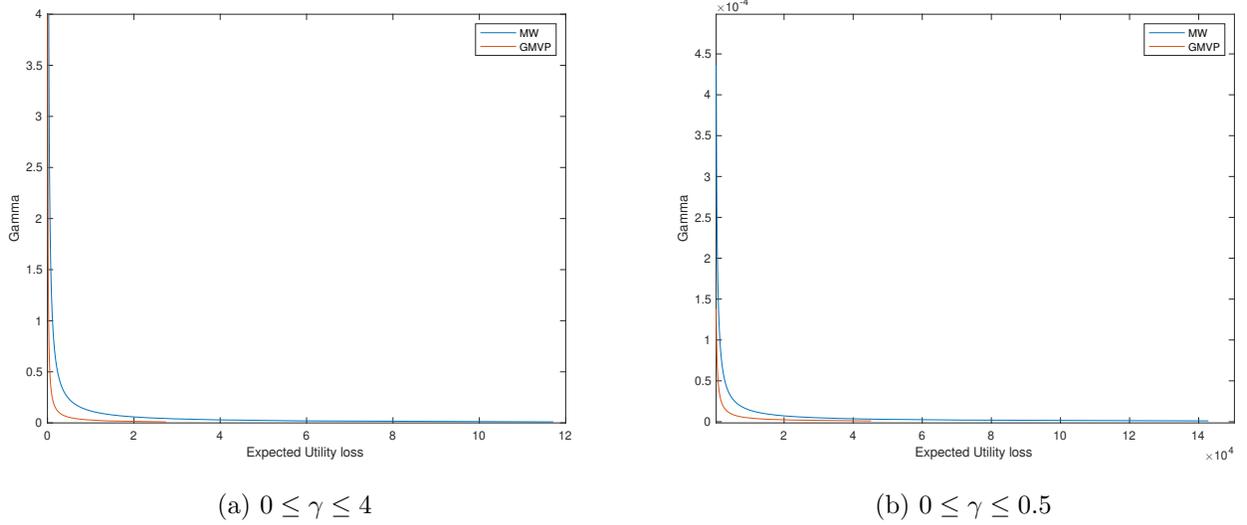
This section serves as a robustness check for the effect of risk aversion on the expected utility losses of the Markowitz and Global Minimum Variance portfolios. In literature, many different values for the risk aversion parameter have been used. As these tend to lie in the range of $[0, 4]$, the plots in Figures 4 and 5 are given for those values. As can be deduced from equation 18, an increasing gamma reduces the expected out-of-sample utility loss. This paper takes an investor risk aversion value of 1, the smallest throughout literature (Ang, 2014), such that the investor is still risk averse, and the effect of parameter uncertainty can be clearly shown. Figures 4 and 5 show that expected losses are lower for the Global Minimum Variance Portfolio than for the Markowitz portfolio. This difference is expected as the Global Minimum Variance Portfolio disregards μ in its estimation and therefore also the uncertainty losses associated with it. Furthermore, observe that these differences increase for smaller γ . It is therefore expected that the Global Minimum Variance Portfolio will perform better and be able to produce more stable portfolios over time in both data sets. This matter is further investigated throughout the paper.

Figure 4: Expected Utility Loss for Varying Risk Aversion Parameters for 48 Industry Portfolios



Note: This figure shows the expected utility losses associated with estimation errors, as characterised in section 2.1, for the Markowitz (MW) and Global Minimum Variance (GMVP) portfolios for varying values of risk aversion parameter γ . The parameter estimates used to calculate the utility losses are the sample mean and covariance of the daily excess returns of the 48 Industry portfolios over the entire sample period July 7, 2004, until September 19, 2012.

Figure 5: Expected Utility Loss for Varying Risk Aversion Parameters for Fama-French 100 Portfolios



Note: This figure shows the expected utility losses associated with estimation errors, as characterised in section 2.1, for the Markowitz (MW) and Global Minimum Variance (GMVP) portfolios for varying values of risk aversion parameter γ . The parameter estimates used to calculate the utility losses are the sample mean and covariance of the daily excess returns of the Fama-French 100 portfolios over the entire sample period July 7, 2004, until September 19, 2012.

B Derivation Alternative Three-Fund Rule, Markowitz and Equally Weighted Portfolio

In this section, the optimal shrinkage intensities for portfolio combinations of the Markowitz and equally weighted portfolios are derived. The derivation is similar to that of the three-fund rule of Kan and Zhou (2007). The shrinkage intensities are determined by minimising expected utility loss given the portfolio combination,

$$\hat{w}_t^{3f} = \frac{1}{\gamma}(c_2 \hat{\Sigma}^{-1} \hat{\mu}) + d_2 \frac{\iota_N}{N}. \quad (82)$$

Here c_2 and d_2 represent the shrinkage intensities towards both portfolios and are to be optimally determined. The expected out-of-sample utility is shown to be,

$$\begin{aligned} E[\tilde{U}(\hat{w}_t^{3f})] &= E[\hat{w}_t^{3f}]' \mu - \frac{\gamma}{2} E[\hat{w}_t^{3f} \Sigma \hat{w}_t^{3f}], \\ &= E\left[\frac{1}{\gamma}(c_2 \hat{\Sigma}^{-1} \hat{\mu}) + d_2 \frac{\iota_N}{N}\right]' \mu - \frac{\gamma}{2} E\left[\left(\frac{1}{\gamma}(c_2 \hat{\Sigma}^{-1} \hat{\mu}) + d_2 \frac{\iota_N}{N}\right)' \Sigma \left(\frac{1}{\gamma}(c_2 \hat{\Sigma}^{-1} \hat{\mu}) + d_2 \frac{\iota_N}{N}\right)\right], \\ &= \frac{c_2}{\gamma} E[\hat{\mu}' \hat{\Sigma}^{-1} \mu] + E\left[\frac{d_2}{N} \iota_N' \mu\right] - \frac{c_2^2}{2\gamma} E[\hat{\mu}' \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \hat{\mu}] - \frac{c_2 d_2}{N} E[\hat{\mu}' \hat{\Sigma}^{-1} \Sigma \iota_N] - \frac{\gamma d_2^2}{2N^2} E[\iota_N' \Sigma \iota_N]. \end{aligned} \quad (83)$$

Where the properties of a variable $W = \Sigma^{-\frac{1}{2}} \hat{\Sigma} \Sigma^{-\frac{1}{2}}$ are used to simplify computations. Haff (1979), show that $W \sim \frac{\text{Wishart}_N(T-1, I_N)}{T}$, such that

$$E[W^{-1}] = E[\Sigma^{\frac{1}{2}} \hat{\Sigma}^{-1} \Sigma^{\frac{1}{2}}] = \left(\frac{T}{T-N-2} \right) I_N \quad (84)$$

$$E[W^{-2}] = E[\Sigma^{\frac{1}{2}} \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \Sigma^{\frac{1}{2}}] = \left(\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \right) I_N. \quad (85)$$

These properties can be used to simplify equation 83,

$$\begin{aligned} E[\tilde{U}(\hat{w}_t^{3f})] &= \frac{c_2 \theta^2}{\gamma} \left(\frac{T}{T-N-2} \right) + \frac{d_2}{N} \iota'_N \mu - \frac{c_2^2}{2\gamma} \left(\frac{N}{T} + \theta^2 \right) \left(\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \right), \\ &- \frac{c_2 d_2}{N} \left(\frac{T}{T-N-2} \right) \iota'_N \mu - \frac{d_2^2 \gamma}{2N^2} \iota'_N \Sigma \iota_N. \end{aligned} \quad (86)$$

Next, apply the first order conditions for c_2 and d_2 . For tractability define $\theta^2 = \mu' \Sigma^{-1} \mu$. The partial derivative of the expected utility with respect to c_2 is taken and set to zero,

$$\begin{aligned} \frac{\partial E[\tilde{U}(\hat{w}_t^{3f})]}{\partial c_2} &= \left(\frac{T}{T-N-2} \right) \frac{\theta^2}{\gamma} - \frac{c_2}{\gamma} \left(\frac{N}{T} + \theta^2 \right) \left(\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \right), \\ &- \frac{d_2}{N} \left(\frac{T}{T-N-2} \right) \iota'_N \mu = 0. \end{aligned} \quad (87)$$

This can be rewritten to,

$$c_2 = \frac{\gamma(T-N-1)(T-N-4)}{(N/T + \theta^2)(T(T-2))} \left[\frac{\theta^2}{\gamma} - \frac{d_2}{N} \iota'_N \mu \right]. \quad (88)$$

Now, the process is repeated but for d_2 . By taking the partial derivative of the expected utility with respect to d_2 and equating to zero,

$$\frac{\partial E[\tilde{U}(\hat{w}_t^{3f})]}{\partial d_2} = \frac{1}{N} \iota'_N \mu - \frac{c_2}{N} \left(\frac{T}{T-N-2} \right) \iota'_N \mu - \frac{d_2 \gamma}{N^2} \iota'_N \Sigma \iota_N = 0. \quad (89)$$

This can be shown to equal,

$$d_2 = \frac{N \iota'_N \mu}{\gamma \iota'_N \Sigma \iota_N} - \frac{N \iota'_N \mu}{\gamma \iota'_N \Sigma \iota_N} \left(\frac{T}{T-N-2} \right) c_2. \quad (90)$$

The expression of d_2 is substituted into c_2 to obtain the following estimate for the optimal shrinkage intensity towards the Markowitz portfolio,

$$c_2^* = \frac{\theta^2 \iota'_N \Sigma \iota_N - (\iota'_N \mu)^2}{\frac{T(T-2)(N/T + \theta^2) \iota'_N \Sigma \iota_N}{(T-N-1)(T-N-4)} - \frac{T(\iota'_N \mu)^2}{T-N-2}}, \quad (91)$$

Finally, substituting c_2^* back into d_2 provides an expression for the optimal shrinkage intensity towards the equally weighted portfolio,

$$d_2^* = \frac{N \iota'_N \mu}{\gamma \iota'_N \Sigma \iota_N} \left(1 - \frac{(T-N-1)(T-N-2)(T-N-4)(\theta^2 - (\iota'_N \mu)^2)}{T^2(T-2)(N/T + \theta^2) \iota'_N \Sigma \iota_N - ((T-N-1)(T-N-2)(T-N-4))(\iota'_N \mu)^2} \right). \quad (92)$$

C Derivation Alternative Three-Fund Rule, Equally Weighted and Global Minimum Variance Portfolio

In this section, the optimal shrinkage intensities for portfolio combinations of the equally weighted and Global Minimum Variance portfolios are derived. The derivation is similar to that of the three-fund rule of Kan and Zhou (2007). The shrinkage intensities are determined by minimising expected utility loss given the portfolio combination,

$$\hat{w}_t^{3fEG} = c_3 \frac{1}{\gamma} \hat{\Sigma}^{-1} \iota_N + d_3 \frac{1}{N} \iota_N. \quad (93)$$

Here c_3 and d_3 represent the shrinkage intensities towards both portfolios and are to be optimally determined. The expected out-of-sample utility is shown to be,

$$\begin{aligned} E[\tilde{U}(\hat{w}_t^{3fEG})] &= E[\hat{w}_t^{3fEG} \mu - \frac{\gamma}{2} E[\hat{w}_t^{3fEG} \Sigma \hat{w}_t^{3fEG}]], \\ &= E[c_3 \frac{1}{\gamma} \hat{\Sigma}^{-1} \iota_N + d_3 \frac{1}{N} \iota_N] \mu + \frac{\gamma}{2} E[(\frac{1}{\gamma} (c_3 \frac{1}{\gamma} \hat{\Sigma}^{-1} \iota_N + d_3 \frac{1}{N} \iota_N)' \Sigma (\frac{1}{\gamma} (c_3 \frac{1}{\gamma} \hat{\Sigma}^{-1} \iota_N + d_3 \frac{1}{N} \iota_N)), \\ &= \frac{c_3}{\gamma} E[\iota_N' \hat{\Sigma}^{-1} \mu] + \frac{d_3}{N} E[\iota_N' \mu] - \frac{\gamma}{2} E[\frac{c_3^2}{\gamma^2} \iota_N' \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \iota_N] - \frac{c_3 d_3}{N} E[\iota_N' \hat{\Sigma}^{-1} \Sigma \iota_N] - \frac{\gamma}{2} E[\frac{d_3^2}{N^2} \iota_N' \Sigma \iota_N]. \end{aligned} \quad (94)$$

Where the properties of a variable $W = \Sigma^{-\frac{1}{2}} \hat{\Sigma} \Sigma^{-\frac{1}{2}}$ are used to simplify computations. Haff (1979) show that $W \sim \frac{\text{Wishart}_N(T-1, I_N)}{T}$, such that

$$E[W^{-1}] = E[\Sigma^{\frac{1}{2}} \hat{\Sigma}^{-1} \Sigma^{\frac{1}{2}}] = \left(\frac{T}{T-N-2} \right) I_N \quad (95)$$

$$E[W^{-2}] = E[\Sigma^{\frac{1}{2}} \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \Sigma^{\frac{1}{2}}] = \left(\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \right) I_N. \quad (96)$$

These properties can be used to simplify equation 94,

$$\begin{aligned} E[\tilde{U}(\hat{w}_t^{3fEG})] &= \frac{c_3}{\gamma} \left(\frac{T}{T-N-2} \right) \iota_N' \Sigma^{-1} \mu + \frac{d_3}{N} \iota_N' \mu - \frac{c_3^2}{2\gamma} \left(\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \right) \iota_N' \Sigma^{-1} \iota_N \\ &\quad - \frac{\gamma d_3^2}{2N^2} \iota_N' \Sigma \iota_N - c_3 d_3 \left(\frac{T}{T-N-2} \right). \end{aligned} \quad (97)$$

Next, apply the first order conditions for c_3 and d_3 . The partial derivative of the expected utility with respect to c_3 is taken and set to zero,

$$\begin{aligned} \frac{\partial E[\tilde{U}(\hat{w}_t^{3fEG})]}{\partial c_3} &= \frac{1}{\gamma} \left(\frac{T}{T-N-2} \right) \iota_N' \Sigma^{-1} \mu - \frac{c_3}{\gamma} \left(\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \right) \iota_N' \Sigma^{-1} \iota_N \\ &\quad - d_3 \left(\frac{T}{T-N-2} \right) = 0. \end{aligned} \quad (98)$$

Which can be rewritten to,

$$c_3 = \left(\frac{\gamma(T-N-1)(T-N-4)}{T(T-2)} \right) \cdot \frac{(\frac{1}{\gamma} \iota_N' \Sigma^{-1} \mu - d_3)}{\iota_N' \Sigma^{-1} \iota_N}. \quad (99)$$

Now, the process is repeated but for d_2 . By taking the partial derivative of the expected utility with respect to d_3 and equating to zero,

$$\frac{\partial E[\tilde{U}(\hat{w}_t^{3fEG})]}{\partial d_3} = \frac{1}{N} \iota'_N \mu - \frac{\gamma d_3}{N^2} \iota'_N \Sigma \iota_N - c_3 \left(\frac{T}{T-N-2} \right) = 0. \quad (100)$$

This can be shown to equal,

$$d_3 = \frac{\iota'_N \mu}{\gamma \iota'_N \Sigma \iota_N} - \left(\frac{T}{T-N-2} \right) \cdot \frac{N}{\gamma} \cdot \frac{c_3}{\iota'_N \Sigma \iota_N}. \quad (101)$$

The expression of d_3 is substituted into c_3 to obtain the following estimate for the optimal shrinkage intensity towards the Global Minimum Variance portfolio,

$$c_3^* = \frac{\iota'_N \Sigma \iota_N \cdot \iota'_N \Sigma^{-1} \mu - \iota'_N \mu}{\frac{T(T-2)(\iota'_N \Sigma \iota_N)(\iota'_N \Sigma^{-1} \iota_N)}{(T-N-1)(T-N-4)} - \frac{NT}{T-N-2}}. \quad (102)$$

Finally, substituting c_3^* back into d_3 provides an expression for the optimal shrinkage intensity towards the equally weighted portfolio,

$$d_3^* = \frac{1}{\gamma \iota'_N \Sigma \iota_N} \left[\iota'_N \mu - \left(\frac{NT}{T-N-2} \right) \cdot \frac{\iota'_N \Sigma \iota_N \cdot \iota'_N \Sigma^{-1} \mu - \iota'_N \mu}{\frac{T(T-2)(\iota'_N \Sigma \iota_N)(\iota'_N \Sigma^{-1} \iota_N)}{(T-N-1)(T-N-4)} - \frac{NT}{T-N-2}} \right]. \quad (103)$$

D Bayesian Inference for the Diffuse Prior

In this section an alternative derivation of the predictive posterior distribution of returns given a diffuse prior is presented. The results correspond with and are provided in the works of Klein and Bawa (1976), Bauder et al. (2020) and Avramov et al. (2009).

First, derive the likelihood function of the returns. As before, let r_t denote a vector of returns for the N assets. The likelihood of the N -dimensional multivariate normal population over sample $t = 1, \dots, T$ is given by the product over t ,

$$\begin{aligned} p(r|\mu, \Sigma) &= \prod_{t=1}^T p(r_t|\Phi_t), \\ &\propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \sum_{t=1}^T (r_t - \mu)' \Sigma^{-1} (r_t - \mu)\right), \\ &\propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} tr(S \Sigma^{-1}) - \frac{T}{2} (\mu - \hat{\mu})' \Sigma^{-1} (\mu - \hat{\mu})\right), \end{aligned} \quad (104)$$

with $S = \sum_{t=1}^T (r_t - \hat{\mu})(r_t - \hat{\mu})'$, the sample covariance multiplied by T , and $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T r_t$. In the second step the decomposition rule¹³ was used along with the trace operator $tr(x)$. Next, define Jeffrey's diffuse prior as,

$$p(\mu, \Sigma) \propto |\Sigma|^{-\frac{N+1}{2}}. \quad (105)$$

¹³ $(Y - X\Pi)'(Y - X\Pi) = (Y - X\hat{\Pi})'(Y - X\hat{\Pi}) + (\Pi - \hat{\Pi})'(X'X)(\Pi - \hat{\Pi})$.

The joint posterior can be found by applying bayes rule,

$$\begin{aligned} p(\mu, \Sigma|r) &\propto p(r|\mu, \Sigma)p(\mu, \Sigma), \\ &\propto |\Sigma|^{-\frac{N+T+1}{2}} \exp\left(-\frac{1}{2}tr(S\Sigma^{-1}) - \frac{T}{2}(\mu - \hat{\mu})'\Sigma^{-1}(\mu - \hat{\mu})\right). \end{aligned} \quad (106)$$

Next, the marginals posteriors are derived. This is done by integrating the joint posterior with respect to the other parameter. For μ the marginal posterior distribution is,

$$\begin{aligned} p(\mu|r) &\propto \int_{\Sigma} p(\mu, \Sigma|r)d\Sigma, \\ &\propto \int_{\Sigma} |\Sigma|^{-\frac{N+T+1}{2}} \exp\left(-\frac{1}{2}tr(S\Sigma^{-1}) - \frac{T}{2}(\mu - \hat{\mu})'\Sigma^{-1}(\mu - \hat{\mu})\right)d\Sigma, \\ &\propto |T(\mu - \hat{\mu})(\mu - \hat{\mu})' + S|^{-\frac{T}{2}}, \end{aligned} \quad (107)$$

where the last step follows from the integral over the density of an Inverse-Wishart distribution with $(N + T + 1)$ degrees of freedom and $T(\mu - \hat{\mu})(\mu - \hat{\mu})' + S$ as parameter matrix. The resulting distribution is a N-dimensional multivariate t-distribution with $(T - N)$ degrees of freedom, $\hat{\mu}$ as location vector and $\frac{1}{(T-N)}S$ as scale matrix (Bauder et al., 2020). Analogously, the same procedure can be repeated but now integrating over μ to obtain the marginal distribution of Σ ,

$$\begin{aligned} p(\Sigma|r) &\propto \int_{\mu} p(\mu, \Sigma|r)d\mu, \\ &\propto \int_{\mu} |\Sigma|^{-\frac{N+T+1}{2}} \exp\left(-\frac{1}{2}tr(S\Sigma^{-1}) - \frac{T}{2}(\mu - \hat{\mu})'\Sigma^{-1}(\mu - \hat{\mu})\right)d\mu, \\ &\propto |\Sigma|^{-\frac{N+T+1}{2}} \exp\left(-\frac{1}{2}tr(S\Sigma^{-1})\right) \underbrace{\left|\frac{1}{T}\Sigma\right|^{+\frac{1}{2}} \int_{\mu} \left|\frac{1}{T}\Sigma\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mu - \hat{\mu})'\left(\frac{1}{T}\Sigma\right)^{-1}(\mu - \hat{\mu})\right) d\mu}_{\sim \mathcal{N}(\hat{\mu}, \frac{1}{T}\Sigma)}, \\ &\propto |\Sigma|^{-\frac{N+T+1}{2}} \exp\left(-\frac{1}{2}tr(S\Sigma^{-1})\right) \left|\frac{1}{T}\Sigma\right|^{+\frac{1}{2}}, \\ &\propto |\Sigma|^{-\frac{2N+T+1}{2}} \exp\left(-\frac{1}{2}tr(S\Sigma^{-1})\right). \end{aligned} \quad (108)$$

Where the integral over a multivariate normal distribution is equal to 1 was used. The resulting distribution is shown to be Inverse-Wishart with $(T + N)$ degrees of freedom, and parameter $T(\mu - \hat{\mu})(\mu - \hat{\mu})' + S$.

Taking into account that returns are multivariate normal IID, the distribution for one step ahead return r_{t+1} is of that type. Consequently, the predictive distribution can be defined as,

$$p(r_{t+1}|\mu, \Sigma, \Phi_t) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(r_{t+1} - \mu)'\Sigma^{-1}(r_{t+1} - \mu)\right), \quad (109)$$

$$\propto |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}tr(\Sigma^{-1}(r_{t+1} - \hat{\mu})(r_{t+1} - \hat{\mu})') - \frac{1}{2}tr(\Sigma^{-1}(\mu - \hat{\mu})(\mu - \hat{\mu})')\right). \quad (110)$$

Where I have used the decomposition rule. Now the posterior predictive distribution can be calculated by,

$$p(r_{t+1}|\Phi_t) = \int_{\Sigma} \int_{\mu} p(r_{t+1}|\mu, \Sigma, \Phi_t) p(\mu, \Sigma|\Phi_t) d\Sigma d\mu, \quad (111)$$

$$\begin{aligned} &\propto \int_{\mu} \int_{\Sigma} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(r_{t+1} - \mu)' \Sigma^{-1} (r_{t+1} - \mu)\right) \\ &\times |\Sigma|^{-\frac{N+T+1}{2}} \exp\left(-\frac{1}{2} \text{tr}(S\Sigma^{-1}) - \frac{T}{2}(\mu - \hat{\mu})' \Sigma^{-1} (\mu - \hat{\mu})\right) d\mu d\Sigma. \end{aligned} \quad (112)$$

The expression can be simplified and rearranged to show that,

$$\begin{aligned} p(r_{t+1}|\Phi_t) &\propto \int_{\Sigma} \int_{\mu} |\Sigma|^{-\frac{N+T+2}{2}} \exp\left(-\frac{1}{2} \text{tr}((r_{t+1} - \hat{\mu})(r_{t+1} - \hat{\mu})\Sigma^{-1}) - \frac{1}{2} \text{tr}(S\Sigma^{-1}) - \frac{T+1}{2} \text{tr}((\mu - \hat{\mu})(\mu - \hat{\mu})\Sigma^{-1})\right) d\mu d\Sigma, \\ &\propto \int_{\Sigma} |\Sigma|^{-\frac{N+T+2}{2}} \exp\left(-\frac{1}{2}(\text{tr}((r_{t+1} - \hat{\mu})(r_{t+1} - \hat{\mu})\Sigma^{-1}) - \frac{1}{2} \text{tr}(S\Sigma^{-1}))\right) \left|\frac{1}{T+1}\Sigma\right|^{\frac{1}{2}} \\ &\times \underbrace{\int_{\mu} \left|\frac{1}{T+1}\Sigma\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mu - \hat{\mu})\left(\frac{1}{T+1}\Sigma\right)^{-1}(\mu - \hat{\mu})\right) d\mu}_{\sim \mathcal{N}(\hat{\mu}, \frac{1}{T+1}\Sigma)} d\mu d\Sigma. \end{aligned} \quad (113)$$

By using that the integral over a (Normal) distribution sums to one, the expression simplifies to,

$$\begin{aligned} p(r_{t+1}|\Phi_t) &\propto \int_{\Sigma} |\Sigma|^{-\frac{N+T+1}{2}} \exp\left(-\frac{1}{2}(\text{tr}((r_{t+1} - \hat{\mu})(r_{t+1} - \hat{\mu})\Sigma^{-1}) - \frac{1}{2} \text{tr}(S\Sigma^{-1}))\right) d\Sigma, \\ &\propto |S + (r_{t+1} - \hat{\mu})(r_{t+1} - \hat{\mu})'|^{-\frac{T}{2}}. \end{aligned} \quad (114)$$

Where the integration over an Inverse-Wishart distribution was used. Finally, using the assumption of IID multivariate returns, the solution can be rewritten in the same form as provided in Avramov et al. (2009),

$$\propto |\hat{\Sigma} + (r_{t+1} - \hat{\mu})(r_{t+1} - \hat{\mu})' / (T + 1)|^{-\frac{T}{2}}. \quad (115)$$

The posterior predictive distribution is recognised as a multivariate t with $(T - N)$ degrees of freedom, $\hat{\mu}$ as scale parameter and $\frac{T+1}{T-N}\hat{\Sigma}$ as scale matrix.

E Bayesian Inference for the Conjugate Prior

In this section, a derivation of the predictive posterior distributions of returns given a conjugate prior is presented. As returns are often assumed to follow a normal distribution, such an analysis is often used in portfolio optimisation literature. The derivation follows from the works of Murphy (2007).

Recall that for a conjugate prior, the posterior distribution $p(\mu, \Sigma|r_t)$ is of the same family of distributions as the priors $p(\mu, \Sigma)$. As is common in financial econometrics, a Normal-Inverse-Wishart conjugate prior for the mean and covariance parameters is opted for. Furthermore the returns are assumed to follow a multivariate normal distribution. As in Section D,

the likelihood of the N-dimensional multivariate normal population over sample $t = 1, \dots, T$ is given by,

$$\begin{aligned} p(r|\mu, \Sigma) &= \prod_{t=1}^T p(r_t|\Phi_t), \\ &\propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \sum_{t=1}^T (r_t - \mu)' \Sigma^{-1} (r_t - \mu)\right). \\ &\propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \text{tr}(S \Sigma^{-1}) - \frac{T}{2} (\mu - \hat{\mu})' \Sigma^{-1} (\mu - \hat{\mu})\right), \end{aligned} \quad (116)$$

with $S = \sum_{t=1}^T (r_t - \hat{\mu})(r_t - \hat{\mu})'$, the sample covariance and $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T r_t$. In the second step the decomposition rule (as shown in 13) was used along with the trace operator, $\text{tr}(x)$. Next, specify the NIW priors for μ and Σ as,

$$\Sigma \sim IW_{\nu_0}(\Sigma_0), \quad (117)$$

$$\mu|\Sigma \sim \mathcal{N}(\mu_0, \frac{1}{\kappa_0} \Sigma). \quad (118)$$

These can be taken together to form a joint prior for μ and Σ ,

$$\begin{aligned} p(\mu, \Sigma) &:= NIW(\mu_0, \kappa_0, \Sigma_0, \nu_0), \\ &= \frac{1}{Z} |\Sigma|^{-((\nu_0+N+2)/2)} \exp\left(-\frac{1}{2} \text{tr}(\Sigma_0 \Sigma^{-1}) - \frac{\kappa_0}{2} (\mu - \mu_0)' \Sigma^{-1} (\mu - \mu_0)\right). \end{aligned} \quad (119)$$

With $Z = \frac{2^{\nu_0 N/2} \Gamma_N(\nu_0/2) (2\pi/\kappa_0)^{N/2}}{|\Sigma_0|^{\nu_0/2}}$. Bayes rule can then be applied to derive the posterior distribution,

$$\begin{aligned} p(\mu, \Sigma|r) &\propto p(r|\mu, \Sigma) p(\mu, \Sigma), \\ &\propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \text{tr}(S \Sigma^{-1}) - \frac{T}{2} (\mu - \hat{\mu})' \Sigma^{-1} (\mu - \hat{\mu})\right) \\ &\times |\Sigma|^{-((\nu_0+N+2)/2)} \exp\left(-\frac{1}{2} \text{tr}(\Sigma_0 \Sigma^{-1}) - \frac{\kappa_0}{2} (\mu - \mu_0)' \Sigma^{-1} (\mu - \mu_0)\right), \\ &\propto |\Sigma|^{-((\nu_0+N+T+2)/2)} \exp\left(-\frac{1}{2} \text{tr}(S \Sigma^{-1}) - \frac{T}{2} (\mu - \hat{\mu})' \Sigma^{-1} (\mu - \hat{\mu}) - \frac{1}{2} \text{tr}(\Sigma_0 \Sigma^{-1}) - \frac{\kappa_0}{2} (\mu - \mu_0)' \Sigma^{-1} (\mu - \mu_0)\right). \end{aligned} \quad (120)$$

This can be shown to be Normal Inverse-Wishart distributed (recall the conjugate nature). The joint posterior is thus,

$$p(\mu, \Sigma|r_t, \mu_0, \kappa_0, \Sigma_0, \nu_0) = NIW(\mu, \Sigma|\mu_T, \kappa_T, \Sigma_T, \nu_T), \quad (121)$$

with parameters defined as,

$$\mu_T = \frac{\kappa_0 \mu_0 + T \hat{\mu}}{\kappa_T + T} = \frac{\kappa_0}{\kappa_0 + T} \mu_0 + \frac{T}{\kappa_0 + T} \hat{\mu}, \quad (122)$$

$$\kappa_T = \kappa_0 + T, \quad (123)$$

$$\nu_T = \nu_0 + T, \quad (124)$$

$$\Sigma_T = \Sigma_0 + S + \frac{\kappa_0 T}{\kappa_0 + T} (\mu_0 - \hat{\mu})(\mu_0 - \hat{\mu})'. \quad (125)$$

By integrating out the other parameter, the marginal posteriors can be derived. Taking the joint posterior as the defined

Normal Inverse-Wishart simplifies the notation. The marginal posterior for μ becomes,

$$\begin{aligned}
p(\mu|r) &\propto \int_{\Sigma} p(\mu, \Sigma|r) d\Sigma, \\
&\propto \int_{\Sigma} |\Sigma|^{-((\nu_T+N+2)/2)} \exp\left(-\frac{1}{2}tr(\Sigma_T \Sigma^{-1}) - \frac{\kappa_T}{2}(\mu - \mu_T)' \Sigma^{-1}(\mu - \mu_T)\right) d\Sigma, \\
&\propto |\Sigma_T + \kappa_T(\mu - \mu_T)(\mu - \mu_T)'|^{-\frac{1}{2}(\nu_T+N-T+1)} \\
&\propto |1 + (\kappa_T + N)(\mu - \mu_T)' \Sigma_T^{-1}(\mu - \mu_T)|^{-\frac{1}{2}(\nu_T-T+N+1)}.
\end{aligned} \tag{126}$$

Integration over an Inverse-Wishart distribution was used in the second step. The resulting expression can be recognised as a multivariate t distribution with $(\nu_T + N - T + 1)$ degrees of freedom, μ_T as scale parameter, and $\Sigma_T \frac{1}{\kappa_T(\nu_T - N + 1)}$ as scale matrix. Similarly, the marginal posterior for Σ can be found by integrating over μ ,

$$\begin{aligned}
p(\Sigma|r) &\propto \int_{\mu} p(\mu, \Sigma|r) d\mu, \\
&\propto \int_{\mu} |\Sigma|^{-((\nu_T+N+2)/2)} \exp\left(-\frac{1}{2}tr(\Sigma_T \Sigma^{-1}) - \frac{\kappa_T}{2}(\mu - \mu_T)' \Sigma^{-1}(\mu - \mu_T)\right) d\mu, \\
&\propto |\Sigma|^{-((\nu_T+N+2)/2)} \exp\left(-\frac{1}{2}tr(\Sigma_T \Sigma^{-1})\right) \underbrace{\left| \frac{1}{\kappa_T} \Sigma \right|^{+\frac{1}{2}} \int_{\mu} \left| \frac{1}{\kappa_T} \Sigma \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mu - \mu_T)' \left(\frac{1}{\kappa_T} \Sigma\right)^{-1}(\mu - \mu_T)\right) d\mu}_{\sim \mathcal{N}(\mu_T, \frac{1}{\kappa_T} \Sigma)}, \\
&\propto |\Sigma|^{-((\nu_T+N+1)/2)} \exp\left(-\frac{1}{2}tr(\Sigma_T \Sigma^{-1})\right).
\end{aligned} \tag{127}$$

This can be recognised as an Inverse-Wishart distribution with distribution with ν_T degrees of freedom and parameter Σ_T .

Finally, the posterior predictive distribution is derived. For that use the definition of the predictive distribution as given in Avramov et al. (2009) and Zellner and Chetty (1965),

$$p(r_{t+1}|\Phi_t) = \int_{\mu} \int_{\Sigma} p(r_{t+1}|\mu, \Sigma, \Phi_t) p(\mu, \Sigma|\Phi_t) d\Sigma d\mu. \tag{128}$$

That is, integrate out the uncertainty in the parameters over the product of the predictive distribution $p(r_{t+1}|\mu, \Sigma, \Phi_t)$, that is multivariate normal,

$$p(r_{t+1}|\mu, \Sigma, \Phi_t) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(r_{t+1} - \mu)' \Sigma^{-1}(r_{t+1} - \mu)\right). \tag{129}$$

and the earlier defined joint-posterior, which is Normal Inverse-Wishart distributed (see (121)). The product of the two results in another Normal Inverse-Wishart, but with differently defined parameters. It is given by,

$$p(r_{t+1}|\mu, \Sigma, \Phi_t) p(\mu, \Sigma|\Phi_t) = NIW(\mu, \Sigma | \mu_T^*, \kappa_T^*, \Sigma_T^*, \nu_T^*), \tag{130}$$

with parameters,

$$\mu_T^* = \frac{\kappa_T \mu_T + \hat{\mu}}{\kappa_T + 1} = \frac{\kappa_T}{\kappa_T + 1} \mu_T + \frac{1}{\kappa_T + 1} \hat{\mu}, \quad (131)$$

$$\kappa_T^* = \kappa_T + 1, \quad (132)$$

$$\nu_T^* = \nu_T + 1, \quad (133)$$

$$\Sigma_T^* = \Sigma_T + S + \frac{\kappa_T}{\kappa_T + 1} (\mu_T - \hat{\mu})(\mu_T - \hat{\mu})'. \quad (134)$$

Where the first and second moments $\hat{\mu}$ and S are the same due to the assumption of IID normal returns. Taking the integral with regards to μ first can simplify the derivation,

$$\begin{aligned} p(r_{t+1} | \Phi_t) &= \int_{\Sigma} \int_{\mu} |\Sigma|^{-((\mu_T^* + N + 2)/2)} \exp\left(-\frac{1}{2} \text{tr}(\Sigma_T^* \Sigma^{-1}) - \frac{\kappa_T^*}{2} (\mu - \mu_T^*)' \Sigma^{-1} (\mu - \mu_T^*)\right) d\mu d\Sigma, \\ &\propto \int_{\Sigma} |\Sigma|^{-((\mu_T^* + N + 2)/2)} \exp\left(-\frac{1}{2} \text{tr}(\Sigma_T^* \Sigma^{-1})\right) \underbrace{\left| \frac{1}{\kappa_T^*} \Sigma \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\mu - \mu_T^*)' \left(\frac{1}{\kappa_T^*} \Sigma\right)^{-1} (\mu - \mu_T^*)\right)}_{\sim \mathcal{N}(\mu_T^*, \frac{1}{\kappa_T^*} \Sigma)} d\mu d\Sigma, \\ &\propto \int_{\Sigma} |\Sigma|^{-((\nu_T + N + 1)/2)} \exp\left(-\frac{1}{2} \text{tr}(\Sigma_T \Sigma^{-1})\right) d\Sigma, \\ &\propto |\Sigma_T + S + \frac{\kappa_T}{\kappa_T + 1} (\mu_T - \hat{\mu})(\mu_T - \hat{\mu})'|^{-\frac{1}{2}(\nu_T^* + N)}. \end{aligned} \quad (135)$$

Where the integration step over an Inverse-Wishart distribution was used in the last step and integration over a Normal density equals to one. Accordingly, as shown in the works of Murphy (2007), the posterior predictive distribution of excess returns can then be shown to be t distributed,

$$p(r_{t+1} | \Phi_T) = t_{\nu_T - N + 1}(\mu_T, \frac{\Sigma_T (\kappa_T + 1)}{\kappa_T (\nu_T - N + 1)}). \quad (136)$$

F Bayesian Inference for the Hierarchical Bayes Gibbs Sampler

In this section, the derivation of the hierarchical Bayes Gibbs sampler is given. It is included for completeness and further specification of the methodology used in this investigation. The Gibbs sampler is derived as in Wolf (2018). I refer to this work along with that of Greyserman et al. (2006) for more detailed derivations.

As is common in financial economics, the likelihood of excess returns is taken to be multivariate normal. For the hierarchical model, the following priors are defined,

$$p(\mu | \tilde{\mu}_0, \Sigma) \sim \mathcal{N}(\tilde{\mu}_0, \Sigma / \kappa_0), \quad (137)$$

$$\Sigma \sim \text{Inv-Wishart}_{\nu_0}(\delta_0^{-2} P_0^{-1}), \quad (138)$$

$$\tilde{\mu}_0 \sim \text{Uniform}(-\infty, \infty), \quad (139)$$

$$\delta_0^2 \sim \text{Inv-Gamma}(\epsilon_1, \epsilon_2). \quad (140)$$

Observe that the priors for parameters μ and Σ are normal and Inverse-Wishart, analogous to the conjugate case. In the hierarchical specification however, distributions are adopted for the hyperparameters as well. Rather than constant, a uniform distribution is adopted for $\tilde{\mu}_0$, and an Inverse-Gamma distribution is taken for δ_0^2 .

The Bayes rule can be applied to obtain the joint posterior distribution. Multiplying the priors and likelihood the complete joint posterior is given by,

$$\begin{aligned}
p(\mu, \Sigma | r_t, \tilde{\mu}_0, \kappa_0, \Sigma_0, \nu_0) &= |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \text{tr}(S\Sigma^{-1}) - \frac{T}{2}(\mu - \hat{\mu})' \Sigma^{-1}(\mu - \hat{\mu})\right) \times \left|\frac{1}{\kappa_0} \Sigma\right|^{-\frac{1}{2}} \exp\left(-\frac{\kappa_0}{2}(\mu - \tilde{\mu}_0)' \Sigma^{-1}(\mu - \tilde{\mu}_0)\right) \\
&\times \text{unif}(-\infty, \infty) \times |\delta_0^{-2} P_0^{-1}|^{\frac{\nu_0}{2}} |\Sigma|^{-\frac{\nu_0 + N + 1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\delta_0^{-2} P_0^{-1} \Sigma^{-1})\right) \times (\delta_0^2)^{-(\epsilon_1 + 1)} \exp\left(-\frac{\epsilon_2}{\delta_0^2}\right).
\end{aligned} \tag{141}$$

The expression for the joint posterior is equivalent to that of the conjugate specification (equation 121) but multiplied by the ‘extra’ uniform and inverse-gamma distributions for the hyperparameters. Consequently, the joint posterior is not of a known type and one must rely on MCMC sampling to approximate it. This is done by iterative sampling from the individual conditional posterior distributions. I present the results of the conditionals as obtained in Greyserman et al. (2006).

The conditional of $\tilde{\mu}_0$ is normal distributed,

$$\begin{aligned}
p(\tilde{\mu}_0 | \mu, \Sigma, \delta_0^2, r_t) &\propto p(\mu | \tilde{\mu}_0, \frac{1}{\kappa_0} \Sigma) p(\tilde{\mu}_0), \\
&\propto \exp\left(-\frac{1}{2}(\mu - \tilde{\mu}_0)' \left(\frac{1}{\kappa_0} \Sigma\right)^{-1}(\mu - \tilde{\mu}_0)\right) \times \text{unif}(-\infty, \infty) \\
&\sim N\left(\mu, \frac{1}{\kappa_0} \Sigma\right).
\end{aligned} \tag{142}$$

The conditional of δ_0^2 is Inverse-Gamma distributed,

$$\begin{aligned}
p(\delta_0^2 | \mu, \Sigma, \tilde{\mu}_0, r_t) &\propto p(\Sigma | \delta_0^{-2} P_0^{-1}, \nu_0) p(\delta_0^2) \\
&\propto |\delta_0^{-2} P_0^{-1}|^{\frac{\nu_0}{2}} \exp\left(-\frac{1}{2} \text{tr}(\delta_0^{-2} P_0^{-1} \Sigma^{-1})\right) \times (\delta_0^2)^{-(\epsilon_1 + 1)} \exp\left(-\frac{\epsilon_2}{\delta_0^2}\right), \\
&\propto |\delta_0^{-2}|^{\frac{N\nu_0}{2} + \epsilon_1 + 1} \exp\left(-\frac{1}{2} \text{tr}(\delta_0^{-2} P_0^{-1} \Sigma^{-1}) - \frac{\epsilon_2}{\delta_0^2}\right),
\end{aligned} \tag{143}$$

where the determinant trick¹⁴ was used in the second line. Rearranging shows,

$$\begin{aligned}
p(\delta_0^2 | \mu, \Sigma, \tilde{\mu}_0, r_t) &\propto |\delta_0^{-2}|^{\frac{N\nu_0}{2} + \epsilon_1 + 1} \exp\left(-(\delta_0^{-2}) \left(\frac{\text{tr}(P_0^{-1} \Sigma^{-1})}{2} + \epsilon_2\right)\right) \\
&\sim \text{Inv-Gamma}\left(\epsilon_1 + \frac{N\nu_0}{2}, \epsilon_2 + \frac{\text{tr}(P_0^{-1} \Sigma^{-1})}{2}\right).
\end{aligned} \tag{144}$$

The conditional of Σ is Inverse-Wishart distributed,

$$\begin{aligned}
p(\Sigma | \mu, \delta_0^2, \tilde{\mu}_0, r_t) &\propto p(r_t | \Phi_T) p(\mu | \tilde{\mu}_0, \frac{1}{\kappa_0} \Sigma) p(\Sigma | \delta_0^{-2} P_0^{-1}, \nu_0) \\
&\propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \text{tr}(S\Sigma^{-1}) - \frac{T}{2}(\mu - \hat{\mu})' \Sigma^{-1}(\mu - \hat{\mu})\right) \times \left|\frac{1}{\kappa_0} \Sigma\right|^{-\frac{1}{2}} \exp\left(-\frac{\kappa_0}{2}(\mu - \tilde{\mu}_0)' \Sigma^{-1}(\mu - \tilde{\mu}_0)\right) \\
&\times |\Sigma|^{-\frac{\nu_0 + N + 1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\delta_0^{-2} P_0^{-1} \Sigma^{-1})\right) \\
&\propto |\Sigma|^{-\frac{\nu_0 + N + 2 + T}{2}} \exp\left(-\frac{1}{2} \text{tr}(S + T(\mu - \hat{\mu})(\mu - \hat{\mu})' + \kappa_0(\mu - \tilde{\mu}_0)(\mu - \tilde{\mu}_0)' + \delta_0^{-2} P_0^{-1}) \Sigma^{-1}\right) \\
&\sim \text{Inv-Wishart}\left(\nu_0 + T + 1, \frac{1}{\delta_0^2} P_0^{-1} + S + T(\mu - \hat{\mu})(\mu - \hat{\mu})' + \kappa_0(\mu - \tilde{\mu}_0)(\mu - \tilde{\mu}_0)'\right).
\end{aligned} \tag{145}$$

¹⁴The determinant trick refers to, $|cA| = c^n |A|$ in the case of $n \in \mathcal{R}$ and $A \in \mathcal{R}^{n \times n}$.

Finally, the conditional of μ is normal distributed,

$$\begin{aligned}
p(\mu|\Sigma, \delta_0^2, \tilde{\mu}_0, r_t) &\propto p(r_t|\Phi_T)p(\mu|\tilde{\mu}_0 \frac{1}{\kappa_0}\Sigma), \\
&\propto \exp\left(-\frac{T}{2}(\mu - \hat{\mu})'\Sigma(\mu - \hat{\mu}) - \frac{\kappa_0}{2}(\mu - \tilde{\mu}_0\iota)'\Sigma^{-1}(\mu - \tilde{\mu}_0\iota)\right) \\
&\propto \exp\left(-\frac{1}{2}\text{tr}(T(\mu - \hat{\mu})(\mu - \hat{\mu})' + \kappa_0(\mu - \tilde{\mu}_0\iota)(\mu - \tilde{\mu}_0\iota)')\Sigma^{-1}\right)
\end{aligned} \tag{146}$$

The brackets can be expanded and rearranged as follows,

$$\begin{aligned}
p(\mu|\Sigma, \delta_0^2, \tilde{\mu}_0, r_t) &\propto \exp\left(-\frac{1}{2}\text{tr}(T(\mu'\mu - 2\mu'\hat{\mu} + \hat{\mu}'\hat{\mu}) + \kappa_0(\mu'\mu - 2\mu'\tilde{\mu}_0\iota + (\tilde{\mu}_0\iota)'\tilde{\mu}_0\iota))\Sigma^{-1}\right) \\
&\propto \exp\left(-\frac{1}{2}\text{tr}((T + \kappa_0)\mu'\mu - 2\mu'\left(\frac{T}{T + \kappa_0}\hat{\mu} + \frac{\kappa_0}{T + \kappa_0}\tilde{\mu}_0\iota\right) + \frac{T}{T + \kappa_0}\hat{\mu}'\hat{\mu} + \frac{\kappa_0}{T + \kappa_0}(\tilde{\mu}_0\iota)'\tilde{\mu}_0\iota)\Sigma^{-1}\right) \\
&\propto \exp\left(-\frac{1}{2}(\mu - \mu_T)'\left(\frac{1}{T + \kappa_0}\Sigma\right)^{-1}(\mu - \mu_T)\right) \\
&\sim \mathcal{N}\left(\mu_T, \frac{\Sigma}{T + \kappa_0}\right),
\end{aligned} \tag{147}$$

with,

$$\mu_T = \frac{\kappa_0\tilde{\mu}_0\iota + T\hat{\mu}}{\kappa_0 + T} = \frac{\kappa_0}{\kappa_0 + T}\tilde{\mu}_0\iota + \frac{T}{\kappa_0 + T}\hat{\mu}. \tag{148}$$

Using the conditional posteriors, the Gibbs sampler can then be run as follows,

1. Set starting values for the parameters, $\theta_0 = (\Sigma^{(0)}\mu^{(0)}, \delta_0^{2(0)}, \tilde{\mu}_0^{(0)},)$
2. Update the parameters by means of simulating from the conditional distributions in order of hierarchy:
 - $\delta_0^{2(i+1)}$ from $p(\delta_0^2|\mu, \Sigma, \tilde{\mu}_0, r_t)$
 - $\tilde{\mu}_0^{(i+1)}$ from $p(\tilde{\mu}_0|\mu, \Sigma, \delta_0^2, r_t)$
 - $\Sigma^{(i+1)}$ from $p(\Sigma|\mu, \delta_0^2, \tilde{\mu}_0, r_t)$
 - $\mu^{(i+1)}$ from $p(\mu|\Sigma, \delta_0^2, \tilde{\mu}_0, r_t)$.
3. After convergence of the Markov Chain, all draws are consider draws of the joint posterior distribution, also for the parameters.

The draws of the joint posterior distribution can then be manipulated and used in the analysis.

G Specifications of the Markov Switching Gaussian Mixture Model

In this section model specifications of the Markov Switching Gaussian Mixture model as shown in Qian (2011) are presented. The provided specifications are included to provide the parameter definitions and support the intuition behind the model. Further derivations can be found in the Appendix of Qian (2011).

First, the parameters are introduced. As before, take asset returns as $r_t = (r_{1,t}, \dots, r_{N,t})'$ with $t = 1, \dots, T$, but now let these be driven by a latent Markov process with S regimes at every time t . Denote these regimes by $\tau_t \in \{1, \dots, S\}$. Also, for conciseness in notation, define the collection of returns over time as $r_1^T = \{r_t\}_{t=1}^T$. Lastly, a starting distribution for every

state S is defined as $\pi = (\pi_1, \dots, \pi_S)'$, with corresponding Markov chain regime-transition matrix,

$$Q = \begin{pmatrix} Q_1 \\ \vdots \\ Q_S \end{pmatrix} = \begin{pmatrix} Q_{1,1} & \dots & Q_{1,S} \\ \vdots & & \\ Q_{S,1} & \dots & Q_{S,S} \end{pmatrix}. \quad (149)$$

Conditional on the latent regime τ_t , r_t follows a Gaussian mixture with K latent states. The latent states $\lambda_t \in \{1, \dots, K\}$ are made to follow a multinomial distribution with regime specific probabilities equal to $\eta_S = (\eta_{S,1}, \dots, \eta_{S,K})'$. By taking the latent regime, τ_t , and latent state, λ_t , as known, the distribution of returns, r_t , can be defined. For that, take $\phi(\cdot)$ as the multivariate normal probability density function and $I(\cdot)$ as an indicator function that equals one when its input holds true. The distribution of returns can then be defined as,

$$P(r_t | \tau_t, \lambda_t) = \prod_{s=1}^S \prod_{k=1}^K [\phi(r_t; \mu_{s,k}, \Sigma_{s,k})]^{(I(\tau_t=s)) \cdot (I(\lambda_t=k))}. \quad (150)$$

The model allows for different means and covariances of excess returns dependent on the regime and state at time t . This is key in improving model robustness for changing economic conditions. Next, specify the priors for the model parameters. These are taken to be conjugate,

$$\mu_{s,k} \sim \mathcal{N}(b_{s,k}, V_{s,k}), \quad (151)$$

$$\Sigma_{s,k}^{-1} \sim \text{Wishart}_{\nu_{s,k}}(\Omega_{s,k}), \quad (152)$$

$$Q_s \sim \text{Dirichlet}(a_{s,1}, \dots, a_{s,S}), \quad (153)$$

$$\pi \sim \text{Dirichlet}(c_1, \dots, c_S), \quad (154)$$

$$\eta_s \sim \text{Dirichlet}(f_{s,1}, \dots, f_{s,K}), \quad (155)$$

where $s = 1, \dots, S$ and $k = 1, \dots, K$. The Dirichlet distribution is a conjugate distribution for categorical distributions, making it suitable for the choice of latent regimes, states and Markov probabilities. In this section I provide the full posterior conditionals for the model parameters as are given in the works of Qian (2011).

First, the full posterior conditional distribution of $\mu_{s,k}$ is shown to be normal distributed,

$$\mu_{s,k} | \cdot \sim \mathcal{N}(D_{s,k} d_{s,k}, D_{s,k}), \quad \text{where,} \quad (156)$$

$$D_{s,k} = [T_{s,k} \Sigma_{s,k}^{-1} + V_{s,k}^{-1}]^{-1}, \quad (157)$$

$$d_{s,k} = \Sigma_{s,k}^{-1} \sum_{t=1}^T [r_t \cdot I(\tau_t = s, \lambda_t = k)] + V_{s,k}^{-1} b_{s,k}, \quad (158)$$

$$T_{s,k} = \sum_{t=1}^T I(\tau_t = s, \lambda_t = k). \quad (159)$$

The posterior of $\mu_{s,k}$ is determined by its prior and the number of observations that are in regime and state, s and k , respectively.

Subsequently, the posterior of $\Sigma_{s,k}^{-1}$ (inverse covariance) is shown to be Wishart distributed. Mathematically,

$$\begin{aligned} \Sigma_{s,k}^{-1} | \cdot &\sim \text{Wishart}_{\tilde{\nu}_{s,k}}(\tilde{\Omega}_{s,k}), \quad \text{where,} \\ \tilde{\Omega}_{s,k} &= \{\Omega_{s,k}^{-1} + \sum_{t=1}^T [(r_t - \mu_{s,k})(r_t - \mu_{s,k})' \cdot I(\tau_t = s, \lambda_t = k)]\}^{-1}, \\ \tilde{\nu}_{s,k} &= \nu_{s,k} + T_{s,k}. \end{aligned} \tag{160}$$

It follows that the posterior for mixture probability η_s is Dirichlet as well,

$$\eta_s | \cdot \sim \text{Dirichlet}(f_{s,1} + T_{s,1}, \dots, f_{s,K} + T_{s,K}). \tag{161}$$

The latent states take on one of the discrete states running up until K . The posterior is shown to be,

$$P(\lambda_t = k | \cdot) \propto \prod_{s=1}^S [\eta_{s,k} \phi(r_t; \mu_{s,k}, \Sigma_{s,k})]^{I(\tau_t = s)}, \tag{162}$$

such that $\lambda_t = k | \cdot$ is a K multinomial distribution with the corresponding probabilities, η_S , defined above.

Next the posteriors relating to the Markov chain are provided. The initial Markov chain has posterior equal to,

$$\pi | \cdot \sim \text{Dirichlet}(c_1 + I(\tau_1 = 1), \dots, c_S + I(\tau_1 = S)), \tag{163}$$

and the posterior of the transition matrix of the Markov chain is,

$$Q_s | \cdot \sim \text{Dirichlet}(a_{s,1} + \tilde{T}_{s,1}, \dots, a_{s,S} + \tilde{T}_{s,S}). \tag{164}$$

Where the number of observations from regime s to regime j are captured in the term $\tilde{T}_{s,j} = \sum_{t=1}^T I(\tau_{t-1} = s, \tau_t = j)$ for all $j = 1, \dots, S$.

To sample latent regimes τ_t , Qian (2011) makes use of the Baum-Welch algorithm. The regimes are sampled forward and backward, similar to Chib (1996). Again, I provide their findings. The latent regimes are sampled in a forward sequence. A parameter $\theta = (\mu_{s,k}, \Sigma_{s,k}, \eta_s, Q_s, \pi)$ is defined which captures all model parameters. Given θ one can define the forward and backward variables.

Define the forward variable as

$$F_{t,s} = P(\tau_t = s, r_t | \theta, \lambda_1^T), t = 1, \dots, T \quad s = 1, \dots, S. \tag{165}$$

which can be computed by means of forward induction,

$$F_{t,s} = \left[\prod_{k=1}^K \phi(r_t; \mu_{s,k}, \Sigma_{s,k})^{I(\lambda_t = k)} \right] \cdot \sum_{l=1}^S F_{t-1,l} Q_{l,s}. \tag{166}$$

The backward variable is defined as,

$$B_{t,s} = P(r_{t+1}^T | \theta, \lambda_1^T, \tau_t = s), \quad (167)$$

which can be evaluated using backward induction,

$$B_{t,s} = \sum_{l=1}^S Q_{s,l} \cdot \left[\prod_{k=1}^K \phi(r_{t+1}; \mu_{l,k}, \Sigma_{l,k})^{I(\lambda_{t+1}=k)} \right] \cdot B_{t+1,l}. \quad (168)$$

Using that,

$$P(\tau_1^T | r_1^T, \theta, \lambda_1^T) = P(\tau_1 | r_1^T, \theta, \lambda_1^T) \cdot \prod_{t=2}^T P(\tau_t | \tau_{t-1}, r_1^T, \theta, \lambda_1^T), \quad (169)$$

τ_1^T can be sampled by the method of composition. This means that for each $s, l = 1, \dots, S$,

$$P(\tau_1 = s | r_1^T, \theta) \propto F_{1,s} \cdot B_{1,s}, \quad (170)$$

$$P(\tau_t = s | \tau_{t-1} = l, r_1^T, \theta) \propto Q_{l,s} \cdot \left[\prod_{k=1}^K \phi(r_t; \mu_{s,k}, \Sigma_{s,k})^{I(\lambda_t=k)} \right] \cdot B_{t,s}. \quad (171)$$

The above sampler makes use of the normal distribution and therefore does not capture the fat tailed nature off returns. To account for this, Qian (2011) pose a method to extend the Hidden Markov model to a Gaussian mixture form. Note how τ_1^T is sampled from conditional posteriors, where realisations of the latent states are used to make the term $\prod_{k=1}^K \phi(r_t; \mu_{s,k}, \Sigma_{s,k})^{I(\lambda_t=k)}$ simplify to a Gaussian distribution. If one where to the regimes and states in one block, that is τ_1^T and λ_1^T , and sample these together using the method of composition, nodes of the MCMC will be shortened and mixing property is improved (Qian, 2011). Consequently, in the block sampler, τ_1^T is first sampled from its posterior distribution without conditioning on the state, and then λ_1^T is sampled from its full conditional posterior. The forward and backward variables can be updated accordingly to give,

$$F_{t,s} = \left[\sum_{k=1}^K \eta_{s,k} \phi(r_t; \mu_{s,k}, \Sigma_{s,k}) \right] \cdot \sum_{l=1}^S F_{t-1,l} Q_{l,s}, \quad (172)$$

$$B_{t,s} = \sum_{l=1}^S Q_{s,l} \cdot \left[\sum_{k=1}^K \eta_{l,k} \phi(r_{t+1}; \mu_{l,k}, \Sigma_{l,k}) \right] \cdot B_{t+1,l}, \quad \text{and} \quad (173)$$

$$P(\tau_t = s | \tau_{t-1} = l, r_1^T, \theta) \propto Q_{l,s} \cdot \left[\sum_{k=1}^K \eta_{s,k} \phi(r_t; \mu_{s,k}, \Sigma_{s,k}) \right] \cdot B_{t,s}. \quad (174)$$

To address the identification problem of latent Markov models and Gaussian mixture models, I refer to Geweke and John (2007). In this paper, Geweke points out that a Gibbs sampler can be used for posterior inference when the posterior of interest is invariant to permutation. As I am interested in modeling the posterior predictive distribution of asset returns, which is not dependent on the exact regime/state label but rather an 'averaged' value over states and regimes, the identification issue does not cause a problem (Qian, 2011).

Finally, the order of sampling is presented. Given that

$$P(r_{t+1}, \tau_{t+1}, \lambda_{t+1}, \theta, \tau_T | r_1^T) = P(\theta, \tau_T | r_1^T) \cdot P(\tau_{t+1} | \theta, \tau_T) \cdot P(\lambda_{t+1} | \tau_{t+1}, \theta) \cdot P(r_{t+1} | \lambda_{t+1}, \tau_{t+1}, \theta), \quad (175)$$

it makes sense to define the sampling procedure as follows¹⁵:

1. Sample τ_{t+1} from $p(\tau_{t+1} | \tau_t, \theta)$,
2. Sample λ_{t+1} from $p(\lambda_{t+1} | \tau_{t+1}, \theta)$,
3. Sample Asset prices r_{t+1} from $P(r_{t+1} | \lambda_{t+1}, \tau_{t+1}, \theta)$.

The resulting predictive posterior distribution of excess asset returns is used to determine trading strategies that can account for parameter uncertainty, switching-regimes and non-normal asset returns.

H Models Out-of-Sample Performance for Fama-French 100 Portfolios

Table 9: Out-of-sample Performance of Direct Parameters Shrinkage models on Fama-French 100 Portfolios

	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
MW	0.004	0.742	0.089	-0.243	52.138	0.220
MW μ_{sh}	0.006	0.692	0.145*	-0.206	48.351	0.228
MW Σ_{sh}	0.005	0.658	0.118	-0.178	45.137	0.205
MW μ_{sh}, Σ_{sh}	0.007	0.615	0.171	-0.152	41.918	0.211
MW 3FF	0.007	0.792	0.148	-0.262	54.009	0.206
MW PCA	0.042	1.009	0.659	-0.486	39.451	0.248
GMV	0.034	0.693	0.779	-0.350	38.643	1.000
BL	0.007	0.790	0.148	-0.260	53.898	0.206
MW-BL	0.007	0.791	0.148	-0.261	53.899	0.206
EW	0.035	1.756	0.321	-1.209	5.116	0.990

Note: This table presents the performance measures for the parameter shrinking methods explained in section 2.2.1. The presented models regard the Markowitz portfolio (MW) with improved input parameters. MW μ_{sh} uses a shrunk estimate of expected returns, MW Σ_{sh} uses a shrunk estimate for covariance and MW μ_{sh}, Σ_{sh} uses both shrunk estimates. MW 3FF and MW PCA use input parameters based on estimates of the Fama-French 3-factor and PCA factor models, respectively. Furthermore, alternative portfolio optimisation models are considered. Provided are the Global Minimum Variance Portfolio (GMVP), Black-Litterman (BL), and a shrinkage variant of the Black-Litterman (MW-BL). The 1/N, equally weighted portfolio is also included as benchmark. The evaluation criteria are calculated for each strategy on daily returns of the Fama-French 100 portfolios data set over time frame June 27, 2006, until September 19, 2012. The reported performance measures for each strategy are: mean return, standard deviation, Sharpe ratio (SR), certainty equivalence (CEQ), turnover and Mean Weight invested in Risky assets (MWR). P-values for difference in Sharpe ratios with respect to Markowitz (MW) portfolio are derived from methods of Ledoit and Wolf (2008). *, **, *** represent 10%, 5% and 1% significance levels, respectively.

¹⁵Matlab code for the MCMC sampler of Qian (2011) can be found on <http://hangqian.weebly.com/research.html>.

Table 10: Out-of-Sample Performance of Weight Shrinkage Models on Fama-French 100 Portfolios

	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
2f	0.002	0.097	0.283	-0.007	6.773	-0.009
3f	0.015	0.353	0.665*	-0.060	22.714	0.379
3f EW	0.002	0.273	0.112	-0.021	18.150	0.092
3f EW x GMV	0.001	0.017	1.137**	0.000	0.865	0.017
EW	0.035	1.756	0.321	-1.209	5.116	0.990

Note: This table presents the performance measures for the weight shrinking methods explained in Section 2.3. 2f and 3f are the two- and three- fund rules of Kan and Zhou (2007). 3f EW is an optimal three-fund combination between the Markowitz portfolio and the equally weighted, and 3f EW x GMVP represents the optimal three-fund combination between the equally weighted and Global Minimum Variance portfolios. The equally weighted (EW) portfolio is also included as benchmark. The evaluation criteria are calculated for each strategy on daily returns of the Fama-French 100 portfolios data set over time frame June 27, 2006, until September 19, 2012. The reported performance measures for each strategy are: mean return, standard deviation, Sharpe ratio (SR), certainty equivalence (CEQ), turnover and Mean Weight invested in Risky assets (MWR). P-values for difference in Sharpe ratios with respect to Markowitz (MW) are derived from methods of Ledoit and Wolf (2008). *, **, *** represent 10%, 5% and 1% significance levels, respectively.

Table 11: Out-of-Sample Performance of Bayesian Models on Fama-French 100 Portfolios

	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
MW diff	0.003	0.591	0.089	-0.144	41.523	0.176
MW conj	0.006	0.152	0.589**	-0.007	9.805	0.099
MW Jor	0.007	0.320	0.339**	-0.043	20.618	0.176
MW hier	0.007	0.721	0.146*	-0.229	49.280	0.228
MW MSGM	0.077	0.564	2.155***	-0.077	65.012	0.166
GMVP conj	0.034	0.694	0.770	-0.350	38.592	1.000
GMVP hier	0.032	0.684	0.743	-0.343	36.269	1.000
GMVP MSGM	0.049	0.671	1.165*	-0.155	48.285	1.000
BL conj	0.017	0.522	0.508*	-0.136	32.052	0.232
BL hier	0.009	0.789	0.174	-0.261	54.024	0.206
BL MSGM	0.092	0.754	1.930***	-0.358	77.964	0.281
EW	0.035	1.756	0.321	-1.209	5.116	0.990

Note: This table presents the performance measures for the Bayesian models explained in section 2.4. Different priors are adopted for the optimal portfolios, Markowitz (MW), Global Minimum Variance Portfolio (GMVP) and the Black-Litterman (BL). The prior specifications are denoted by diffuse (diff), conjugate (conj), Jorion (Jor) and hierarchical (hier). Furthermore, MSGM refers to the Markov Switching Gaussian Mixture specification of the model. The equally weighted (EW) portfolio is included as a benchmark. The evaluation criteria are calculated for each strategy on daily returns of the Fama-French 100 portfolios data set over time frame June 27, 2006, until September 19, 2012. The reported performance measures for each strategy are: mean return, standard deviation, Sharpe ratio (SR), certainty equivalence (CEQ), turnover and Mean Weight invested in Risky assets (MWR). P-values for difference in Sharpe ratios with respect to Markowitz (MW) are derived from methods of Ledoit and Wolf (2008). *, **, *** represent 10%, 5% and 1% significance levels, respectively.

Table 12: Out-of-Sample Performance of Combination Models on Fama-French 100 Portfolios

	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
U EWC hier	0.052	1.572	0.524*	-3.962	77.302	2.381
U MWC hier	0.009	0.737	0.193*	-0.242	50.750	0.217
U WC hier	0.009	0.737	0.193*	-0.242	50.750	0.217
U EWC MSGM	0.073	0.492	2.338***	-0.077	50.356	0.482
U MWC MSGM	0.064	0.536	1.904***	-0.065	59.779	0.141
U WC MSGM	0.064	0.536	1.904***	-0.065	59.779	0.141
3f hier	0.016	0.311	0.796***	-0.052	18.807	0.396
3f MSGM	0.040	0.307	2.077***	-0.009	27.970	0.375
EW	0.035	1.756	0.321	-1.209	5.116	0.990

Note: This table presents the performance measures for the Bayesian portfolio combination methods explained in Section 2.5.2. The combinations are constructed using Bayesian implementations of Markowitz, Global Minimum Variance Portfolio and Black-Litterman. The Bayesian implementations are based on a hierarchical prior specification (hier) and a Markov Switching Gaussian Mixture (MSGM) model. The hierarchical combinations include the equally weighted (U EWC hier), maximum utility (U MWC hier) and utility weighted (U WC hier) combinations. The MSGM model combinations include the equally weighted (U EWC MSGM), maximum utility (U MWC MSGM) and utility weighted (U WC MSGM) combinations. Additionally, Bayesian implementations of the optimal three-fund rule of Kan and Zhou (2007) are also presented for hierarchical (3f hier) and MSGM specifications (3f MSGM). The equally weighted (EW) portfolio is included as a benchmark. Evaluation criteria are calculated for each strategy on daily returns of the Fama-French 100 portfolios data set over time frame June 27, 2006, until September 19, 2012. The reported performance measures for each strategy are: mean return, standard deviation, Sharpe ratio (SR), certainty equivalence (CEQ), turnover and Mean Weight invested in Risky assets (MWR). P-values for difference in Sharpe ratios with respect to Markowitz (MW) are derived from methods of Ledoit and Wolf (2008). *, **, *** represent 10%, 5% and 1% significance levels, respectively.

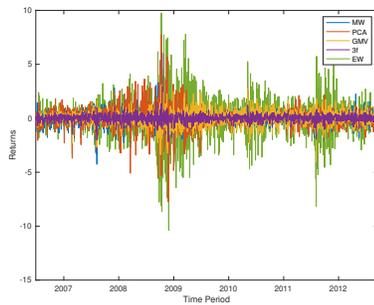
I Robustness Analysis of the Winner Markowitz Modifications for Fama-French 100 Portfolios

Table 13: Out-of-Sample Performance of Winner Markowitz Modifications for Subperiods of Fama-French 100 Portfolios

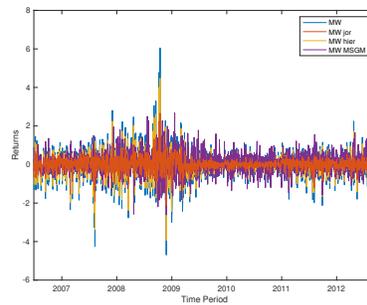
Pre Crisis	Strategy	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
	MW	-0.030	0.629	-0.755	-0.554	11.154	0.504
	GMV	0.061	0.646	1.489	-1.553	7.650	0.990
	MW PCA	0.098	0.296	5.252*	-0.551	3.701	1.000
	EW	0.102	0.817	1.987	-1.209	0.703	0.990
	3f	0.013	0.270	0.739	-0.143	4.424	0.446
	MW Jor	0.006	0.292	0.339	-0.142	4.918	0.402
	MW hier	-0.013	0.493	-0.404	-0.349	8.655	0.440
	MW MSGM	0.063	0.370	2.714	-0.078	13.137	0.168
	U EWC hier	0.476	1.592	4.744	-16.866	18.595	5.360
	U EWC MSGM	0.097	0.337	4.573	-0.150	11.157	0.631
	3f hier	0.023	0.209	1.755	-0.117	3.245	0.441
	3f MSGM	0.049	0.172	4.544	-0.010	5.285	0.376
Crisis	Strategy	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
	MW	0.032	0.915	0.553	-0.299	15.935	0.193
	GMV	0.076	1.332	0.910	-0.584	12.681	0.093
	MW PCA	-0.022	0.764	-0.456	-0.337	11.203	1.000
	EW	0.004	2.066	0.033	-1.209	1.544	0.990
	3f	0.009	0.400	0.349	-0.068	6.841	0.356
	MW Jor	0.014	0.401	0.566	-0.058	6.571	0.154
	MW hier	0.028	0.731	0.600	-0.180	12.367	0.162
	MW MSGM	0.094	0.654	2.277***	-0.077	18.098	0.166
	U EWC hier	-0.051	1.928	-0.420	-4.659	24.831	2.726
	U EWC MSGM	0.078	0.559	2.214***	-0.076	14.027	0.475
	3f hier	0.005	0.340	0.221	-0.056	5.542	0.373
	3f MSGM	0.036	0.347	1.645***	-0.009	7.905	0.375
Post Crisis	Strategy	Mean Return	Std. Dev	SR	CEQ	Turnover	MWR
	MW	-0.019	0.510	-0.580	-0.122	10.865	0.197
	GMV	0.000	0.544	-0.011	-0.176	7.682	0.279
	MW PCA	0.085	0.660	2.041**	-0.327	9.600	1.000
	EW	0.061	1.495	0.643	-1.209	1.163	0.990
	3f	0.023	0.308	1.167	-0.035	4.929	0.393
	MW Jor	-0.001	0.204	-0.059	-0.008	3.901	0.157
	MW hier	-0.015	0.399	-0.590	-0.066	8.361	0.167
	MW MSGM	0.061	0.481	2.012**	-0.076	15.677	0.164
	U EWC hier	0.087	1.021	1.354	-0.758	14.384	1.439
	U EWC MSGM	0.063	0.436	2.293**	-0.064	11.940	0.462
	3f hier	0.027	0.295	1.448*	-0.036	4.297	0.413
	3f MSGM	0.044	0.279	2.505***	-0.009	6.682	0.375

Note: This table presents the performance measures for the chosen winner MW portfolio modifications over subperiods of the Fama-French 100 portfolios data set. The presented frequentist models are Markowitz (MW), Global Minimum Variance Portfolio (GMVP), Markowitz based on Principal Component Analysis estimates (MW PCA) and the optimal three-fund rule (3f). The Markowitz portfolio and three-fund rule are also considered in a Bayesian setting. Markowitz using a Jorion (MW Jor) prior and both using hierarchical (MW hier, 3f hier) and Markov Switching Gaussian Mixture (MW MSGM, 3f MSGM) prior specifications. Equally weighted Bayesian portfolio combinations of Markowitz, Global Minimum Variance Portfolio and Black-Litterman are also considered for hierarchical (U EWC hier) and MSGM specifications (U EWC MSGM). The equally weighted (EW) portfolio is also included as benchmark. The out-of-sample period spans from June 27, 2006, until September 19, 2012. The time period is split up into pre-crisis (2006), crisis (2007-2009) and post-crisis (2010-2012) periods. The reported performance measures for each strategy are: mean return, standard deviation, Sharpe ratio (SR), certainty equivalence (CEQ), turnover and Mean Weight invested in Risky assets (MWR). P-values for difference in Sharpe ratios with respect to the EW are derived from methods of Ledoit and Wolf (2008). *, **, *** represent 10%, 5% and 1% significance levels respectively.

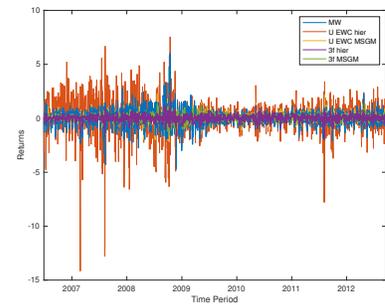
Figure 6: Performance Measures for Winner Markowitz Modifications over Time for Fama-French 100 portfolios



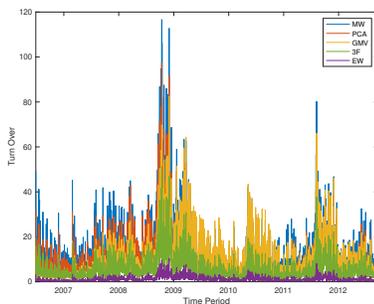
(a) Returns A



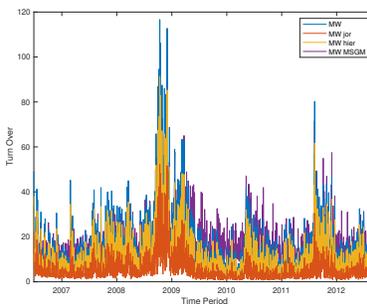
(b) Returns B



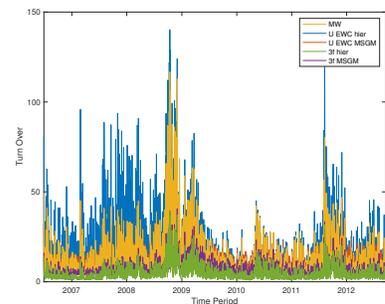
(c) Returns C



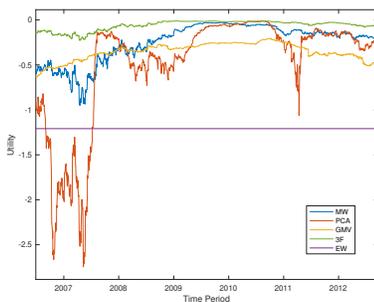
(d) Turnover A



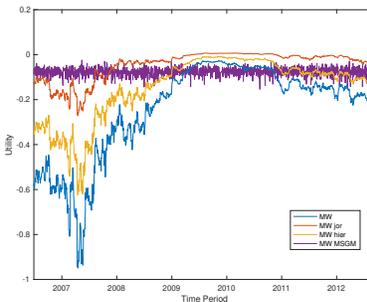
(e) Turnover B



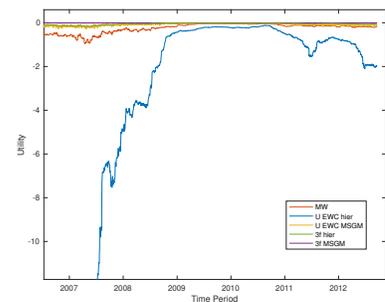
(f) Turnover C



(g) Utility A



(h) Utility B



(i) Utility C

Note: This figure shows the evaluation criteria for the winner Markowitz modification models over time. It presents the out-of-sample returns, turnover and corresponding utility for the period June 27, 2006 until September 19, 2012 on returns of assets from the Fama-French 100 portfolios (100FF) data set. Figures A provide performance of the frequentist Markowitz (MW) alternatives: Global Minimum Variance (GMV), equally weighted (EW) and three-fund rule (3f). Figures B provide the performance for Bayesian MW implementations: Jorion prior (Jor), hierarchical MW (MW hier) and Markov Switching Gaussian Mixture MW (MSGM MW). Figures C report the proposed Bayesian combination methods: Equally weighted combination of hierarchical MW, GMVP and Black-Litterman (U EWC hier), equally weighted combination of MSGM MW, GMVP and Black-Litterman (U EWC MSGM), hierarchical three-fund rule (3f hier) and MSGM three-fund (3f MSGM).