Robust Portfolio Selection by Bayesian Estimation Methods

ADNE VAN ENGELEN*

Abstract

This research proposes several performance-based measures to evaluate the estimation of a high-dimensional covariance matrix, where the number of variables can be close to the number of observations. In case of high-dimensional data, the sample covariance contains a lot of estimation error, and is not even invertible in some cases. This research makes a comparative analysis of covariance estimation based on different priors for the Bayesian estimation methodology. The performances of the resulting portfolios are judged in the realistic capital asset environment of the New York Stock Exchange. The sample covariance points out to be an unfit estimator in case of high-dimensional data. When a proper prior is chosen for the Bayesian estimation method, this method does not suffer from this dimensionality issue and clearly outperforms the natural estimator.

Nowadays the amount of money invested in investment funds is larger than ever. Most of this capital is invested in the financial market and the goal of large institutional investors like pension funds, mutual funds and insurance companies, is to invest their capital in such a way that the aims of their investors are fulfilled as good as possible. For different companies the investment strategies may vary, but something every company wants, is to predict the returns and riskiness of its investment portfolio properly. Therefore a lot of research is done during the last decades in order to try to determine and understand the characteristics of asset returns. Natural estimation methods are applied to determine sample moments of asset returns and then it became clear that it is difficult to determine the covariance matrix of asset returns properly. Bai and Shi (2011) pointed out that the natural way to estimate the covariance, by means of the sample covariance estimator, in some situations leads to estimation error in the parameters. Particularly if there are not much more observations (T) than assets (N).

So although this estimator is unbiased, it might contain estimation error and since the construction of a investment portfolio is mainly based on the assumed covariance matrix, this may lead to suboptimal asset allocation. Not only the expected return of the portfolio may differ from the realized return, also the riskiness of the portfolio may be estimated incorrectly, since wrong

*Van Engelen was an undergraduate student of Econometrics and Operations Research at the Erasmus School of Economics of the Erasmus University Rotterdam. I am grateful to Bart Keijsers of the Econometric Institute for his supervision of this project.

correlations between assets are assumed. This means that if the goal is to forecast the performance and riskiness of an investment portfolio in an appropriate way, it is necessary to use an estimation method that does not suffer from this dimensionality issue. A nice way to do this is by applying Bayesian estimation methods. Bayesian estimation methods namely not only tackle the dimensionality issue, but also give the investor the opportunity to incorporate beliefs about the assets in the estimation process. In the Bayesian estimation method a certain model is assumed to be the data generating process and the parameters in this model are considered as random variables. So beside realized returns on the assets in the past, this methodology also requires a specification for the distribution of the parameters in the underlying model for the returns; the so-called prior distribution. A detailed introduction to Bayesian statistics and an explanation of the Bayesian estimation method applied in the field of portfolio management is given in section I.

In this research the performance of the Bayesian estimation method is evaluated under different circumstances. Different prior distributions are considered and the effect on the estimation of the covariance matrix and the performances of the resulting minimum variance portfolios (Markowitz, 1952) are evaluated. The portfolio returns are also compared with the returns of portfolios based on the sample mean and sample covariance. This is done by using several econometric and economic evaluation criteria as in the paper of van Engelen et al. (2014).

The paper proceeds as follows. Section I gives an introduction to Bayesian statistics. Section II introduces the data used in this research. In section III both the natural and the Bayesian estimation methodology for the mean and covariance is presented. At the end of the section also Markowitz's Portfolio Theory is discussed for the determination of the portfolio weights. Section IV introduces the evaluation criteria. Section V presents the primary results. Section VI concludes.

I. Introduction to Bayesian Statistics

To be able to understand the theory behind the Bayesian estimation method, it is necessary to be familiar with Bayesian statistics. In Bayesian statistics population parameters θ are considered as random variables that have probability distributions. These probabilities measure the degree of belief in certain values of the variables and as a result the degree of belief in all states of nature are specified; these are non-negative and the total belief in all states of nature is fixed to one. Subsequently via the rules of probability - known as Bayes' Theorem (Bayes, 1763) - beliefs are revised, given the observed data: y. Bayes' Theorem is given by the following formula:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{1}$$

where $p(.)$ denotes a probability distribution and $p(.|.)$ a conditional distribution. When y represents the data and θ the parameters in a statistical model, Bayes' Theorem provides the basis for Bayesian inference. Within the context of Bayesian statistics, the assumed probability distribution of the parameters: $p(\theta)$ is called the prior distribution and the revised beliefs are denoted by the conditional distribtuion $p(\theta|y)$, which is called the posterior distribution.

To translate this formal definition to the field of portfolio management this research considers; assume there is just one asset and the following model is assumed to be the data generating process:

$$y_t = \theta + \varepsilon_t. \tag{2}$$

In this model y_t denotes the return at time t, θ is the 'population parameter' that can be seen as the expected return and ε_t is the idiosyncratic return at time t. Theoretically, the return on this asset in the next period can be any value between minus infinity and infinity. Altough it is not certain what the return will be, an investor may have so-called prior beliefs about it. For example the investor may belief that the average return on this asset is $\bar{m}$ and that the volatility is about $\bar{s}$. To specify beliefs about all 'states of nature' a certain distribution for the parameter θ must be chosen, such that all possible returns have a non-negative probability and all probabilities sum up to one. Therefore a possible probability specification for θ might be a normal distribution with mean $\bar{m}$ and variance $\bar{s}^2$. Subsequently observed data is used to revise these prior beliefs to get posterior beliefs. The revision of the beliefs is graphically illustrated in figure 1.

A. Graphical Illustration

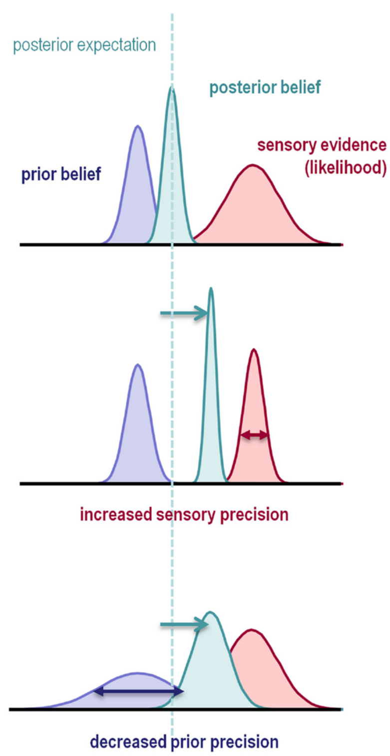

Figure 1. Graphical illustration of Bayesian estimation method

Since the Bayesian estimation method is not the most trivial way to estimate a parameter, a graphical illustration of the estimation process is presented in figure 1. In this figure three different situations are sketched. In the most upper part a quite informative prior is chosen. This means that the user is quite sure that the true parameter value is equal to or at least close to the value believed beforehand and is illustrated by a spiky prior distribution. According to the fat tails of the likelihood function it is clear that there is relatively much variation in the data. As discussed before the resulting posterior distribution is some kind of a combination of the prior distribution and the likelihood function and lies somewhere in between. Finally, from this posterior distribution a point estimate for the parameter of interest can be made, so in this research from the posterior of μ (Σ) an estimate for μ (Σ) is derived by using the moments of these distributions.

In the second situation the same prior is chosen, but there is less variation in the data. This is illustrated by thinner tails of the likelihood function. As a result the mode of the posterior distribution lies closer to the mode of the likelihood function and since there is less uncertainty about the true parameter value, the posterior distribution is more peaked as well.

The third situation shows a situation wherein the prior specification is less informative than in situation 1 and the variation in the data is similar. Since there is more uncertainty about the correctness of the prior specification, the posterior moves towards the likelihood function again. However due to this uncertainty about the prior and the variation in the data, the tails of the posterior distribution are big in this situation.

To summarize: in the Bayesian estimation process first a prior distribution for the parameters of interest is specified and subsequently this distribution is 'updated' by information contained in the data. This way an optimal combination between bias and variance is found by combining the investor's prior beliefs about the assets - which are mainly biased, but have very little variance - and the information in the past returns - which is captured in unbiased estimators with a relatively high variance. A step-by-step guide through this process is given by Idzorak (2004). The resulting distribution is called the posterior distribution and from this distribution characteristics of the parameters can be derived.

II. Data

The primary data used in this research comes from the Center for Research in Security Prices database. It consists of monthly returns on stocks that were continuously listed on the New York Stock Exchange (NYSE) between january 1998 and december 2012. From all stocks that met these requirements, at random 50 stocks were selected. To be able to apply the Capital Asset Pricing Model implied by Sharpe (1964) and the Fama & French Three Factor Model (Fama and French, 1993) also data on the market portfolio, on firms with small and large market capitalizations and on firms with high and low book-to-market ratios is collected. For ease and simplicity the riskfree interest rate is ignored by setting it equal to zero.

During this research a moving window approach is applied to evaluate the out-of-sample performance of the portfolios implied by the different estimation methods. To check for the influence of the dimensionality issue with the sample mean and covariance discussed by Bai and Shi (2011), different lengths of the moving window are considered. Window lengths of 60 and 100 are used to create situations in which $T \approx N$ and $T \gg N$, respectively.

III. Estimation Methodology

In this section a brief overview of the methodology applied in this research is given. In the first part of this section the natural way to estimate the mean and covariance is presented. The next subsection contains a detailed explanation of the Bayesian estimation method. To conclude the construction of the optimal portfolio based on Markowitz's Portfolio Theory is discussed.

A. Sample Mean and Sample Covariance Estimation Method

Before the more advanced Bayesian estimation method is discussed, the natural way to estimate the covariance is shown. Let R_t be a Nx1 vector of asset returns in period t. Then the estimates of the sample mean and sample covariance based on T time periods are respectively given by

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t, \tag{3}$$

$$\hat{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})'. \tag{4}$$

B. Bayesian Estimation Method

As discussed extensively in section I, the Bayesian estimation method is based on conditional probabilities and offers the user the opportunity to incorporate his or her prior beliefs in the estimation process. This is done by specifying the prior distribution of the parameters in the model, so the first step in the process is to specify the model that is assumed to be the underlying data generating process for the stock returns. In this research the following model is assumed:

$$R_t = \mu + \varepsilon_t, \tag{5}$$

where R_t is the vector of realized returns at time t, μ is the expected return vector and ε_t contains the specific returns on the assets at time t. Within this model the parameter μ is assumed to follow a certain distribution and so prior beliefs can be incorporated by chosing this distribution and setting certain 'starting values' for the parameters. In this research it is assumed that μ is normally distributed with parameters b and B:

$$\mu \sim N(b, B). \tag{6}$$

B.1. Prior Specifications

To be able to determine the posterior distribution analytically, in this research is chosen to work with conjugate priors for the expected return vector μ and the covariance matrix of the assets Σ. Consequently two informative priors are chosen, but to account for the uncertainty about Σ and the aim to find a proper posterior of it that is mainly based on the data, the least informative prior is used for Σ. This is a N-dimensional Wishart distribution for the precision matrix Σ^{-1}, with $v_0 = N$ degrees of freedom and scale matrix $R_0 = \frac{1}{v_0}\widetilde{\Sigma}^{-1}$. Here $\widetilde{\Sigma}$ might be any prior guess for the real covariance and therefore in this research three different priors are used[1]. The first is an identity matrix multiplied by a constant based on empirical research: λI_N. Empirical research shows an average volatility of about 20% on annual stock returns, which results in $\lambda = 3.333e^{-3}$ for monthly returns. For ease and simplicity no correlations between assets are incorporated in the prior in this situation. The second and third priors are more advanced and do incorporate correlations. They are based on the covariances implied by the Capital Asset Pricing Model and the Fama & French Three Factor Model, respectively. Both models are presented later in this section and are seen as descriptive models for asset returns. Since both models are factor models, the implied covariances should not suffer from the dimensionality issue and might be good priors for Σ.

For μ three more informative priors are considered. The first is based on empirical research again. That shows an average annual return on stocks of about 8%. This results in a vector in which each element equals $6.434e^{-3}$, because this research deals with monthly data. The second prior is the expected return vector based on the Capital Asset Pricing Model (CAPM) and the last is the expected return vector based on the Fama & French Three Factor Model. To regulate the influence of the prior specifications on the posterior distribution, the certainty about the prior for μ can be incorporated. This is done by defining B in equation (6) as cI_N and chosing a certain value for c. To ensure a nice trade-off between bias and variance this research uses $c = 1.000e^{-3}$.

B.2. Capital Asset Pricing Model

In the previous paragraphs different priors for μ and Σ are presented. The latter two are based on the CAPM and the Fama & French Three Factor Model, respectively and to show how these priors are found, these two models are presented in the next subsections. This paper starts with the CAPM. The CAPM is a univariate regression model in which stock returns are regressed on the market return. In this model it is assumed that investors are risk-averse and that they make their one-period investment decisions based solely on the expected returns and the volatility of the returns. The CAPM is written as

$$R_t = \alpha + \beta R_t^{market} + \varepsilon_t, \tag{7}$$

where R_t is the vector of realized returns, R_t^{market} is the realized return on the market portfolio

[1] This prior specification for Σ^{-1} corresponds with a N-dimensional inverse Wishart distribution for Σ with scale parameter $v^* = v_0 + N + 1$ and scale matrix $R^* = R_0^{-1}$.

and ε_t contains the idiosyncratic returns on the assets at time t. The priors for μ and Σ implied by this model therefore are given by:

$$\begin{aligned}\widetilde{\mu}_{CAPM} &= \hat{\alpha} + \hat{\beta}\bar{R}^{market} &(8)\\ \widetilde{\Sigma}_{CAPM} &= \hat{\beta}\hat{\beta}'\sigma^2_{market} &(9)\end{aligned}$$

respectively, in which $\hat{\alpha}$ and $\hat{\beta}$ denote the regression coefficients of equation (7), $\bar{R}^{market}$ is the average market return and σ^2_{market} denotes the variance of the market return.

B.3. Fama & French Three Factor Model

The next model that is applied to get prior specifications, is the Fama & French Three Factor Model. This is an extension of the CAPM. Beside the market return, also the return differentials between companies with small and big market capitalizations (R^{SMB}) and between companies with high and low book-to-market ratios (R^{HML}) are used as regressors in this model:

$$R_t = \alpha + \beta_1 R_t^{market} + \beta_2 R_t^{SMB} + \beta_3 R_t^{HML} + \varepsilon_t. \tag{10}$$

The priors for μ and Σ implied by this three factor model are:

$$\begin{aligned}\widetilde{\mu}_{FFM} &= \hat{\alpha} + \hat{\beta}_1\bar{R}^{market} + \hat{\beta}_2\bar{R}^{SMB} + \hat{\beta}_3\bar{R}^{HML} &(11)\\ \widetilde{\Sigma}_{FFM} &= \hat{\boldsymbol{\beta}}\Omega\hat{\boldsymbol{\beta}}' &(12)\end{aligned}$$

where $\hat{\boldsymbol{\beta}} = [\beta_1, \beta_2, \beta_3]$ denotes the Nx3 matrix of parameter estimates and Ω is the covariance matrix of the three factors.

B.4. Posterior distributions

Once the priors are specified, the next step is to update the prior beliefs about the parameters with information in the stock returns. After applying Bayes' Theorem it can be seen that the posterior distribution $\mu|R$ is proportional to the prior distribution of μ multiplied by the conditional distribution of $R|\mu$:

$$\begin{aligned}p(\mu|R) &= \frac{p(\mu, R)}{p(R)}\\ &= \frac{p(\mu)p(R|\mu)}{p(y)}\\ &\propto p(\mu)p(R|\mu). &(13)\end{aligned}$$

Since conjugate priors are chosen for μ and Σ, this is a known density function and it is possible to obtain the marginal posterior density functions of the parameters given the data: $p(\mu|R)$ and $p(\Sigma|R)$ analytically. This is done by integrating over all possible values of the other parameter and these distributions are of the same class as the prior distributions were, so

$$\begin{aligned} p(\mu|R) &= \int_{\Sigma} p(\mu, \Sigma|R) d\Sigma \qquad (14) \\ p(\Sigma|R) &= \int_{\mu} p(\mu, \Sigma|R) d\mu \qquad (15) \end{aligned}$$

denote the marginal posterior densities of μ and Σ, respectively. From (14) we can derive the 'updated' parameters of the distribution of μ: $b^{(1)}$ and $B^{(1)}$. An estimation of the covariance matrix of the assets, $\Sigma^{(1)}$, can be derived from (15). Together these estimates lead to the predictive density of R: $p(R|\mu^{(1)}, \Sigma^{(1)})$. To find the updated parameters different techniques can be applied. In this research the first moments of the posterior distributions for μ and Σ are used as point estimates. The posterior distribution for μ is

$$\mu|R, \Sigma \sim N_N(b^{(1)}, B^{(1)}), \qquad (16)$$

where

$$\begin{aligned} B^{(1)} &= \left[\sum_{t=1}^{T} I_N' \Sigma^{-1} I_N + B_0^{-1} \right]^{-1}, \qquad (17) \\ b^{(1)} &= B^{(1)} \left[\sum_{t=1}^{T} I_N' \Sigma^{-1} R_t + B_0^{-1} b_0 \right]^{-1}. \qquad (18) \end{aligned}$$

In definition (17) and (18) b_0 and B_0 denote the parameters of the prior distribution for μ (see equation 6). The posterior distribution for Σ is

$$\Sigma|R, \mu \sim W_N(v^{(1)}, R^{(1)}), \qquad (19)$$

where

$$\begin{aligned} v^{(1)} &= v_0 + T, \qquad (20) \\ R^{(1)} &= \left[R_0^{-1} + \sum_{t=1}^{T} (R_t - b^{(1)})(R_t - b^{(1)})' \right]^{-1}. \qquad (21) \end{aligned}$$

Using this results, point estimates for μ and Σ are given by

$$\hat{\mu} = b^{(1)}, \tag{22}$$

$$\hat{\Sigma} = v^{(1)} R^{(1)}. \tag{23}$$

The predictive density itself will not be used in this paper, because no forecasts will be made. In the portfolio optimization context discussed here, only the estimates of $\mu^{(1)}$ and $\Sigma^{(1)}$ are relevant as they are the inputs for Markowitz's theory to determine the optimal portfolio weights.

C. Markowitz's Mean-Variance Portfolio Theory

In the previous sections a detailed explanation is given how to estimate the covariance matrix of a set of asset returns. Now the next step is to find optimal portfolio weights corresponding to this covariance. The method used to determine these portfolio weights is based on Markowitz's Mean-Variance Portfolio Theory (Markowitz, 1952). The required inputs are the expected returns, the covariance matrix and a set of constraints. Constraints may be no short-selling, limitations on portfolio weights or what so ever.

The mean-variance principal can be expressed by the following problem

$$\begin{aligned} \underset{\omega}{min} \quad & \omega' \Sigma_t \omega \\ s.t. \quad & \omega' \mu_t \geq \mu^* && (24) \\ & \omega' \iota \ = 1 && (25) \end{aligned}$$

in which the variance of the portfolio is minimized for a given target return (24) and under the constraint that all capital is invested (25). In this optimization problem ω denotes the vector of portfolio weights, Σ_t the covariance matrix estimated at time t, μ_t the expected returns at time t, μ^* the target return and ι a vector of ones. Solving this problem results in the optimal weights stated in the following formula:

$$\omega^* = \frac{\Sigma_t^{-1} \mu_t}{\iota' \Sigma_t^{-1} \mu_t}. \tag{26}$$

C.1. Minimum Variance Portfolio

A special case of a mean-variance optimal portfolio is the minimum variance portfolio. Best and Grauer (1992) pointed out that estimation errors in the expected return vector are about ten times more influential than estimation errors in the covariance matrix. Since this research aims to focus on differences in estimation error in the covariance instead of the expected return, using the minimum variance portfolio helps to avoid this problem. The asset weights for this portfolio are found by solving the same minimizing problem as above, but without contraint (24) and are given by:

$$\omega^*_{MVP} = \frac{\Sigma_t^{-1}\iota}{\iota'\Sigma_t^{-1}\iota}. \tag{27}$$

IV. Evaluation Methodology

In this section the evaluation criteria for the minimum variance portfolios are presented. The evaluation criteria used in this research are the realized return, realized volatility and the ex post Sharpe Ratio.

A. Realized Return and Realized Volatility

The first evaluation criterion that is considered, is the realized return. In the end this what matters most for investors, because it determines the profit or loss they have made. The realized return of a portfolio is calculated by multiplying the portfolio weights chosen for a certain period t by the realized returns on the assets in that period:

$$RR_t^{port} = \omega_t' R_t. \tag{28}$$

This may differ from the expected return, which is given by $E[R_t^{port}] = \omega_t' E[R_t]$, and therefore the expected and realized return can be compared to check if predictions before investing were valid. The same can be done with the volatility of the portfolio returns. Before investing a volatility forecast of the portfolio is made by using the portfolio weights and the estimated covariance matrix:

$$E[\sigma_t] = \sqrt{\omega_t' \widehat{\Sigma}_t \omega_t}. \tag{29}$$

Again the realized volatility can - and in most situations does - differ from the expected volatility due to unexpected price changes in the assets. The realized volatility can easily be computed by formula (4) and so it is possible to compare the realized volatility over a time period with for example the average expected volatility over this period.

B. Ex Post Sharpe Ratio

Combining the two perfomance measures discussed above results in the ex post Sharpe ratio (Sharpe, 1966). This is a measure of the risk-adjusted return on a portfolio. A higher ratio means that a higher return is realized per unit of risk. Since the performance of an investment portfolio is mainly based on the composition of it, and for the minimum variance portfolio considered in this research holds that the composition solely depends on the estimated covariance matrix, the ex post Sharpe ratio is a good evaluation criterion for the estimation of the covariance. Under the assumpion of a riskfree interest rate of zero, the Sharpe ratio is given by

$$SR = \frac{1}{T} \sum_{t=1}^{T} \frac{RR_t^{port}}{\sigma_T}, \tag{30}$$

where RR_t^{port} denotes the realized return of the portfolio at time t and σ_T the realized volatility of the portfolio over the time period $t = 1, 2, ..., T$.

V. Results

After introducing the estimation methods for the covariance matrix and the portfolio weights, and showing the different evaluation criteria, in this section the results of the portfolios are discussed. However, before showing the out-of-sample performance of the portfolios an overview of the different priors that are used in the Bayesian estimation method is given.

A. *Prior Specifications*

One thing that is the same for all three prior specifications, are the prior distributions. To ensure known distributions for the posteriors, the following conjugate priors are chosen for μ and Σ, respectively:

$$\mu \sim N_N(b_0, B_0), \tag{31}$$

$$\Sigma \sim W_N(v_0, R_0). \tag{32}$$

The differences between the priors are in the 'start values' of the parameters b_0 and R_0. The certainty about the expected return vector b_0 is captured in B_0. This parameter is set equal to cI_N with $c = 1.000e^{-3}$ in all prior specifications. The value of c is chosen in such a way that a nice trade-off between bias and variance is ensured in the posterior. Because in this research is chosen to use the least informative prior for Σ to limit the influence of the prior as much as possible, but to ensure a known distribution of the posterior, v_0 is the same in each situation as well and equals N. The starting values for b_0 and R_0 can be found in table I, where a and σ^2_{month} are constants based on empiral research[2]. For monthly data that gives $a = 6.434e^{-3}$ and $\sigma^2_{month} = 3.333e^{-3}$.

Table I Prior specifications

	b_0	B_0	v_0	R_0
Situation 1	$a\iota_N$	cI_N	N	$\frac{1}{N}[\sigma^2_{month} I_N]^{-1}$
Situation 2	$\widetilde{\mu}_{CAPM}$	cI_N	N	$\frac{1}{N}\widetilde{\Sigma}^{-1}_{CAPM}$
Situation 3	$\widetilde{\mu}_{FFM}$	cI_N	N	$\frac{1}{N}\widetilde{\Sigma}^{-1}_{FFM}$

[2]Empirical research shows an average annual return of 8% on stocks. The average volatility is about 20%.

B. *Minimum Variance Portfolio Evaluation*

After specifying the priors the Bayesian estimation method can be applied to find the posteriors and finally get estimates for the expected returns and the covariance matrix. For this estimation process a moving window approach is applied. Each iteration a subsample of 60 (or 100) successive observations is used as dataset to find the posterior. From this posterior an estimation of the covariance is derived and this covariance is used to determine the portfolio weights of the minimum variance portfolio. Subsequently with these weights the expected return, realized return and expected volatility of the portfolio in the next period are determined. Since the data came from the period January 1998 - December 2012, 180 observations on each asset were available. This resulted in 120 (80) out-of-sample forecasts for a window length of 60 (100). From these forecasts the realized volatility and ex post Sharpe ratio are derived.

This procedure is followed for the three different prior specifications presented earlier in this section, as well as for the natural sample estimators presented in definition (3) and (4). To give a clear overview of the performances of the portfolios implied by the different methods, also the compounded return is given. In case a window length of 60 (100) is used, the compounded return is the amount of euros an investor that initially invested one euro in the minimum variance portfolio will have after 120 (80) months. This is under the assumption that after each month the current saldo is invested in the minimum variance portfolio again. As a side note it has to be mentioned that no transaction costs are considered. The results with window lengths of 60 and 100 are presented in tables II and III, respectively. All percentages in these tables are annualized.

Table II Performance Minimum Variance Portfolio

Window length: 60	Bayes Prior 1	Bayes Prior 2	Bayes Prior 3	Sample Estimators
Expected Return	10.18%	10.34%	10.10%	10.03%
Realized Return	5.788%	-0.063%	1.151%	-0.488%
Expected Volatility	6.176%	2.992%	2.982%	3.953%
Realized Volatility	11.24%	20.15%	20.40%	21.59%
Sharpe Ratio	0.1448	-0.0009	0.0162	-0.0065
Compounded Return	1.6467	0.8106	0.9108	0.7540

When we take a look at table II the first thing that points out, is that the expected return on the minimum variance portfolio is approximately the same for all estimation methods. However, the realized returns differ a lot. Only the portfolio implied by the Bayes estimation method with the prior from situation 1 shows a nice positive return that approximates the expected return to some extend. The other three portfolios show very low or even negative returns, which seems to be quite strange, since the realized return on the market index is about 15% in the same period.

While considering the expected and realized volatility, practically the same tendence is noticed. The only portfolio which volatility approximates the expected volatility a little bit, is the one based on the first prior. The remaining methods clearly underestimate the volatility and show very high realized volatilities. Especially if it is mentioned that the realized volatility on the market index is just 17% in the same time period and the aim was to find the portfolio with the lowest possible volatility: the minimum variance portfolio. These results are not striking for the portfolio based on the sample estimators. It is well known that these estimators suffer from the dimensionality issue implied by Bai and Shi (2011) if the number of assets is close to the number of observations. However, for the portfolios based on the Bayesian estimation method it is strange, since this method should not have this problem.

A possible explanation is found by inspecting the priors chosen for Σ. In both situation 2 and 3 the parameter R_0 is based on a covariance matrix implied by a factor model. Below the problem in situation 2 is discussed extensively. Situation 3 will not be discussed explicitly, but the reasoning why situation 3 does not lead to the desired outcome is quite similar. When the expressions for R_0 in table I and $\widetilde{\Sigma}_{CAPM}$ in equation (9) are investigated, it becomes clear that the vector $\hat{\beta}$ is very influential. In case this vector has two elements that are approximately the same, this will lead to a situation in which two columns of $\widetilde{\Sigma}_{CAPM}$ are approximately the same. A well known theorem of matrix algebra states that a square matrix is invertible if and only if all its columns are linearly independent. In this situation the columns are independent, since they are not exactly the same and therefore $\widetilde{\Sigma}_{CAPM}$ can be inverted and used as prior, but the inverse will contain a lot of estimation error. As can be seen in equations (17) and (18) this inverse occurs quite often in calculations in the Bayesian estimation process and subsequently resulting matrices are inverted as well. Due to this errors are compounded and the final estimate of the covariance becomes worthless.

Turning back to the results in the table and taking a look at the Sharpe ratios and the compounded returns, the same conclusion can be drawn. The first prior leads to a nice positive Sharpe ratio of 0.14, which means a good reward per unit of risk. The other methods have ratios close to or even smaller than zero. The compounded return gives approximately the same information as the realized return, but it points out that only the prior specification of situation 1 is not capital destructive. This seems to be in contradiction with the positive average return of prior 3, but can simply be explained by the fact that $(1-\alpha)(1+\alpha) < 1$, although the average return is zero.

Table III Performance Minimum Variance Portfolio

Window length: 100	Bayes Prior 1	Bayes Prior 2	Bayes Prior 3	Sample Estimators
Expected Return	9.894%	9.237%	9.523%	9.519%
Realized Return	3.590%	0.777%	2.123%	2.068%
Expected Volatility	6.923%	5.390%	5.351%	6.502%
Realized Volatility	12.15%	13.74%	14.13%	14.06%
Sharpe Ratio	0.0839	0.0163	0.0430	0.0420
Compounded Return	1.2033	0.9874	1.0748	1.0715

Next the results of portfolios based on the 100 most recent observations are discussed. These results are presented in table III and it is important to notice that these values cannot be compared one-to-one with values from table II, because the forecast period is not the same. The data in table II is based on returns of portfolios that are constructed over the period January 2003 - December 2012, while the data in table III is based solely on portfolio returns over the period May 2006 - December 2012.

When the first row of the table - with expected returns - is inspected, the conclusion is the same as for a window length of 60. There are no significant differences between the estimation methods. The realized returns and volatilities show again that prior specification 1 leads to the best portfolio performance. The difference with the other methods is not that big, though. In comparison with the situation in which $T \approx N$, it can be seen that the natural estimation method performs much better. Results are comparable with the other methods now. The second prior specification for the Bayesian estimation method shows the previously discussed dependency issue with the columns again. For the third prior this seems not to be dramatically any more. A reason can be that a larger sample to estimate the parameter matrix $\boldsymbol{\beta}$ from equation (12) leads to less uncertainty in the parameter estimates.

The Sharpe ratios by definition show approximately the same pattern as the realized returns and volatilities. Prior 1 performs best and the reward per unit of risk is almost twice as high as of the portfolio based on the sample estimators. Prior 2 shows a positive ratio this time, but this prior specification does not seem to be very useful, since it is even outperformed by the natural estimation method. The same tendence holds for the compounded returns, with that difference that the portfolio based on prior 2 is still capital destructive.

C. Adjusted Prior Specifications

While overlooking the results, the general conclusion is that the more advanced prior specifications of situation 2 and 3 do not lead to better results than the simple specification of situation 1. Therefore it is chosen to create two new priors that combine best of both situations. Priors 4 and 5 have the same prior specifications for μ as in situations 2 and 3, respectively, but the start values for R_0 differ. All elements that are not on the main diagonal are set equal to zero. This is to avoid a situation where the columns of R_0 are nearly dependent, while the advanced estimates of the volatilities can be used. Again window lengths of 60 and 100 are used to estimate the covariances and portfolio performances are compared with that of prior 1 and the sample estimator. Annualized results are presented in tables IV and V.

When table IV is analyzed, it points out that the results of the portfolios based on the new priors are approximately the same as of the portfolio based on prior 1. The expected returns are a bit higher and the expected volatilities lower, but these expectation are not proven. In table V though, it can be seen that for a window length of 100, not only the expected returns, but also the realized returns are higher for the portfolios based on the new priors. This might be due to more accurate parameter estimates in equations (7) and (10), because more data is available while performing ordinary least squares to get these estimates. As a consequence this leads to better priors for μ and Σ and in the end to better posteriors and covariance estimates.

Table IV Performance Minimum Variance Portfolio

Window length: 60	Bayes Prior 1	Bayes Prior 4	Bayes Prior 5	Sample Estimators
Expected Return	10.18%	11.89%	11.58%	10.03%
Realized Return	5.788%	5.483%	5.780%	-0.488%
Expected Volatility	6.176%	4.998%	5.230%	3.953%
Realized Volatility	11.24%	11.75%	11.46%	21.59%
Sharpe Ratio	0.1448	0.1315	0.1423	-0.0065
Compounded Return	1.6467	1.5901	1.6437	0.7540

Table V Performance Minimum Variance Portfolio

Window length: 100	Bayes Prior 1	Bayes Prior 4	Bayes Prior 5	Sample Estimators
Expected Return	9.894%	10.31%	10.28%	9.519%
Realized Return	3.590%	4.040%	4.055%	2.068%
Expected Volatility	6.923%	6.103%	6.213%	6.502%
Realized Volatility	12.15%	12.28%	12.26%	14.06%
Sharpe Ratio	0.0839	0.0933	0.0938	0.0420
Compounded Return	1.2033	1.2371	1.2383	1.0715

It can also be seen that the highest Sharpe ratios and compounded returns so far are reached with the new priors and therefore it seems to be useful to apply factor models for prior specifications. Now an interesting thing might be to consider a prior for Σ that is more informative to see which influence that has on the portfolio performance.

D. More Informative Prior for Σ

As discussed in section III this research so far used the least informative prior for Σ to ensure a posterior distribution for Σ that is mainly based on the data. However, in the previous subsection it is pointed out that the portfolios implied by the Bayesian estimation method clearly outperform the portfolios implied by the sample estimators. Therefore it might be interesting to slightly adjust the priors for Σ, such that they become more influential.

In the least informative prior for Σ (or Σ^{-1}) the prior distribution for Σ^{-1} is given by a Wishart distribution with $v_0 = N$ degrees of freedom and scale matrix $R_0 = \frac{1}{v_0}\widetilde{\Sigma}^{-1}$. As graphically illustrated in figure 1 the certainty about the prior to some extend influences the posterior distribution and therefore it might be interesting to change the degrees of freedom in the prior for Σ. The degrees of freedom can be seen as the number of observations added to the dataset by using this prior. To illustrate the effect of changing v_0, v_0 is specified as pN and different values for p are considered. The prior becomes more influential if $p > 1$ and less influential if $p < 1$. Consequently setting $p = \infty$ leads to a situation in which the posterior will be the same as the prior and setting

$p \approx 0$ will result in approximately the same estimate for Σ as the sample estimator gives.

Since the effect of changes in v_0 is evaluated for the Bayesian priors of situation 1, 4 and 5, it does not make sense to chose a very high value for p. This is because all off-diagonal elements of the covariance will remain very close to zero in that situation and no correlations between assets, that are necessary to achieve the diversification needed to create the minimum variance portfolio, are found. Therefore in this research values for p are chosen between 0.5 and 2.0. The results for a window length of 60 are presented in table VI. A window length of 100 results in quite comparable results and therefore the table is shown in appendix A.

Table VI Performance Minimum Variance Portfolio

Window length: 60		Exp Ret	Real Ret	Exp Vol	Real Vol	SR	Comp Ret
Situation 1	p=0.5	10.49%	4.768%	6.114%	11.81%	0.1140	1.484
	p=1.0	10.18%	5.789%	6.176%	11.24%	0.1448	1.647
	p=1.5	10.02%	6.444%	6.075%	10.99%	0.1644	1.757
	p=2.0	9.930%	6.940%	5.943%	10.88%	0.1786	1.843
Situation 4	p=0.5	12.12%	4.630%	5.198%	12.58%	0.1040	1.451
	p=1.0	11.89%	5.483%	4.998%	11.75%	0.1315	1.590
	p=1.5	11.71%	6.010%	4.747%	11.31%	0.1493	1.680
	p=2.0	11.56%	6.389%	4.516%	11.04%	0.1623	1.747
Situation 5	p=0.5	11.92%	4.831%	5.400%	12.27%	0.1112	1.485
	p=1.0	11.58%	5.798%	5.230%	11.46%	0.1423	1.644
	p=1.5	11.34%	6.385%	4.993%	11.05%	0.1621	1.746
	p=2.0	11.16%	6.803%	4.770%	10.80%	0.1764	1.821

When table VI is examined the first thing that points out is that in all situations the realized returns go up if p increases and that the realized volatilities go down at the same time. This might be interesting, but can be declared by the following two observations. The first is based on the trade-off between bias and variance in the Bayesian estimation method. The sample covariance namely contains a lot of estimation error when $N \approx T$ and for small values of p the posterior distribution, and therefore also the estimation of Σ, probably contains this error, which results in bad portfolio performance. The second is based on the analysis of the data set and the market return. It turns out that the average return on the assets, and so the market return, in the investment period is about 15%, which is very high, and has of volatility of just 17%. When p increases, the estimate of Σ becomes more like a diagonal matrix and therefore no correlations can be used for diversification purposes. So in situation 1 the minimum variance portfolio will converge to the market portfolio if p increases too much, because all assets are assumed to have the same expected return and variance, and in situation 4 and 5 the assets with the lowest variances will get huge weights, while others are ignored. Since the aim was to determine the influence of p on the estimate of Σ and these estimates become worthless if p is too large, the maximum value of p

is set equal to 2.0. It has to be mentioned though, that when a more advanced matrix is used for $\widetilde{\Sigma}$, such that correlations are already included in the prior, probably a nicer pattern will show up when p changes.

To come back to the results; the next remarkable thing is that both the expected and realized return and the expected and realized volatility converge to each other when p increases. So this increase in p probably leads to a decrease in estimation error in the parameters and therefore better expectations. When the latter two columns with the summary statistics are considered, the conclusion obviously is the same. For all situations both the Sharpe ratio and the compounded return increase when p increases.

VI. Conclusion

Taking everything into consideration the Bayesian estimation method seems to be a good method to estimate a covariance matrix for investment purposes. The sample estimators are clearly outperformed; even in situations they do not suffer from the dimensionality issue. However, the specification of the prior distribution in the Bayesian estimation method points out to be very important. In this paper it is shown that more advanced specifications of the prior distributions do not necessarily lead to better results. Particularly when not so much data is available, it can be more efficient to use simple properties found in empiral research to set priors. A drawback from these simple specifications is that they do not incorporate correlations between assets in the priors, though and this leads to missed diversification possibilities.

Another thing that might be very interesting to consider, is to chose other prior distributions for μ and/or Σ. As Cont (2000) discusses in his paper about empirical properties of asset returns, it points out that asset returns are not normally distributed. Outliers seem to occur more often than a normal distribution can anticipate on and large negative returns occur more often than large positive ones. Therefore a (lightly) skewed distribution with fatter tails might be better as prior distribution for μ.

REFERENCES

Jushan Bai and Shuzhong Shi. Estimating high dimensional covariance matrices and its applications. *Annals of Economics and Finance*, pages 199–215, 2011.

Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, pages 370–418, 1763.

Michael J. Best and Robert R. Grauer. The analytics of sensitivity analysis for mean-variance portfolio problems. *International Review of Financial Analysis*, 1992.

Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Institute of Physics Publishing*, 2000.

Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 1993.

Thomas M. Idzorak. A step-by-step guide to the black-litterman model, 2004.

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

William F. Sharpe. Capital asset prices - a theory of market equilibrium under conditions of risk. *Journal of Finance*, 1964.

William F. Sharpe. Mutual fund performance. *The Journal of Business*, pages 119–138, 1966.

Adne W. van Engelen, Max J.H. van Keulen, Diederik W.F. Mathol, and Sierk B.J. Wiegers. Estimating a high dimensional covariance matrix. *Financial Case Studies Econometrics and Operations Research*, 2014.

Appendix A.

In table VII the results of the minimum variance portfolios based on a window length of 100 are shown. In the out-of-sample forecast period the average return on the market index was 11% and had a volatility of about 19%.

Table VII Performance Minimum Variance Portfolio

Window length: 100		Exp Ret	Real Ret	Exp Vol	Real Vol	SR	Comp Ret
Situation 1	*p=0.5*	9.843%	2.841%	6.974%	12.52%	0.0647	1.143
	p=1.0	9.894%	3.590%	6.923%	12.15%	0.0839	1.203
	p=1.5	9.928%	4.117%	6.792%	11.96%	0.0975	1.247
	p=2.0	9.964%	4.515%	6.647%	11.85%	0.1078	1.280
Situation 4	*p=0.5*	10.19%	3.417%	6.406%	12.67%	0.0767	1.185
	p=1.0	10.31%	4.040%	6.103%	12.28%	0.0933	1.237
	p=1.5	10.38%	4.454%	5.811%	12.05%	0.1046	1.273
	p=2.0	10.42%	4.761%	5.552%	11.90%	0.1130	1.299
Situation 5	*p=0.5*	10.18%	3.423%	6.485%	12.64%	0.0770	1.185
	p=1.0	10.28%	4.055%	6.213%	12.26%	0.0938	1.238
	p=1.5	10.32%	4.474%	5.940%	12.05%	0.1051	1.274
	p=2.0	10.35%	4.784%	5.693%	11.91%	0.1135	1.301