The Impact of Estimation Error on Portfolio Selection for Investors with Constant Relative Risk Aversion

Bengtsson, Christoffer

2003

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

The Impact of Estimation Error on Portfolio Selection for Investors with Constant Relative Risk Aversion

Christoffer Bengtsson*
Department of Economics
Lund University

29th April 2004

Abstract

This paper examines the impact of estimation error in a simple single-period portfolio choice problem when the investor has power utility and asset returns are jointly lognormally distributed. These assumptions imply that such an investor selects portfolios using a modified mean-variance framework where the parameters that he has to estimate are the mean vector of log returns and the covariance matrix of log returns. Following Chopra and Ziemba (1993), I simulate estimation error in what are assumed to be the true mean vector and the true covariance matrix and the impact of estimation error is measured in terms of percentage cash equivalence loss for the investor. To obtain estimation error sizes that are similar to the estimation error sizes in actual estimates, I use a Bayesian approach and Markov Chain Monte Carlo Methods. The empirical results differ significantly from Chopra and Ziemba (1993), suggesting that the effect of estimation error may have been overestimated in the past. Furthermore, the results tend to question the traditional viewpoint that estimating the covariance matrix correctly is strictly less important than estimating the mean vector correctly.

Keywords: Portfolio selection; Estimation risk; Markov Chain Monte Carlo.
JEL classification: G11.

1 Introduction

The mean-variance (MV) theory for portfolio selection has been the subject of much debate since it was proposed by Markowitz (1952), and in spite of its theoretical appeal, it has had some trouble being fully accepted by practitioners (see e.g. Michaud, 1989, for a discussion). Much of the controversy has arisen because MV optimization is thought to be sensitive to errors in the two input parameters; the expected return vector and the covariance matrix for the returns. Chopra and Ziemba (1993) examine the effect of estimation error on the investor's utility in terms of cash equivalence loss under the assumptions that the investor has preferences with constant absolute risk aversion, simple returns are jointly normally distributed, and short selling is not allowed. In a simulation study they find that the loss resulting from estimation error in the mean vector can be substantial, whereas estimating variances and covariances correctly seems to be of less importance.

This paper examines the effect of estimation error in a slightly different setting. I assume instead that gross returns are jointly lognormally distributed, implying that log returns are jointly normally distributed, and that the investor has preferences with constant *relative* risk aversion.

*E-mail: christoffer.bengtsson@nek.lu.se, phone: +46 (0)46 - 222 79 11, postal address: P.O. Box 7082, SE-220 07 Lund, Sweden. The discussions and comments provided by Björn Hansson are greatly appreciated.

The paper also examines what effect portfolio weight constraints have on the impact of estimation error. To obtain estimation error sizes that are consistent with those of actual estimates, I use a Bayesian approach and Markov Chain Monte Carlo methods.

Using the same simulation approach as Chopra and Ziemba (1993) combined with their values for the estimation error size, I find that the effect of estimation error is smaller than expected, much smaller than in Chopra and Ziemba (1993). The certainty equivalence loss resulting from estimation error in the mean vector is for instance at most 0.1 percent when the investor is very aggressive and no short selling is allowed, and it is even smaller for errors in the covariance matrix. Using the estimation error sizes from the Bayesian approach provides a more plausible situation and it increases the loss resulting from errors in means. When a limited amount of short selling is permitted, errors in the covariance matrix become increasingly important compared to errors in means. For the estimation error sizes of Chopra and Ziemba (1993), the cash equivalence loss resulting from errors in the covariance matrix can for instance be up to 4.5 percent when errors are large and the investor is very conservative, while errors in the mean vector never result in a cash equivalence loss greater than 0.6 percent. Even for moderate levels of relative risk aversion and for the estimation error sizes from the Bayesian approach, it appears that errors in the covariance matrix are in some situations just as important as errors in the mean vector which goes against the conventional wisdom that the effect of errors in means dominate over the effect of errors in the covariance matrix in all situations.

The rest of the paper is organized as follows: The next section, Section 2, derives the portfolio optimization problem, first in the setting when absolute risk aversion is constant and returns are jointly normally distributed, and then in the setting when relative risk aversion is constant and returns are jointly lognormal. Details on the latter derivation are found in Appendix A. Section 3 then describes the data and the methodology and Section 4 presents the empirical results and the analysis. Section 5 concludes the paper. All tables and figures are contained in Appendix C, Appendix B gives a short introduction to, and some details on, the Bayesian estimation method, and Appendix D explains why the results in this paper differ to such a large extent from the results in Chopra and Ziemba (1993).

2 Portfolio Optimization

This section begins by showing that the MV criterion is consistent with single-period utility maximization when absolute risk aversion is constant and asset returns are jointly normally distributed[1]. This approach has, however, two major drawbacks: (1) Constant absolute risk aversion implies that relative risk aversion is increasing in wealth, which is not consistent with observed asset prices. (2) If returns on common stocks, e.g., were truly normally distributed, there would exist a positive probability of infinitely negative return, whereas, in reality, stock returns can at most be -100 percent. I therefore proceed with the more relevant case when *relative* risk aversion is constant and gross returns are jointly lognormal, and I again arrive at an optimization problem similar to the traditional MV optimization problem. The assumption of lognormality is appealing since lognormal variables can never be negative. The downside is that the portfolio return will not itself be strictly lognormal since the product, and not the sum, of lognormal variables is itself lognormal. However, by considering returns in the limit of continuous time, the non-lognormality disappears. So, inspired by Campbell and Viceira (2002), I make a discrete time approximation of the log portfolio return in continuous time. The approximation is exact in continuous time and it is accurate over sufficiently short time intervals.

[1]Strictly speaking, the MV criterion is consistent with utility maximization for any concave utility function when returns are normally distributed, but constant absolute risk aversion utility provides particularly simple calculations.

2.1 The Case of Constant Absolute Risk Aversion

Assume first that the investor derives utility from his wealth at the end of the period and that he, as in Chopra and Ziemba (1993), has negative exponential utility

$$u(x) = -\mathrm{e}^{-ax}, \tag{1}$$

where $a > 0$ is the coefficient of constant absolute risk aversion. The reciprocal of a is referred to as the coefficient of constant absolute risk tolerance. Assume also that returns are jointly normally distributed.

The investor's wealth at the end of the period is determined by his wealth today, W_t, and by the portfolio of N risky assets that the wealth is invested in. The portfolio is described by a vector of weights, $\mathbf{w}_t = [w_{1t}, w_{2t}, ..., w_{Nt}]^{\mathrm{T}}$, where w_{it} is the fraction of total wealth at time t that is placed in asset i. For now the portfolio weights are allowed to be both positive and negative, just as long as they sum up to one. In practice, however, negative portfolio weights come at a high cost. Partly because there is a price tag on short selling[2], and partly because short selling permits the possibility of negative wealth at time $t+1$. Short selling is therefore generally only permitted for investors with a good credit rating. In the empirical part of the paper I will restrict attention to two levels of short selling by adding the additional constraint $\mathbf{w}_t \geq \ell_i \mathbf{1}$, where $i = 1, 2$, $\ell_1 = 0$, $\ell_2 = -0.10$, and $\mathbf{1}$ is a $N \times 1$ vector of ones.

Let $\mathbf{R}_{t+1}$ be a $N \times 1$ random vector where element i, $R_{i,t+1}$, is the (net) simple return of asset i over one period. The portfolio return is then $R_{p,t+1} = \mathbf{w}_t^{\mathrm{T}} \mathbf{R}_{t+1}$, and since the sum of normally distributed variables is also normally distributed, $R_{p,t+1}$ will be normally distributed.

The utility maximization problem of the investor is

$$\max \left\{ \mathrm{E}_t \left[u(W_{t+1}) \right] \ \text{ subject to } \ W_{t+1} = (1 + R_{p,t+1}) W_t \right\},$$

which in this case becomes

$$\max \left\{ \mathrm{E}_t \left[-\mathrm{e}^{-aW_{t+1}} \right] \ \text{ subject to } \ W_{t+1} = (1 + R_{p,t+1}) W_t \right\}. \tag{2}$$

The notation $\mathrm{E}_t[\cdot]$ is short for $\mathrm{E}[\cdot \mid \mathcal{F}_t]$, where $\mathcal{F}_t$ is the information set or filtration available at time t. The same notation is used on variances. The assumption of normality allows the expectation in equation (2) to be rewritten as

$$-\exp \left\{ -aW_t \left(1 + \mathrm{E}_t[R_{p,t+1}] - \frac{aW_t}{2} \mathrm{Var}_t[R_{p,t+1}] \right) \right\}. \tag{3}$$

Since maximizing a function is equivalent to minimizing the negative of the function and since maximizing the log of a function is equivalent to maximizing the function itself, the optimization problem boils down to

$$\max_{\mathbf{w}_t} \left\{ \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\eta}_t - \frac{aW_t}{2} \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\Omega}_t \mathbf{w}_t \ \text{ subject to } \ \mathbf{w}_t^{\mathrm{T}} \mathbf{1} = 1 \right\}, \tag{4}$$

where $\boldsymbol{\eta}_t$ is the vector of conditional expected returns, $\boldsymbol{\Omega}_t$ is the conditional covariance matrix of the returns, and $\mathbf{1}$ is a $N \times 1$ vector of ones. Equation (4) is exactly equal to the traditional MV optimization problem; the investor makes a linear trade-off between the expected portfolio return and the portfolio variance. The degree to which the investor is willing to trade off mean for variance is determined by the constant aW_t, which is the relative risk aversion at time t of an investor with constant absolute risk aversion. The wealthier the investor is, the more emphasis is put on minimizing the variance of the portfolio. This is because the same relative change in

[2] Usually something like 1 or 2 percent per year of the stock's market value.

wealth is larger in absolute terms for a wealthy investor than it is for a poor investor, and it is aversion against changes in absolute terms that is central for an investor with constant absolute risk aversion.

Often, equation (4) is given in terms of risk tolerance. Normalizing today's wealth to 1 gives the expression

$$\max_{\mathbf{w}_t} \left\{ \tau \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\eta}_t - \frac{1}{2} \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\Omega}_t \mathbf{w}_t \ \text{ subject to } \ \mathbf{w}_t^{\mathrm{T}} \mathbf{1} = 1 \right\},$$

where $\tau = 1/a$. Chopra and Ziemba (1993) claim that most institutional investors have risk tolerance in the range 0.2-0.3, implying that relative risk aversion is approximately in the range 3-5. This is well inside the range of what is commonly thought of as plausible (see e.g. Mehra and Prescott, 1985; Rietz, 1988).

2.2 The Case of Constant Relative Risk Aversion

Assume now that gross returns are instead jointly lognormal and that the investor has power utility

$$u(x) = \frac{x^{1-\gamma}}{1-\gamma}, \tag{5}$$

where $\gamma > 0$ is the coefficient of constant relative risk aversion[3]. The utility maximization problem of the investor in this case becomes

$$\max \left\{ \mathrm{E}_t \left[\frac{W_{t+1}^{1-\gamma}}{1-\gamma} \right] \ \text{ subject to } \ W_{t+1} = (1 + R_{p,t+1}) W_t \right\}. \tag{6}$$

In Appendix A, I derive an expression of the portfolio dynamics in continuous time which is then discretized in order to obtain a loglinear approximation of the log portfolio return in discrete time. The approximation is accurate over sufficiently small time intervals and the portfolio return is therefore treated as a lognormal variable.

Let $r_{p,t+1}$ denote the log of the gross simple return $(1 + R_{p,t+1})$. By normality of $r_{p,t+1}$, the expectation in equation (6) can be rewritten as

$$\frac{W_t^{1-\gamma}}{1-\gamma} \exp \left\{ (1-\gamma) \mathrm{E}_t[r_{p,t+1}] + \frac{(1-\gamma)^2}{2} \mathrm{Var}_t[r_{p,t+1}] \right\}, \tag{7}$$

and thus, maximizing expected utility is equivalent to

$$\max \left\{ \mathrm{E}_t[r_{p,t+1}] + \frac{1-\gamma}{2} \mathrm{Var}_t[r_{p,t+1}] \right\}. \tag{8}$$

The approximations in Appendix A of the expected log portfolio return and the variance of the log portfolio return are

$$\mathrm{E}_t[r_{p,t+1}] = \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\mu}_t + \frac{1}{2} \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\sigma}_t^2 - \frac{1}{2} \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\Sigma}_t \mathbf{w}_t, \tag{9}$$

$$\mathrm{Var}_t[r_{p,t+1}] = \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\Sigma}_t \mathbf{w}_t, \tag{10}$$

where $\boldsymbol{\mu}_t$ is a $N \times 1$ vector of conditional expected log returns, $\boldsymbol{\Sigma}_t$ is the conditional covariance matrix of the log returns, and $\boldsymbol{\sigma}_t^2$ is a vector consisting of the diagonal elements of $\boldsymbol{\Sigma}_t$. Inserting equations (9) and (10) into equation (8) yields

$$\max_{\mathbf{w}_t} \left\{ \mathbf{w}_t^{\mathrm{T}} \left(\boldsymbol{\mu}_t + \frac{1}{2} \boldsymbol{\sigma}_t^2 \right) - \frac{\gamma}{2} \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\Sigma}_t \mathbf{w}_t \ \text{ subject to } \ \mathbf{w}_t^{\mathrm{T}} \mathbf{1} = 1 \right\}, \tag{11}$$

[3] In the limit as $\gamma \rightarrow 1$, the utility function becomes $u(x) = \log(x)$.

which is a maximization problem similar to the usual MV optimization problem. The difference between equation (11) and equation (4) is the corrective term $\frac{1}{2}\boldsymbol{\sigma}_t^2$ and that the expected return vector and the covariance matrix of the simple returns are replaced by the return vector and the covariance matrix of the log returns. However, as in equation (4), it is still relative risk aversion that is the factor in front of the quadratic (variance) term. In this case, however, relative risk aversion, and thereby also the optimal portfolio composition, does not depend on the investor's level of wealth.

3 Data and Methodology

The data set consists of monthly returns during the ten year period between January 1986 and December 1995 for the 30 companies that belonged to the DJIA at the end of 1995. This is three times the number of stocks considered by Chopra and Ziemba (1993)[4]. The data is extracted from the CRSP data base. For these stocks I estimate the sample mean vector and the sample covariance matrix of the log returns and these estimates are assumed to be the true, and to the investor unknown, parameters $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$.

The main objective of this paper is to examine what effect estimation error in $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ has on optimal portfolio choice. Since the investor does not know the true input parameters, he will base the portfolio optimization problem of equation (11) on input parameters that he has estimated in some way, i.e., he will replace $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ with some $\hat{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{\Sigma}}_t$, where $\hat{\boldsymbol{\mu}}_t \neq \boldsymbol{\mu}_t$ and $\hat{\boldsymbol{\Sigma}}_t \neq \boldsymbol{\Sigma}_t$. So consequently, although the portfolio that he chooses is optimal for the estimated input parameters, it is suboptimal for the true input parameters, and hence, it does not truly maximize expected utility.

Denote by $\mathbf{w}_t^*$ the portfolio that is the true optimal portfolio and by $\hat{\mathbf{w}}_t$ the suboptimal portfolio based on the estimated input parameters. As a measure of how suboptimal $\hat{\mathbf{w}}_t$ is compared to $\mathbf{w}_t^*$, I follow Chopra and Ziemba (1993) and compare the cash equivalent (CE) values of the two portfolios. The CE of a portfolio is defined as the risk free amount of cash that gives the same (expected) utility as the risky portfolio, i.e.

$$CE(\mathbf{w}_t) = u^{-1}(\mathcal{U}(\mathbf{w}_t)),$$

where $\mathcal{U}(\mathbf{w}_t)$ is the expected utility of some portfolio $\mathbf{w}_t$. Since $\mathbf{w}_t^*$ provides the maximum expected utility for the investor, it will correspond to the maximum CE. The CE is, as opposed to units of utility, independent of an affine transformation of the utility function and is instead measured in dollars, the same unit as consumption is measured in. It is straightforward to show, using equations (7), (9), and (10), that the CE implied by a lognormal portfolio return and power utility is given by

$$CE(\mathbf{w}_t) = W_t \exp\left\{\mathbf{w}_t^{\mathrm{T}}\left(\boldsymbol{\mu}_t + \frac{1}{2}\boldsymbol{\sigma}_t^2\right) - \frac{\gamma}{2}\mathbf{w}_t^{\mathrm{T}}\boldsymbol{\Sigma}_t\mathbf{w}_t\right\}.$$

The percentage cash equivalent loss (CEL) suffered by an investor that holds a portfolio other than $\mathbf{w}_t^*$ is then simply

$$CEL = 100 \cdot \frac{CE(\mathbf{w}_t^*) - CE(\mathbf{w}_t)}{CE(\mathbf{w}_t^*)}.$$

In order to examine which type of error that has most impact, I will distinguish between errors in the mean vector, the covariance matrix, variances alone, and covariances alone. Chopra and Ziemba (1993) only consider errors in the mean vector, variances alone, and covariances alone. Not errors in the entire covariance matrix.

[4] Chopra and Ziemba (1993) consider $N = 10$ randomly selected stocks, also from the DJIA.

I begin by calculating $\mathbf{w}_t^*$ given $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$. The impact of errors in means is estimated by first replacing μ_{it} with

$$\hat{\mu}_{it} = (1 + k_i \varepsilon_i)\mu_{it},\ i = 1, 2, ..., N,$$

where ε_i is a standard normal random number and k_i is the size of the estimation error of μ_{it}. $\boldsymbol{\Sigma}_t$ is left unchanged. Given $\hat{\boldsymbol{\mu}}_t$, $\boldsymbol{\Sigma}_t$, and the restrictions on the portfolio weights, I can then calculate the (sub-) optimal portfolio $\hat{\mathbf{w}}_t$ and the resulting CEL. This procedure is repeated 10,000 times and the impact of errors in means is taken as the average CEL over all these 10,000 iterations. Chopra and Ziemba (1993) perform 100 iterations.

The same procedure is then used to examine the effect of errors in the covariance matrix. To examine the impact of errors in the entire covariance matrix, I replace σ_{ijt} with

$$\hat{\sigma}_{ijt} = (1 + K_{ij}\varepsilon_{ij})\sigma_{ijt},\ i, j = 1, 2, ..., N,$$

where, again, ε_{ij} is a standard normal random number[5], K_{ij} is the estimation error size of σ_{ijt}, and, of course, $x_{ij} = x_{ji}$. $\boldsymbol{\mu}_t$ is left unchanged. To examine the effect of errors in variances alone, I put $x_{ij} = 0$ when $i \neq j$, and to examine the effect of errors in covariances alone, I put $x_{ij} = 0$ when $i = j$.

The entire exercise is repeated for different values of the relative risk aversion, for different values of the error sizes, and for different portfolio weight restrictions. The constant relative risk aversion, γ, takes on the values $1, 3, 5, 10$, and 25. $\gamma = 3$, 5, 10 represents investors with relative risk aversion consistent with what is commonly considered plausible, whereas γ = 1 represents a very aggressive investor and γ = 25 represents a very conservative investor.

In Chopra and Ziemba (1993), the estimation error sizes take on the values $k_i = K_{ij} = k = 0.05, 0.10, 0.15, 0.20$, $i, j = 1, 2, ..., N$. However, nothing is said about whether these values are reasonable or not. In order to say what is most damaging, estimation error in the mean vector or in the covariance matrix, it is important to know how large the estimation error is in actual estimates. For instance, the mean vector is known to be notoriously hard to estimate, especially from past returns alone, and hence, the size of the estimation error in the mean vector may be expected to be larger than the size of the estimation error in the covariance matrix. Consequently, it can be misguiding to compare the CELs resulting from estimation error in the two parameters when the sizes of their estimation errors are the same, i.e. when $k_i = K_{ij}$, $i, j = 1, 2, ..., N$.

To obtain values for the estimation error sizes that are consistent with the estimation error sizes in actual estimates, I estimate the mean vector and the covariance matrix with the Markov Chain Monte Carlo method (MCMC), which is a Bayesian estimation method. As opposed to traditional methods of estimation, the MCMC estimate is not a point estimate, but rather a sample from the joint distribution of the parameters conditional on the data. This joint conditional distribution is referred to as the posterior distribution and by using the sample from it is straightforward to calculate the standard deviations of the individual elements of the estimated mean vector end the estimated covariance matrix.

Therefore, in addition to the values for the size of the estimation error used by Chopra and Ziemba (1993), I also perform the above simulation study when the estimation error size for an element (in either the mean vector or the covariance matrix) is taken as (1) its posterior standard deviation divided by the absolute value of its assumed true value, or (2) half its posterior standard

[5]That the ε_is and ε_{ij}s are mean zero random numbers imply that the estimates are assumed not to have any systematic biases. In reality, however, many estimators contain some amount of specification error.

deviation divided by the absolute value of its assumed true value[6]. That is,

$$k_i = \sqrt{\frac{s_i^2}{\mu_{it}^2}},\ K_{ij} = \sqrt{\frac{S_{ij}^2}{\sigma_{ijt}^2}},\ i,j = 1,2,...,N,$$

or

$$k_i = 0.5\sqrt{\frac{s_i^2}{\mu_{it}^2}},\ K_{ij} = 0.5\sqrt{\frac{S_{ij}^2}{\sigma_{ijt}^2}},\ i,j = 1,2,...,N,$$

where s_i is the posterior standard deviation of element i of the mean vector and S_{ij} is the posterior standard deviation of element ij of the covariance matrix. Case (1), referred to as *large errors*, is meant to represent the situation when the input parameters are poorly estimated and case (2), referred to as *small errors*, is meant to represent the situation when the input parameters are better estimated. Details on how the MCMC estimates are obtained can be found in Appendix B.

Finally, Chopra and Ziemba (1993) only examine the case when portfolio weights are restricted to be non-negative. This is a valid restriction for most individuals, but not for large institutional investors with a good credit rating. I therefore also examine the case when portfolio weights are restricted not to fall below -10 percent[7].

4 Empirical Results

The (assumed true) means, standard deviations, and correlations for the 30 DJIA stocks are found in Tables 1 and 2, and Tables 3 and 4 summarizes the optimal portfolio weights in $\mathbf{w}_t^*$ for the different combinations of relative risk aversion and portfolio weight restrictions. The empirical results from the simulation study described in the previous section are found in Tables 5-8. Tables 5 and 6 represent the cases when the estimation error sizes are the same as in Chopra and Ziemba (1993) and Tables 7 and 8 represents the cases when the estimation sizes are obtained from the MCMC estimation. All tables and figures are found in Appendix C.

4.1 Estimation Error Sizes the Same as in Chopra and Ziemba (1993)

Consider first the cases when the CELs are calculated using the same estimation error sizes as in Chopra and Ziemba (1993). An initial and important observation is that the CELs are generally very low compared to those in Chopra and Ziemba (1993) (see Table 5). The effect of errors in means when short selling is not allowed is at most 0.1469 percent, not several percent as in Chopra and Ziemba (1993), and errors in the covariance matrix have even smaller effects. In Appendix D, I discuss the cause of this difference, seemingly the result of an error in Chopra and Ziemba (1993).

Table 5 reveals that when portfolio weight are restricted to be non-negative, errors in means generally result in CELs much larger than the CELs resulting from errors the covariance matrix. For example, when $\gamma = 5$ and $k = 0.10$, errors in means are about 7.6 times more important than errors in the entire covariance matrix. It is only when the investor is very risk averse that estimation error in the covariance matrix is more important than estimation error in the mean vector. Intuitively, this result can be understood by the following reasoning: Since negative portfolio weights are not allowed and since stock returns generally are positively correlated, the optimal portfolio

[6]Four elements of the mean vector are quite small in absolute terms and the corresponding estimation error sizes of these elements become much larger than the estimation error sizes of the remaining elements. I therefore impose that $k_i = 1$ for these elements, which is about the same magnitude as the fifth largest estimation error size of an element in the mean vector.

[7]The case when portfolio weights are restricted not to fall below -5 percent has been omitted to save space. The results fall exactly in between the remaining two cases and they are available upon request.

will consist only of stocks that have high expected returns, low standard deviations, and low correlations with other stocks. That is, stocks that are represented by large elements in the mean vector and small elements in the covariance matrix. Consequently, because of the multiplicative nature of the errors, the absolute size of the errors in the relevant parts of the mean vector will be much larger than the absolute size of the errors in the relevant parts of the covariance matrix, and hence, errors in means will have a much larger effect on the objective function for a given $\mathbf{w}_t^*$.

When the investor is permitted to sell stocks short, the CEL increases (see Table 6). For $\gamma = 5$, and $k = 0.10$, the CEL ratio between errors in means and errors in the covariance matrix is now only 1.3. Even for a relatively aggressive investor with $\gamma = 3$ are errors in means now, on average, only twice as important as errors in the covariance matrix, compared to about 15 times more important in Table 5. When $\gamma = 5$, errors in the covariance matrix are, on average, more important than errors in means, and when $\gamma = 25$ and $k = 0.20$, the CEL is now as large as 4.5 percent for errors in the covariance matrix. Also, when short selling is allowed, errors in covariances alone are often of an equal or greater importance than errors in variances alone. This may add validity to covariance matrix estimation techniques such as the one proposed by Ledoit and Wolf (2003) that essentially only improves the estimation of the off-diagonal elements of the covariance matrix.

4.2 Estimation Error Sizes Estimated with the MCMC Method

Consider now the cases when the estimation error sizes are obtained from the MCMC estimation in order for them to resemble the sizes of the estimation error in actual estimates. Figures 1, 2, and 3 show the histograms of the estimation error sizes for the means, the variances, and the covariances, respectively[8]. The average estimation error size for means is 0.6110[9], the average estimation error size for variances is 0.0663, and the average estimation error size for covariances is 0.1361. This implies that, for instance, $k = 0.20$ above represents (relative sample estimates) good estimates of the means, while it represents very poor estimates of the variances and the covariances.

Table 7 shows that when short selling is not allowed, then the CELs are all quite small for both the large and the small errors. No CEL is ever greater than 1 percent and errors in means are always more important than errors in the covariance matrix, even for a very conservative investor. In Table 8 when the investor is permitted to short sell 10 percent of each stock, however, errors in means are no longer always more important than errors in the covariance matrix. When $\gamma = 10$ and errors are small, the CEL for the covariance matrix and CEL for the mean vector are approximately equal, and for $\gamma = 25$, errors in the covariance matrix are most important. The CEL for the covariance matrix is in the latter case approximately five times the CEL of the mean vector, both when errors are small and when they are large. So, again, when short selling is allowed it is not the case that errors in the mean vector are always most important.

Summing up the conclusions from Tables 5-8, if the investor cannot short sell, then estimation error has very little effect. Only when the investor is allowed to short sell can estimation error have substantial effects; either if the size of the estimation error is relatively large and/or if the investor is very conservative, and the more short selling is allowed, the more important are errors in the covariance matrix. For instance, although not reported, when 50% short selling per stock is allowed, then errors in the covariance matrix are more important than errors in means even for a quite aggressive investor with $\gamma = 3$. Jagannathan and Ma (2003) analytically show how non-negativity constraints can help control for estimation error in (especially) the covariance matrix. This paper points at a similar conclusion: When non-negativity constraints are in place, the sensitivity of the portfolio weights to changes in both input parameters is reduced. Of course,

[8] For the error sizes of the means, the four values of k_i that are restricted to be equal to 1 are omitted.

[9] The average estimation error size of means is 0.5511 if the values that are restricted to equal 1 are omitted.

since monthly data is used, I have implicitly assumed an investment horizon of a month. If the input parameters were estimated on a yearly basis, e.g., then the elements in both $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ would grow by roughly a factor 12. Consequently, the impact of estimation error will grow also with the investment horizon.

4.3 Further Empirical Results

It might be interesting to examine how large the CEL is for an investor with power utility facing lognormal stock returns that knows his relative risk aversion, but who still uses the traditional MV optimization problem when choosing his portfolio. Suppose that such an investor, as is implied by equation (4), uses his relative risk aversion to determine the trade-off between mean and variance. This means that the investor solves

$$\max_{\mathbf{w}_t} \left\{ \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\eta}_t - \frac{\gamma}{2} \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\Omega}_t \mathbf{w}_t \ \text{ subject to } \ \mathbf{w}_t^{\mathrm{T}} \mathbf{1} = 1, \mathbf{w}_t \geq \ell \mathbf{1} \right\}, \tag{12}$$

where ℓ is the lower bound for the portfolio weights, instead of his correct optimization problem defined by equation (11) with the additional constraint that $\mathbf{w}_t \geq \ell\mathbf{1}$. The result, presented in Table 9, shows that the differences between the optimal portfolios are quite small for all combinations of γ and ℓ. So, although the investor uses the wrong optimization problem, he still obtains a portfolio that is very close to the one that truly maximizes his expected utility. As a comparison, in Table 10 I present the CEL when the investor does not optimize his portfolio at all, but simply invests an equal amount in each stock.

Finally, the modified MV optimization problem of equation (11) relies on an approximation of the log portfolio return which is claimed to be accurate over sufficiently small time intervals. But is a month a sufficiently time small interval? In order to examine how accurate the approximation is, I use the following procedure: Denote by $\mathbf{Y}$ the $K \times N$ matrix containing the data set described in Section 3 where element ki of $\mathbf{Y}$ is return observation number k of stock number i. I then calculate the historical mean and standard deviation of the log return of some portfolio $\mathbf{w}_t$ using the formulas

$$\begin{aligned} m &= \frac{1}{K} \sum_{k=1}^{K} \log\left(1 + \mathbf{Y}_k \mathbf{w}_t\right), \\ s &= \sqrt{\frac{1}{K-1} \sum_{k=1}^{K} \left(\log\left(1 + \mathbf{Y}_k \mathbf{w}_t\right) - m\right)^2}, \end{aligned}$$

where $\mathbf{Y}_k$ is row number k of $\mathbf{Y}$. I then compare these values with what I obtain using equations (9) and (10). Just to take two examples, an equally weighted portfolio yields $m = 3.32\%$, $s = 6.27\%$, $\mathrm{E}_t[r_{p,t+1}] = 3.32\%$, and $\mathrm{Std}_t[r_{p,t+1}] = 6.25\%$, and a portfolio with $\gamma = 5$ and a lower bound of $\ell = -10\%$ yields $m = 1.30\%$, $s = 4.85\%$, $\mathrm{E}_t[r_{p,t+1}] = 1.30\%$, and $\mathrm{Std}_t[r_{p,t+1}] = 4.88\%$. The approximations are in other words quite accurate.

5 Conclusions

In this paper I have examined the effect of estimation error on optimal portfolio choice when asset returns are jointly lognormally distributed and the investor has power utility. Such an investor faces a modified MV optimization problem and the parameters that he has to estimate are the vector of expected log returns and the covariance matrix of the log returns. As a measure of loss, I use the percentage reduction in cash equivalence that the investor experiences since his portfolio is based on estimated input parameters containing estimation error, rather than the true input parameters.

By simulating estimation error in what are assumed to be the true input parameters, I found that when the estimation error sizes of Chopra and Ziemba (1993) are used, the loss is generally very small (especially when short selling is not allowed). Much smaller than in Chopra and Ziemba (1993). I provide an explanation to this difference and I show that the results are not due to the use of another utility function together with different distributional assumptions, but rather the result of an error in Chopra and Ziemba (1993). I have also performed the simulation study when the estimation error sizes are estimated using a Bayesian approach in order to get values that are more in tune with the estimation error sizes in actual estimates. This exercise shows that the size of the estimation error in the mean vector is larger than the size of the estimation error of the covariance matrix.

When short selling is not allowed, I find, in line with Chopra and Ziemba (1993), that errors in means result in the largest loss, especially for the estimation error sizes obtained by the Bayesian approach. The loss is, however, as mentioned above generally quite small. When the investor is allowed a limited amount of short selling (10 percent of each stock), the effect of estimation error increases. In particular, the loss due to estimation error in the covariance matrix increases to the extent that it is no longer the case that errors in means always result in the largest loss. In fact, with the estimation error sizes of Chopra and Ziemba (1993), the only investor that is significantly affected by estimation error, or estimation risk, is an investor that is already very averse towards risk in the traditional sense of a high relative risk aversion coefficient. With the estimation error sizes obtained from the Bayesian approach, an investor that is conservative, but not extremely so, experiences approximately the same loss from estimation error in the mean vector as from estimation error in the covariance matrix. This result to some extent goes against the received wisdom that estimating the mean vector correctly is strictly more important than estimating the covariance matrix correctly.

What this paper shows is primarily that the situation may not be as simple as implied by Chopra and Ziemba (1993) and that portfolio weight constraints have significant effects on the impact of estimation error. Estimating the expected returns may be a more difficult task than estimating the covariance matrix, and indeed, when short selling is not allowed (the reality faced by most investors), then the loss from errors in means is larger than the loss from errors in the covariance matrix. At the same time, however, the loss resulting from estimation error in means may be limited. Sharing similarities with the conclusion of Jagannathan and Ma (2003), the empirical results of this paper show that short selling constraints reduce the effect of errors in both the mean vector and the covariance matrix. When the level of short selling in increased, then the loss from errors in the covariance matrix can become larger than the loss from errors in means even for relatively aggressive investors. Consequently, an investor that is permitted short selling and that believes that estimating the covariance matrix correctly is unimportant, runs the risk of seriously reducing the quality of his portfolio.

References

Best, M. and R. Grauer, 1991, On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results, *The Review of Financial Studies* 4, 315-342.

Campbell, J.Y. and L.M. Viceira, 2002, *Strategic Asset Allocation*, Oxford University Press.

Chopra, V. and W. Ziemba, 1993, The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice, *Journal of Portfolio Management* 19, 6-12.

Freund, R., 1956, The Indtroduction of Risk into a Programming Model, *Econometrica* 24, 253-263.

Hammersley, J. and P. Clifford, 1970, Markov fields on finite graphs and lattices, Unpublished manuscript.

Jagannathan, R., and T. Ma, 2003, Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps, *Journal of Finance* 58, 1651-1684.

Ledoit, O. and M. Wolf, 2003, Improved Estimation of the Covariance Matrix of Stock Returns With an Application to Portfolio Selection, *Journal of Empirical Finance* 10, 603-621.

Markowitz, H., 1952, Portfolio Selection, *Journal of Finance* 7, 77-91.

Markowitz, H., 1959, *Portfolio Selection: Efficient Diversification of Investment*, Yale University Press.

Mehra, R. and E. Prescott, 1985, The Equity Premium: A Puzzle, *Journal of Monetary Economics* 15, 145-162.

Michaud, R.O., 1989, The Markowitz Optimization Enigma: Is 'Optimized' Optimal?, *Financial Analysts Journal* 45, 31-43.

Ohlson J. and W. Ziemba, 1976, Portfolio Selection in a Lognormal Market When the Investor has a Power Utility Function, *Journal of Financial and Quantitative Analysis* 11, 57-71.

Rietz, T., 1988, The Equity Risk Premium: A Solution, *Journal of Monetary Economics* 22, 117-136.

A Approximation of the Log Portfolio Return

Following the methodology of Campbell and Viceira (2002), I assume that the continuous time price process for the N risky assets, in this paper common stocks, is

$$\frac{d\mathbf{S}_t}{\mathbf{S}_t} = \boldsymbol{\alpha}_t dt + \boldsymbol{\sigma}_t d\mathbf{Z}_t,$$

where $\boldsymbol{\alpha}_t$ is a $N \times 1$ vector, $\boldsymbol{\sigma}_t$ is a $N \times M$ matrix, and $\mathbf{Z}_t$ is a $M \times 1$ vector of uncorrelated standard Wiener processes. $d\mathbf{S}_t/\mathbf{S}_t$ is just bad notation for $[dS_{1t}/S_{1t}, dS_{2t}/S_{2t}, ..., dS_{Nt}/S_{Nt}]^{\mathrm{T}}$. For an individual asset the dynamics are

$$\frac{dS_{it}}{S_{it}} = \alpha_{it} dt + \boldsymbol{\sigma}_{it} d\mathbf{Z}_t,$$

where α_{it} and $\boldsymbol{\sigma}_{it}$ are the ith rows of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\sigma}_t$, respectively.

The main objective is a description of the log price change of the portfolio (i.e. the log return) as a function of the log price changes of the individual assets. Denote by V_t the value of the portfolio at time t. The dynamics of the log price change of the portfolio and the log price change of asset i follows from Itô's Lemma as

$$d\log V_t = \frac{dV_t}{V_t} - \frac{1}{2}\left(\frac{dV_t}{V_t}\right)^2, \tag{13}$$

and

$$d\log S_{it} = \left(\alpha_{it} - \frac{1}{2}\boldsymbol{\sigma}_{it}\boldsymbol{\sigma}_{it}^{\mathrm{T}}\right) dt + \boldsymbol{\sigma}_{it} d\mathbf{Z}_t = \frac{dS_{it}}{S_{it}} - \frac{1}{2}\boldsymbol{\sigma}_{it}\boldsymbol{\sigma}_{it}^{\mathrm{T}} dt.$$

The first term in equation (13) is

$$\frac{dV_t}{V_t} = \mathbf{w}_t^{\mathrm{T}}\left(\frac{d\mathbf{S}_t}{\mathbf{S}_t}\right) = \mathbf{w}_t^{\mathrm{T}}\left(d\log\mathbf{S}_t + \frac{1}{2}[\boldsymbol{\sigma}_{it}\boldsymbol{\sigma}_{it}^{\mathrm{T}}]dt\right),$$

where $[\boldsymbol{\sigma}_{it}\boldsymbol{\sigma}_{it}^{\mathrm{T}}]$ denotes an $N \times 1$ vector with element i equal to $\boldsymbol{\sigma}_{it}\boldsymbol{\sigma}_{it}^{\mathrm{T}}$. Because $(dt)^2 = 0$ and $dt \cdot dZ_{jt} = 0,\ j = 1, ..., M$, the second term of equation (13) is

$$\left(\frac{dV_t}{V_t}\right)^2 = \mathbf{w}_t^{\mathrm{T}} \left(d \log \mathbf{S}_t\right) \left(d \log \mathbf{S}_t\right)^{\mathrm{T}} \mathbf{w}_t. \tag{14}$$

Since $dZ_{kt} \cdot dZ_{\ell t} = 0, k \neq \ell,\ k, \ell = 1, ...M$, and $(dZ_{jt})^2 = dt, j = 1, ..., M$, the following expression holds

$$\left(d \log \mathbf{S}_t\right) \left(d \log \mathbf{S}_t\right)^{\mathrm{T}} = \boldsymbol{\sigma}_t \left(d\mathbf{Z}_t\right) \left(d\mathbf{Z}_t\right)^{\mathrm{T}} \boldsymbol{\sigma}_t^{\mathrm{T}} = \boldsymbol{\sigma}_t \boldsymbol{\sigma}_t^{\mathrm{T}} dt,$$

and the portfolio dynamics are thus

$$d \log V_t = \mathbf{w}_t^{\mathrm{T}} \left(d \log \mathbf{S}_t + \frac{1}{2}[\boldsymbol{\sigma}_{it}\boldsymbol{\sigma}_{it}^{\mathrm{T}}]dt\right) - \frac{1}{2}\mathbf{w}_t^{\mathrm{T}} \boldsymbol{\sigma}_t \boldsymbol{\sigma}_t^{\mathrm{T}} \mathbf{w}_t dt. \tag{15}$$

Finally, an Euler approximation of equation (15), where dt is replaced with some small but not infinitesimal time interval Δt, yields the approximate log portfolio return as

$$r_{p,t+1} = \mathbf{w}_t^{\mathrm{T}} \mathbf{r}_{t+1} + \frac{1}{2}\mathbf{w}_t^{\mathrm{T}} \boldsymbol{\sigma}_t^2 - \frac{1}{2}\mathbf{w}_t^{\mathrm{T}} \boldsymbol{\Sigma}_t \mathbf{w}_t,$$

where $r_{p,t+1} = \Delta \log V_{t+1} = \log V_{t+1} - \log V_t,\ \mathbf{r}_{t+1} = \Delta \log \mathbf{S}_{t+1},\ \Delta t = 1,\ \boldsymbol{\sigma}_t \boldsymbol{\sigma}_t^{\mathrm{T}} = \mathrm{Var}_t[\mathbf{r}_{t+1}] = \boldsymbol{\Sigma}_t$, and $[\boldsymbol{\sigma}_{it}\boldsymbol{\sigma}_{it}^{\mathrm{T}}] = [\mathrm{Var}_t[r_{1,t+1}], ..., \mathrm{Var}_t[r_{N,t+1}]]^{\mathrm{T}} = \boldsymbol{\sigma}_t^2$. The expected log portfolio return and the variance of the log portfolio return can then be written as

$$\begin{aligned} \mathrm{E}_t[r_{p,t+1}] &= \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\mu}_t + \frac{1}{2}\mathbf{w}_t^{\mathrm{T}} \boldsymbol{\sigma}_t^2 - \frac{1}{2}\mathbf{w}_t^{\mathrm{T}} \boldsymbol{\Sigma}_t \mathbf{w}_t, \\ \mathrm{Var}_t[r_{p,t+1}] &= \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\Sigma}_t \mathbf{w}_t, \end{aligned}$$

where $\boldsymbol{\mu}_t = \mathrm{E}_t[\mathbf{r}_{t+1}]$.

B Markov Chain Monte Carlo Estimation

The MCMC method for inference and parameter estimation is a Bayesian and simulation based estimation method. Traditional methods such as maximum likelihood (ML) treats the parameters of the model at hand as unknown constants, whereas the Bayesian approach is to treat the parameter vector $\boldsymbol{\vartheta}$ as an outcome of the random variable $\boldsymbol{\Theta}$. So, while other methods produce a point estimate, the Bayesian estimate is the joint distribution of the parameters conditional on the data[10]. This joint conditional distribution, referred to as the posterior distribution, can be derived via Bayes' formula as

$$p(\boldsymbol{\vartheta}|\mathbf{Y}) \propto p(\mathbf{Y}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta}),$$

where $\mathbf{Y}$ is a matrix of observations, $p(\mathbf{Y}|\boldsymbol{\vartheta})$ is the likelihood of the data, and $p(\boldsymbol{\vartheta})$ is the so called prior distribution. Knowledge about the normalizing constant is generally not required. The prior distribution has to be specified unconditional of the data by the researcher, and it can be thought of as a natural way to impose non-sample information, if there is any, and to impose stationarity and non-negativity where it is needed. If there is no non-sample information to be imposed, the prior is usually chosen so that it is as uninformative as possible, typically with a very large variance over the relevant parameter space, or it is chosen to be diffuse which means that it is completely uninformative. Diffuse priors, however, do not integrate to unity and they are therefore not well suited for all situations.

[10]The Bayesian point estimate is typically taken as the posterior mean.

The posterior distribution is often very complex and non-standard, and to explore it, the MCMC method can be used. The MCMC method samples from the posterior distribution by generating a Markov Chain $\{\boldsymbol{\vartheta}^{(j)}\}_{j=1}^{n}$ over $\boldsymbol{\vartheta}$ such that its equilibrium distribution is $p(\boldsymbol{\vartheta}|\mathbf{Y})$. The MCMC method is based on the Clifford-Hammersley theorem (Hammersley and Clifford, 1970) which states that a joint distribution $p(a, b|c)$ is completely characterized by the two conditional marginal distributions $p(a|b, c)$ and $p(b|a, c)$. In this section, $a = \boldsymbol{\mu}$, $b = \boldsymbol{\Sigma}$, and $c = \mathbf{Y}$. When the two conditional marginal distributions are standard distributions that can be easily sampled from, the simplest MCMC algorithm, the Gibbs sampler, can be used. The Gibbs sampler iteratively first updates the first parameter by drawing from its marginal posterior, keeping all the other parameters constant, then updates the second parameter by drawing from its marginal posterior, keeping all the other parameters constant and using the updated value of the first parameter, and so on. When all p parameters have been updated, the process starts over again and it does so until each parameter has been updated n times, thereby obtaining the sequence $\{\boldsymbol{\vartheta}^{(j)}\}_{j=1}^{n}$, where $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, ..., \boldsymbol{\vartheta}_p)$.

To obtain the Bayesian estimates of the mean vector of the log returns and the covariance matrix of the log returns, assume as in the paper that the log returns are jointly normally distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. This corresponds to assuming the linear regression model

$$\mathbf{Y} = \mathbf{1}\boldsymbol{\mu}^{\mathrm{T}} + \boldsymbol{\varepsilon}$$

for the returns where where element ki of $\mathbf{Y}$ is return observation k of stock i, $i = 1, 2, ..., N$, $k = 1, ..., K$, and $\boldsymbol{\varepsilon}$ is a $K \times N$ matrix of error terms whose rows are independently normally distributed with a mean vector equal to an $N \times 1$ vector of zeros and positive definite covariance matrix $\boldsymbol{\Sigma}$. With these specifications, the likelihood of the data can be written as

$$\begin{aligned} p(Y|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-K/2} \exp\left\{-\frac{1}{2}\left(Y - \mathbf{1}\boldsymbol{\mu}^{\mathrm{T}}\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left(Y - \mathbf{1}\boldsymbol{\mu}^{\mathrm{T}}\right)\right\} \qquad (16) \\ &= |\boldsymbol{\Sigma}|^{-K/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\left(Y - \mathbf{1}\boldsymbol{\mu}^{\mathrm{T}}\right)^{\mathrm{T}} \left(Y - \mathbf{1}\boldsymbol{\mu}^{\mathrm{T}}\right) \boldsymbol{\Sigma}^{-1}\right)\right\}. \end{aligned}$$

By noting that

$$\left(Y - \mathbf{1}\boldsymbol{\mu}^{\mathrm{T}}\right)^{\mathrm{T}} \left(Y - \mathbf{1}\boldsymbol{\mu}^{\mathrm{T}}\right) = K\hat{\boldsymbol{\Sigma}} + K\left(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\right)^{\mathrm{T}},$$

where

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \left(\frac{1}{K}\mathbf{1}^{\mathrm{T}} Y\right)^{\mathrm{T}}, \\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{K}\left(Y - \mathbf{1}\hat{\boldsymbol{\mu}}^{\mathrm{T}}\right)^{\mathrm{T}} \left(Y - \mathbf{1}\hat{\boldsymbol{\mu}}^{\mathrm{T}}\right), \end{aligned}$$

equation (16) can be simplified into

$$p(Y|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-K/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(K\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right) - \frac{1}{2}\mathrm{tr}\left(K\left(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}\right)\right\}.$$

I assume standard conjugate priors according to

$$\begin{aligned} \boldsymbol{\mu} &\in \mathcal{N}(a, A), \\ \boldsymbol{\Sigma} &\in \mathcal{IW}(c, C), \end{aligned}$$

where $\mathcal{N}$ denotes the multivariate normal distribution and $\mathcal{IW}$ denotes the inverted Wishart distribution. A conjugate prior is a distribution under which the prior and posterior of a parameter is the same type of distribution, but with different hyperparameters. These assumption imply that the posterior of μ is

$$\begin{aligned}
p(\mu|Y,\Sigma) &\propto p(Y|\mu,\Sigma)p(\mu) \\
&\propto \exp\left\{-\frac{1}{2}\left(\mu-a\right)^{\mathrm{T}}A^{-1}\left(\mu-a\right)\right\} \\
&\times \exp\left\{-\frac{1}{2}\left(\mu-\hat{\mu}\right)^{\mathrm{T}}\left(K\Sigma^{-1}\right)\left(\mu-\hat{\mu}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\mu^{\mathrm{T}}\left(K\Sigma^{-1}+A^{-1}\right)\mu-2\mu^{\mathrm{T}}\left(\left(K\Sigma^{-1}\right)\hat{\mu}+A^{-1}a\right)\right)\right\},
\end{aligned}$$

which implies that the posterior of μ is $\mathcal{N}(a^*, A^*)$, where

$$\begin{aligned}
a^* &= A^*\left(K\Sigma^{-1}\hat{\mu}+A^{-1}a\right), \\
A^* &= \left(K\Sigma^{-1}+A^{-1}\right)^{-1}.
\end{aligned}$$

The posterior of Σ is derived as

$$\begin{aligned}
p(\Sigma|Y,\mu) &\propto p(Y|\mu,\Sigma)p(\Sigma) \\
&\propto |\Sigma|^{-K/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(K\hat{\Sigma}\Sigma^{-1}\right)-\frac{1}{2}\mathrm{tr}\left(K\left(\mu-\hat{\mu}\right)\left(\mu-\hat{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\right)\right\} \\
&\times |\Sigma|^{-(c+N+2)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(C\Sigma^{-1}\right)\right\} \\
&= |\Sigma|^{-(c+K+N+2)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\left(C+K\hat{\Sigma}+K\left(\mu-\hat{\mu}\right)\left(\mu-\hat{\mu}\right)^{\mathrm{T}}\right)\Sigma^{-1}\right)\right\},
\end{aligned}$$

implying that the posterior of Σ is $\mathcal{IW}(c^*, C^*)$, where

$$\begin{aligned}
c^* &= c+K, \\
C^* &= C+K\hat{\Sigma}+K\left(\mu-\hat{\mu}\right)\left(\mu-\hat{\mu}\right)^{\mathrm{T}}.
\end{aligned}$$

When I estimate this model I use the following hyperparameters in the prior distributions: $a_i = 0, A_{ij} = 0.25$ when $i = j$ and $A_{ij} = 0$ when $i \neq j$, $c_i = N + 2 = 32$, and $C_{ij} = 0.2$ when $i = j$ and $C_{ij} = 0$ when $i \neq j$, $i, j = 1, 2, ..., N$. I run the Gibbs sampler 1100 times, discarding the first 100 iteration as burn in. The standard deviations of the elements of the Bayesian estimates of μ and Σ are then simply the standard deviations of the elements in the resulting sequence $\{\mu^{(j)}, \Sigma^{(j)}\}_{j=1}^{n}$.

C Tables and Figures

	Mean	Standard deviation
Allied Signal Inc.	0.99	6.63
Aluminum Co. of America	1.08	7.94
American Express Co.	0.73	8.55
AT&T Corp.	1.08	6.28
Bethlehem Steel	-0.04	14.32
Boeing Co.	1.21	7.41
Caterpillar Inc.	0.98	8.89
Chevron	1.23	6.06
Coca Cola Co.	2.15	5.67
Du Pont	1.25	6.78
Eastman Kodak	1.09	6.67
Exxon Corp.	1.30	4.38
General Electric Co.	1.39	6.14
General Motors Corp.	0.73	8.17
Goodyear	1.13	9.77
Int. Business Machines	-0.15	7.38
Int. Paper Company	1.14	7.42
J.P Morgan & Co.	1.09	7.22
McDonald s Corp.	1.43	6.24
Merck & Company, Inc.	1.99	6.42
Minnesota Mining & Mfg.	1.18	5.60
Philip Morris Co.'s Inc.	2.07	7.50
Procter & Gamble Co.	1.52	6.06
Sears Roebuck & Co.	1.16	7.90
Texaco Inc.	1.36	5.48
Union Carbide	1.86	8.81
United Technologies Corp.	0.92	7.81
Walt Disney Co.	1.82	8.12
Westinghouse Electric	0.02	8.69
Woolworth Corp.	0.15	9.72

Table 1: Means and standard deviations in percents per month for the 30 companies that belonged to the DJIA at the end of 1995 calculated from the monthly returns between January 1986 and December 1995.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.	28.	29.	30.
1. Allied Signal Inc.	1																													
2. Aluminum Co. of America	0.44	1																												
3. American Express Co.	0.39	0.43	1																											
4. AT&T Corp.	0.38	0.30	0.33	1																										
5. Bethlehem Steel	0.31	0.55	0.30	0.20	1																									
6. Boeing Co.	0.49	0.30	0.46	0.37	0.27	1																								
7. Caterpillar Inc.	0.39	0.65	0.29	0.27	0.55	0.36	1																							
8. Chevron	0.39	0.27	0.38	0.38	0.21	0.29	0.24	1																						
9. Coca Cola Co.	0.43	0.24	0.44	0.39	0.19	0.53	0.25	0.25	1																					
10. Du Pont	0.52	0.60	0.44	0.46	0.43	0.45	0.52	0.53	0.38	1																				
11. Eastman Kodak	0.36	0.31	0.28	0.31	0.19	0.23	0.26	0.22	0.40	0.43	1																			
12. Exxon Corp.	0.42	0.26	0.42	0.38	0.21	0.31	0.22	0.73	0.39	0.56	0.33	1																		
13. Electric Co.	0.58	0.54	0.59	0.52	0.47	0.57	0.49	0.43	0.56	0.63	0.41	0.53	1																	
14. General Motors Corp.	0.41	0.52	0.41	0.28	0.52	0.38	0.50	0.30	0.23	0.55	0.31	0.31	0.54	1																
15. Goodyear	0.45	0.46	0.41	0.21	0.46	0.28	0.47	0.22	0.31	0.45	0.24	0.28	0.47	0.43	1															
16. Int. Business Machines	0.28	0.39	0.21	0.10	0.33	0.21	0.39	0.19	0.15	0.42	0.37	0.27	0.32	0.47	0.15	1														
17. Int. Paper Company	0.48	0.62	0.41	0.34	0.40	0.42	0.49	0.27	0.37	0.74	0.37	0.42	0.64	0.56	0.45	0.37	1													
18. J.P Morgan & Co.	0.39	0.29	0.44	0.45	0.22	0.45	0.26	0.43	0.49	0.54	0.36	0.56	0.56	0.20	0.33	0.13	0.49	1												
19. McDonald s Corp.	0.49	0.31	0.49	0.36	0.34	0.47	0.37	0.30	0.52	0.52	0.46	0.47	0.61	0.41	0.48	0.34	0.38	0.48	1											
20. Merck & Company, Inc.	0.31	0.33	0.43	0.26	0.27	0.41	0.27	0.17	0.57	0.50	0.37	0.40	0.52	0.24	0.33	0.30	0.46	0.43	0.48	1										
21. Minnesota Mining & Mfg.	0.45	0.51	0.46	0.34	0.37	0.51	0.51	0.34	0.51	0.66	0.48	0.47	0.71	0.47	0.45	0.38	0.69	0.56	0.53	0.52	1									
22. Philip Morris Co.'s Inc.	0.30	0.30	0.42	0.32	0.26	0.31	0.23	0.29	0.60	0.39	0.40	0.43	0.48	0.28	0.24	0.24	0.40	0.38	0.50	0.57	0.49	1								
23. Procter & Gamble Co.	0.37	0.24	0.44	0.36	0.21	0.39	0.33	0.30	0.59	0.49	0.36	0.44	0.50	0.28	0.26	0.24	0.45	0.48	0.63	0.55	0.53	0.57	1							
24. Sears Roebuck & Co.	0.53	0.45	0.62	0.37	0.37	0.48	0.43	0.37	0.44	0.59	0.51	0.36	0.61	0.56	0.45	0.30	0.51	0.43	0.59	0.38	0.52	0.47	0.38	1						
25. Texaco Inc.	0.31	0.36	0.24	0.27	0.30	0.29	0.42	0.63	0.29	0.51	0.25	0.57	0.36	0.23	0.30	0.18	0.35	0.30	0.27	0.22	0.40	0.30	0.29	0.31	1					
26. Union Carbide	0.41	0.45	0.38	0.25	0.35	0.28	0.47	0.16	0.27	0.54	0.28	0.17	0.37	0.48	0.30	0.32	0.55	0.28	0.25	0.28	0.43	0.29	0.36	0.49	0.16	1				
27. United Technologies Corp.	0.50	0.54	0.55	0.45	0.46	0.64	0.54	0.42	0.56	0.62	0.43	0.44	0.68	0.51	0.47	0.33	0.55	0.49	0.50	0.51	0.63	0.50	0.46	0.58	0.45	0.45	1			
28. Walt Disney Co.	0.56	0.44	0.44	0.34	0.43	0.44	0.47	0.33	0.58	0.52	0.39	0.36	0.59	0.48	0.47	0.33	0.51	0.42	0.66	0.47	0.55	0.54	0.56	0.52	0.28	0.42	0.53	1		
29. Westinghouse Electric	0.38	0.47	0.58	0.30	0.32	0.42	0.40	0.33	0.39	0.48	0.39	0.43	0.66	0.49	0.42	0.30	0.50	0.38	0.53	0.39	0.63	0.45	0.44	0.57	0.37	0.34	0.58	0.48	1	
30. Woolworth Corp.	0.38	0.35	0.48	0.31	0.36	0.48	0.30	0.25	0.42	0.48	0.26	0.31	0.47	0.43	0.40	0.28	0.52	0.36	0.47	0.41	0.52	0.45	0.43	0.54	0.26	0.47	0.53	0.52	0.42	1

Table 2: Correlations for the 30 companies that belonged to the DJIA at the end of 1995 calculated from the monthly returns between January 1986 and December 1995.

	Optimal Portfolio Weight				
	Relative Risk Aversion				
	$\gamma = 1$	$\gamma = 3$	$\gamma = 5$	$\gamma = 10$	$\gamma = 25$
Allied Signal Inc.	0	0	0	0	0
Aluminum Co. of America	0	0	0	0	0
American Express Co.	0	0	0	0	0
AT&T Corp.	0	0	0	0	0.0464
Bethlehem Steel	0	0	0	0	0
Boeing Co.	0	0	0	0	0
Caterpillar Inc.	0	0	0	0	0
Chevron	0	0	0	0	0
Coca Cola Co.	0.6199	0.5902	0.5439	0.3977	0.2502
Du Pont	0	0	0	0	0
Eastman Kodak	0	0	0	0	0.0204
Exxon Corp.	0	0	0	0.1543	0.3248
General Electric Co.	0	0	0	0	0
General Motors Corp.	0	0	0	0	0
Goodyear	0	0	0	0	0
Int. Business Machines	0	0	0	0	0
Int. Paper Co.	0	0	0	0	0
J.P. Morgan & Co.	0	0	0	0	0
McDonald's Corp.	0	0	0	0	0
Merck & Co., Inc.	0	0.1260	0.1758	0.1512	0.0993
Minnesota Mining & Mfg.	0	0	0	0	0
Philip Morris Co.'s Inc.	0.2709	0.1236	0.0775	0.0070	0
Procter & Gamble Co.	0	0	0	0	0
Sears Roebuck & Co.	0	0	0	0	0
Texaco Inc.	0	0	0.0433	0.1622	0.1646
Union Carbide	0.1092	0.1602	0.1596	0.1276	0.0943
United Technologies Corp.	0	0	0	0	0
Walt Disney Co.	0	0	0	0	0
Westinghouse Electric	0	0	0	0	0
Woolworth Corp.	0	0	0	0	0

Table 3: Optimal portfolio weights when the portfolio weights are restricted to be non-negative.

	Optimal Portfolio Weight				
	Relative Risk Aversion				
	$\gamma = 1$	$\gamma = 3$	$\gamma = 5$	$\gamma = 10$	$\gamma = 25$
Allied Signal Inc.	-0.1	-0.1	-0.1	-0.1	-0.0911
Aluminum Co. of America	-0.1	0.0110	0.1169	0.1548	0.1538
American Express Co.	-0.1	-0.1	-0.1	-0.1	-0.1
AT&T Corp.	-0.1	-0.1	-0.1	-0.0231	0.0695
Bethlehem Steel	-0.1	-0.1	-0.1	-0.0827	-0.0482
Boeing Co.	-0.1	-0.1	-0.0771	0.0332	0.0835
Caterpillar Inc.	-0.1	-0.1	-0.0831	-0.0489	-0.0513
Chevron	-0.1	0.1228	0.2102	0.1126	0.0510
Coca Cola Co.	1.5597	0.9799	0.7223	0.4564	0.2791
Du Pont	-0.1	-0.1	-0.1	-0.1	-0.1
Eastman Kodak	-0.1	-0.1	-0.1	-0.0219	0.0282
Exxon Corp.	-0.1	-0.1	-0.0956	0.2185	0.3702
General Electric Co.	-0.1	-0.0141	0.1470	0.1216	0.0286
General Motors Corp.	-0.1	-0.1	-0.0711	0.0295	0.0592
Goodyear	-0.1	0.0764	0.0606	0.0491	0.0445
Int. Business Machines	-0.1	-0.1	-0.1	-0.1	-0.0596
Int. Paper Co.	-0.1	-0.1	-0.1	-0.1	-0.0835
J.P. Morgan & Co.	-0.1	-0.1	-0.1	-0.1	-0.1
McDonald's Corp.	-0.1	-0.0306	0.1885	0.2209	0.1695
Merck & Co., Inc.	0.4540	0.5027	0.4015	0.2773	0.1720
Minnesota Mining & Mfg.	-0.1	-0.1	-0.1	-0.0736	0.0861
Philip Morris Co.'s Inc.	0.6131	0.2796	0.1719	0.0917	0.0144
Procter & Gamble Co.	-0.1	-0.1	-0.1	-0.1	-0.0183
Sears Roebuck & Co.	-0.1	-0.1	-0.1	-0.0893	-0.0226
Texaco Inc.	-0.1	0.2138	0.2752	0.2337	0.1832
Union Carbide	0.7726	0.5869	0.4816	0.3492	0.2287
United Technologies Corp.	-0.1	-0.1	-0.1	-0.1	-0.1
Walt Disney Co.	0.1006	0.1717	0.0513	-0.0091	-0.0471
Westinghouse Electric	-0.1	-0.1	-0.1	-0.1	-0.1
Woolworth Corp.	-0.1	-0.1	-0.1	-0.1	-0.1

Table 4: Optimal portfolio weights when the portfolio weights are restricted to be greater than or equal to -10 percent.

Relative Risk Aversion	Errors in	Mean CEL			
		Size of Errors			
		$k = 0.05$	$k = 0.10$	$k = 0.15$	$k = 0.20$
$\gamma = 1$					
	Means	0.0303	0.0618	0.0860	0.1095
	Covariance Matrix	0.0003	0.0014	0.0034	0.0055
	Variances	0.0002	0.0010	0.0024	0.0041
	Covariances	0.0001	0.0002	0.0007	0.0015
$\gamma = 3$					
	Means	0.0177	0.0573	0.1026	0.1459
	Covariance Matrix	0.0009	0.0037	0.0075	0.0119
	Variances	0.0003	0.0016	0.0040	0.0068
	Covariances	0.0005	0.0021	0.0041	0.0064
$\gamma = 5$					
	Means	0.0145	0.0502	0.0984	0.1462
	Covariance Matrix	0.0016	0.0066	0.0139	0.0246
	Variances	0.0008	0.0035	0.0081	0.0159
	Covariances	0.0008	0.0032	0.0068	0.0114
$\gamma = 10$					
	Means	0.0103	0.0382	0.0777	0.1231
	Covariance Matrix	0.0022	0.0102	0.0263	0.0537
	Variances	0.0011	0.0051	0.0142	0.0334
	Covariances	0.0011	0.0047	0.0116	0.0221
$\gamma = 25$					
	Means	0.0040	0.0180	0.0431	0.0752
	Covariance Matrix	0.0048	0.0239	0.0643	0.1393
	Variances	0.0025	0.0119	0.0341	0.0804
	Covariances	0.0021	0.0108	0.0277	0.0534

Table 5: Average percentage cash equivalent loss when the portfolio weights are restricted to be non-negative.

Relative Risk Aversion	Errors in	Mean CEL			
		Size of Errors			
		$k = 0.05$	$k = 0.10$	$k = 0.15$	$k = 0.20$
$\gamma = 1$					
	Means	0.0714	0.2158	0.3947	0.5732
	Covariance Matrix	0.0039	0.0163	0.0400	0.0757
	Variances	0.0011	0.0053	0.0154	0.0362
	Covariances	0.0027	0.0108	0.0238	0.0409
$\gamma = 3$					
	Means	0.0496	0.1770	0.3271	0.4871
	Covariance Matrix	0.0136	0.0673	0.1859	0.3878
	Variances	0.0034	0.0176	0.0617	0.1782
	Covariances	0.0099	0.0434	0.1062	0.1990
$\gamma = 5$					
	Means	0.0394	0.1479	0.2859	0.4224
	Covariance Matrix	0.0226	0.1172	0.3395	0.7291
	Variances	0.0045	0.0223	0.0954	0.3345
	Covariances	0.0172	0.0805	0.2027	0.3730
$\gamma = 10$					
	Means	0.0275	0.1009	0.2050	0.3201
	Covariance Matrix	0.0442	0.2582	0.7869	1.6942
	Variances	0.0071	0.0419	0.2279	0.7957
	Covariances	0.0347	0.1724	0.4442	0.8134
$\gamma = 25$					
	Means	0.0130	0.0507	0.1082	0.1832
	Covariance Matrix	0.0987	0.6864	2.1625	4.5316
	Variances	0.0138	0.0976	0.5646	2.1785
	Covariances	0.0780	0.4373	1.2083	2.1831

Table 6: Average percentage cash equivalent loss when the portfolio weights are restricted to be greater than or equal to -10 percent.

Relative Risk	Errors in	Mean CEL	
		Large Errors	Small Errors
$\gamma = 1$			
	Means	0.3582	0.1281
	Covariance Matrix	0.0039	0.0008
	Variances	0.0019	0.0004
	Covariances	0.0019	0.0004
$\gamma = 3$			
	Means	0.4034	0.1476
	Covariance Matrix	0.0105	0.0031
	Variances	0.0033	0.0007
	Covariances	0.0078	0.0023
$\gamma = 5$			
	Means	0.3831	0.1397
	Covariance Matrix	0.0216	0.0058
	Variances	0.0069	0.0015
	Covariances	0.0156	0.0043
$\gamma = 10$			
	Means	0.3321	0.1116
	Covariance Matrix	0.0521	0.0115
	Variances	0.0110	0.0021
	Covariances	0.0391	0.0089
$\gamma = 25$			
	Means	0.2416	0.0658
	Covariance Matrix	0.1513	0.0279
	Variances	0.0267	0.0049
	Covariances	0.1130	0.0217

Table 7: Average percentage cash equivalent loss when the portfolio weights are restricted to be non-negative. *Large Errors* means that the size of the simulated estimation error for a specific element (an element in the mean vector or in the covariance matrix) is the posterior standard deviation of that parameter when estimated with the MCMC method divided by its assumed true value which is the ordinary sample estimate. *Small Errors* means that the corresponding large error has simply been cut in half.

Relative Risk	Errors in	Mean CEL	
		Large Errors	Small Errors
$\gamma = 1$			
	Means	1.5405	0.6191
	Covariance Matrix	0.0689	0.0162
	Variances	0.0115	0.0022
	Covariances	0.0551	0.0137
$\gamma = 3$			
	Means	1.2758	0.5139
	Covariance Matrix	0.3803	0.0839
	Variances	0.0437	0.0068
	Covariances	0.3069	0.0731
$\gamma = 5$			
	Means	1.1176	0.4473
	Covariance Matrix	0.7093	0.1596
	Variances	0.0631	0.0085
	Covariances	0.5711	0.1413
$\gamma = 10$			
	Means	0.9374	0.3534
	Covariance Matrix	1.5261	0.3443
	Variances	0.1473	0.0138
	Covariances	1.2023	0.3016
$\gamma = 25$			
	Means	0.6720	0.2089
	Covariance Matrix	3.9741	0.9295
	Variances	0.3356	0.0291
	Covariances	3.1518	0.8093

Table 8: Average percentage cash equivalent loss when the portfolio weights are restricted to be greater than or equal to -10 percent. *Large Errors* means that the size of the simulated estimation error for a specific element (an element in the mean vector or in the covariance matrix) is the posterior standard deviation of that parameter when estimated with the MCMC method divided by its assumed true value which is the ordinary sample estimate. *Small Errors* means that the corresponding large error has simply been cut in half.

Relative Risk Aversion	CEL		
	Lower Bound		
	$\ell = 0$	$\ell = -0.05$	$\ell = -0.1$
$\gamma = 1$	0.0002	0.0006	0.0012
$\gamma = 3$	0.0014	0.0013	0.0012
$\gamma = 5$	0.0019	0.0012	0.0009
$\gamma = 10$	0.0012	0.0015	0.0037
$\gamma = 25$	0.0030	0.0039	0.0075

Table 9: The percentage cash equivalent loss for the investor when he solves the traditional MV optimization problem of equation (12) instead of equation (11) with the additional constraint that $\mathbf{w}_t \geq \ell\mathbf{1}$.

Relative Risk Aversion	CEL		
	Lower Bound		
	$\ell = 0$	$\ell = -0.05$	$\ell = -0.1$
$\gamma = 1$	0.6842	1.5115	2.1673
$\gamma = 3$	0.5779	1.1361	1.5096
$\gamma = 5$	0.5819	1.0578	1.3640
$\gamma = 10$	0.7608	1.1801	1.3880
$\gamma = 25$	1.4838	2.0122	2.1468

Table 10: The percentage cash equivalent loss for the investor when he simply invests an equal amount in each stock instead of solving equation (11) with the additional constraint that $\mathbf{w}_t \geq \ell\mathbf{1}$.

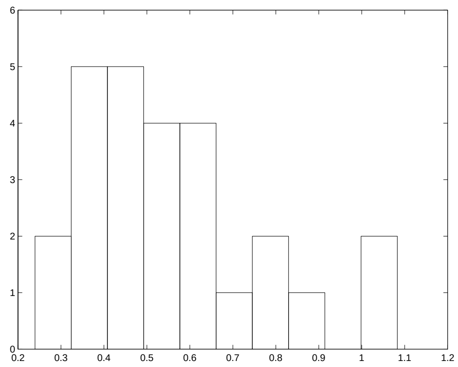

Figure 1: Histogram of the estimation error sizes $k_{\mu_{it}}, i = 1, 2, ..., N$, for the elements of the mean vector, obtained using the MCMC estimation. (The four values of $k_{\mu_{it}}$ that are restricted to be equal to 1 are omitted.)

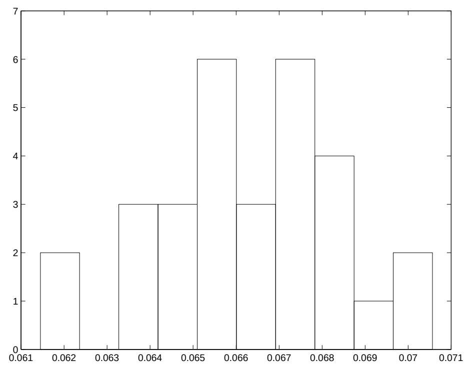

Figure 2: Histogram of the estimation error sizes $k_{\sigma_{iit}}, i = 1, 2, ..., N$, for the diagonal elements of the covariance matrix, obtained using the MCMC estimation.

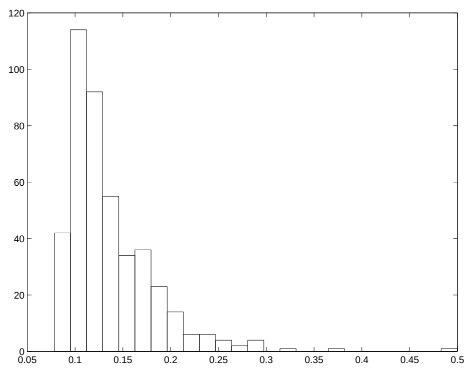

Figure 3: Histogram of the estimation error sizes $k_{\sigma_{ijt}}, i = 1, 2, ..., N-1, j = i+1, i+2, ..., N$, for the off-diagonal elements of the covariance matrix, obtained using the MCMC estimation.

D Discussion on the Results in Chopra and Ziemba (1993)

What is the cause of the large difference between the results in this paper and in Chopra and Ziemba (1993)? To find an explanation I begin by deriving the CE under the assumptions of Chopra and Ziemba (1993). Using equation (3) and $W_t = 1$, I get

$$CE(\mathbf{w}_t) = 1 + \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\eta}_t - \frac{a}{2} \mathbf{w}_t^{\mathrm{T}} \boldsymbol{\Omega}_t \mathbf{w}_t. \tag{17}$$

This definition of the CE is then plugged into the simulation exercise described in the Section 3, together with non-negativity constraints on the portfolio weights and their estimation error sizes. The data set is also changed to exactly match theirs. The result is presented in Table 11. This exercise shows that the CEL values are still of the same magnitude as in Table 5. Only if the 1-term in equation (17) is omitted do I get the same results as Chopra and Ziemba (1993) (see Table 12). When Chopra and Ziemba (1993) calculate the CE, they refer to a paper by Freund (1956), in which utility is defined over net returns, not next-period wealth. In that case the 1-term would disappear from equation (17). However, since next-period wealth is the gross return of the portfolio if $W_t = 1$, the 1-term should not be left out from the expression for the CE. Consequently, with the CE defined as the solution to $u(CE) = \mathcal{U}(\mathbf{w}_t)$ and since $\mathbf{w}_t^{*\mathrm{T}} \boldsymbol{\eta}_t - (a/2)\mathbf{w}_t^{*\mathrm{T}} \boldsymbol{\Omega}_t \mathbf{w}_t^*$ is in the order of a few percents, it seems that the numbers in Exhibit 3 of Chopra and Ziemba (1993) may be almost 100 times too large. Alternatively, if the measure was cash equivalence return, i.e., $u(W_t(1 + r_{CE})) = \mathcal{U}(\mathbf{w}_t)$, the 1-term would disappear. In any case, it is not the use of the power utility function together with lognormality that cause the difference. When the sensitivity measure is the CEL, the effect of estimation error is the more or less the same, irrespective of if we assume negative exponential utility together with jointly normally distributed returns or power utility together with jointly lognormal returns, and the result that errors in the covariance matrix are more important when short selling is allowed is still valid. However, the loss under the certainty equivalence return measure would for even quite small amounts of short selling be very very large.

Absolute Risk Aversion	Errors in	Mean CEL			
		Size of Errors			
		$k = 0.05$	$k = 0.10$	$k = 0.15$	$k = 0.20$
$a = 2.6667$					
	Means	0.0089	0.0428	0.0960	0.1507
	Covariance Matrix	0.0007	0.0028	0.0062	0.0110
	Variances	0.0005	0.0020	0.0044	0.0075
	Covariances	0.0002	0.0007	0.0016	0.0030
$a = 4$					
	Means	0.0089	0.0391	0.0856	0.1321
	Covariance Matrix	0.0012	0.0052	0.0114	0.0198
	Variances	0.0009	0.0034	0.0076	0.0135
	Covariances	0.0003	0.0016	0.0040	0.0070
$a = 8$					
	Means	0.0081	0.0278	0.0581	0.0945
	Covariance Matrix	0.0037	0.0124	0.0252	0.0447
	Variances	0.0023	0.0088	0.0186	0.0345
	Covariances	0.0016	0.0053	0.0101	0.0159

Table 11: Average percentage cash equivalent loss when the data set is the same as in Chopra and Ziemba (1993), utility is negative exponential, returns are jointly normally distributed, and the portfolio weights are restricted to be non-negative. The values for the parameter a are the same as in Chopra and Ziemba (1993).

Absolute Risk Aversion	Errors in	Mean CEL			
		Size of Errors			
		$k = 0.05$	$k = 0.10$	$k = 0.15$	$k = 0.20$
$a = 2.6667$					
	Means	0.5355	2.5638	5.6733	9.1277
	Covariance Matrix	0.0399	0.1615	0.3723	0.6787
	Variances	0.0308	0.1182	0.2607	0.4680
	Covariances	0.0090	0.0409	0.1000	0.1860
$a = 4$					
	Means	0.5872	2.6385	5.6308	8.9548
	Covariance Matrix	0.0804	0.3457	0.7716	1.3014
	Variances	0.0585	0.2267	0.5176	0.9030
	Covariances	0.0207	0.1100	0.2638	0.4598
$a = 8$					
	Means	0.7843	2.6372	5.5470	9.1764
	Covariance Matrix	0.3626	1.2111	2.5020	4.4827
	Variances	0.2243	0.8563	1.8427	3.4361
	Covariances	0.1553	0.5242	0.9961	1.5948

Table 12: Average percentage ”cash equivalent” loss when the data set is the same as in Chopra and Ziemba (1993), utility is negative exponential, returns are jointly normally distributed, the 1-term is omitted from equation (17), and the portfolio weights are restricted to be non-negative. The values for the parameter a are the same as in Chopra and Ziemba (1993).