DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 1/08

A General Approach to Bayesian Portfolio Optimization

by

Alexander Bade

Gabriel Frahm

Uwe Jaekel

3rd version
November 12, 2008

DISKUSSIONSBEITRÄGE ZUR STATISTIK UND ÖKONOMETRIE

SEMINAR FÜR WIRTSCHAFTS- UND SOZIALSTATISTIK
UNIVERSITÄT ZU KÖLN

Albertus-Magnus-Platz, D-50923 Köln, Deutschland

This page intentionally left blank.

DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 1/08

A General Approach to Bayesian Portfolio Optimization

by

Alexander Bade[1]
Gabriel Frahm[2]
Uwe Jaekel[3]

3rd version
November 12, 2008

Abstract

We develop a general approach to portfolio optimization taking account of estimation risk and stylized facts of empirical finance. This is done within a Bayesian framework. The approximation of the posterior distribution of the unknown model parameters is based on a parallel tempering algorithm. The portfolio optimization is done using the first two moments of the predictive discrete asset return distribution. For illustration purposes we apply our method to empirical stock market data where daily asset log-returns are assumed to follow an orthogonal MGARCH process with t-distributed perturbations. Our results are compared with other portfolios suggested by popular optimization strategies.

Keywords: Bayesian portfolio optimization, Gordin's condition, Markov chain Monte Carlo, Stylized facts.

AMS Subject Classification: Primary 62F15, Secondary 91B28.

[1]Corresponding author. University of Cologne, Graduate School of Risk Management, Albertus-Magnus-Platz, 50923 Cologne, Germany. Phone: +49 221 470-7709, email: bade@wiso.uni-koeln.de.

[2]University of Cologne, Statistics & Econometrics, Department of Economic and Social Statistics, Albertus-Magnus-Platz, 50923 Cologne, Germany. Phone: +49 221 470-4267, email: frahm@statistik.uni-koeln.de.

[3]University of Applied Sciences Koblenz, RheinAhrCampus, Department of Mathematics and Technology, Südallee 2, 53424 Remagen, Germany. Phone: +49 2642 932-334, email: jaekel@rheinahrcampus.de.

This page intentionally left blank.

A General Approach to Bayesian Portfolio Optimization*

Alexander Bade†
University of Cologne
Graduate School of Risk Management
Albertus-Magnus-Platz, 50923 Cologne, Germany

Gabriel Frahm‡
Department of Economic and Social Statistics, University of Cologne
Albertus-Magnus-Platz, 50923 Cologne, Germany

Uwe Jaekel§
University of Applied Sciences Koblenz, RheinAhrCampus
Südallee 2, 53424 Remagen, Germany

12th November 2008

Abstract

We develop a general approach to portfolio optimization taking account of estimation risk and stylized facts of empirical finance. This is done within a Bayesian framework. The approximation of the posterior distribution of the unknown model parameters is based on a parallel tempering algorithm. The portfolio optimization is done using the first two moments of the predictive discrete asset return distribution. For illustration purposes we apply our method to empirical stock market data where daily asset log-returns are assumed to follow an orthogonal MGARCH process with t-distributed perturbations. Our results are compared with other portfolios suggested by popular optimization strategies.

Keywords: Bayesian portfolio optimization, Gordin's condition, Markov chain Monte Carlo, Stylized facts.

AMS Subject Classification: Primary 62F15, Secondary 91B28.

---

*Published online first 2008 in *Mathematical Methods of Operations Research*. The original publication is available at http://www.springerlink.com.

†Corresponding author. Phone: +49 221 470-3535, email: bade@wiso.uni-koeln.de.

‡Phone: +49 221 470-4267, email: frahm@statistik.uni-koeln.de.

§Phone: +49 2642 932-334, email: jaekel@rheinahrcampus.de.

1 Motivation

Traditional portfolio optimization strategies are susceptible to parameter uncertainty (Jorion, 1986, Kalymon, 1971, Klein and Bawa, 1976, Markowitz, 1952, Michaud, 1989). Estimation risk is mainly driven by the uncertainty regarding the expected asset returns rather than their variances and covariances (Chopra and Ziemba, 1993). However, it can be shown that estimating the covariance matrix is also problematic if the sample size is small compared to the number of assets (Frahm, 2007, Kempf and Memmel, 2006). Many portfolio optimization approaches rely on rather simple assumptions about the distribution of asset returns. However, it is well-known that short-term financial data can be heavy-tailed or at least leptokurtic, tail-dependent, skewed or possessing other kinds of asymmetries. Financial time series typically exhibit volatility clusters or even long-memory which holds especially if log-price changes (so-called *log-returns*) of stocks, stock indices, and foreign exchange rates are considered. Moreover, high-frequency data generally are non-stationary, have jumps, and are strongly dependent.

One might argue that the stylized facts do not matter for long investment horizons since Gordin's central limit theorem (Hayashi, 2000, p. 404) takes effect even for ergodic stationary processes. For example, many applications in finance rely on the normal distribution assumption and so low-frequency data are used to estimate the expected values of long-term, such as monthly or quarterly, asset returns. Indeed, Merton (1980) showed that the estimation of expected returns generally cannot be improved by increasing the sampling frequency. However, decreasing the sampling frequency leads to a loss of statistical efficiency since relevant information about the variances and covariances of asset returns get lost. Today's availability of high-frequency data offers new opportunities for statistical analysis, since these data include much more information than samples of low-frequency data. Nevertheless, by using high-frequency data and *ignoring* the stylized facts of empirical finance we would also obtain inaccurate estimates of the optimal portfolio weights. That means when working with high-frequency-data we need an appropriate model which accounts for the specific characteristics of the data generating process. The principal goal of this paper is to present a general approach which takes account of both estimation risks and stylized facts. Such kind of approach nowadays is feasible due to the permanent rise of computational power, especially the facilities of high-performance computing.

In order to incorporate estimation risk we rely on the Bayesian framework. This will be described in detail in Section 2. The Bayesian framework has several advantages. First of all we are able to make *finite-sample* inferences. This is important even for a large number of observations since the effective sample size strongly depends on the number of observations relative to the number of assets (Frahm and Jaekel, 2007). Further, Bayesian analysis allows us to consider not only historical data but also to incorporate prior information such as expert knowledge. This can lead to more reasonable and well-diversified portfolios rather than relying on pure statistical portfolio optimization methods (Black and Litterman, 1992, Herold and Maurer, 2006, Scherer and Martin, 2007, Ch. 7). The dynamics of high-frequency data might become very complicated so that traditional estimation procedures such as maximum-likelihood estimation quickly hit the wall. In contrast, by using contemporary methods of numerical integration such as Markov chain Monte Carlo or importance sampling, calculating the Bayesian posterior distribution of some parameter is possible even for very complicated time series models (Geweke, 1989, 1995).

For the purpose of portfolio optimization we are interested in the *predictive distribution* of asset returns. The predictive distribution combines both estimation risk and market risk. Many Bayesian approaches to portfolio optimization are based on a purely analytical fundament (Garlappi et al., 2007, Jorion, 1986, Klein and Bawa, 1976, Polson and Tew, 2000, Meucci, 2005, Ch. 7). However, this is not suitable if we want to take stylized facts into account and then generally it is not possible to find the predictive distribution analytically. To avoid limitations of such kind, we suggest a Metropolis-Hastings-like algorithm for simulating the posterior distribution of the unknown parameters. This is derived on the basis of empirical information obtained from time series data and prior information possibly given by an expert. The Markov chain Monte Carlo method belongs to the broad class of tempering algorithms which have been frequently used in natural sciences and proven to be able to simulate high-order distributions. It is therefore natural to apply them to high-order financial problems like portfolio optimization. By choosing a numerical framework, principally we can use almost any probabilistic model for the data and parameters. In Section 4 we will present a realistic portfolio optimization problem which has been performed on a standard PC in reasonable time.

2 The General Approach

2.1 Portfolio Optimization Problem

In the following we consider the *discrete predictive returns* of several assets after some long investment horizon. We specifically concentrate on discrete or, say, simple returns instead of log-returns for two reasons:

(1) Traditional portfolio theory is based on and can work only with discrete returns rather than, e.g., log-returns.

(2) Moreover, discrete returns usually differ substantially from log-returns if the investment horizon is long.

The latter is often neglected in literature. Moreover, we concentrate on long investment horizons since in practice investors usually do not want to liquidate or re-balance a portfolio each day or week. In contrast, we can think of, e.g., quarterly or yearly investment periods. The meaning of 'predictive' asset returns is to be understood in the Bayesian sense and will be explained later on in more detail. Roughly speaking, the distribution of predictive asset returns do not only account for market risk but also for the parameter uncertainty which is always present if the parameters of some model for the asset returns are unknown.

Let $R = (R_1, \ldots, R_d)$ be a d-dimensional vector of discrete predictive asset returns, $\mu = \mathbb{E}(R)$ the $d \times 1$ vector of predictive expected returns and $\Sigma = \mathbb{V}\mathrm{ar}(R) < \infty$ the corresponding $d \times d$ matrix of predictive variances and covariances. We are searching for

$$w = \arg\max_{v} \varphi(v'\mu, v'\Sigma\, v)\,, \qquad \text{s.t. } v \in \mathcal{C} \subset \mathbb{R}^d\,, \tag{1}$$

where v represents a portfolio, i.e. a vector of asset weights and φ is an appropriate objective function (i.e. φ is strongly increasing in the first and decreasing in the second argument) such as the well-known mean-variance certainty equivalent

$$\varphi(v'\mu, v'\Sigma\, v) = v'\mu - \frac{\alpha}{2} \cdot v'\Sigma\, v \tag{2}$$

with $\alpha \geq 0$. Note that $v'\mu$ represents the expectation and $v'\Sigma\, v$ is the variance of the predictive portfolio return of a buy-and-hold portfolio after the given investment period. The principal goal of this work is to show how the predictive moments μ and Σ (which incorporate both market and estimation risk) can be calculated if short-term asset log-returns are not normally distributed, possibly serially dependent, or exhibit other kinds of stylized facts (see below).

2.2 Gordin's Central Limit Theorem

Now let $(X_t \,|\, \theta)$ ($t \in \mathbb{Z}$) be a strongly stationary process representing the short-term *log-returns* of some asset with $\mathbb{E}(X_t \,|\, \theta) = \eta(\theta)$. Note that here we consider a stochastic process under some unknown parameter $\theta \in \Theta \subset \mathbb{R}^p$. We assume also that $(X_t \,|\, \theta)$ is ergodic. Ergodicity means that any existing and finite moment of $X_t \,|\, \theta$ can be consistently estimated by using the corresponding sample moment of the time series $X_1, \ldots, X_n$ ($n \to \infty$). This is guaranteed if $(X_t, \ldots, X_{t+k} \,|\, \theta)$ is asymptotically independent of $(X_{t-n}, \ldots, X_{t-n+l} \,|\, \theta)$ as $n \to \infty$ for all $k, l \in \mathbb{N}$ (Hayashi, 2000, p. 101). Further, we suppose that the second moments of $X_t \,|\, \theta$ exist and are finite.

However, for the central limit theorem (CLT) we need some additional assumption. More precisely, the CLT holds for the sample mean of $(X_t \,|\, \theta)$ if the centered process $(X_t - \eta(\theta) \,|\, \theta)$ satisfies Gordin's condition. Let $\mathcal{H}_t := (X_t, X_{t-1}, \ldots \,|\, \theta)$ be the history of $(X_t \,|\, \theta)$ at time $t \in \mathbb{Z}$. Roughly speaking, Gordin's condition implies that the impact of $\mathcal{H}_{t-n}$ on the conditional expectation of $X_t \,|\, \theta$ vanishes as $n \to \infty$ and also that the conditional expectations of $X_t \,|\, \theta$ do not vary too much in time (Hayashi, 2000, p. 403). In that case it is guaranteed that the CLT holds with an asymptotic or, say, *long-run variance*

$$\sigma_{\mathrm{L}}^2(\theta) := \sum_{k=-\infty}^{\infty} \gamma_\theta(k)\,,$$

where γ_θ is the autocovariance function of $(X_t \,|\, \theta)$ (Hayashi, 2000, p. 401) given the unknown parameter θ . This result can be easily extended to any d-dimensional stochastic process (Hayashi, 2000, p. 405). Hence, in the following let $(X_t \,|\, \theta)$ be an ergodic stationary d-dimensional process satisfying Gordin's condition.

¿From Gordin's CLT it follows that long-term asset log-returns typically tend to be normally distributed even if the short-term log-returns are serially dependent and heavy tailed. A broad class of time series models satisfy Gordin's condition. Hence, long-term asset log-return vectors are approximately normally distributed, i.e.

$$\log(\mathbf{1} + R) \,|\, \theta = \sum_{t=1}^{T} X_t \,|\, \theta =: X \,|\, \theta \sim \mathcal{N}_d\big\{T\eta(\theta), T\Upsilon_{\mathrm{L}}(\theta)\big\}\,, \tag{3}$$

where $\mathbf{1}$ represents a column vector of ones and $\log(\cdot)$ is understood as taking the logarithm of each component separately. Here $\Upsilon_{\mathrm{L}}(\theta)$ denotes the long-run covariance matrix of the stochastic process (Hayashi, 2000, p. 404) and $T \in \mathbb{N}$ represents the number of aggregated short-term log-returns or, say, the investment horizon. For example, if $X_1, \ldots, X_T$ represent daily log-returns, the sum given by Eq. 3 denotes a quarterly log-return if $T = 63$ and a yearly log-return in case $T = 252$.

Of course, the Gaussian distribution hypothesis holds only approximately. However, in the following the additional suffix 'approximately' or any corresponding symbol are suppressed for convenience. It is worth to mention that we generally suppose that both $\eta(\theta)$ and $\Upsilon_{\mathrm{L}}(\theta)$ can be computed either numerically or analytically under the specific time series model which is used for the short-term asset log-returns provided the model parameter θ is known. Specifically, if $(X_t - \eta(\theta) \,|\, \theta)$ is a martingale difference sequence (Hayashi, 2000, p. 104), that means if

$$\mathbb{E}(X_t \,|\, \mathcal{H}_{t-1}, \theta) = \eta(\theta)\,, \qquad \forall\, t \in \mathbb{Z}\,,$$

the components of $(X_t \,|\, \theta)$ are serially uncorrelated. In that case the long-run covariance matrix $\Upsilon_{\mathrm{L}}(\theta)$ turns out to be the *stationary* variance $\Upsilon(\theta)$ of $(X_t \,|\, \theta)$. The martingale difference property is satisfied for a broad class of time series models, such as the family of multivariate GARCH processes (Bauwens et al., 2006).

As elucidated in the introduction, estimating the moments $T\eta(\theta)$ and $T\Upsilon_{\mathrm{L}}(\theta)$ from long-term asset returns is inefficient. For example, we could estimate the quantity $T\Upsilon_{\mathrm{L}}(\theta)$ simply by applying the sample covariance matrix to the corresponding long-term asset log-returns. However in that case we would ignore a large part of the data and the resulting standard error would increase roughly by a factor of

$\sqrt{T}$ relative to the approach based on high-frequency data. Hence, decreasing the sampling frequency leads to a loss of statistical efficiency.

2.3 Bayesian Framework

In the Bayesian framework the model parameter θ is not assumed to be fixed but it is considered as a random quantity possessing some prior distribution $p(\theta)$. The posterior distribution $p(\theta \,|\, x)$ corresponds to the distribution of θ given some observed data x. More specifically, in the following we shall interpret x as historical short-term asset log-return data. The likelihood function $\mathcal{L}(\theta\,;x) = p(x \,|\, \theta)$ represents some pre-defined probabilistic model for x. Now the posterior distribution of θ can be obtained by the Bayes formula

$$p(\theta \,|\, x) = \mathcal{L}(\theta\,;x)\, p(\theta)/p(x)\,,$$

so that the posterior involves both empirical and subjective information.

However, in Bayesian analysis the posterior distribution is not always the desired object. Instead, one can be interested in the predictive distribution of the data. Let y be some unobserved data where x and y are conditionally independent given θ. Then

$$p(y \,|\, x) = \int p(y \,|\, \theta)\, p(\theta \,|\, x)\, d\theta$$

represents the predictive distribution of y. In the following discussion this can be interpreted as the distribution of a long-term asset log-return if we take the parameter uncertainty additionally into account. Each parameter is weighted by its posterior probability, i.e. the probability of θ given the historical observations and some expert knowledge. Notice that analytical solutions for the portfolio optimization problem which are based on the predictive distribution are only available for relatively simple expressions for the prior $p(\theta)$ and the likelihood $\mathcal{L}(\theta\,;x)$.

The prior $p(\theta)$ can be either 'diffuse' or 'informative'. If the prior is *diffuse* the model parameter is assumed to possess some ad-hoc distribution such as the uniform distribution or the standard normal distribution. The prior is called *informative* if some subjective information is necessary to determine $p(\theta)$. The chosen terminology is somewhat misleading since we do not mean that diffuse priors in general are non-informative in the probabilistic sense since the posterior distribution might drastically depend on the chosen diffuse prior. Hence, we believe that Bayesian analysis is inherently subjective and since most practitioners have some basic opinions about the evolution of asset prices they might want to include that information in the optimization process (Black and Litterman, 1992). The present work heavily relies on the idea of using subjective information whenever it is possible.

One popular example of Bayesian portfolio optimization is the approach of Black and Litterman (1992). They show how to distill implicit information about the distribution of asset returns from the market by using standard results of portfolio theory. This is combined with the investor's own belief which typically leads to optimal portfolios being more robust against estimation errors than solutions obtained by pure statistical methods. However, in order to be analytically tractable, the Black-Litterman approach assumes that asset returns are normally distributed. Other Bayesian portfolio optimization techniques are given by the work of Frost and Savarino (1986) and Jorion (1986). They all share the same disadvantage, namely that an analytic expression of the predictive distribution or optimal portfolio is only available by imposing unrealistic assumptions on the underlying data or otherwise being inefficient, since they have to be applied by using low-frequency data.

Scherer and Martin (2007, Ch. 7) suggest to apply so-called *conjugate* priors in Bayesian portfolio optimization. These are informative priors which, after multiplying with the likelihood, lead to a posterior distribution that is of the same type as the chosen likelihood function. Again, this limitation can be motivated by the requirement to obtain analytically tractable expressions for the posterior distribution. However, unrealistic assumptions about the distribution of empirical data are necessary in general and the set of possible prior distributions is substantially restricted. In particular, conjugate priors often are not available if the assumption of normally distributed asset returns is relaxed. Scherer and Martin (2007, Ch. 7) refer to a Markov chain Monte Carlo method (which will be discussed later on in Chapter 3) to simulate the posterior distribution of the mean and variance of a single asset return. In this work we will show how this idea can be extended to incorporate arbitrary prior information given the asset returns are not normally distributed.

For choosing some likelihood function for θ we have to consider an appropriate model for the data, that means to take account for the stylized facts of empirical finance. These can be subsumed by the following anomalies (see McNeil et al., 2005, p. 117):

(1) Short-term asset returns are heavy-tailed and particularly not Gaussian.

(2) Asset returns are not independent and identically distributed although they show little serial correlation.

(3) In contrast, squared asset returns show strong serial correlation.

(4) Asset volatility varies over time and appears in clusters.

There are several alternatives to deal with these phenomena. For instance, GARCH processes (Bollerslev, 1986, Engle, 1982) can be used to model volatility clusters. Another possibility is to work with stochastic volatility models (Barndorff-Nielsen et al., 2002, Jacquier et al., 1994, 2004).

2.4 Predictive Moments

In the last section we mentioned that the parameter θ is considered as a random quantity and from Section 2.2 we know that

$$X \mid \theta \sim \mathcal{N}_d\big\{T\eta(\theta), T\Upsilon_{\mathrm{L}}(\theta)\big\} ,$$

where $X \mid \theta$ denotes a long-term log-return vector given the unknown parameter θ. Hence, the vector of long-term discrete returns is given by

$$R \mid \theta = \exp(X \mid \theta) - \mathbf{1} ,$$

where $\exp(\cdot)$ shall be interpreted as a component-wise function. Thus each component of $R \mid \theta$ is log-normally distributed and it can be easily shown that

$$\mathbb{E}(R \mid \theta) = \exp\big[T\big\{\eta(\theta) + \mathrm{diag}(\Upsilon_{\mathrm{L}}(\theta))/2\big\}\big] - \mathbf{1}$$

and

$$\mathbb{V}\mathrm{ar}(R \mid \theta) = \exp\Big[T\big\{\eta(\theta)\mathbf{1}' + \mathbf{1}\eta(\theta)' + D(\theta)\big\}\Big] \odot \Big[\exp\big\{T\Upsilon_{\mathrm{L}}(\theta)\big\} - \mathbf{1}\mathbf{1}'\Big] ,$$

where $\odot$ denotes the Hadamard (i.e. component-wise) product, and

$$D(\theta) = \frac{\mathrm{diag}\{\Upsilon_{\mathrm{L}}(\theta)\}\mathbf{1}' + \mathbf{1}\mathrm{diag}\{\Upsilon_{\mathrm{L}}(\theta)\}'}{2} .$$

Finally, we obtain the predictive moments of the long-term log-return vector by the law of total expectations and the variance decomposition theorem, viz

$$\mu = \mathbb{E}(R) = \mathbb{E}\big\{\mathbb{E}(R \mid \theta)\big\}$$

and

$$\Sigma = \mathbb{E}\big\{\mathbb{V}\mathrm{ar}(R \mid \theta)\big\} + \mathbb{V}\mathrm{ar}\big\{\mathbb{E}(R \mid \theta)\big\} .$$

Interestingly, the conditional means of the discrete returns are also determined by the long-run variances. Moreover, predictive expectations and variances of discrete returns are *nonlinear* functions of the investment horizon T. Hence, the investment horizon can have a substantial impact on the optimal portfolio. In Section 3 we will see how the predictive moments can be approximated by Monte Carlo simulation.

3 Numerical Implementation

Now we will discuss several Markov chain Monte Carlo algorithms for simulating the posterior distribution $p(\theta \mid x)$ even if this has a rather complicated analytical structure. There is a big number of different simulation techniques like for instance *importance sampling* (Gamerman and Lopes, 2006, Ch. 3.4). However, we got the best simulation results in reasonable time using a Markov chain Monte Carlo algorithm, which will be presented in the following sections. In our case we want to use Markov chains only to sample from a complex posterior distribution. Hence, we have to guarantee that the stationary distribution of the considered Markov chain corresponds to $p(\theta \mid x)$.

3.1 Gibbs Sampling

A simple approach is known as *Gibbs sampling*. That means for simulating θ we could principally start with some initial parameter vector $\theta = (\theta_1, \ldots, \theta_p)$ and draw a new realization θ'_1 of the first component from the conditional distribution of θ_1 given $\theta_2, \ldots, \theta_p$. Then we can take the new parameter vector $(\theta'_1, \theta_2, \ldots, \theta_p)$ into consideration and simulate the second component of θ by drawing from the distribution of θ_2 under the new condition $\theta'_1, \theta_3, \ldots, \theta_p$, etc., until we obtain the parameter vector $\theta' = (\theta'_1, \ldots, \theta'_p)$. If the same procedure is repeated with θ' and so on we obtain a Markov chain whose stationary distribution corresponds to the posterior distribution of θ. Scherer and Martin (2007, Ch. 7) give an example of how to use Gibbs sampling for simulating the posterior distribution of the mean and variance of a normally distributed single asset return by using a conjugate prior. However, in our case this is not a useful approach since drawing from the conditional posterior distributions of θ is not substantially easier than drawing directly from $p(\theta \mid x)$.

3.2 Metropolis-Hastings Algorithm

Another MCMC scheme which is frequently used in Bayesian statistics is the Metropolis-Hastings algorithm (Hastings, 1970, Metropolis et al., 1953). An application to the Bayesian analysis of stochastic volatility models is presented by Jacquier et al. (2004). The Metropolis-Hastings algorithm is very similar to the Gibbs sampler, but unlike that, it does not require to sample from the conditional stationary distribution. In contrast, the sampling part is completely reduced to sampling from an arbitrary *proposal distribution* which is easy to draw from. The stationary distribution is then only needed to calculate the *acceptance probability* of each new state in the chain, which comes from the proposal distribution. This is why we choose a Metropolis-Hastings-like algorithm to simulate the distribution of $\theta|x$. First, we will present the Metropolis-Hastings algorithm and after that an extension called *parallel tempering* will be discussed.

Assume there exists some *target distribution* $\pi(\theta)$ which shall be simulated. The current state of the chain will be denoted by ϕ . In case of the Metropolis-Hastings algorithm, the simulation is done by introducing an ‘easy to draw from’ *proposal distribution* $q(\phi, \phi')$ which denotes the distribution of a proposal to move from state ϕ to state ϕ'. However, the actual probability to move from ϕ to ϕ' is determined by the acceptance probability

$$\alpha(\phi, \phi') = \min \left\{ 1, \frac{\pi(\phi')\, q(\phi', \phi)}{\pi(\phi)\, q(\phi, \phi')} \right\} . \tag{4}$$

Note that if we have a symmetric proposal distribution, the acceptance probability is simply given by $\alpha = \min\{1, \pi(\phi')/\pi(\phi)\}$.

The probability density of a new state ϕ' given an old state ϕ, that is the so-called *transition kernel* $K(\phi, \phi')$ (Gamerman and Lopes, 2006, p. 194) of the Markov chain, is given by

$$K(\phi, \phi') = q(\phi, \phi')\, \alpha(\phi, \phi') + \delta(\phi' - \phi)\left(1 - \int q(\phi, \xi)\, \alpha(\phi, \xi)\, d\xi\right),$$

where δ is the Dirac distribution. It can be shown that for the acceptance probability given by Eq. 4, the *detailed balance condition*

$$\pi(\phi) K(\phi, \phi') = \pi(\phi') K(\phi', \phi)$$

is satisfied for all ϕ and ϕ'. Thus we obtain a reversible Markov chain (Gamerman and Lopes, 2006, Ch. 4.6). That means by the presented Metropolis-Hastings algorithm in fact we are able to simulate realizations from the target distribution π.

3.3 Parallel Tempering

Though the Metropolis-Hastings algorithm is very powerful, one big problem can easily occur: The Markov chain can get stuck in local optima for a very long time. Assume for instance a univariate bi-modal distribution. If the chain is currently in a region around one of the two modes, there is almost no incentive to move to the region around the other mode, since the acceptance probability $\alpha(\phi, \phi')$ approaches zero if $\pi(\phi')$ is much smaller than $\pi(\phi)$. To avoid this problem, the idea of *heated* equilibrium distributions has been introduced. Instead of simulating only one stationary distribution $\pi(\theta)$ at a time, m parallel chains are used, each having an equilibrium distribution

$$\pi_i(\theta) \propto \pi_1(\theta)^{(1/T_i)}, \qquad \forall\, i = 1, \ldots, m\,,$$

where T_i is the *temperature* of the distribution $\pi_i(\theta)$. The temperature of the desired stationary distribution $\pi_1(\theta)$ is $T_1 = 1$. At each iteration of the algorithm, an exchange between the states ϕ_i and ϕ_j of chain i and j is proposed. The acceptance probability of this swap is

$$\alpha_{ij}(\phi_i, \phi_j) = \min\left\{1, \frac{\pi_i(\phi_j)\, \pi_j(\phi_i)}{\pi_i(\phi_i)\, \pi_j(\phi_j)}\right\}.$$

One disadvantage of this method is that only the outcome of chain 1 contains samples from the desired distribution and all the other samples are dropped. However, especially for very complex distributions the advantage of not getting stuck in local modes overcomes the disadvantage of high computational effort. For further details and applications of tempering algorithms see for instance Gamerman and Lopes (2006, Ch. 6 and Ch. 7).

In our case the stationary distribution which has to be simulated is the posterior distribution of the model parameters which can become very complex. In our empirical study we will use $m = 2$ different chains. For the proposal distribution we choose a composite distribution $q(\theta, \theta')$ by taking account of the specific domains of the different components of θ. Of course we could also choose a proposal distribution which probably leads to realizations outside of Θ but, however, if some parameter is proposed to exceed the parameter set, the prior probability and thus also the acceptance probability becomes zero. Hence, it cannot happen that we get some realizations of θ such that $\theta \notin \Theta$.

Our implementation of the parallel tempering algorithm is as follows:

1. Create the initial parameter vectors θ_1 and θ_2.
2. Repeat the following steps very often:
   (a) Generate θ_1' and θ_2' by randomly drawing from the proposal distributions.
   (b) Calculate $p(\theta_1' \,|\, x) \propto \mathcal{L}(\theta_1'\,; x)\, p(\theta_1')$ and $p(\theta_2' \,|\, x) \propto \mathcal{L}(\theta_2'\,; x)\, p(\theta_2')$.
   (c) Calculate
   $$\alpha_1 = \min\left\{1, \frac{p(\theta_1' \,|\, x)\, q(\theta_1', \theta_1)}{p(\theta_1 \,|\, x)\, q(\theta_1, \theta_1')}\right\}$$
   and
   $$\alpha_2 = \min\left\{1, \frac{p(\theta_2' \,|\, x)^{(1/T_2)} q(\theta_2', \theta_2)}{p(\theta_2 \,|\, x)^{(1/T_2)} q(\theta_2, \theta_2')}\right\}.$$
   (d) Set $\theta_1 = \theta_1'$ with probability α_1, and $\theta_2 = \theta_2'$ with probability α_2, otherwise keep the old θ_1 or θ_2, respectively.
   (e) Swap the states θ_1 and θ_2 of the chains with probability
   $$\alpha_{12}(\theta_1, \theta_2) = \min\left\{1, \frac{p(\theta_2 \,|\, x)\, p(\theta_1 \,|\, x)^{(1/T_2)}}{p(\theta_1 \,|\, x)\, p(\theta_2 \,|\, x)^{(1/T_2)}}\right\}.$$

As mentioned above we only consider the realizations of the first chain which are obtained after some burning-in phase.

4 Empirical Study

In this section we will present an empirical study based on the framework developed in the previous sections. First, we create a model for high-frequency asset log-returns by taking account of stylized facts. It is a multivariate extension of the GARCH model developed by Bollerslev (1986). A comprehensive overview on different multivariate GARCH (MGARCH) models is given in Bauwens et al.

(2006). MGARCH processes are martingale difference sequences and so Gordin's condition (see Section 2.2) is automatically satisfied. Further, the predictive moments (see Section 2.4) can be easily calculated by the MCMC algorithm discussed in Section 3. After the data generating process is developed, we present the chosen prior information for the unknown model parameter θ. Then we will apply our method to time series data to find optimal portfolios.

4.1 Modeling the Distribution of Asset Log-Returns

In this section we will describe a way for modeling the distribution of daily asset log-returns. We will concentrate on risky assets. The risk-free asset or, say, money market account does not possess any market risk per definition. That means we do not need any stochastic model and so there exists no parameter uncertainty.

In order be provide a flexible framework for the asset returns, we rely on the broad class of elliptically symmetric distributions. A d-dimensional random vector X is said to be *elliptically symmetric distributed* (Cambanis et al., 1981) if and only if

$$X \stackrel{\mathrm{d}}{=} \eta + \Gamma \mathcal{R} U$$

with $\eta \in \mathbb{R}^d$ being a location vector, $\Gamma \in \mathbb{R}^{d \times k}$ is a transformation matrix, U a k-dimensional random vector uniformly distributed on the unit hypersphere, and $\mathcal{R}$ is a non-negative random variable stochastically independent of U. The positive semi-definite matrix $\Omega := \Gamma\Gamma'$ is referred to as the *dispersion matrix* of X and $\mathcal{R}$ is called its *generating variate*. By choosing $\mathcal{R}$ properly, we are able to account for stylized facts like heavy tails. Further, it can be shown that

$$V := \mathbb{V}\mathrm{ar}(X) = \mathbb{E}(\mathcal{R}^2)/k \cdot \Omega$$

is the covariance matrix of X provided $\mathbb{E}(\mathcal{R}^2) < \infty$.

A d-dimensional MGARCH process (X_t) is characterized by

$$X_t \,|\, \mathcal{H}_{t-1} \stackrel{\mathrm{d}}{=} \eta + V_t^{\frac{1}{2}} \epsilon_t\,,$$

where η is a $d \times 1$ vector of time-independent expected log-returns, V_t is a function only of $\mathcal{H}_{t-1}$ and denotes the $d \times d$ positive definite conditional covariance matrix of the log-return vector X_t, and ϵ_t is an independent and identically distributed $d \times 1$ vector of perturbations with $\mathbb{E}(\epsilon_t) = 0$ and covariance matrix $\mathbb{V}\mathrm{ar}(\epsilon_t) = I_d$. If ϵ_t is assumed to be spherically distributed, i.e. elliptically symmetric with location 0 and dispersion proportional to I_d, then the MGARCH model perfectly fits into the class of elliptically symmetric distributions.

There are various specifications of the time-dependent covariance matrix V_t. For a thorough discussion of MGARCH processes see Bauwens et al. (2006). Since MGARCH specifications often require a huge number of parameters and are hardly applicable to practical problems, for complexity reduction we suggest to use a principal components model for the asset log-returns. The underlying idea of principal components is that most of the dynamics of the observed data can be explained by a small number of uncorrelated factors. The spectral decomposition theorem assures that the covariance matrix V of an elliptically symmetric distributed random vector X can be decomposed into $V = \mathcal{O}\Lambda\mathcal{O}'$, where

- Λ is the diagonal matrix of the eigenvalues $\lambda_1, \ldots, \lambda_d$ of V and
- $\mathcal{O}$ is an orthogonal $d \times d$ matrix containing the associated eigenvectors.

By applying this decomposition for the vector of asset log-returns we can specify the MGARCH model as

$$X_t \,|\, \mathcal{H}_{t-1} \stackrel{\mathrm{d}}{=} \eta + \mathcal{O}\Lambda_t^{\frac{1}{2}}\epsilon_t$$

and define

$$Y_t := \Lambda_t^{\frac{1}{2}}\epsilon_t = \mathcal{O}'(X_t - \eta)\,.$$

This reduces the number of required model parameters tremendously, since the elements of Y_t are uncorrelated per definition. However, we have to presume that the eigenvectors do not change over time. Speaking economically, the factors which drive the dynamics of the asset log-returns do not change but the *impact* of each factor can vary over time. For modeling the components of Λ_t we can simply assume that Y_t consists of d unrelated univariate GARCH(1,1) processes. The resulting process is sometimes called *orthogonal* GARCH (Bauwens et al., 2006).

Principally, we can choose any elliptically symmetric distribution for modeling the perturbation ϵ_t as long as the corresponding density function can be computed either numerically or analytically. However, here we assume that ϵ_t is multivariate t-distributed, i.e.

$$\epsilon_t \sim t_d\!\left(0\,, \frac{\nu - 2}{\nu} \cdot I_d\,, \nu\right)$$

with $\nu > 2$ degrees of freedom and the dispersion matrix is such that $\mathbb{V}\mathrm{ar}(\epsilon_t) = I_d$. Hence, the random vector $X_t \,|\, \mathcal{H}_{t-1}$ possesses the density

$$p(x_t \,|\, \mathcal{H}_{t-1}) = \frac{\Gamma(\frac{d+\nu}{2})}{\Gamma(\frac{\nu}{2})} \cdot \sqrt{\frac{\det \Lambda_t^{-1}}{(\nu\pi)^d}} \cdot \left(1 + \frac{(x_t - \eta)'\mathcal{O}\Lambda_t^{-1}\mathcal{O}'(x_t - \eta)}{\nu - 2}\right)^{-\frac{d+\nu}{2}},$$

where Λ_t is a diagonal $d \times d$ matrix with main diagonal elements

$$\lambda_{it} = \gamma_i + \alpha_i Y_{i,t-1}^2 + \beta_i \lambda_{i,t-1}\,, \qquad i = 1, \ldots, d\,, \tag{5}$$

representing the conditional variances of the d principal components. Note that the orthogonal matrix $\mathcal{O}$ $(d \times d)$ contains $\binom{d}{2}$ free parameters and there are $3d$ GARCH parameters. Altogether, the resulting data generating process contains only $d\,(d + 7)/2 + 1$ parameters.

4.2 Modeling the Prior Information

There are several ways to implement prior information. In case of a diffuse prior there is no explicit information that is incorporated into the prior distribution. This is often done to get an analytical expression for the posterior distribution and so to obtain an analytical result for the optimal portfolio. However, it can be shown that the diffuse prior approach can lead to paradox results (Berger, 2006) and the concrete choice of the diffuse prior can have a substantial impact on the optimal decision. Therefore, as already mentioned, it is suggested to use informative priors whenever it is possible.

Our hierarchical approach is very general. First of all note that our model parameters are given by $\eta, \alpha, \beta, \lambda, \mathcal{O}, \nu$. Here η $(d \times 1)$ is the vector of expected asset log-returns, α $(d \times 1)$ and β $(d \times 1)$ contain the GARCH(1,1) parameters according to (5) and the $d \times 1$ vector λ contains the *unconditional* variances $\lambda_1, \ldots, \lambda_d$, i.e.

$$\lambda_i = \frac{\gamma_i}{1 - \alpha_i - \beta_i}, \qquad i = 1, \ldots, d.$$

Note that the parameters $\gamma_i = \lambda_i (1 - \alpha_i - \beta_i)$ $(i = 1, \ldots, d)$ follow implicitly from α, β, and λ. That means we use the following re-parameterization of Eq. 5:

$$\lambda_{it} = \lambda_i (1 - \alpha_i - \beta_i) + \alpha_i Y_{i,t-1}^2 + \beta_i \lambda_{i,t-1}, \qquad i = 1, \ldots, d.$$

We will substitute $\mathcal{O}$ by an estimate based on the sample covariance matrix of the time series data. That means $\mathcal{O}$ is fixed for the sake of simplicity. Finally, the number of degrees of freedom ν is set to 3 to account for the typical heavy tails of daily log-returns. We did not observe any improvements by introducing some prior distribution for ν. Hence, we obtain the parameter vector $\theta = (\eta, \alpha, \beta, \lambda)$ and suppose that they are a priori stochastically independent, i.e.

$$p(\theta) = p(\eta)\, p(\alpha)\, p(\beta)\, p(\lambda).$$

Since $\alpha, \beta \in (0, 1)$ we decided to use flat priors for α and β where the components of α and β are assumed to be mutually independent. So the prior for θ can be simply expressed as $p(\theta) = p(\eta)\, p(\lambda)$.

Also the components of λ are assumed to be mutually independent but each one follows a gamma distribution, i.e. $\lambda_i \sim \Gamma(\kappa_2, \lambda_0/\kappa_2)$ $(i = 1, \ldots, d)$ and $\lambda_0, \kappa_2 > 0$. Hence, we expect a priori that each principal component has the same proportion of total variation. Note that $\mathbb{E}(\lambda_i) = \lambda_0$ is constant but $\mathbb{V}\mathrm{ar}(\lambda_i) = \lambda_0^2/\kappa_2$. That means κ_2 can be interpreted as the investor's confidence that the unconditional variances of the principal components indeed correspond to λ_0. In our empirical study we choose $\lambda_0 = 0.2^2/T$ and $\kappa_2 = 2$.

For the expected values of the daily log-returns we use the prior proposed by Jorion (1986), i.e.

$$\eta \,|\, V \sim \mathcal{N}_d(\eta_0, V/\kappa_1),$$

	USA	UK	JPN	ITA	GER	FRA	CAN
$\hat{\mu}$	5.98%	12.50%	12.78%	17.63%	14.27%	14.53%	20.97%
$\hat{\sigma}$	16.07%	13.70%	21.55%	14.68%	23.44%	17.90%	22.10%

Table 1: Descriptive statistics of yearly discrete returns.

where η_0 is a vector of prior expected returns. We decided to choose $\eta_0 = 0$ since sample means of daily log-returns are typically close to zero (McNeil et al., 2005, p. 117). The scale parameter κ_1 represents the confidence of the investor in their a priori assumption concerning η and can be seen as a virtual sample size. For instance, if there are $n = 1260$ observations (i.e. 5 trading years) then $\kappa_1 = 1260$ would mean that the investor trusts in their own belief about η as much as the empirical evidence given by the time series.

Note that $V = \mathcal{O}\Lambda\mathcal{O}'$ where $\mathcal{O}$ is fixed and Λ is random. Hence, we can write Jorion's prior equivalently as

$$\eta \,|\, \Lambda \sim \mathcal{N}_d(0\,, \mathcal{O}\Lambda\mathcal{O}'/\kappa_1)$$

such that $p(\eta) = p(\eta \,|\, \Lambda)\, p(\Lambda)$ can be easily calculated, since

$$p(\Lambda) = p(\lambda) \propto \prod_{i=1}^{d} \lambda_i^{\kappa_2 - 1} \, \exp\left(-\frac{\kappa_2 \lambda_i}{\lambda_0}\right)$$

and

$$p(\eta \,|\, \Lambda) \propto \exp\left(-\frac{\kappa_1}{2} \cdot \eta'\mathcal{O}\Lambda^{-1}\mathcal{O}'\eta\right).$$

4.3 Data Description

In our empirical study we use daily log-returns of seven MSCI stock indices of the countries USA, UK, Japan, Italy, Germany, France, and Canada. The indices are adjusted by dividends, splits, etc. and are calculated on the basis of USD stock prices. We have $n = 1260$ daily observations ranging from 2001-12-03 to 2006-09-29 and the whole sample is divided chronologically into 5 subsets where each subset contains 252 observations. In Table 1 we can see the sample means and standard deviations of the yearly discrete returns of each country. In our study we assume that the investment horizon corresponds to 1 year, i.e. $T = 252$ and the quantities given in Table 1 are based on the available 5 observations of yearly discrete returns. Of course, since the sample size is very small, these values are strongly affected by estimation errors.

The process $(X_t \,|\, \theta)$ of daily log-returns is assumed to be an ergodic stationary martingale difference sequence as described in Section 2.2. Hence, both the sample mean $\hat{\eta}$ and the sample covariance matrix $\widehat{\Upsilon}$ of the daily log-returns are strongly

	USA	UK	JPN	ITA	GER	FRA	CAN
$\hat{\mu}$	6.29%	13.41%	13.46%	18.54%	15.23%	15.73%	20.71%
$\hat{\sigma}$	17.42%	19.60%	24.13%	20.42%	28.14%	24.52%	19.29%

Table 2: Descriptive statistics of yearly discrete returns based on daily log-returns.

$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\lambda}_5$	$\hat{\lambda}_6$	$\hat{\lambda}_7$
6.45 e-4	1.64 e-4	0.94 e-4	0.48 e-4	0.32 e-4	0.21 e-4	0.14 e-4
63.35 %	16.09 %	9.21 %	4.75 %	3.17 %	2.09 %	1.33 %

Table 3: Eigenvalues of the sample covariance matrix of daily log-returns.

consistent estimators for $\eta(\theta)$ and $\Upsilon_{\mathrm{L}}(\theta)$, respectively. Now we can also estimate the first and second moments of yearly discrete returns by using the formulas given in Section 2.4 based on daily log-returns, viz

$$\widehat{\mathbb{E}}(R \mid \theta) = \exp\Big[252\,\big\{\hat{\eta} + \operatorname{diag}\big(\widehat{\Upsilon}\big)/2\big\}\Big] - \mathbf{1}$$

and

$$\widehat{\mathbb{Var}}(R \mid \theta) = \exp\Big[252\,\big\{\hat{\eta}\mathbf{1}' + \mathbf{1}\hat{\eta}' + \widehat{D}\big\}\Big] \odot \Big[\exp\big\{252\,\widehat{\Upsilon}\big\} - \mathbf{1}\mathbf{1}'\Big]\,,$$

where

$$\widehat{D} = \frac{\operatorname{diag}\big\{\widehat{\Upsilon}\big\}\mathbf{1}' + \mathbf{1}\operatorname{diag}\big\{\widehat{\Upsilon}\big\}'}{2}\,.$$

The corresponding values are given in Table 2. Note that there are only slight differences between the results in Table 1 and Table 2 regarding the means but for the standard deviations the results can differ substantially.

Table 3 contains the eigenvalues of $\widehat{\Upsilon}$ as well as their proportions of the total variation. As described earlier, each eigenvalue can be interpreted as the unconditional variance of a principal component. In our case, the first component (i.e. the systematic risk of the market) almost explains two third of the total variation and the impact of the other components are relatively small. Similar results for financial data have been frequently observed in literature (see, e.g., Plerou et al., 1999). Note that our prior expectation for λ_i corresponds to $\lambda_0 = 0.2^2/252$ =1.59 e-4, which reflects a relatively conservative assumption relative to the empirical results. For the confidence in λ_0 we choose the parameter $\kappa_2 = 2$ which leads to an a priori standard deviation of λ_i roughly corresponding to 1.12 e-4 ($i = 1, \ldots, d$).

4.4 Results

In this section we present the results of our simulation. Our main objective is to demonstrate the practical applicability of our approach. We want to show how prior information can be used to account for estimation risk – even if the underlying model is complex – and to obtain well-diversified portfolios. The parameter κ_1, which reflects the investor's confidence in their prior assumption about the expected log-returns, is varied in order to see how expert knowledge determines the optimal portfolio. Asset return variances and covariances can be estimated quite good by using short-term asset returns. In contrast, it is well-known that portfolio selection is very sensitive to expected asset returns which cannot be estimated accurately. Hence, investors preferably have a strong confidence about expected asset returns in order to reduce estimation risk. This is the reason why we kept $\kappa_2 = 2$ fixed, which indicates that there is only little confidence in the prior information about the eigenvalues.

We performed standard Markowitz portfolio selection (Markowitz, 1952). Our objective function is the traditional mean-variance certainty equivalent given by Eq. 2 where we choose a risk aversion of $\alpha = 1$. In many practical situations constraints are included in the optimization problem. For instance, investors might be willing to forbid short-selling. Other constraints might be given by legal issues and so on. We do not want to provide optimal portfolios for each imaginable investor, but instead we present a flexible framework which can be adapted to most kinds of situations.

Each additional constraint limits the space of alternatives. Therefore, in the first part of the study (P1) we have only one constraint, namely the budget constraint $\mathcal{C}_{\mathrm{B}} : w'\mathbf{1} = 1$. The short-selling constraint $\mathcal{C}_{\mathrm{S}} : w \geq 0$ is additionally considered in the second part of the study (P2). In our study we are searching for the optimal portfolio given by (1) using the objective function

$$\varphi(v) = v'\mu - \frac{1}{2} \cdot v'\Sigma\, v\,, \qquad \text{s.t. } v \in \mathcal{C}\,,$$

where $\mathcal{C} = \mathcal{C}_1 = \mathcal{C}_{\mathrm{B}}$ in P1 and $\mathcal{C} = \mathcal{C}_2 = \mathcal{C}_{\mathrm{B}} \cap \mathcal{C}_{\mathrm{S}}$ in P2.

Table 4 contains our results of the portfolio optimization. These can be compared with the portfolio weights obtained by traditional Markowitz optimization, i.e. searching for the *Markowitz portfolio* (MP), viz

$$\mathrm{MP} = \arg\max_v v'\widehat{\mathbb{E}}(R \,|\, \theta) - \frac{1}{2} \cdot v'\widehat{\mathbb{V}\mathrm{ar}}(R \,|\, \theta)v\,, \qquad \text{s.t. } v \in \mathcal{C}\,,$$

and the so-called *global minimum variance portfolio* (MVP), i.e.

$$\mathrm{MVP} = \arg\min_v v'\widehat{\mathbb{V}\mathrm{ar}}(R \,|\, \theta)v\,, \qquad \text{s.t. } v \in \mathcal{C}\,.$$

The MVP has been advocated by many authors as an alternative to the traditional mean-variance optimal portfolio since there are no expected asset returns which have to be estimated and thus the impact of estimation errors can be substantially reduced (Frahm, 2007).

As we can see in Table 4 the Markowitz portfolios tend to overrate assets with large expected returns. In P1 the MP suggests a short-selling of 484.01% of USA

empirical	USA	UK	JPN	ITA	GER	FRA	CAN
$\hat{\mu}$	6.29%	13.41%	13.46%	18.54%	15.23%	15.73%	20.71%
$\hat{\sigma}$	17.42%	19.60%	24.13%	20.42%	28.14%	24.52%	19.29%
MP_1	−484.01%	−195.93%	−13.18%	373.62%	−2.45%	−65.10%	487.05%
MP_2	0%	0%	0%	0%	0%	0%	100%
MVP_1	50.37%	37.72%	20.13%	43.27%	−28.86%	−28.46%	5.84%
MVP_2	42.49%	19.17%	23.88%	4.64%	0%	0%	9.83%
$\kappa_1 = 1$	USA	UK	JPN	ITA	GER	FRA	CAN
μ	5.25%	12.44%	19.57%	16.29%	13.32%	14.52%	25.29%
σ	17.96%	20.66%	27.29%	21.39%	29.02%	25.80%	21.53%
w_1	−553.87%	−208.89%	54.80%	200.23%	−25.00%	2.05%	630.68%
w_2	0%	0%	0%	0%	0%	0%	100%
1260	USA	UK	JPN	ITA	GER	FRA	CAN
μ	4.76%	9.06%	14.02%	11.77%	11.94%	11.52%	15.26%
σ	17.57%	19.39%	24.97%	19.97%	27.98%	24.39%	18.83%
w_1	−373.32%	−190.63%	62.17%	98.56%	64.67%	33.79%	404.76%
w_2	0%	0%	14.70%	0%	0%	0%	85.30%
2520	USA	UK	JPN	ITA	GER	FRA	CAN
μ	4.51%	7.73%	10.81%	9.67%	10.74%	10.01%	11.67%
σ	17.31%	18.93%	23.91%	19.33%	27.41%	23.81%	17.93%
w_1	−278.63%	−159.50%	52.44%	55.36%	79.64%	44.51%	306.18%
w_2	0%	0%	18.69%	0%	0%	0%	81.31%
6300	USA	UK	JPN	ITA	GER	FRA	CAN
μ	4.06%	5.86%	7.00%	7.05%	9.05%	7.91%	7.55%
σ	17.14%	18.36%	22.48%	18.66%	26.53%	22.96%	17.11%
w_1	−147.49%	−108.20%	40.34%	8.29%	88.43%	49.15%	169.47%
w_2	0%	0%	19.46%	0%	36.04%	0%	44.50%
12600	USA	UK	JPN	ITA	GER	FRA	CAN
μ	2.96%	4.15%	5.11%	4.84%	6.72%	5.70%	4.88%
σ	16.80%	18.00%	22.10%	18.15%	25.83%	23.37%	16.55%
w_1	−76.83%	−60.97%	43.01%	−6.15%	71.30%	35.85%	93.78%
w_2	0%	0%	31.10%	0%	43.28%	0%	25.62%

Table 4: Empirical and predictive moments of yearly discrete returns as well as the corresponding portfolio weights for the constraints $\mathcal{C}_1$ and $\mathcal{C}_2$.

and investing 487.05% in CAN - a strategy which would certainly not be pursued in practice. When short-selling is forbidden, all the available capital is invested in CAN. Compared to that the two minimum variance portfolios are far more diversified. However, it can be clearly seen that these portfolios are not optimal in the sense of expected return maximization, since the asset with the smallest estimated return, USA, possesses the highest weight in both minimum variance portfolios.

The optimal portfolios in case $\kappa_1 = 1$, which almost corresponds to a diffuse prior information about the expected asset returns, are similar to the Markowitz portfolios. However, using an appropriate model for high-frequency data apparently leads to slight changes of the expected returns, variances, and covariances which alters the optimal portfolios. Nevertheless, the optimal portfolio for $\kappa_1 = 1$ in P2 is the same as in the empirical case, where all the capital is invested in CAN.

The more confident the investor is about the expected asset returns, the more the optimal portfolios tend to be diversified. In case $\kappa_1 = 1260$ the investor relies on their prior assumption about the expected returns as much as on the empirical information. The optimal portfolio in P1 does not possess weights which are such excessive as for traditional Markowitz optimization or in the case $\kappa_1 = 1$. For instance the amount of capital invested in CAN reduces to 404.76%. In P2 not all the capital is put into CAN anymore. Instead, 14.70% is invested in JPN now. The reason for that is that the expected predictive asset returns are shrunk towards the prior assumption $\eta_0 = 0$. So increasing the confidence in prior information clearly reduces estimation risk. This effect even strengthens when κ_1 is further increased.

In fact, $\kappa_1 = 6300$ is a configuration which can be seen as typical for practical investment problems. Here the investor trusts their own assumption about the expected returns 5 times more than the empirical information. Recall that we use a time series of daily log-returns lasting 5 years, which means that the estimation of yearly expected returns is based on 5 observations. So from a practical point of view, when it comes to estimating expected returns it makes sense to trust far more in expert knowledge than in time series information. The optimal portfolio in P2 is more diversified than the Markowitz portfolio on the one hand. On the other hand, in contrast to the MVP, it also takes account for the expected predictive returns and the investor's will to reap the profit.

The optimal portfolios for $\kappa_1 = 12600$ are even more diversified. However, here almost all of the empirical information about the expected returns is lost, since the confidence in the corresponding prior assumption is 10 times higher than the empirical evidence.

5 Conclusion

We develop an approach to incorporate the stylized facts of high-frequency financial data and arbitrary prior information into the portfolio optimization process. Our approach is characterized by rather weak assumptions about the underlying stochastic process. Using Gordin's central limit theorem, we are able to approximate the distribution of asset log-returns of long investment horizons by the normal distribution. In order to avoid estimation risk, we rely on the Bayesian framework which allows us to include subjective prior information such as expert knowledge. By using a Markov chain Monte Carlo algorithm we simulate the posterior distribution of the unknown model parameters and after that we calculate the first two moments of the discrete predictive asset returns after the given investment period. In a last step, we perform a standard portfolio optimization using these predictive moments, which incorporate both empirical information contained in the data and subjective prior information of the investor.

We give a practical example to demonstrate the applicability of our approach to real-world problems. For that purpose, we use 7 time series of daily log-returns. For the data generating process, we propose an orthogonal MGARCH model. The investor's subjective prior information about expected asset returns and eigenvalues of the covariance matrix is modeled using a hierarchical approach. The suggested portfolios show that prior assumptions have a substantial impact on the optimal decision. Our portfolios become well-diversified compared to the outcomes of traditional portfolio optimization strategies and reflect the investor's assessment about the market. The computational performance of our algorithm encourages applying our approach to higher-dimensional problems in practice, where both empirical information contained in time series and expert knowledge are available.

References

O.E. Barndorff-Nielsen, E. Nicolato, and N. Shepard (2002), 'Some recent developments in stochastic volatility modelling', *Quantitative Finance* **2**, pp. 11–23.

L. Bauwens, S. Laurent, and J.V.K. Rombouts (2006), 'Multivariate GARCH models: a survey', *Journal of Applied Econometrics* **21**, pp. 79–109.

J. Berger (2006), 'The case for objective Bayesian analysis', *Bayesian Analysis* **1**, pp. 385–402.

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

T. Bollerslev (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, pp. 307–327.

S. Cambanis, S. Huang, and G. Simons (1981), 'On the theory of elliptically contoured distributions', *Journal of Multivariate Analysis* **11**, pp. 368–385.

V.K. Chopra and W.T. Ziemba (1993), 'The effect of errors in means, variances, and covariances on optimal portfolio choice', *Journal of Portfolio Management* **19**, pp. 6–11.

R.F. Engle (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation', *Econometrica* **50**, pp. 987–1007.

G. Frahm (2007), 'Linear statistical inference for global and local minimum variance portfolios', Discussion paper, University of Cologne, Department of Economic and Social Statistics, Germany.

G. Frahm and U. Jaekel (2007), 'Tyler's M-estimator, random matrix theory, and generalized elliptical distributions with applications to finance', Discussion paper, University of Cologne, Department of Economic and Social Statistics, Germany.

P.A. Frost and J.E. Savarino (1986), 'An empirical Bayes approach to efficient portfolio selection', *Journal of Financial and Quantitative Analysis* **21**, pp. 293–305.

D. Gamerman and H.F. Lopes (2006), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman & Hall, 2nd edition.

L. Garlappi, R. Uppal, and T. Wang (2007), 'Portfolio selection with parameter and model uncertainty: A multi-prior approach', *Review of Financial Studies* **20**, pp. 41–81.

J. Geweke (1989), 'Bayesian inference in econometric models using Monte Carlo integration', *Econometrica* **57**, pp. 1317–1339.

J. Geweke (1995), 'Bayesian comparison of econometric models', Working paper 532, Federal Reserve Bank of Minneapolis, Minnesota.

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

F. Hayashi (2000), *Econometrics*, Princeton University Press.

U. Herold and R. Maurer (2006), 'Portfolio choice and estimation risk: A comparison of Bayesian to heuristic approaches', *ASTIN Bulletin* **36**, pp. 135–160.

E. Jacquier, N.G. Polson, and P.E. Rossi (1994), 'Bayesian analysis of stochastic volatility models', *Journal of Business and Economic Statistics* **12**, pp. 371–389.

E. Jacquier, N.G. Polson, and P.E. Rossi (2004), 'Bayesian analysis of stochastic volatility models with fat-tails and correlated errors', *Journal of Econometrics* **122**, pp. 185–212.

P. Jorion (1986), 'Bayes-Stein estimation for portfolio analysis', *Journal of Financial and Quantitative Analysis* **21**, pp. 279–292.

B.A. Kalymon (1971), 'Estimation risk in the portfolio selection model', *Journal of Financial and Quantitative Analysis* **6**, pp. 559–582.

A. Kempf and C. Memmel (2006), 'Estimating the global minimum variance portfolio', *Schmalenbach Business Review* **58**, pp. 332–348.

R.W. Klein and V.S. Bawa (1976), 'The effect of estimation risk on optimal portfolio choice', *Journal of Financial Economics* **3**, pp. 215–231.

H. Markowitz (1952), 'Portfolio Selection', *Journal of Finance* **7**, no. 1, pp. 77–91.

A.J. McNeil, R. Frey, and P. Embrechts (2005), *Quantitative Risk Management*, Princeton University Press.

R.C. Merton (1980), 'On estimating the expected return on the market: An exploratory investigation', *Journal of Financial Economics* **8**, pp. 323–361.

N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, et al. (1953), 'Equation of state calculations by fast computing machines', *The Journal of Chemical Physics* **21**, pp. 1087–1092.

A. Meucci (2005), *Risk and Asset Allocation*, Springer.

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

V. Plerou, P. Gopikrishnan, B. Rosenow, et al. (1999), 'Universal and nonuniversal properties of cross correlations in financial time series', *Physical Review Letters* **83**, pp. 1471–1474.

N.G. Polson and B.V. Tew (2000), 'Bayesian portfolio selection: An empirical analysis of the S&P 500 index 1970-1996', *Journal of Business and Economic Statistics* **18**, pp. 164–173.

B. Scherer and R.D. Martin (2007), *Introduction to Modern Portfolio Optimization with NuOPT, S-PLUS and S+Bayes*, Springer, 2nd edition.