An error formula for Monte Carlo based portfolio optimization

J. Abad and T. R. Hurd*
Dept. of Mathematics and Statistics
McMaster University
Hamilton ON L8S 4K1
Canada

November 12, 2003

1 Introduction

One–period utility based portfolio optimization is a classic method in financial economics [13]. The theory is extremely well developed, and extends in many directions, for example to dynamic portfolio optimization in continuous time [12]. Nonetheless, the solutions to these problems, characterized theoretically, often remain dismayingly difficult to compute in practise. Very often, finance practitioners retreat from the economically sound methods of utility based optimization to a computationally simple but economically unjustified mean–variance method.

This paper will examine a straightforward Monte Carlo method for computing utility-optimal portfolios for problems of investing in M assets over a single period. After reviewing the basics of portfolio theory in sections 2 and 3, in section 4 we propose a formula for the error between the true solution and the Monte Carlo estimate and give a theoretical justification for it (but not a proof). As is typical in Monte Carlo methods, the error term is difficult to estimate accurately. In this paper we give rough estimates of the theoretical error in two ways: first by using the Monte Carlo simulation itself, and second by using an approximate formula derived under a Gaussian assumption. We find these two approaches to be mutually consistent, and consistent with the observed convergence of the Monte Carlo estimates.

*Research supported by the Natural Sciences and Engineering Research Council of Canada and Mathematics of Information Technology and Complex Systems, Canada

In section 5 we test the efficiency of the method in a benchmark problem which admits an exact solution, namely a four dimensional problem with asset returns modeled by a mixture of two Gaussians. Here we observe that the convergence of the method to the true portfolio as the sample size increases is completely consistent with the error formula.

In section 6, we apply the method to determine the optimal four asset portfolio in a market modeled by the multivariate Student t distribution with five degrees of freedom. This distribution is a favourite example of a fat tailed distribution which has often been used for asset modeling. To keep the model relevant, the model parameters are calibrated to a data set comprising the historical joint daily returns of four large cap equities, GE, Exxon-Mobil, Coke and Intel. The risk aversion parameter in the optimization is taken over a range appropriate to the financial context. In the resulting investigation, we find that the error formula implies that simulation sizes in the range of 10^6 to 10^8 are needed to compute portfolios with an accuracy of 1%. By choosing a range of simulation sizes from 2^{10} to 2^{20}, we are able to see an agreement between the predicted convergence rate and the observed rate.

The long history of applying Monte Carlo methods in finance began in [4] with the pricing of derivatives. An excellent recent review of Monte Carlo methods for pricing and hedging in finance is [7]. In this book the focus is on univariate problems. Only recently have Monte Carlo methods been applied to the more general problem of portfolio optimization, beginning with [6] and [5]. In particular, even the one period portfolio problem posed in the present paper has not been widely studied. No doubt our method has often been used in industry, but as far as the authors are aware, no works on this subject have been published in the academic literature. In particular

In summary, this paper extends the huge existing literature on portfolio selection by providing a robust, theoretically justified, numerical method for computing one period utility optimal portfolios. We have tested the method on a practical four dimensional problem and found that the observed numerical behaviour of the approximation is compatible with the theoretical error formula. For this example, adequate performance is achievable in several minutes on a desktop computer.

The primary motivation for developing results of this type is to gain confidence and proficiency in computing optimal portfolios for financially realistic problems. In particular we wish to compare the performance of utility optimal portfolios to those produced by the industry standard, but not economically justified mean-variance method. In the present paper, we observe in our example that the efficient frontiers of the utility problem and mean–variance problem coincide to a high degree of accuracy. This somewhat surprising result will be studied in much greater detail in a companion paper [1] which surveys one period portfolio problems for a range of probabilistic models.

2 Setup

We begin our modeling with a data set of multivariate asset prices sampled at discrete times $t_i = i\Delta t, i \in \mathbb{Z}$ where $t_0 = 0$ denotes the present time. Using data from the past $t_i < 0$, we will select a probabilistic model which captures the important stylized facts of the time series as well as matching the important statistical parameters. Based on the selected model, we then provide the means to compute the optimal allocation of assets for an investor with given risk preferences (specified by a utility function) and initial wealth, over a time period $[0, T], T = t_N$.

In section 4 we shall review the two basic methods for portfolio selection over a single time period: the industry standard mean–variance optimization method and the economically preferable utility based method.

Let $\tilde{Y}_i = \{\tilde{Y}_i^\alpha\}_{\alpha=1}^M$ where $\tilde{Y}_i^\alpha = \log S_t^\alpha$ is the log price of the αth asset at time t_i. Let $Y_i = \tilde{Y}_i - \tilde{Y}_{i-1}$ denote the vector of log returns over the period $(t_{i-1}, t_i]$. We also consider $X_i = [S_i - S_{i-1}]/S_{i-1} = e^{Y_i} - 1$, the vector of relative returns over this period.

The basic model for returns is the *geometric Gaussian model* G1 (attributable to Samuelson) which assumes that the *log returns* $\{Y_i\}_{i\in\mathbb{Z}}$ form a sequence of iid multivariate Gaussian random variables $Y_i \sim N(\mu\Delta t, \Sigma'\Sigma\Delta t)$ where the vector μ is called the mean rate of return and the symmetric positive matrix $C = \Sigma'\Sigma$ is called the covariance. The *arithmetic Gaussian model* A1 is attributable to Bachelier [3] and assumes that the relative returns X_i are iid multivariate Gaussian. The A1 model is very tractable, but leads to negative values of stock prices with non-zero probability. Note that for Δt sufficiently small, the relative returns X_i become indistinguishable from the log returns Y_i.

The experience of decades of quantitative analysis has shown some specific ways in which observed financial time series fail to satisfy the properties of the geometric Gaussian model G1.

1. fat tails: the marginal distributions of observed time series have a typical shape which differs from the Gaussian. Figure 1 shows the historical daily log returns for Exxon, together with the probability density function for the Gaussian distribution which best approximates it. We note that the historical distribution is higher in the tail and central areas, while lower in the shoulders, characteristic of the leptokurtic nature of financial returns.

2. stochastic volatility: it is observed (and expected by no arbitrage considerations) that the autocorrelation of log returns decays quickly to near zero over the time scale of minutes. However, higher order autocorrelations exhibit memory effects which can be interpreted as serial correlation in the covariance matrix.

3. skewness: left tails tend to be slightly fatter than right tails.

4. scaling effects: [8] have argued that the marginal distributions of asset returns measured over two different time increments are related by a simple scaling transformation.

5. multivariate effects: observed multivariate data has "tail dependence" [9]. The observed tail dependence corresponds to the econometric statement that joint extreme moves are systematically more frequent than is consistent with multivariate Gaussian models.

The final effect listed above has important implications for portfolio theory. It implies that the strategy of portfolio diversification, the most important principle in financial risk management, has less power than otherwise expected to mitigate risk under scenarios of market distress.

In the present paper, we shall use an extension of the G1 model in which log returns $\{Y_i\}_{i\in\mathbb{Z}}$ form a sequence of iid multivariate Student t random variables. Such random variables can be written (see e.g. [7]) as

$$Y = \mu + \sqrt{\frac{\nu}{\nu-2}}\Sigma Z \tag{1}$$

where $Z = (Z_1, \ldots, Z_M)$ is a vector of iid Student t random variables with $\nu > 2$ degrees of freedom and $\mu, C = \Sigma'\Sigma$ are the mean and covariance of Y. This model has fat tails and tail dependence, but does not exhibit stochastic volatility, skewness or scaling effects.

3 Review of one period optimization

We now discuss the problem of determining at time $t = 0$ the optimal allocation of the agent's wealth into the assets for the period $T = \Delta t$. Let θ^α denote the fraction of wealth W_0 invested in asset α at time 0. It follows that the wealth at time T will be

$$W = W_0\left(1 + \sum_\alpha \theta^\alpha X^\alpha\right) \tag{2}$$

where X denotes the relative return vector for the period. For each possible vector θ we define the mean return to be $\mu(\theta) = \Delta t^{-1}E[(W - W_0)/W_0]$ and the variance to be $\sigma^2(\theta) = \Delta t^{-1}\mathrm{Var}[(W - W_0)/W_0]$.

1. Mean-variance optimization [11]: For a given mean return value μ (selected by the investor), the MV portfolio is defined by

$$\theta^{MV}(\mu) = \mathrm{argmin}_{\substack{\theta:\mu(\theta)\geq\mu \\ \theta\cdot e=1}}\sigma^2(\theta) \tag{3}$$

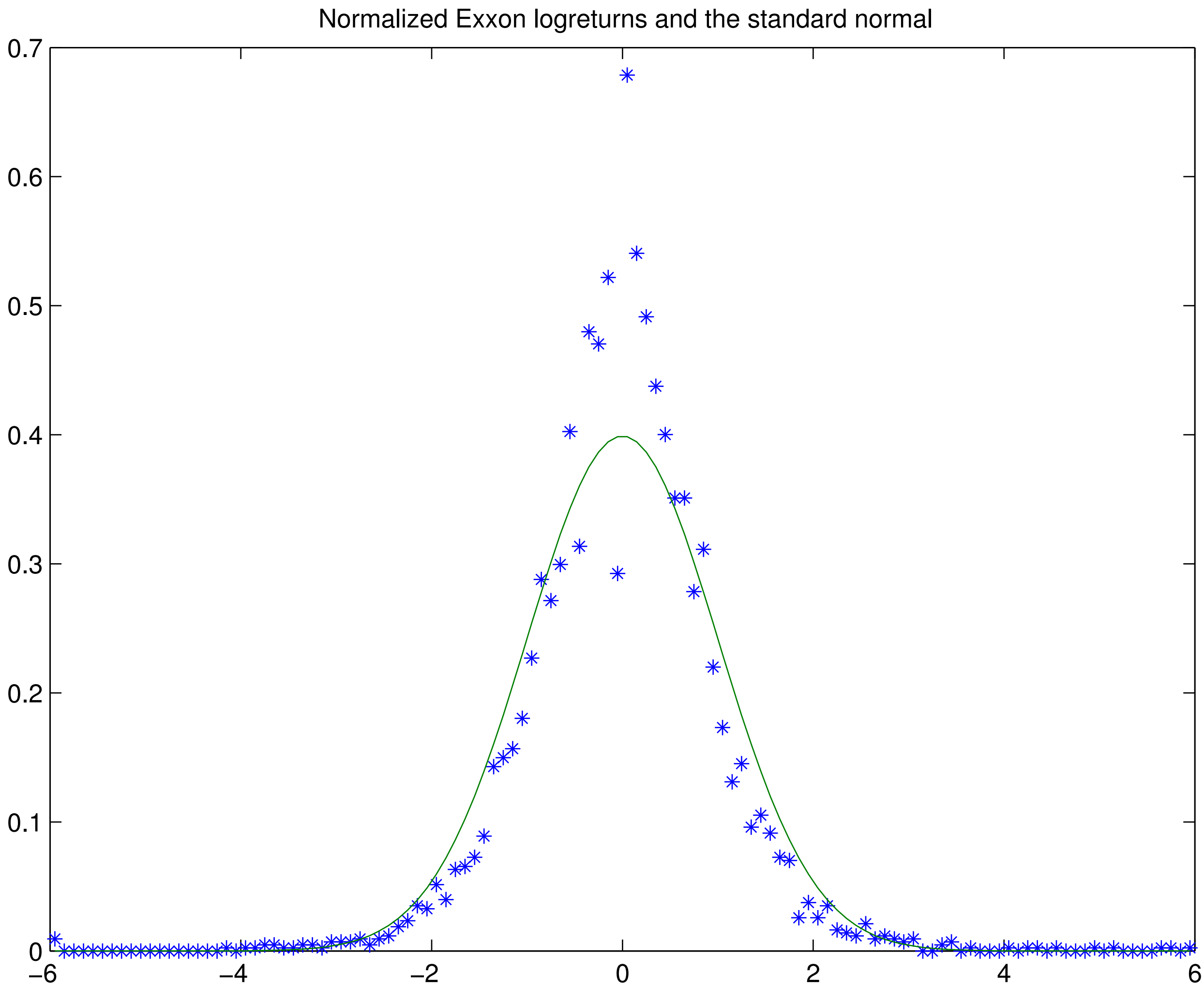

Figure 1: : Normalized sample probability distribution of daily log returns for Exxon over the period 1950-2003, compared to standard Gaussian distribution function

The solution, as a function of μ, is dependent only on the mean $\mu_X = (\Delta t)^{-1}E[X]$ and covariance $C_X = (\Delta t)^{-1}\mathrm{Cov}[X]$ of X. When C_X is nondegenerate, then

$$\theta(\mu) = C_X^{-1}\left[\lambda_1 \mu_X + \lambda_2 \mathbf{e}\right] \tag{4}$$

where

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} \mu_X' C_X^{-1} \mu_X & \mu_X' C_X^{-1} e \\ e' C_X^{-1} \mu_X & e' C_X^{-1} e \end{bmatrix}^{-1} \begin{bmatrix} \mu \\ 1 \end{bmatrix} \tag{5}$$

The curve $\{(\sigma^2(\theta^{MV}(\mu)), \mu), \mu \geq \mu^*\}$ is called the *mean variance efficient frontier.*

2. Utility based optimization:

   "Utility" is a measure of an agent's attitude to wealth: more wealth gives the agent more "happiness" or utility. A utility function $U : \mathbb{R} \to \bar{\mathbb{R}}$ is defined to be a concave, increasing differentiable function taking values in the extended real line. For each U, the utility–optimal portfolio $\theta(U)$ is that allocation which gives the agent the maximal expected utility of terminal wealth, that is:

   $$\theta(U) = \mathrm{argmax}_{\theta:\theta\cdot e=1} E[U(W(\theta))] \tag{6}$$

   It can be proved under broad hypotheses on the model and the utility function that there exists a unique solution to this problem.

   Examples:

   (a) Exponential utility: This is the family $U(x) = -\gamma^{-1}e^{-\gamma x}$ with one parameter $\gamma > 0$ called the absolute risk aversion coefficient. In the arithmetic Gaussian model A1, we find

   $$E[U(\theta \cdot X)] = -\gamma^{-1}\Phi_X(\gamma\theta) \tag{7}$$

   where Φ_X is the characteristic function of X:

   $$\Phi_X(a) = e^{-ia'\mu_X - a'C_X a/2} \tag{8}$$

   If X is a multivariate Student t distribution, which has power law tails in all directions, one can see that

   $$E[U(\theta \cdot X)] = \begin{cases} -\gamma^{-1} & \theta = 0 \\ -\infty & \theta \neq 0 \end{cases} \tag{9}$$

   We see that the exponential utility does not admit investment in the risky assets in this case. When $Y = \log(1 + X)$ is a multivariate Student t distribution with all correlations $|\rho_{ij}| < 1$, one can show that $E[U(\theta \cdot X)] = -\infty$ if and only if at least one component of θ is negative.

(b) Power utility: This family has the form $U(x) = p^{-1}(x^p - 1)$ (including the limiting form $U(x) = \log x$ when $p = 0$) for values of the relative risk aversion coefficient $p < 1$. We extend with $U(x) = -\infty$ when $x < 0$. For this utility we have no analytic expressions for $E[U(\theta \cdot X)]$ in any of the above models for returns. We can see that the arithmetic models which allow returns which are unbounded from above and below all give $E[U(\theta \cdot X)] = -\infty$ unless $\theta = 0$. For the geometric models, one can see again that $E[U(\theta \cdot X)] = -\infty$ if and only if at least one component of θ is negative.

In the arithmetic model A1, utility optimization coincides with mean–variance optimization:

Proposition 1 (Stein's Lemma) *Suppose the market model is of type A1 and let U be any utility function. Then the utility optimal portfolio lies on the mean–variance efficient frontier.*

Remark: This result was proved in [10] using Stein's results in [14].

4 Monte Carlo optimization

Analytical formulas are usually impossible to obtain for expectations such as the one in (6), and thus a numerical method of some sort is necessary. In this section we introduce a simple optimization method based on Monte Carlo and estimate its accuracy. Let $\vec{X} \doteq \{X_1, \ldots, X_N\}$ be a size N Monte Carlo sample of the relative return vector X.

We define the random variable $U_\theta(X) = U([\theta, 1 - \theta \cdot e]'X)$, where we have solved the portfolio constraint $\theta \cdot e = 1$ by writing each portfolio allocation vector as $[\theta, 1 - \theta \cdot e]$ with $\theta \in \mathbb{R}^{M-1}$. The optimal allocation of an agent with initial wealth 1 and utility U is the solution of

$$\hat{\theta} = \mathrm{argmax}_\theta F(\theta), \quad F(\theta) \doteq E[U_\theta] \tag{10}$$

The Monte Carlo approximation $\tilde{\theta}(\vec{X})$ generated by $\vec{X}$ is defined to be

$$\tilde{\theta}(\vec{X}) = \mathrm{argmax}_\theta f(\theta|\vec{X}), \quad f(\theta|\vec{X}) \doteq \frac{1}{N} \sum_{i \le N} U_\theta(X_i) \tag{11}$$

Now $f(\theta|\vec{X})$ is certainly concave in θ. If N is large enough then, with near certainty, for any number $P > 0$ there is R such that $\|\theta\| > R$ implies $f(\theta) \le -P$. In this case it is easy to see that the function $f(\theta|\vec{X})$ achieves its maximum value at a unique finite point $\tilde{\theta} \in \mathbb{R}^{N-1}$. Therefore, for N large enough, the method produces a finite approximation $\tilde{\theta}$ to the exact optimum $\hat{\theta}$ with near certainty.

The following formula gives an estimate of the accuracy of this approximation.

Error formula: Let $\hat{K} = \hat{K}_1^2 \hat{K}_2$ be defined by (20) and (16) respectively. Provided the technical assumptions of Lemma 1 and Lemma 2 hold, then for any $\alpha < 1, \epsilon > 0$

$$\text{Prob}\left[\|\tilde{\theta}(\vec{X}) - \hat{\theta}\| > \epsilon\right] < \alpha \tag{12}$$

provided $N \geq \frac{\hat{K}}{\alpha\epsilon^2}$ where N is the size of the simulation.

Our heuristic derivation of this error estimate is based on noting that the approximate solution $\tilde{\theta}$ satisfies $\sum_i \nabla U_\theta(X_i) = 0$ while $\hat{\theta}$ is the root of $\nabla F(\theta) = 0$. Two lemmas proved in the Appendix will be used: we will see that $\nabla F(\tilde{\theta})$ will be small by Lemma 1, and then by Lemma 2 $\tilde{\theta} - \hat{\theta}$ will be small.

The first lemma applies to the $\mathbb{R}^M$ random variable X, the random sample $\vec{X}$ and utility function U and involves further random variables defined in terms of X:

$$\begin{aligned} \nabla U_\theta(X) &= U'([\theta, 1 - \theta \cdot e]'X)[X^{<M} - eX^M] \in \mathbb{R}^{M-1} && (13) \\ \nabla^2 U_\theta(X) &= U''([\theta, 1 - \theta \cdot e]'X)[X^{<M} - eX^M][X^{<M} - eX^M] \in \mathbb{R}^{(M-1)\times(M-1)} && (14) \end{aligned}$$

Note that $E[\nabla U_\theta] = \nabla F(\theta), E[\nabla^2 U_\theta] = \nabla^2 F(\theta)$.

Lemma 1 *Let $\vec{X} = \{X_1, \ldots, X_N\}$ be a size N Monte Carlo sample of the relative return vector X and $\theta \in \mathcal{V} \subset \mathbb{R}^{M-1}$. Then for any $\epsilon > 0$ and confidence level $1 - \alpha, \alpha \in (0, 1)$*

$$\text{Prob}\left[\|\frac{1}{N}\sum_i \nabla U_\theta(X_i) - E[\nabla F(\theta)]\| > \epsilon\right] \leq \alpha \tag{15}$$

provided

$$\hat{K}_2 \doteq \max_{\theta \in \mathcal{V}} \text{tr}[\text{Cov}(\nabla U_\theta)] < \infty \tag{16}$$

and N is chosen larger than $\frac{\hat{K}_2}{\alpha\epsilon^2}$.

We suppose that $\mathcal{V} \subset \mathbb{R}^{M-1}$ is chosen convex and large enough that it contains both $\tilde{\theta}$ and $\hat{\theta}$. For the specific value of $\tilde{\theta}(\vec{X})$ we assume[1] we can apply Lemma 1. For any constant K_1, if we take

$$N \geq \frac{K_1^2 \hat{K}_2}{\alpha\epsilon^2} \tag{17}$$

[1] This point is not rigorous because $\tilde{\theta}$ is actually a function of the random vector $\vec{X}$. Fixing $\tilde{\theta}$ imposes a nonlinear constraint on $\vec{X}$ violating the independence assumption underlying 1. We must keep this caveat in mind when applying the error estimate.

with $\hat{K}_2$ defined by (16), then with probability $1-\alpha$,

$$\begin{aligned}\|\nabla F(\tilde{\theta})]\| &\leq \|\frac{1}{N}\sum_i \nabla U_{\tilde{\theta}}(X_i)\| + \|\frac{1}{N}\sum_i \nabla U_{\tilde{\theta}}(X_i) - \nabla F(\tilde{\theta})]\| \\ &\leq \frac{\epsilon}{K_1}. \end{aligned} \tag{18}$$

Finally we apply the following Lemma to the function $g = -F$:

Lemma 2 *Let $g : \mathcal{V} \subset \mathbb{R}^n \to \mathbb{R}$ be two times differentiable and convex on the convex domain $\mathcal{V}$. Suppose $\hat{\theta} \in \mathcal{V}$ satisfies $\nabla g(\theta) = 0$ and thus minimizes g. Then for any $\theta \in \mathcal{V}$*

$$\|\theta - \hat{\theta}\| \leq \hat{K}_1 \|\nabla g(\theta)\| \tag{19}$$

provided

$$\hat{K}_1 \doteq \sup_{\theta \in \mathcal{V}} \|(\nabla^2 g(\theta))^{-1}\| < \infty \tag{20}$$

(this is the reciprocal of the smallest singular value the Hessian matrix achieves on $\mathcal{V}$.)

This leads to the desired estimate

$$\|\tilde{\theta} - \hat{\theta}\| \leq \hat{K}_1 \|\nabla F(\tilde{\theta})\| \leq \epsilon \tag{21}$$

with probability $1-\alpha$ provided

$$N \geq \frac{\hat{K}_1^2 \hat{K}_2}{\alpha \epsilon^2}. \tag{22}$$

Remark: It is clear that an important difficulty in applying this error estimate lies in determining the quantity $\hat{K} \doteq \hat{K}_1^2 \hat{K}_2$. In our simulation work, we estimate $\hat{K}$ simply by evaluating the quantities $\mathrm{Cov}(\nabla U_\theta)$, $\|(\nabla^2 g(\theta))^{-1}\|$ at the values of θ which arise during the iterations of the optimization algorithm: when these quantities are observed to be relatively stable we view their average value as a somewhat optimistic approximation to $\hat{K}_1$ and $\hat{K}_2$.

We can also understand the size of $\hat{K}$ by an exact computation in the A1 model of

$$\tilde{K} = \|\nabla^2 F^{-1}\| \times \mathrm{tr}\left(E[\nabla U_\theta \nabla U_\theta] - E[\nabla U_\theta] E[\nabla U_\theta]\right) \tag{23}$$

at $\hat{\theta}$, the minimizer of $F(\theta) = E[U_\theta]$. In terms of $\theta \in \mathbb{R}^{M-1}$ we have

$$F(\theta) = -\gamma^{-1} \exp\left[-\gamma[A\theta + e_M]'\mu_X + \frac{\gamma^2}{2}[A\theta + e_M]' C_X [A\theta + e_M]\right] \tag{24}$$

where A is the $M \times (M-1)$ matrix such that $[A\theta]' = [\theta', -\theta \cdot e]$ and $e_M = [0, \ldots, 0, 1]'$. The minimizer satisfies

$$A'\mu_X = \gamma A' C_X [A\hat{\theta} + e_M] \tag{25}$$

Then at $\theta = \hat{\theta}$ we find

$$E[\nabla U_\theta \nabla U_\theta] = \frac{\partial}{\partial \theta_1}\frac{\partial}{\partial \theta_2} E\left[\exp[-\gamma(\theta_1+\theta_2)' A' X - 2\gamma e_M' X]\right]\Big|_{\theta_1=\theta_2=\hat{\theta}} \tag{26}$$

$$= [\gamma^2 A' C_X A + \gamma^2 (A'\mu_X) \otimes (A'\mu_X)] \tag{27}$$

$$\times F(\hat{\theta})^2 \exp\left[\gamma^2 [A\hat{\theta} + e_M]' C_X [A\hat{\theta} + e_M]\right] \tag{28}$$

$$\tag{29}$$

and

$$\nabla^2 F(\hat{\theta}) = \gamma^2 A' C_X A \; F(\hat{\theta}) \tag{30}$$

Therefore

$$\tilde{K} = \gamma^{-2} \|(A' C_X A)^{-1}\|^2 \left[\mathrm{tr}(A' C_X A) + \|A'\mu_X\|^2\right) \exp\left[\gamma^2 [A\hat{\theta} + e_M]' C_X [A\hat{\theta} + e_M]\right] \tag{31}$$

When γ is not too large, we find

$$\tilde{K} \sim M\gamma^{-2}, \qquad M = \|(A' C_X A)^{-1}\|^2 \left[\mathrm{tr}(A' C_X A) + \|A'\mu_X\|^2\right) \tag{32}$$

5 A Benchmark Problem

Before applying the method to a real–life problem in the next section we consider here a simpler portfolio problem where the objective function can be computed in closed form. This will allow us to determine if the Monte Carlo approximants converge to the true solution in the predicted fashion. We are also able to choose the model so that the resulting portfolios lie quite far from the mean–variance frontier.

We model X by a simple mixture of two Gaussians $X = \xi N_1 + (1-\xi) N_2$ where the independent random variables are ξ, binomial with probabilities $p_1, p_2 = 1 - p_2$, and $N_i \sim N(\mu_i, C_i)$. If we choose $p_1 = .9$ and mixture components

stock	$\hat{\mu}_1$	$\hat{C}_1(\cdot, \mathrm{GE})$	$\hat{C}_1(\cdot, \mathrm{XOM})$	$\hat{C}_1(\cdot, \mathrm{KO})$	$\hat{C}_1(\cdot, \mathrm{INTC})$
GE	0.00031	0.00017	0.00010	0.00009	0.00008
XOM	0.00036	0.00010	0.00022	0.00012	0.00011
KO	0.00048	0.00009	0.00016	0.00019	0.00013
INTC	0.00069	0.00008	0.00011	0.00013	0.00060

stock	$\hat{\mu}_2$	$\hat{C}_2(\cdot, \mathrm{GE})$	$\hat{C}_2(\cdot, \mathrm{XOM})$	$\hat{C}_2(\cdot, \mathrm{KO})$	$\hat{C}_2(\cdot, \mathrm{INTC})$
GE	0.00224	0.00087	0.00025	0.00027	0.00025
XOM	0.00419	0.00025	0.00118	0.00039	0.00021
KO	0.00349	0.00027	0.00039	0.00142	0.00090
INTC	0.00736	0.00025	0.00021	0.00090	0.00347

then μ_X, C_X agree with the values for the Student t model found in the next section. Thus the mean–variance optimal portfolios coincide. However, the mixture of Gaussians here is quite different both from an A1 model and the Student t distribution and thus will lead to utility–optimal portfolio curves which are quite different from the mean–variance efficient frontier.

The theoretical optimizer $\hat{\theta}(\gamma)$ is obtainable by a numerical maximization of the objective function

$$F(\theta) = -\gamma^{-1} \sum_{i=1}^{2} p_i \exp\left[-\gamma[A\theta + e_M]'\mu_i + \frac{\gamma^2}{2}[A\theta + e_M]'C_i[A\theta + e_M]\right] \tag{33}$$

The one period optimization algorithm was computed on Monte Carlo simulations of the relative return vector X with a range of sample sizes $N = 2^n, n = 12, 14, 16, 18, 20$ and with the exponential utility with a range of risk aversion parameters $\gamma \in [1, 32]$. For each simulated data set and choice of γ, a straightforward Newton–Raphson iteration algorithm was used to compute $\tilde{\theta}^{(n)}(\gamma)$, the solution of (11). Our interest is in observing whether in all cases the expected convergence of the Monte Carlo approximants to the true solutions $\hat{\theta}(\gamma)$ is exhibited as n increases.

To test whether the errors $\epsilon^{(n)}(\gamma) \doteq \|\tilde{\theta}^{(n)} - \hat{\theta}\|$ behave consistently with the theoretical error formula we also need to obtain approximate values for $\hat{K} = \hat{K}_1^2\hat{K}_2$. For this we computed values using the simulation itself. Since $\tilde{\theta}$ varies randomly with n during the computation, we thereby observe K on a random set of points near $\hat{\theta}$. Table 6 shows these values as well as the estimated value $\tilde{K}$ given by (23). This table shows that K tends to be stable in n over the range of γ values, and furthermore these values agree with the estimate $\tilde{K}$. Therefore we have confidence that these values serve as a reliable proxy for the theoretical best value $\hat{K}$.

$N\backslash\gamma$	2^1	2^2	2^3	2^4	2^5
2^{10}	55988	14053	3534.9	887.34	215.14
2^{12}	48319	12091	3016.6	739.54	167.77
2^{14}	54087	13537	3378.2	828.14	186.72
2^{16}	51699	12929	3221.4	787.45	176.79
2^{18}	52372	13094	3261	796.05	177.85
2^{20}	52552	13143	3274.8	800.31	179.29
$\tilde{K}$	32520	8147	2053.9	530.93	151.73

Table 3: The values of $K_1^2 K_2$ for the mixture of Gaussians evaluated at the values $\tilde{\theta}^{(n)}$ for a range of values of $N = 2^n$ and γ. Also shown are the values of $\tilde{K}$.

Figure 2 shows that for this problem the observed error is completely consistent with the theoretical formula.

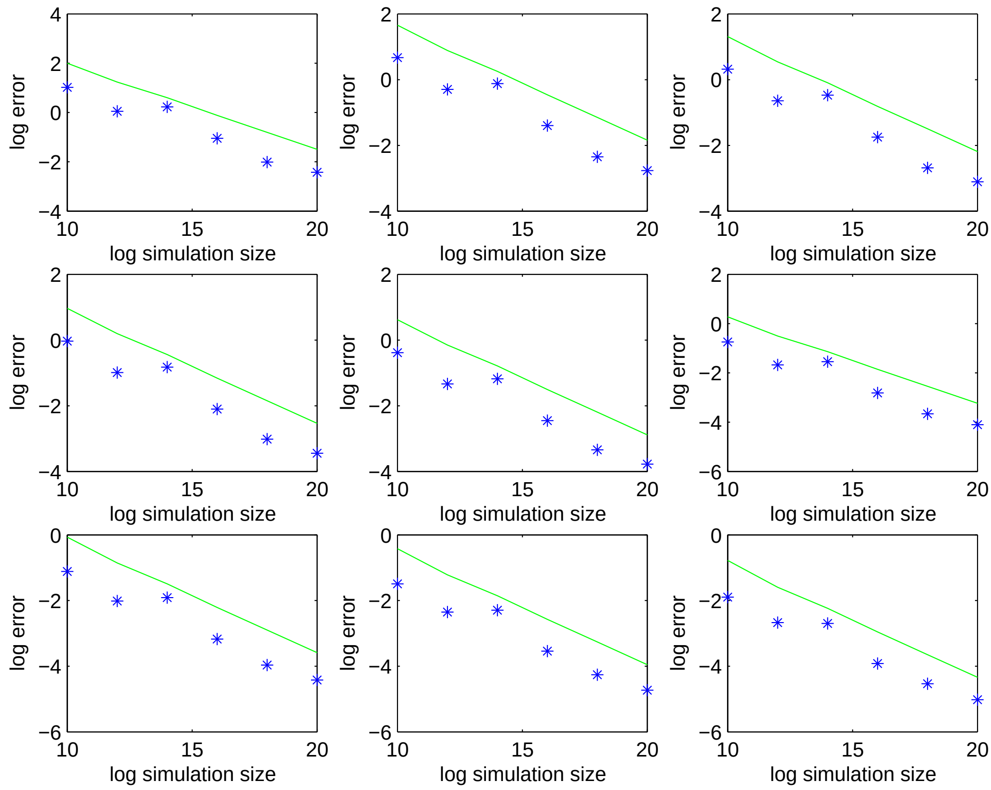

Figure 2: Observed errors and theoretical errors as functions of the log simulation size for the mixture of Gaussians model. Starting from the top left, these correspond to the values $\log_2(\gamma) = 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5$.

6 Real–life Application

With the method verified in the previous section, we consider a financially relevant portfolio problem where no analytical results are known, and so where a numerical method such as the Monte Carlo optimization is needed. We focused on a data set obtained from `http://finance.yahoo.com/` consisting of closing daily share prices of four large cap companies trading on NASDAQ and NYSE, namely GE, Exxon (XOM), Coke (KO), and Intel (INTC) for the period July 9/86 to July 18/03, and consists of 4274 trading days. Then we addressed the problem of an investor trying to determine the portfolio made up of these four assets which maximizes expected exponential utility over the next day to the future, based on the historical data.

By the general properties of the exponential utility, the solution depends on the risk aversion parameter γ but not on the investor's wealth. Before we begin, we expect that for small values of γ, the investor will be excessively risk tolerant, and the corresponding solution will involve shortselling of the worst performing asset. Disallowing shortselling provides a natural lower bound to the values of γ. We will also observe that for large values of γ the portfolios become more difficult to compute, and results turn out to be less and less dependent on γ. Thus in practice only a narrow range of γ corresponds to acceptable risk preferences.

We assume that the log returns $Y = \log(X + 1) \in \mathbb{R}^4$ follow a multivariate Student t distribution as described in section 2. For simplicity we take the degrees of freedom parameter ν to be 5, which is a typical value for daily asset returns. The method of moments applied to the data sample log returns leads to the following values for $\hat{\mu}_Y, \hat{C}_Y$:

stock	$\hat{\mu}$	$\hat{C}(\cdot, \mathrm{GE})$	$\hat{C}(\cdot, \mathrm{XOM})$	$\hat{C}(\cdot, \mathrm{KO})$	$\hat{C}(\cdot, \mathrm{INTC})$
GE	0.00038	0.00024	0.00012	0.00011	0.00010
XOM	0.00058	0.00012	0.00032	0.00015	0.00013
KO	0.00062	0.00011	0.00015	0.00031	0.00021
INTC	0.00091	0.00010	0.00013	0.00021	0.00090

The one period optimization algorithm was computed on Monte Carlo simulations of the relative return vector $X = e^Y - 1$ based on the historically calibrated parameters $\hat{\mu}_Y, \hat{C}_Y$ with a range of sample sizes $N = 2^n$ and with the exponential utility with a range of risk aversion parameters $\gamma \in [1, 32]$. For each simulated data set and choice of γ, a straightforward Newton–Raphson iteration algorithm was used to compute $\tilde{\theta}^{(n)}(\gamma)$, the solution of (11). Our interest is in observing whether in all cases the expected convergence properties as n increases is exhibited.

Since the true optimal portfolio is not known, we ran preliminary computations based on a very large sample size of 2^{24} random vectors, which was the largest sample size our desktop machine could easily handle. If the resulting portfolios are denoted by $\tilde{\theta}^{(24)}(\gamma)$, we obtain the theoretical errors $\epsilon^{(th)}(\gamma) \doteq \|\tilde{\theta}^{(24)} - \hat{\theta}\|$ based on the observed values of K_1, K_2 as well as $\delta^{(mv)} \doteq \|\tilde{\theta}^{(24)}(\gamma) - \theta^{(mv)}(\mu(\gamma))\|$. These values are tabulated below, from which we see that the theoretical error is small enough to clearly resolve the difference between the mean–variance and utility optimal portfolios.

To test the theoretical error formula over the range of sample sizes, we observed the behavior of $\epsilon^{(n)} \doteq \|\tilde{\theta}^{(n)} - \tilde{\theta}^{(24)}\|$) for values of $n = \log_2(N)$ in the range $[10, 20]$. We compare to $\tilde{\theta}^{(24)}$ because $\hat{\theta}$ is not calculable. To obtain approximate values for $\hat{K}_1, \hat{K}_2$ we used the simulation itself. Since $\tilde{\theta}$ varies randomly with n during the computation, we thereby observe K_1, K_2 on a random set of points near $\hat{\theta}$. Table 6 shows that these values tend to be stable in n over the range of acceptable γ values, so it is clear that they serve as a reliable proxy for the theoretical best values $\hat{K}_1, \hat{K}_2$.

$N\backslash\gamma$	2^1	2^2	2^3	2^4	2^5
2^{10}	43581	10880	2689.7	638.54	130.04
2^{12}	41133	10293	2553.9	598.03	103.32
2^{14}	46277	11535	2837.3	661.12	116.13
2^{16}	47566	11912	2961.1	712.93	203.73
2^{18}	47810	11971	2961.2	671.68	190.12
2^{20}	47398	11859	2925.8	663.27	121.77
$\tilde{K}$	32520	8147	2053.9	530.93	151.73

Table 3: The values of $K_1^2 K_2$ evaluated at the values $\tilde{\theta}^{(n)}$ for a range of values of $N = 2^n$ and γ. Also shown are the values of $\tilde{K}$.

Figure 3 graphs the log error between the portfolio vectors versus log simulation sizes, for nine different values of the risk aversion parameter γ. On the same graph, we plot the theoretical error versus the simulation size using the quantity $K_1^2 K_2$ obtained from Table 3. As a rule of thumb, no more than one data point in ten should lie above the curve in these graphs.

Based on the observed values of $\hat{K}$, we estimate that to achieve theoretical error $\epsilon(n) \leq 0.01$ requires sample sizes of approximately $2^{29}, 2^{27}, 2^{25}, 2^{23}, 2^{21}$ for $\gamma = 2^1, 2^2, 2^3, 2^4, 2^5$ respectively.

Figure 4 is a variance–mean plot of the portfolios which result from exponential utility optimization within the multivariate student t model. The Monte Carlo method is used with a range of simulation sizes from 2^{10} to 2^{20}.

7 Summary

The Monte Carlo method provides a feasible method for producing approximate solutions to the one–period utility optimization problem for low dimensional stochastic models calibrated to market data. The method can be applied in any model which can be efficiently simulated. The picture that arises is that in the model we investigated, the Monte Carlo method does indeed converge with the rate predicted by theory. The values $\hat{K}_1, \hat{K}_2$ entering in the error formula can be determined accurately enough to be able to estimate errors simply by computing $E[D^2 U_\theta]$ and $\mathrm{Cov}(DU_\theta)$ for the approximate optimal portfolio vector, since these quantities are observed to vary only slowly over the domain on which the optimization operates.

We notice in our computations that the resulting portfolios do not differ very much from the naive mean–variance portfolios, even when the underlying probability distribution is as

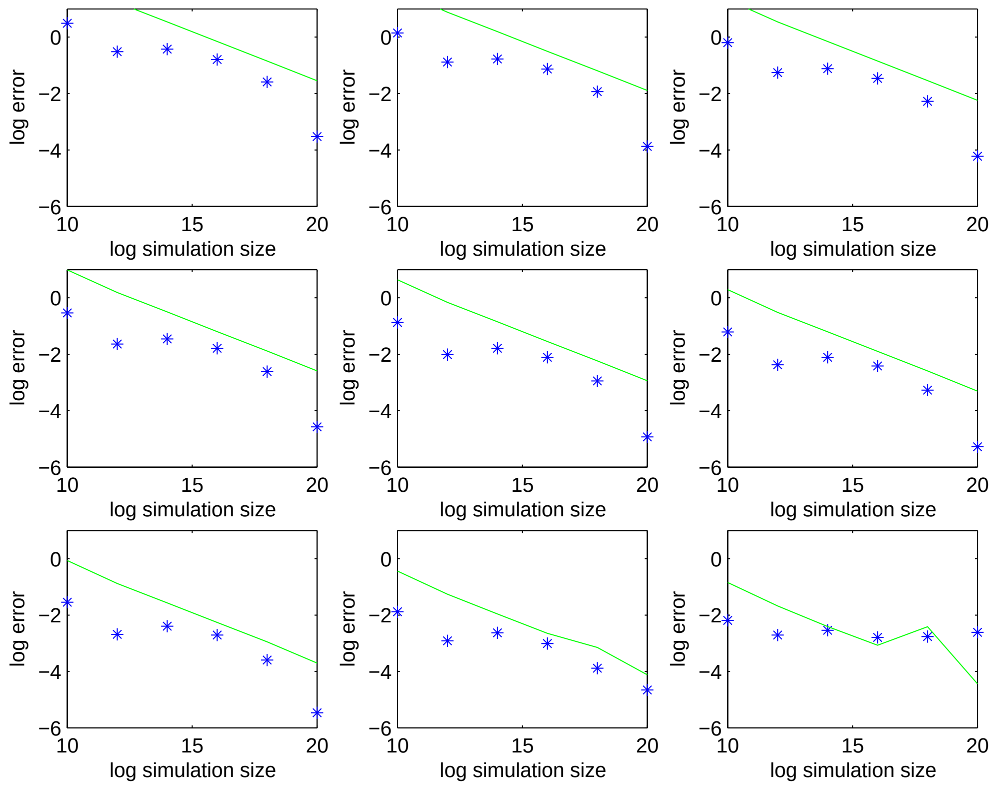

Figure 3: Observed errors and theoretical errors as functions of the log simulation size. Starting from the top left, these correspond to the values $\log_2(\gamma) = 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5$.

fat-tailed as the multivariate Student t with 5 degrees of freedom. This suggests that fat tails in themselves do not greatly influence the choice of optimal portfolio. This somewhat surprising result is the focus of a companion paper [2]. It remains to be seen to what extent tail dependence influences portfolio selection.

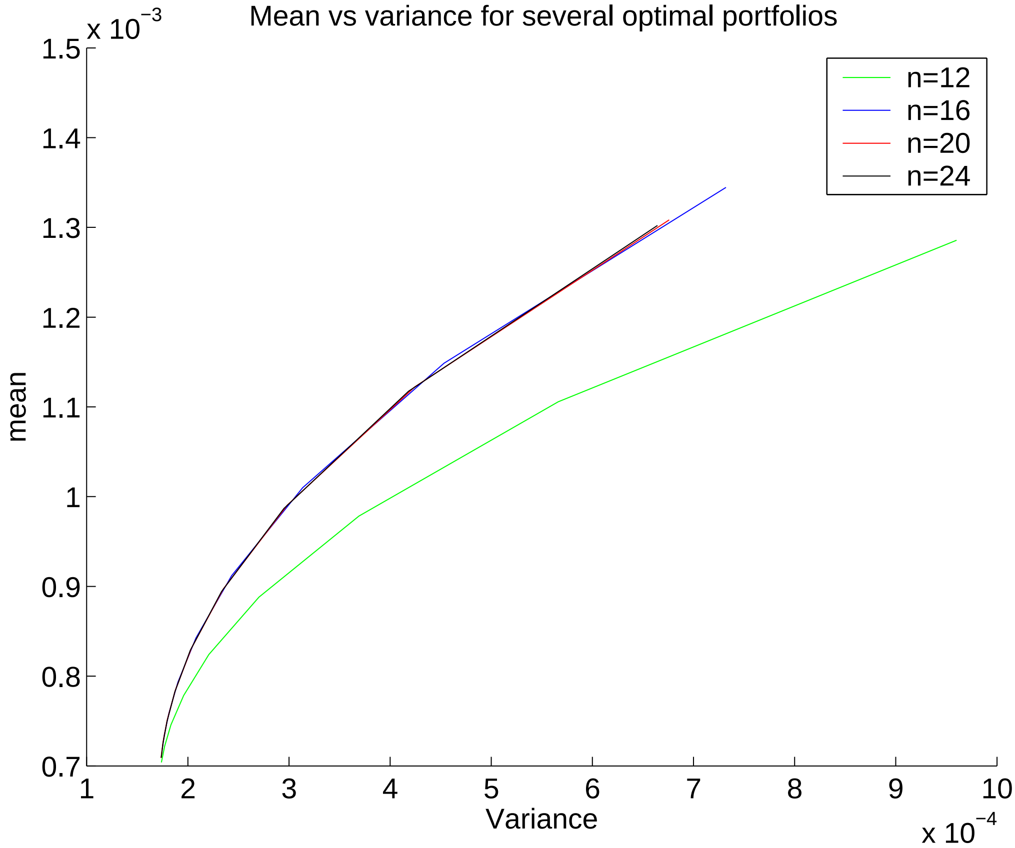

Figure 4: Mean–variance plots of the approximate utility optimal portfolios for a range of γ values. The different curves correspond to the simulation sizes $N = 2^n$ for $n = 12, 16, 20, 24$.

A Appendix

A.1 Proof of Lemma 1

The probability in question equals

$$\mathrm{Prob}\left[\frac{1}{N^2\epsilon^2}\sum_{\alpha=1}^{M-1}\sum_{i,j=1}^{N}\left(\nabla^\alpha U_\theta(X_i) - E[\nabla^\alpha U_\theta]\right)\left(\nabla^\alpha U_\theta(X_j) - E[\nabla^\alpha U_\theta]\right) > 1\right] \tag{34}$$

which is bounded from above by

$$E\left[\frac{1}{N^2\epsilon^2}\sum_{\alpha,i,j}\left(\nabla^\alpha U_\theta(X_i) - E[\nabla^\alpha U_\theta]\right)\left(\nabla^\alpha U_\theta(X_j) - E[\nabla^\alpha U_\theta]\right)\right]$$

$$
\begin{aligned}
&= \frac{1}{N^2\epsilon^2}\sum_{\alpha,i} E\left[(\nabla^\alpha U_\theta(X_i) - E[\nabla^\alpha U_\theta])^2\right] \\
&= \frac{1}{N\epsilon^2}\mathrm{tr}\left[\mathrm{Cov}(\nabla U_\theta)\right] \\
&\leq \frac{K_2}{N\epsilon^2} \leq \alpha.
\end{aligned} \tag{35}
$$

A.2 Proof of Lemma 2

Let $\gamma : [0,1] \to \mathcal{V}$ be the curve of steepest ascent with $\gamma(0) = \hat{\theta}, \gamma(1) = \theta$ and $\|\gamma'(t)\|^2 = \Gamma =$ constant. Then $\int_0^1 \sqrt{\|\gamma'(t)\|^2}\, dt = \Gamma$ is the length of the curve. Now

$$
\begin{aligned}
\|\nabla g(\theta)\|^2 &= \int_0^1 \frac{d}{dt}\|\nabla g(\gamma(t))\|^2 dt \\
&= 2\int_0^1 \left[(\nabla g)\cdot[(\nabla^2 g)\gamma']\right] dt
\end{aligned} \tag{36}
$$

By the steepest ascents condition γ' is parallel to ∇g and from the definition of K_1 this implies $(\nabla g)\cdot[(\nabla^2 g)\gamma'] \geq K_1^{-1}\nabla g\cdot\gamma'$. Thus

$$
\begin{aligned}
\|\nabla g(\theta)\|^2 &\geq 2K_1^{-1}\int_0^1 (\nabla g)\cdot\gamma'(t)dt \\
&= 2K_1^{-1}[g(\theta) - g(\hat{\theta})]
\end{aligned} \tag{37}
$$

Let $G(\lambda) = g(\lambda\theta + (1-\lambda)\hat{\theta})$ and then because $G'(0) = 0$

$$
g(\theta) - g(\hat{\theta}) = G(1) - G(0) = \frac{1}{2}G''(\xi) \tag{38}
$$

for some $\xi \in (0,1)$. But

$$
G''(\xi) = (\theta - \hat{\theta})\cdot[(\nabla^2 g)\ (\theta - \hat{\theta})] \geq K_1^{-1}\|\theta - \hat{\theta}\|^2 \tag{39}
$$

and hence

$$
\|\nabla g(\theta)\|^2 \geq K_1^{-2}\|\theta - \hat{\theta}\|^2 \tag{40}
$$

References

[1] J. Abad and T.R. Hurd. An error formula for monte carlo based portfolio optimization. `http://www.`, 2003.

[2] J. Abad, T.R. Hurd, and L. Seco. Mean-variance versus utility based optimization. `http://www.`, 2003.

[3] L. Bachelier. Théorie de la spéculation. Doctoral dissertation in Mathematical Sciences, Faculté des Sciences de Paris, 1900.

[4] P.P. Boyle. Options: a Monte Carlo approach. *J. of Fin. Economics*, 4:323–338, 1977.

[5] J. Cvitanic, L. Goukasian, and F. Zapatero. Monte carlo simulation of optimal portfolios in complete markets. *Journal of Economic Dynamics and Control*, 27:971–986, 2003.

[6] J. Detemple, R. Garcia, and M. Rindisbacher. A Monte Carlo approach for optimal portfolios. working paper, 2003.

[7] P. Glasserman. *Monte Carlo methods in financial engineering*. Springer Verlag, New York, 2003.

[8] P. Gopikrishnan, V. Plerou, L. Amaral, M. Meyer, and H. E. Stanley. Scaling of the distribution of fluctuations of financial market indices. xxx.lanl.gov/cond-mat/9905305, 1999.

[9] H. Joe. *Multivariate models and dependence concepts*. Chapman and Hall, 1997.

[10] J. G. Kallberg and W. T. Ziemba. Comparison of alternative functions in portfolio selection problems. *Management Science*, 11:1257–1276, 1983.

[11] H. Markowitz. Portfolio selection. *J. Finance*, 8:77–91, 1952.

[12] R. C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *J. Econom. Theory*, 3:373–413, 1971.

[13] John Von Neumann and Oskar Morgenstern. *Theory of games and economic behavior*. Princeton University Press, Princeton, New Jersey, first edition, 1944.

[14] C. Stein. Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9:1135–1151, 1981.