Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

STATISTICAL PROPERTIES OF MEAN-VARIANCE PORTFOLIO OPTIMIZATION

Zhaolin Hu[1]

[1]School of Economics and Management, Tongji University, Shanghai, CHINA

ABSTRACT

We study Markowitz's mean-variance portfolio optimization problem. When practically using this model, the mean vector and the covariance matrix of the assets returns often need to be estimated from the sample data. The sample errors will be propagated to the optimization output. In this paper, we consider three commonly used mean-variance models and build the asymptotic properties for the conventional sample approximations that are widely adopted and studied, by leveraging the stochastic optimization theory. We show that for all three models, under certain conditions the sample approximations have the desired consistency and achieve a convergence rate of square root of sample size, and the asymptotic variance depends on the first four moments of the returns. We conduct numerical experiments to test the asymptotic properties for the estimation. We also conduct experiments to illustrate that the asymptotic normality might not hold when the fourth moments of the returns do not exist.

1 INTRODUCTION

The mean-variance model is a classical portfolio selection model proposed by Markowitz (1952). Its basic idea is to use variance to model the portfolio risk and propose to make a tradeoff between return and risk. The idea opened a door for quantitative financial investment and risk management and Harry Markowitz was awarded the 1990 Nobel prize in economics for his portfolio theory. In this model, decision makers optimize one metric with constraint on the other metric, or optimize a weighted sum of the two metrics. When practically using this model, decision makers need to specify the mean vector and covariance matrix for the random returns/losses of the portfolio assets. However, in practice the parameters are rarely known exactly and often need to be estimated. By using the estimators to replace the true parameters one obtains a sample approximation for the true model and the sample errors may be propagated to that of the decisions. In this paper, we consider the sample errors when formulating and solving this model. We build asymptotic properties for variants of mean-variance models.

While the mean-variance model is easy to solve, the performance of the optimal portfolio returned by the sample approximation typically admits a substantial variability. This issue has been concerned and studied extensively. A vast volume of studies investigated the estimation errors in finance. Britten-Jones (1999) studied the sampling errors in the mean-variance model. They derived inference procedures for the hypotheses about the portfolio weights. DeMiguel et al. (2009b) empirically evaluated the out-of-sample performance of the sample-based mean-variance model against the $1/N$ portfolio and found that the former could not consistently beat the latter over a number of empirical datasets. They showed that a large sample size is required for the sample-based mean-variance strategy to outperform the $1/N$ benchmark. DeMiguel et al. (2009a) and Brodie et al. (2009) studied using regularization to address the sample errors in the mean-variance models. Many of the studies targeted to derive some good estimators for the parameters to achieve better out-of-sample performances for the portfolio optimization model.

The mean-variance model can be viewed as a representative of the mean-risk portfolio optimization models. Lim et al. (2011) studied a portfolio selection model where the risk is measured by the conditional value-at-risk, and showed that the sample solutions are fragile. Shapiro et al. (2014) used the stochastic optimization theory to build statistical properties for the estimation of various risk measures. Hu and

Zhang (2018) studied the statistical properties for computing the utility-based shortfall risk. Wang et al. (2023) studied a mean-variance model that optimizes a weighted sum of the mean and the variance where the portfolio consists of derivative securities. They developed an estimator for the covariance matrix based on simulation and constructed an approximation problem for the mean-variance model. They built statistical properties for the approximation. However, their estimator for the covariance matrix may not be positive semi-definite. They projected the estimator to a positive definite matrix and further solved a new approximation problem. In this paper, we study the sample approximations of the mean-variance models where the mean vector and covariance matrix are estimated by the conventional empirical estimators. This kind of empirical approximations are widely adopted in the financial practice. We analyze the asymptotic properties of the sample approximations.

The variability of the sample approximation is closely related to the perturbation analysis on the mean vector and the covariance matrix. Note, however, that the variance is not in an expectation expression. Actually, it takes a composite form as in the definition of the variance the expectation is embedded in the outer expectation. Directly conducting a perturbation analysis is not an easy task. In this paper, we treat the variance as an optimal value of an expectation and convert the mean-variance model to a stochastic program that only involves expectations. We then use stochastic optimization theory to build the statistical properties of the sample mean-variance model. Based on this approach, our analysis is applicable for various mean-variance models. We show that the sample approximations have the desired consistency and the convergence rate of the sample approximations is in the order of square root of the sample size. The asymptotic variance is affected by the first four moments of the random distribution of the assets. It suggests that while the mean-variance model only depends on the first two moments of the underlying randomness, the sample approximation performance (e.g., the variation) will be affected by the higher moments of the distribution. Furthermore, our analysis shows that adding regularization to the mean-variance model does not affect the asymptotic regimes. Besides building the theoretical results, our work also shows the powerfulness of the stochastic optimization theory via the analysis, which might provide some inspiration on using stochastic optimization to address other sampling-based optimization problems.

The rest of this paper is organized as follows. In Section 2 we study a mean-variance model that optimizes the risk with constraint on the return, and build statistical properties for the sample approximation. In Section 3 we generalize the analysis for more variants of mean-variance models. We conduct some simulation experiments in Section 4 to test the theoretical findings. Section 5 concludes the paper.

2 MEAN-VARIANCE MODEL

Suppose an investor aims to invest on k assets with random return vector $\xi = (\xi^1, \cdots, \xi^k)^{\mathrm{T}}$ where ξ^j is the return rate of asset j. Suppose x_j is the capital invested in asset j. Then the random return of the portfolio is $H(x,\xi) = x^{\mathrm{T}}\xi$ where $x = (x_1, \cdots, x_k)^{\mathrm{T}}$. The mean-variance model proposed by Markowitz (1952) suggests to optimally balance the risk and return. There are alternative expressions for the mean-variance model. We first study the following formulation

$$\begin{aligned} \underset{x\in X}{\text{minimize}} \quad & \mathrm{V}\left[H(x,\xi)\right] \\ \text{subject to} \quad & \mathrm{E}\left[H(x,\xi)\right] \geq r, \end{aligned} \tag{1}$$

where E and V denote the expectation and variance of the random function, r is a prespecified threshold and X is the feasible set of x. Without loss of generality, we assume X is a compact convex set throughout this paper. In Problem (1), the support of ξ, denoted as Ξ, is a closed subset of $\Re^k$. Throughout the paper, it is implicitly assumed that $H(x,\xi)$ has finite first and second moments. Therefore, Problem (1) is well defined. There are other mean-variance variants of Problem (1). We defer the discussion in the next section.

For the linear portfolio $H(x,\xi) = x^{\mathrm{T}}\xi$, Problem (1) is actually a quadratic program

$$\begin{aligned} \underset{x\in X}{\text{minimize}} \quad & x^{\mathrm{T}}\Sigma x \\ \text{subject to} \quad & \mu^{\mathrm{T}}x \geq r, \end{aligned} \tag{2}$$

where μ and Σ are the mean vector and covariance matrix of ξ. Throughout the paper, we assume that Σ is a positive definite matrix. Therefore, the objective function in Problem (2) is strictly convex and Problem (2) has a unique optimal solution. With given parameters μ and Σ, the quadratic program is easy to solve. When practically formulating and solving the problem, however, the parameters μ and Σ are not available directly but need to be estimated. Investors often use a sample from ξ to estimate μ and Σ. Suppose we have n independent and identically distributed (i.i.d.) observations of ξ, denoted as $\{\xi_1, \xi_2, \cdots, \xi_n\}$, which may be the real historical data or the sample simulated from the distribution of ξ. Let

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\xi_i,\ \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \hat{\mu})(\xi_i - \hat{\mu})^{\mathrm{T}}, \tag{3}$$

which are called sample mean and sample covariance matrix. Note that we use $\hat{\Sigma}$ as an estimator of Σ. In statistics, the estimator $\hat{\Sigma}_u = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \hat{\mu})(\xi_i - \hat{\mu})^{\mathrm{T}}$ is also frequently used, which is an unbiased estimator of Σ.

By using the sample estimators in (3) to replace μ and Σ in Problem (2), we obtain the following sample approximation problem

$$\begin{aligned} \underset{x\in X}{\text{minimize}} \quad & x^{\mathrm{T}}\hat{\Sigma} x \\ \text{subject to} \quad & \hat{\mu}^{\mathrm{T}}x \geq r. \end{aligned} \tag{4}$$

Let v_n and v^* denote the optimal values of Problems (4) and (2) respectively. A natural and central problem in this data driven approach is how well v_n approximates v^*. Ban et al. (2018) built that the sample mean-variance model is fragile. They did numerical experiments to check the robustness of the sample based optimal values. Theoretically, they proved that the solution of Problem (4) converges to that of Problem (2) in probability. We show that the solution converges with probability one (w.p.1).

There are substantial difficulties for studying the perturbation of the mean vector and covariance matrix. This structure sets obstacles for the use of the stochastic optimization theory. In this paper, we propose to consider a stochastic program reformulation and analyze the statistical properties of the sample approximation. The consistency of the solution and the optimal value is relatively simple under this stochastic optimization perspective.

2.1 Asymptotic Properties

In this paper, we aim to build the asymptotic properties based on the theory of stochastic optimization. This approach can be directly implemented to build the consistency. We show that the convergence rate is in the order of $n^{-1/2}$. We further derive the asymptotic variance. A critical observation is that Problem (1) can be converted equivalently to the following problem

$$\underset{x\in X, t\in\Re}{\text{minimize}} \quad \mathrm{E}\left[(H(x,\xi) - t)^2\right] \tag{5}$$

$$\text{subject to} \quad \mathrm{E}[H(x,\xi)] \geq r, \tag{6}$$

and Problem (4) can be converted equivalently to the following problem

$$\begin{aligned} \underset{x\in X, t\in\Re}{\text{minimize}} \quad & \frac{1}{n}\sum_{i=1}^{n}(H(x,\xi_i) - t)^2 \\ \text{subject to} \quad & \frac{1}{n}\sum_{i=1}^{n}H(x,\xi_i) \geq r, \end{aligned} \tag{7}$$

where $\{\xi_1, \xi_2, \cdots, \xi_n\}$ is the set of i.i.d. sample that is used in (3). We summarize the results in the following proposition.

Proposition 1 (i) Problem (5) and Problem (1) are equivalent, in the sense that x is an optimal solution of Problem (1) if and only if there exists t such that (x,t) is an optimal solution of Problem (5), and both problems have the same optimal value v^*.

(ii) Problem (7) and Problem (4) are equivalent, in the sense that x is an optimal solution of Problem (4) if and only if there exists t such that (x,t) is an optimal solution of Problem (7), and both problems have the same optimal value v_n.

Proof. (i) The equivalence is built on the fact that

$$\mathrm{V}[H(x,\xi)] = \inf_{t \in \Re} \mathrm{E}\left[(H(x,\xi)-t)^2\right]. \tag{8}$$

To see this, consider the objective function for the infimum in (8), which can be expressed as

$$\mathrm{E}\left[(H(x,\xi)-t)^2\right] = t^2 - 2\mathrm{E}[H(x,\xi)]\,t + \mathrm{E}\left[(H(x,\xi))^2\right].$$

It is a strictly convex function and has a unique minimizer $t^*(x) = \mathrm{E}[H(x,\xi)]$. Plugging $t^*(x)$ into the objective function yields that the infimum is the variance of $H(x,\xi)$. Integrating the minimization and infimum we immediately obtain the equivalence.

(ii) The equivalence can be built based on the argument for (i) by replacing the distribution with its empirical distribution. □

The perspective of treating the variance and the mean of a random variable as the optimal value and the optimal solution of an expectation function is dated back to early analysis, e.g., Huber (1964). Proposition 1 suggests that analyzing the approximation of Problem (4) to Problem (2) is equivalent to analyzing the approximation of Problem (7) to Problem (5). Problem (5) is a standard stochastic optimization problem. The more interesting fact is that Problem (7) is exactly the sample average approximation (SAA) of Problem (5). This bridge enables us to use the stochastic optimization theory to build the statistical properties of the sample mean-variance approximation.

When $H(x,\xi)$ is a linear function, which is the focus of this paper, we can see that the objective function in Problem (5) is jointly convex in (x,t). For the more general nonlinear function, the convexity may not be guaranteed even when $H(x,\xi)$ is convex in x. A partial reason is that the quadratic function is not monotone. However, for the general function, even with the absence of convexity, the reformulation allows us to solve the mean-variance model and conduct analysis by using stochastic optimization techniques. In this paper, we focus on the statistical properties for the linear portfolio.

Note that Σ is positive definite. Problem (2) has a unique optimal solution. Moreover, $\mathrm{E}\left[(H(x,\xi)-t)^2\right]$ is strictly convex in t. Therefore, Problem (5) has a unique optimal solution. We denote it as (x^*,t^*). We make the following assumptions.

Assumption 1 There exists a point $x \in X$ such that $\mathrm{E}[H(x,\xi)] > r$.

Assumption 1 is a Slater condition, which is a standard constraint qualification in optimization.

Assumption 2 There exists a random function $M(\xi)$ with $\mathrm{E}[M(\xi)] < \infty$ such that

$$|H(x,\xi)| \le M(\xi), \forall\, x \in X,\ a.e.\ \xi \in \Xi.$$

Furthermore, there exists a random function $K(\xi)$ with $\mathrm{E}\left[K(\xi)^2\right] < \infty$ such that

$$|H(x_1,\xi) - H(x_2,\xi)| \le K(\xi)\ \|x_1 - x_2\|, \forall\, x_1, x_2 \in X,\ a.e.\ \xi \in \Xi.$$

Assumption 2 requires that $H(x,\xi)$ is dominated by an integrable random variable and satisfies the Lipschitz condition. It is a standard assumption in stochastic optimization literature, see, for instance, Shapiro et al. (2014). It is also a critical assumption for the well-definedness and the differentiability of the expectations that are considered below.

In both Problem (5) and Problem (7), $t \in \Re$. In this paper, we assume that we can specify a compact interval T such that for the optimal solution (x,t) of Problem (5), t is an interior point of T. To analyze the asymptotic properties, we replace $t \in \Re$ with $t \in T$ in Problem (5) and Problem (7). We also make a parallel assumption on $(H(x,\xi)-t)^2$.

Assumption 3 There exists a random function $\tilde{M}(\xi)$ with $\mathrm{E}\left[\tilde{M}(\xi)\right] < \infty$ such that

$$|\,(H(x,\xi)-t)^2\,| \leq \tilde{M}(\xi), \forall\ (x,t) \in X \times T,\ a.e.\ \xi \in \Xi.$$

Furthermore, there exists a random function $\tilde{K}(\xi)$ with $\mathrm{E}\left[\tilde{K}(\xi)^2\right] < \infty$ such that

$$\left|(H(x_1,\xi)-t_1)^2 - (H(x_2,\xi)-t_2)^2\right| \leq \tilde{K}(\xi)\ \|(x_1,t_1)-(x_2,t_2)\|, \forall\ (x_1,t_1),(x_2,t_2) \in X \times T,\ a.e.\ \xi \in \Xi.$$

Recall that (x^*,t^*) is the unique optimal solution of Problem (5). We write the constraint (6) as $\mathrm{E}[-H(x,\xi)+r] \leq 0$ and associate the constraint with a Lagrangian multiplier λ. Let Λ denote the set of optimal multipliers associated with (x^*,t^*). Let (x_n,t_n) denote an optimal solution of Problem (7). Define the random variable

$$Y(x,t,\lambda) = \lim_{n\to\infty} \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\left[(H(x,\xi_i)-t)^2 - \lambda H(x,\xi_i)\right] - \mathrm{E}\left[(H(x,\xi)-t)^2 - \lambda H(x,\xi)\right]\right].$$

By the central limit theorem (CLT), $Y(x,t,\lambda)$ has a normal distribution $N(0,\sigma^2(x,t,\lambda))$ with $\sigma^2(x,t,\lambda) = \mathrm{V}\left[(H(x,\xi)-t)^2 - \lambda H(x,\xi)\right]$. We build the following theorem.

Theorem 1 Suppose that Assumptions 1 and 2 are satisfied. Then $v_n \to v^*$ with probability one (w.p.1), and $(x_n,t_n) \to (x^*,t^*)$ w.p.1. Suppose that Assumptions 1, 2 and 3 are satisfied. Then

$$\sqrt{n}(v_n - v^*) \Rightarrow \sup_{\lambda\in\Lambda}\ [Y(x^*,t^*,\lambda)],$$

where "$\Rightarrow$" denotes the convergence in distribution. If, moreover, $\Lambda = \{\lambda^*\}$ is a singleton, then

$$\sqrt{n}(v_n - v^*) \Rightarrow N(0,\sigma^2),$$

where

$$\sigma^2 = \mathrm{V}\left[(H(x^*,\xi) - \mathrm{E}[H(x^*,\xi)])^2 - \lambda^* H(x^*,\xi)\right]. \tag{9}$$

Proof. As discussed we can impose a compact interval T and use $t \in T$ to replace $t \in \Re$, and consider the compact set $X \times T$. For the i.i.d. sample, the pointwise law of large numbers holds for $(H(x,\xi)-t)^2$ and $H(x,\xi)$. Because $(H(x,\xi)-t)^2$ and $H(x,\xi)$ are convex in (x,t), by Theorem 7.55 of Shapiro et al. (2014), the uniform law of large numbers in $X \times T$ also holds for them. Then, the assumptions in Theorem 5.3 of Shapiro et al. (2014) are satisfied. Following the analysis on Page 181 of Shapiro et al. (2014), condition (a) in Theorem 5.5 of Shapiro et al. (2014) holds. Because Assumption 1 is satisfied, following the analysis on Page 182 of Shapiro et al. (2014), condition (b) in Theorem 5.5 of Shapiro et al. (2014) also holds. It follows from Theorem 5.5 of Shapiro et al. (2014) that $v_n \to v^*$ w.p.1, and $(x_n,t_n) \to (x^*,t^*)$ w.p.1.

Note that the set of optimal solutions $S = \{(x^*,t^*)\}$ is a singleton, and $t^* = \mathrm{E}[H(x^*,\xi)]$. Suppose that Assumptions 1, 2 and 3 are satisfied. It follows from Theorem 5.11 of Shapiro et al. (2014) that the remain asymptotic results hold. This concludes the proof of the theorem. □

An important observation from the theorem is that the variability of the sample mean-variance model depends on the first four moments of the random returns ξ. When the distribution of ξ is sufficiently light-tailed, Assumption 3 typically holds. It can ensure the existence of the asymptotic variance σ^2 in (9). However, when the distribution tail of ξ becomes heavier, Assumption 3 may be violated and σ^2 may blow up. In the numerical experiments, we will conduct SAA for the t-distribution.

The optimal solution of Problem (1) depends on X. In practice, the decision makers may impose different constraints on the portfolio weights. For some X, it is possible to derive analytical solution for Problem (1). We recall some classical results of the mean-variance model. Consider $X = \left\{x : \mathbf{1}^{\mathrm{T}} x = 1\right\}$ where $\mathbf{1}$ is the all one vector. Note that this set is not compact. But we can impose very loose bounds on x to make the set compact. We consider solving Problem (2) based on the Lagrangian approach. Let λ and ν denote the Lagrangian multipliers associated with the constraints $\mu^{\mathrm{T}} x \geq r$ and $\mathbf{1}^{\mathrm{T}} x = 1$, respectively. Let λ^* and ν^* denote the optimal multipliers corresponding to the optimal solution x^*. Consider the following two cases. The results can be found in Burke (2020).

Case 1: $\mu^{\mathrm{T}} x^* = r$. In this case, we exclude the setting where μ and $\mathbf{1}$ are linear dependent. Consider that μ and $\mathbf{1}$ are linear independent. Then it can verified that

$$\delta := \left(\mu^{\mathrm{T}}\Sigma^{-1}\mu\right)\left(\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1}\right) - \left(\mu^{\mathrm{T}}\Sigma^{-1}\mathbf{1}\right)^2 > 0.$$

The optimal solution is

$$x^* = (1-\alpha)\frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1}} + \alpha\frac{\Sigma^{-1}\mu}{\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mu},$$

where

$$\alpha = \delta^{-1}\left[r\left(\mu^{\mathrm{T}}\Sigma^{-1}\mathbf{1}\right)\left(\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1}\right) - \left(\mu^{\mathrm{T}}\Sigma^{-1}\mathbf{1}\right)^2\right],$$

and the corresponding multipliers are

$$\lambda^* = \delta^{-1}\mathbf{1}^{\mathrm{T}}\Sigma^{-1}(r\mathbf{1}-\mu), \quad \nu^* = -\delta^{-1}\mu^{\mathrm{T}}\Sigma^{-1}(r\mathbf{1}-\mu).$$

For this case, we can obtain that the Lagrangian multiplier is unique. Thus, the asymptotic normality in Theorem 1 holds.

Case 2: $\mu^{\mathrm{T}} x^* > r$. In this case, $\lambda^* = 0$, $\nu^* = 1/\left(\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1}\right)$, $x^* = \nu^*\Sigma^{-1}\mathbf{1}$. Then the uniqueness of the Lagrangian multiplier is also guaranteed and the asymptotic normality holds. Especially, we have that the asymptotic variance in (9) becomes $\sigma^2 = \mathrm{V}\left[\left(H(x^*,\xi) - \mathrm{E}\left[H(x^*,\xi)\right]\right)^2\right]$ due to $\lambda^* = 0$.

2.2 Setting Extensions

We now discuss more about the asymptotic properties when the modeling settings are slightly different. We first consider an alternative estimator for the covariance matrix. In the sample approximation discussed above, we used the biased estimator $\hat{\Sigma}$. This estimator exactly matches the SAA formulation. In statistics, the unbiased estimator $\hat{\Sigma}_u$ provided above may sometimes be preferred. If we use the unbiased estimator $\hat{\Sigma}_u$ and obtain the following approximation

$$\begin{aligned} \underset{x\in X}{\text{minimize}} \quad & x^{\mathrm{T}}\hat{\Sigma}_u x \\ \text{subject to} \quad & \hat{\mu}^{\mathrm{T}} x \geq r, \end{aligned} \tag{10}$$

we can still derive the asymptotic properties for the optimal value of the approximation. To see this, note that the optimal value of Problem (10) is $\frac{n}{n-1}v_n$ where v_n is the optimal value of Problem (4). It can be verified that $\sqrt{n}\left(\frac{n}{n-1}v_n - v^*\right)$ has the same asymptotic distribution as $\sqrt{n}\left(v_n - v^*\right)$.

Next we consider the regularization in the mean-variance model. Because there exist sample errors, the performance of the sample optimal portfolio may not be satisfactory. Many studies proposed to impose some regularization for the portfolio weights to enhance the out-of-sample performance for the mean-variance model. A first class of models with regularization takes the following form

$$\begin{aligned} \underset{x\in X}{\text{minimize}} \quad & x^{\mathrm{T}}\Sigma x+\rho(x) \\ \text{subject to} \quad & \mu^{\mathrm{T}}x\geq r. \end{aligned} \tag{11}$$

In Problem (11), $\rho(x)$ is a convex function of x. It is often specified as the L_1 or L_2 norm of x, see, e.g., Brodie et al. (2009). A second type of regularization proposes to add constraints for the weights and consider the following problem

$$\begin{aligned} \underset{x\in X}{\text{minimize}} \quad & x^{\mathrm{T}}\Sigma x \\ \text{subject to} \quad & \mu^{\mathrm{T}}x\geq r,\ \rho(x)\leq\rho_0, \end{aligned} \tag{12}$$

where $\rho(x)$ is also a convex function. For instance, DeMiguel et al. (2009a) imposed norm constraint on x to build the portfolio optimization model. By using the stochastic optimization approach discussed above, we can show that the consistency and the asymptotic normality hold for the sample approximations of Problems (11) and (12). The asymptotic variance has the same expression as (9). The only difference is the optimal solution x^*. This observation suggests that imposing regularization on the portfolio weights will not affect the convergence rate of the sample approximation, but it may affect the optimal solution and thus the asymptotic variance.

3 ALTERNATIVE MEAN-VARIANCE MODELS

There are alternative expressions for the mean-variance model. We now consider a formulation that optimizes the weighted sum of the mean and the variance

$$\underset{x\in X}{\text{minimize}} \quad \mathrm{V}\left[H(x,\xi)\right]-w\mathrm{E}\left[H(x,\xi)\right]. \tag{13}$$

In Problem (13), $w\geq 0$ is a prespecified parameter that adjusts the weights of the return and the risk. When $w=0$, we obtain a minimum-variance problem, which is widely studied in the literature, see, e.g., Jagannathan and Ma (2003). For the linear portfolio, the problem can be rewritten as

$$\underset{x\in X}{\text{minimize}} \quad x^{\mathrm{T}}\Sigma x-w\mu^{\mathrm{T}}x. \tag{14}$$

This model is relatively easier to analyze because it does not involve the constraint. By plugging the estimators in (3) into Problem (14), we can obtain the following sample approximation problem

$$\underset{x\in X}{\text{minimize}} \quad x^{\mathrm{T}}\hat{\Sigma}x-w\hat{\mu}^{\mathrm{T}}x. \tag{15}$$

Let v_n and v^* denote the optimal values of Problems (15) and (14) respectively. We analyze the statistical properties of the sample approximation. Similarly as in the previous section, Problem (14) is equivalent to the following stochastic optimization problem

$$\underset{x\in X,t\in\Re}{\text{minimize}} \quad \mathrm{E}\left[(H(x,\xi)-t)^2-wH(x,\xi)\right]. \tag{16}$$

Problem (15) is equivalent to the following SAA of Problem (16)

$$\underset{x\in X,t\in\Re}{\text{minimize}} \quad \frac{1}{n}\sum_{i=1}^{n}\left[(H(x,\xi_i)-t)^2-wH(x,\xi_i)\right]. \tag{17}$$

Similarly, we specify a compact interval T and use $t\in T$ to replace $t\in\Re$. Problem (16) has a unique optimal solution (x^*,t^*). Let (x_n,t_n) denote an optimal solution of Problem (17). We make the following assumptions.

Assumption 4 For some $(x,t) \in X \times T$, $\mathrm{E}\left[\left[(H(x,\xi)-t)^2 - wH(x,\xi)\right]^2\right] < +\infty$.

Assumption 5 There exists a random function $\tilde{K}(\xi)$ with $\mathrm{E}\left[\tilde{K}(\xi)^2\right] < \infty$ such that

$$\left|\left[(H(x_1,\xi)-t_1)^2 - wH(x_1,\xi)\right] - \left[(H(x_2,\xi)-t_2)^2 - wH(x_2,\xi)\right]\right| \leq \tilde{K}(\xi)\ \|(x_1,t_1)-(x_2,t_2)\|,$$

$\forall\ (x_1,t_1),(x_2,t_2) \in X \times T,\ a.e.\xi \in \Xi$.

Then we have the following proposition.

Proposition 2 Suppose that Assumptions 4 and 5 are satisfied. Then $v_n \to v^*$ w.p.1, $(x_n,t_n) \to (x^*,t^*)$ w.p.1, and

$$\sqrt{n}\,(v_n - v^*) \Rightarrow N(0,\sigma^2),$$

where

$$\sigma^2 = \mathrm{V}\left[(H(x^*,\xi) - \mathrm{E}[H(x^*,\xi)])^2 - wH(x^*,\xi)\right]. \tag{18}$$

Proof. The consistency follows from Theorem 5.4 of Shapiro et al. (2014). Note that Problem (16) has a unique optimal solution (x^*,t^*) where $t^* = \mathrm{E}[H(x^*,\xi)]$. Then it follows from Shapiro et al. (2014) that the asymptotic normality holds. This concludes the proof of the proposition. □

It is interesting to note that the variances in (18) and (9) share a similar structure. The difference lies in whether to use the prespecified weight or the optimal Lagrangian multiplier.

We now consider a third mean-variance model which takes the following expression

$$\begin{aligned} \underset{x\in X}{\text{minimize}} \quad & -\mu^{\mathrm{T}}x \\ \text{subject to} \quad & x^{\mathrm{T}}\Sigma x \leq \sigma^2, \end{aligned} \tag{19}$$

where σ^2 is some prespecified risk threshold, and consider its sample approximation

$$\begin{aligned} \underset{x\in X}{\text{minimize}} \quad & -\hat{\mu}^{\mathrm{T}}x \\ \text{subject to} \quad & x^{\mathrm{T}}\hat{\Sigma} x \leq \sigma^2. \end{aligned} \tag{20}$$

Problem (19) is equivalent to the following problem

$$\underset{x\in X, t\in\Re}{\text{minimize}} \quad \mathrm{E}[-H(x,\xi)] \tag{21}$$

$$\text{subject to} \quad \mathrm{E}\left[(H(x,\xi)-t)^2\right] \leq \sigma^2. \tag{22}$$

Problem (20) is equivalent to the SAA of Problem (21)

$$\begin{aligned} \underset{x\in X, t\in\Re}{\text{minimize}} \quad & \frac{1}{n}\sum_{i=1}^{n}[-H(x,\xi_i)] \\ \text{subject to} \quad & \frac{1}{n}\sum_{i=1}^{n}\left[(H(x,\xi_i)-t)^2\right] \leq \sigma^2. \end{aligned} \tag{23}$$

As in the analysis above, we specify a compact interval T and use $t \in T$ to replace $t \in \Re$ in Problems (21) and (23). We make the following assumptions.

Assumption 6 There exists a point $x \in X$ such that $x^{\mathrm{T}}\Sigma x < \sigma^2$.

This assumption implies that there exists $(x,t) \in X \times T$ such that $\mathrm{E}\left[(H(x,\xi)-t)^2\right] < \sigma^2$, i.e., the Slater condition holds for constraint (22). Different from the previous two models, the uniqueness of the optimal solution of Problem (21) may not be guaranteed. Let S denote the set of optimal solutions of Problem (21), and Λ denote the set of associated Lagrangian multipliers. Let S_n denote the set of optimal solutions of Problem (23). Let $D(S_n,S)$ denote the deviation of S_n to S. That is, $D(S_n,S) = \sup_{x\in S_n} d(x,S)$ where $d(x,S) = \inf_{y\in S}\|x-y\|$. Define

$$Y(x,t,\lambda) = \lim_{n\to\infty} \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\left[-H(x,\xi_i)+\lambda\,(H(x,\xi_i)-t)^2\right] - \mathrm{E}\left[-H(x,\xi)+\lambda\,(H(x,\xi)-t)^2\right]\right]. \tag{24}$$

By the CLT, $Y(x,t,\lambda)$ has a normal distribution $N(0,\sigma^2(x,t,\lambda))$ with

$$\sigma^2(x,t,\lambda) = \mathrm{V}\left[-H(x,\xi)+\lambda\,(H(x,\xi)-t)^2\right].$$

By Theorem 5.11 of Shapiro et al. (2014), we can obtain the following result.

Proposition 3 Suppose that Assumptions 2, 3, 6 are satisfied. Then $v_n \to v^*$ w.p.1, $D(S_n,S) \to 0$ w.p.1, and

$$\sqrt{n}\,(v_n - v^*) \Rightarrow \inf_{(x,t)\in S}\sup_{\lambda\in\Lambda} Y(x,t,\lambda).$$

If, moreover, $S = \{(x^*,t^*)\}$ and $\Lambda = \{\lambda^*\}$ are singletons, then

$$\sqrt{n}\,(v_n - v^*) \Rightarrow N(0,\sigma^2),$$

where

$$\sigma^2 = \mathrm{V}\left[-H(x^*,\xi)+\lambda^*\,(H(x^*,\xi)-\mathrm{E}[H(x^*,\xi)])^2\right]. \tag{25}$$

The proof of Proposition 3 is similar to that of Theorem 1 and thus is omitted. The asymptotic variance in (25) also depends on the first four moments of random vector ξ. If the constraint (22) is not binding at (x^*,t^*), i.e., $\mathrm{E}\left[(H(x^*,\xi)-t^*)^2\right] < \sigma^2$, then the optimal multiplier $\lambda^* = 0$ by the complementary slackness (Boyd and Vandenberghe 2004). In this case, the asymptotic variance $\sigma^2 = \mathrm{V}[-H(x^*,\xi)]$ which only depends on the first two moments of the random returns. An intuitive interpretation of this result is that adding or removing the risk constraint does not affect the optimal solution of the portfolio optimization. The sample approximation in this case may tend to have a smaller variation.

4 NUMERICAL ILLUSTRATIONS

In this section, we conduct some numerical experiments to justify the theoretical results built in the paper. For illustrative purpose, we only solve the first mean-variance model. The experiments for the other models can be conducted accordingly. We consider Problem (1) with the set $X = \left\{x : \mathbf{1}^{\mathrm{T}}x = 1\right\}$ and the number of assets $k = 50$. We first assume that ξ follows a multivariate normal distribution $N(\mu,\Sigma)$. For the parameters of $N(\mu,\Sigma)$, we adopt a configuration in Hong et al. (2014). The elements of $\mu = (\mu_1,\cdots,\mu_k)^{\mathrm{T}}$ evenly spread between 0.04 and 0.50 and increase with the subscript. The standard deviation $\mathrm{std}[\xi_i]$ is equal to $\mu_i + 0.05$ for $i = 1,\cdots,k$ and the correlation between any two elements ξ^i and ξ^j of ξ is 0.35 where $i \neq j$. The true optimal solution and optimal value can be derived analytically based on the discussion in Section 2. They are used as the benchmark in the experiments. The sample approximation Problem (4) is solved by the CVX package (Grant and Boyd 2020).

In Table 1, we report the optimal value of Problem (1) and five typical replications of the sample optimal value for different sample sizes. We can see that the sample optimal value gradually tends to the true value as n increases. For small sample sizes, the sample errors are relatively large. Especially, the

Table 1: Mean-variance optimization with normal distribution.

	$n=10^2$	$n=10^3$	$n=10^4$	$n=10^5$
True Opt	0.0328	0.0328	0.0328	0.0328
1	0.0112	0.0279	0.0326	0.0329
2	0.0150	0.0282	0.0323	0.0329
3	0.0116	0.0325	0.0334	0.0327
4	0.0105	0.0313	0.0325	0.0330
5	0.0121	0.0302	0.0342	0.0328

results exhibit a negative bias to the true value. This is consistent to the fact that the SAA minimization yields negative bias.

We further test the convergence speed of the sample optimal value. Let $S_n = v_n - v^*$ and $S = \sqrt{n}(v_n - v^*)$. We simulate S_n and S for 1000 times for different n, and plot the densities of S_n and S in Figure 1. The left panel of the figure shows the density and trend of S_n. When n becomes larger, S_n gradually concentrates around 0. The right panel of Figure 1 shows the density of S. The density appears to admit an asymptotic normality behavior when n is sufficiently large. These observations support the result built in Theorem 1.

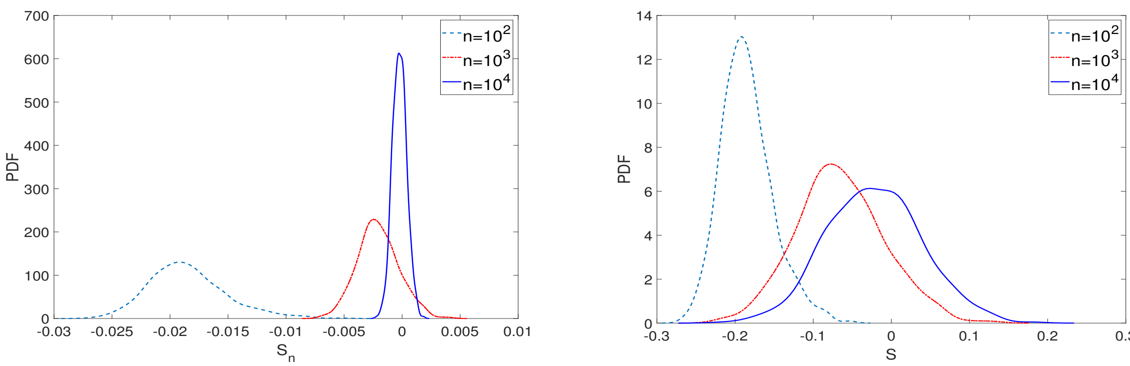

Figure 1: Performance of sample approximation with normal distribution.

We also test the effects of the regularization. Consider Problem (12) with the set $X = \{x : x \geq 0, \mathbf{1}^{\mathrm{T}} x \leq 1\}$, and set $\rho(x) = \|x\|_1$ where $\|\cdot\|_1$ denotes the L_1 norm. The minimum of $\|x\|_1$ over all feasible x is 0.5094. We consider $\rho_0 = 0.55, 0.80$, and solve Problem (12) to obtain v^* for the two settings. We simulate S defined above for 1000 times with $n = 10^4$ and plot the densities of S in Figure 2. From the figure, we observe that with regularization, the sample optimal value still admit an asymptotic normality behavior.

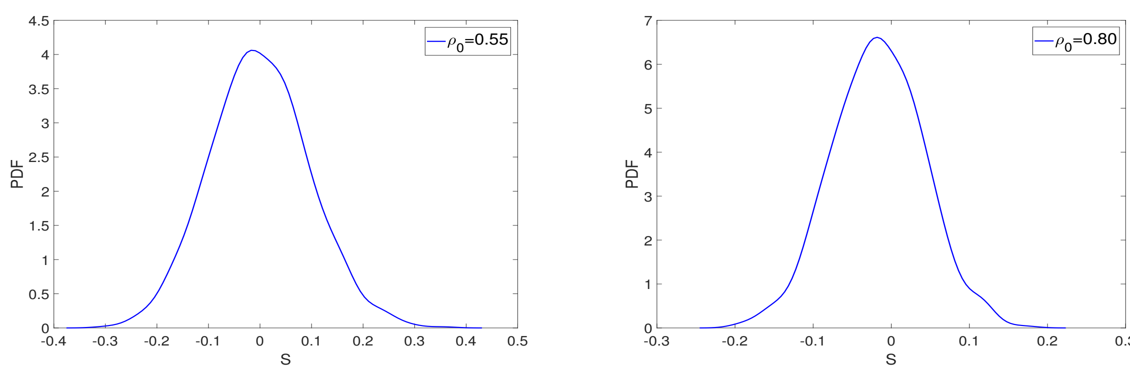

Figure 2: Performance of sample approximation with normal distribution with regularization.

Now we consider a distribution with a heavier tail. Let $R = \text{Chol}(\Sigma)$ denote the Cholesky decomposition of Σ. Then $\Sigma = R^{\mathrm{T}}R$. We assume that $\xi = R^{\mathrm{T}}\left(\sqrt{\frac{\nu-2}{\nu}}\eta\right) + \mu$ where η is a k-dimensional random vector with each element following a t distribution with degree of freedom ν and all elements being independent. We can compute that the mean vector and the covariance matrix of ξ are still μ and Σ provided that $\nu > 2$. Therefore, the true optimization problem remains the same. Recall that for a t distribution with degree of freedom ν, if $m > \nu$, the m-th moment of the t distribution does not exist. In the experiment, we consider $\nu = 5$ and $\nu = 3$. For $\nu = 5$, the fourth moment of η exists. However, when $\nu = 3$, the fourth moment of η does not exist and in this setting the asymptotic variance in (9) may not exist. We conduct experiments and examine it. The other settings are the same as above.

Table 2 summarizes five typical simulation replications for the sample optimal values for the two degrees of freedom. The results essentially shows a consistency behavior. Comparing the two parts of the table, we observe that the results for $\nu = 3$ appear to have relatively larger variability than that for $\nu = 5$ when $n = 10^5$.

Table 2: Mean-variance optimization with t distribution.

$\nu = 5$	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = 10^5$	$\nu = 3$	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = 10^5$
True Opt	0.0328	0.0328	0.0328	0.0328	True Opt	0.0328	0.0328	0.0328	0.0328
1	0.0148	0.0336	0.0329	0.0326	1	0.0103	0.0297	0.0318	0.0319
2	0.0125	0.0299	0.0326	0.0329	2	0.0107	0.0348	0.0322	0.0317
3	0.0151	0.0338	0.0315	0.0332	3	0.0136	0.0284	0.0321	0.0324
4	0.0132	0.0299	0.0328	0.0327	4	0.0098	0.0279	0.0324	0.0322
5	0.0090	0.0289	0.0328	0.0325	5	0.0199	0.0298	0.0319	0.0317

Finally, we test the convergence speed of the sample approximation. We set $n = 10^4$ and simulate S for 1000 times. Based on the simulation we plot the density of S, which are shown in Figure 3 (left panel for $\nu = 5$ and right panel for $\nu = 3$). From the figure, we can see that when $\nu = 5$, S admits an asymptotic normality behavior. However, when $\nu = 3$, the density exhibits non-symmetry and has a long right tail. It appears to have quite different behavior compared to the normal distribution. We will further explore the properties in the future study.

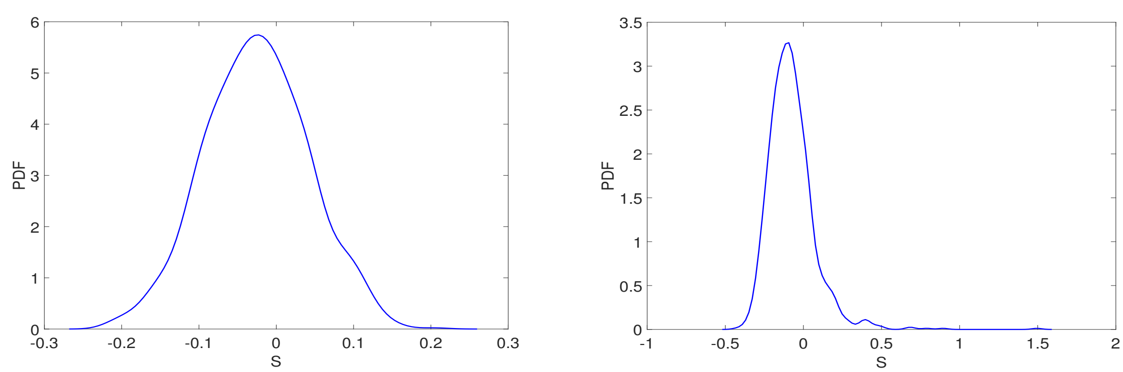

Figure 3: Performance of sample approximation with t distribution.

5 CONCLUSION

In this paper, we have investigated the statistical properties for various mean-variance portfolio optimization models when using sample approximations. We have made an attempt to provide some theoretical analysis to the robustness and variability of the portfolio optimization model. Essentially, our investigation shows that the sample approximation of the mean-variance model typically has a convergence rate of the square root of the sample size, which is the conventional convergence rate of the Monte Carlo simulation, and the asymptotic variance depends on up to the fourth moment. The results shed some light on why the empirical performance of the mean-variance might exhibit high variation. They may also help quantify the sample error and build confidence region for the portfolio performance in the data driven environment.

ACKNOWLEDGEMENTS

This work was partially supported by the National Natural Science Foundation of China [No. 72471177].

REFERENCES

Ban, G.-Y., N. El Karoui, and A. E. B. Lim. 2018. "Machine Learning and Portfolio Optimization". *Management Science*, 64(3):1136-1154.

Britten-Jones, M. 1999. "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights". *The Journal of Finance*, 54(2):655-671.

Burke, J. 2020. Markowitz Mean-Variance Portfolio Theory. http://sites.math.washington.edu/~burke/crs/408/notes/fin/mark1.pdf, accessed February 2025.

Boyd, S. and L. Vandenberghe. 2004. *Convex Optimization*. Cambridge: Cambridge University Press.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris. 2009. "Sparse and Stable Markowitz Portfolios". *Proceedings of the National Academy of Sciences*, 106(30):12267-12272.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal. 2009a. "A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms". *Management Science*, 55(5):798-812.

DeMiguel, V., L. Garlappi, and R. Uppal. 2009b. "Optimal versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?". *The Review of Financial Studies*, 22(5):1915-1953.

Grant, M. and S. Boyd. 2020. CVX: Matlab Software for Disciplined Convex Programming. https://cvxr.com/cvx, accessed March 2025.

Hong, L. J., Z. Hu, and L. Zhang. 2014. "Conditional Value-at-Risk Approximation to Value-at-Risk Constrained Programs: A Remedy via Monte Carlo". *INFORMS Journal on Computing*, 26(2):385-400.

Hu, Z. and D. Zhang. 2018. "Utility-based Shortfall Risk: Efficient Computations via Monte Carlo". *Naval Research Logistics*, 65(5):378-392.

Huber, P. J. 1964. "Robust Estimation of A Location Parameter". *The Annals of Mathematical Statistics*, 35(1):73-101.

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

Lim A. E. B., J. G. Shanthikumar, and G. Y. Vahn. 2011. "Conditional Value at Risk in Portfolio Optimization: Coherent but Fragile". *Operations Research Letters*, 39(3):163-171.

Markowitz H. 1952. "Portfolio Selection". *The Journal of Finance*, 7(1):77-91.

Shapiro, A., D. Dentcheva, and A. Ruszczyński. 2014. *Lectures on Stochastic Programming: Modeling and Theory*, 2nd ed. Philadelphia: SIAM.

Wang, S., G. Cai, P. Yu, G. Liu, and J. Luo. 2023. "Mean-Variance Portfolio Optimization with Nonlinear Derivative Securities". In *2023 Winter Simulation Conference (WSC)*, 576-587 https://doi.org/10.1109/WSC60868.2023.10407819.

AUTHOR BIOGRAPHIES

ZHAOLIN HU is a professor in the School of Economics and Management at Tongji University. His research interests include stochastic optimization, simulation theory and practice, machine learning, and risk management. He is currently an associate editor of *Journal of Management Science and Engineering* and *Journal of the Operations Research Society of China*. His email address is russell@tongji.edu.cn.