The Distribution of the Sample Minimum-Variance Frontier

RAYMOND KAN and DANIEL R. SMITH*

First draft: August 2005
This version: September 2007

*Kan is from the University of Toronto and Smith is from Simon Fraser University. We thank Bob Korkie, Sergei Sarkissian, Wolfgang Schmid, Tim Simin, Chu Zhang and seminar participants at Bank of Canada, Federal Reserve Bank of Atlanta, Simon Fraser University, University of Toronto and participants at the 18th annual Australasian Banking and Finance Conference for helpful discussions and comments. Kan gratefully acknowledges financial support from the National Bank Financial of Canada.

The Distribution of the Sample Minimum-Variance Frontier

Abstract

In this paper, we present a finite sample analysis of the sample minimum-variance frontier under the assumption that the returns are independent and multivariate normally distributed. We show that the sample minimum-variance frontier is a highly biased estimator of the population frontier and we propose an improved estimator of the population frontier. In addition, we provide the exact distribution of the out-of-sample mean and variance of sample minimum-variance portfolios. This allows us to understand the impact of estimation error on the performance of in-sample optimal portfolios

Keywords: Minimum-variance frontier; Efficiency set constants; Finite sample distribution.

1. Introduction

The minimum-variance frontier plays a central role in much of finance. For example, it is the cornerstone of modern portfolio theory and many asset pricing models. In addition, it is closely linked to the Hansen-Jagannathan bound on the variance of admissible stochastic discount factors (Hansen and Jagannathan (1991)). It is a common practice in the finance literature to trace out the minimum-variance frontier using sample estimates of the mean and covariance matrix of returns. However, a sample frontier is subject to estimation error, so it is important to understand its finite sample distribution.

While the sample frontiers are widely calculated and interpreted, their finite sample properties are virtually unknown. Dickinson (1974) shows that the sample variance of the sample global minimum-variance portfolio has a chi-squared distribution.[1] For the two assets case, he also provides the exact distribution of the weights of the sample global minimum-variance portfolio. Jobson and Korkie (1980) provide the exact mean and variance of the three sample efficiency set constants that determine the sample minimum-variance frontier. In addition, they provide approximation formulas for the mean and variance of the weights of the sample tangency portfolio. Jobson (1991) provides the exact distribution for two out of the three random variables that are crucial in determining the location of the sample minimum-variance frontier. Okhrin and Schmid (2006) derive the exact mean and variance of the weights of the sample optimal portfolio that maximizes a given expected quadratic utility function. For the special case of sample global minimum-variance portfolio, they also provide the exact distribution of its weights.

We contribute to this literature by deriving the finite sample distribution and moments of the sample minimum-variance frontier when returns are independent multivariate normal random variables. We find that the sample minimum-variance frontier is a heavily biased estimator of the population frontier even when the length of the estimation window is very long. To correct for this bias, we propose a new adjusted estimator of the population frontier that has a significantly smaller bias than the traditional sample estimator.

For many investors, inference on the population minimum-variance frontier is of little

[1]Dickinson's result contains an error in the degrees of freedom.

interest when they cannot hold the portfolios on the population frontier because the mean and covariance matrix of the returns are unknown. The classical approach to mean-variance portfolio selection (Markowitz (1952)) involves estimating the sample mean and covariance matrix of returns and then treating these as their population values when selecting the optimal portfolio. However, using sample estimates of the means and covariances results in estimation error of the optimal portfolio weights. To an investor who holds an optimized portfolio that is based on sample estimates, he is most concerned with its out-of-sample performance.

Researchers have been questioning the out-of-sample performance of optimal portfolios formed using sample estimates for quite some time. Using a simulation experiment of three stocks, Frankfurter, Phillips and Seagle (1971) find that grossly inefficient portfolios in terms of true parameters may appear to be efficient in-sample for a large proportion of the time. Dickinson (1974) finds that the estimates of the weights and variance of the global minimum-variance portfolio are highly unreliable. Using a simulation experiment of twenty stocks, Jobson and Korkie (1981) conclude that with a typical length of estimation window, the sample mean and variance of the sample tangency portfolio is a very poor estimator of its out-of-sample performance. They find that even a naïve equally weighted portfolio can significantly outperform the sample tangency portfolio. In a simulation experiment using five assets, Broadie (1993) demonstrates that sample frontiers are overly optimistic while the out-of-sample frontiers are inferior to both the true and sample frontiers. Michaud (1989) discusses various problems of using the Markowitz's method of selecting optimal portfolio.

As a result, researchers have started to incorporate estimation risk when solving the portfolio choice problem. Broadly speaking, there are three different approaches that are used to improve the out-of-sample performance of an optimal portfolio. The first approach is to use better statistical estimators of the mean and covariance matrix to form optimal portfolios. For example, Jorion (1986) proposes the use of a shrinkage estimator of the mean, and Ledoit and Wolf (2003) propose the use of a shrinkage estimator of the covariance matrix.[2] The second approach is to impose some structure on the data generating process

[2]Jagannathan and Ma (2003) show that imposing no short-selling constraints on the portfolio is equivalent to using a particular shrinkage estimator of the covariance matrix.

to reduce the number of parameters to be estimated. For example, Sharpe (1963) suggests the use of a single factor model, and Jorion (1985) suggests the use of the capital asset pricing model to improve the estimates of mean and covariance matrix. The third approach explicitly incorporates uncertainty in choosing optimal portfolios. For example, Barry (1974) and Brown (1976) incorporate estimation risk in the problem with Bayesian procedure (see Bawa, Brown, and Klein (1979) for an extensive survey of the early work). Garlappi, Uppal, and Wang (2007) propose optimal portfolio rules that take into account both parameter and model uncertainty for an investor that exhibits uncertainty aversion.

All three approaches result in portfolio rules that appear to have better out-of-sample performance than optimal portfolios formed using the sample mean and covariance matrix. However, the evidence is largely based on simulation or historical data, so it is not entirely clear what is the real reason for the improvement and what are the conditions for the dominance. As a result, it is important to understand analytically what determines the out-of-sample performance of a sample minimum-variance portfolio. The out-of-sample performance of a sample minimum-variance portfolio has received quite a bit of attention lately. TerHorst, DeRoon, and Werkerzx (2002), Mori (2004), and Kan and Zhou (2007) study the average out-of-sample performance (or risk function) associated with holding the sample tangency and global minimum-variance portfolios. They also propose superior portfolio rules that dominate the sample tangency portfolio. Two recent papers that are closely related to our study are Basak, Jagannathan, and Ma (2005) and Siegel and Woodgate (2007). Basak, Jagannathan, and Ma (2005) show that the in-sample minimum-variance frontier is overly optimistic and, in particular, the sample variance can dramatically understate the actual variance of holding such a portfolio out-of-sample. They present a Jackknife estimator for the out-of-sample variance which corrects some of this in-sample optimism and illustrate the economic significance of ignoring the overoptimism. Siegel and Woodgate (2007) make the same point as Basak, Jagannathan, and Ma (2005) and further point out that the in-sample mean of a sample minimum-variance portfolio can also be a highly biased estimator of its out-of-sample mean. In particular, the out-of-sample expected return is shrunk toward the expected return of the global minimum-variance portfolio. In addition, they provide approximate formulas to correct for this overoptimism in the sample mean and variance of the sample minimum-variance portfolio. We extend the results of Basak, Jagannathan, and Ma (2005) and Siegel and Woodgate (2007) by presenting the exact distribution of the out-of-sample mean and variance of a sample minimum-variance portfolio. In addition, we present unbiased forecasts of the expected out-of-sample mean and variance of optimal portfolios based on sample estimates. Our results suggest that for estimation windows normally encountered in practice, estimation error overwhelms the benefits of optimization, and the sample tangency portfolio does not perform much better than the sample global minimum-variance portfolio.[3]

The remainder of the paper is organized as follows. Section 2 presents the joint distribution of three sample efficiency set constants that are the key elements for constructing the sample minimum-variance frontier. Section 3 discusses the distribution of the sample minimum-variance frontier and presents an improved estimator of the population minimum-variance frontier. Section 4 discusses the distribution of the out-of-sample performance of a sample minimum-variance portfolio. It also suggests unbiased estimators of the average out-of-sample mean and variance of such a portfolio. Section 5 concludes the paper. Proofs and some additional materials are in the appendix.

2. Analysis of sample efficiency set constants

In order for us to study the distribution of the sample minimum-variance frontier, we need to first study the joint distribution of three random variables that together determine the sample minimum-variance frontier. Let R_t be the vector of returns on $N \geq 2$ risky assets at time t for $t = 1, \ldots, T$ where T denotes the length of the time series. Depending on the context, R_t can be gross returns, net returns, or excess returns (in excess of either the risk-free or some other rate). For finite sample inference, we assume R_t is independent and identically distributed (i.i.d.) as multivariate normal with mean μ and covariance matrix V. In addition, we assume μ is not proportional to 1_N and V is nonsingular. To address the issue of the finite

[3] Jobson, Korkie, and Ratti (1979) suggest investing in only the sample global minimum-variance portfolio. Haugen and Baker (1991) and Clarke, de Silva and Thorley (2006) find that sample global minimum-variance portfolios dominate the value-weighted market portfolio. Jorion (1986) suggests using a shrinkage estimator of expected return, which effectively suggests moving the optimal portfolio from the sample tangency portfolio to the sample global minimum-variance portfolio. Kan and Zhou (2007) suggest a portfolio rule that optimally combines the risk-free asset, the sample tangency portfolio, and the sample global minimum-variance portfolio to maximize the average out-of-sample performance.

sample distribution of the minimum-variance frontier, we must make an assumption about the distribution of returns. Although it is well known that daily stock returns are generally nonnormal and heterogeneous, the evidence of predictability and nonnormality in monthly returns is much weaker. Since we mainly deal with monthly stock returns in our empirical applications, we treat the i.i.d. normality assumption as a reasonable working approximation. In addition, we make an attempt in the paper to understand the impact of departure from normality on our results by studying the distribution of the minimum-variance frontier when returns follow a multivariate-t distribution. While our finite sample distribution results are not exact when returns are not normally distributed, they generally provide a much better approximation to the actual distribution than asymptotic results would.

For any target expected return μ_p, the minimum-variance portfolio solves the following minimization problem

$$\min_{w \in \Re^N} w'Vw \quad \text{s.t.} \;\; w'\mu = \mu_p \;\text{ and }\; w'1_N = 1, \tag{1}$$

where 1_N denotes an N-vector of ones. The minimum-variance frontier traces out the variance of the minimum-variance portfolios for a range of target expected returns μ_p, and it plays a central role in portfolio theory and asset pricing. As is well known, (see e.g., Merton (1972)) the minimum-variance frontier can be characterized using the following three so-called efficiency set constants:

$$a = \mu' V^{-1} \mu, \qquad b = \mu' V^{-1} 1_N, \qquad c = 1_N' V^{-1} 1_N. \tag{2}$$

For our analysis of the minimum-variance frontier, it is useful to consider a remapping of these three constants to another set of three constants (see also Jobson (1991)):

$$\psi^2 = a - \frac{b^2}{c}, \qquad \mu_g = \frac{b}{c}, \qquad \sigma_g^2 = \frac{1}{c}. \tag{3}$$

It can be easily shown that ψ^2 is the square of the slope of the asymptote to the minimum-variance frontier in the (σ, μ) space, and μ_g and σ_g^2 are the mean and variance of the global minimum-variance portfolio. There exists a one-to-one mapping between $(\psi^2, \mu_g, \sigma_g^2)$ and (a, b, c). Knowing $(\psi^2, \mu_g, \sigma_g^2)$, we can determine (a, b, c) using the following relations:

$$a = \psi^2 + \frac{\mu_g^2}{\sigma_g^2}, \qquad b = \frac{\mu_g}{\sigma_g^2}, \qquad c = \frac{1}{\sigma_g^2}. \tag{4}$$

Suppose we have T observations on R_t, the maximum likelihood estimates of μ and V are given by

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t, \tag{5}$$

$$\hat{V} = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})'. \tag{6}$$

It is well-known that under the i.i.d. normality assumption, $\hat{\mu}$ and $\hat{V}$ are independent of each other and they have the following exact distributions (see, e.g., Muirhead (1982))

$$\hat{\mu} \sim N(\mu, V/T), \tag{7}$$

$$\hat{V} \sim W_N(T-1, V/T), \tag{8}$$

where $W_N(T-1, V/T)$ denotes an N-dimensional Wishart distribution with $T-1$ degrees of freedom and covariance matrix V/T. We assume $T > N$ so that $\hat{V}$ is invertible.

The sample efficiency set constants are defined as[4]

$$\hat{a} = \hat{\mu}'\hat{V}^{-1}\hat{\mu}, \qquad \hat{b} = \hat{\mu}'\hat{V}^{-1}1_N, \qquad \hat{c} = 1_N'\hat{V}^{-1}1_N. \tag{9}$$

We also define $\hat{\psi}^2 = \hat{a} - \hat{b}^2/\hat{c}$, $\hat{\mu}_g = \hat{b}/\hat{c}$, $\hat{\sigma}_g^2 = 1/\hat{c}$ as the sample counterparts of ψ^2, μ_g and σ_g^2.[5] We are interested in the finite sample joint distribution of $(\hat{a}, \hat{b}, \hat{c})$, or equivalently the finite sample joint distribution of $(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2)$. The following Proposition presents the joint distribution of $(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2)$.[6]

Proposition 1 *Let $r \sim (N-1)F_{N-1,T-N+1}(T\psi^2)/(T-N+1)$ where $F_{m,n}(\delta)$ stands for a noncentral F-distribution with m and n degrees of freedom and a noncentrality parameter of*

[4] We can also define the sample efficiency set constants using the unbiased estimator of V (i.e., divided by $T-1$ instead of by T) or $\tilde{V} = T\hat{V}/(T-N-2)$ (such that $E[\tilde{V}^{-1}] = V^{-1}$). It requires only multiplying a different scaling factor in our analysis to deal with these alternative estimators of V.

[5] Note that $\hat{\psi}^2$ can be written as

$$\hat{\psi}^2 = \hat{\mu}'[\hat{V}^{-1} - \hat{V}^{-1}1_N(1_N'\hat{V}^{-1}1_N)^{-1}1_N'\hat{V}^{-1}]\hat{\mu},$$

and it is proportional to the Hotelling's T^2-statistic of testing $H_0: \mu = k1_N$ for some scalar k.

[6] Knight and Satchell (2006) present the finite sample distribution of $\hat{\psi}^2$. Jobson (1991) presents the finite sample distribution of $\hat{\psi}^2$ and $\hat{\sigma}_g^2$, and Bodnar and Schmid (2006a) independently derive the joint distribution of $(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2)$.

δ, $q \sim \chi^2_{T-N}$, and $x \sim N(0,1)$, independent of each other. $\hat{\psi}^2$, $\hat{\mu}_g$, and $\hat{\sigma}_g^2$ can be written as functions of r, q, and x as follows:

$$\hat{\psi}^2 = r, \qquad \hat{\mu}_g = \mu_g + \left(\frac{1+r}{T}\right)^{\frac{1}{2}} \sigma_g x, \qquad \hat{\sigma}_g^2 = \frac{\sigma_g^2 q}{T}. \tag{10}$$

Conditional on $\hat{\psi}^2$, $\hat{\mu}_g \sim N(\mu_g, \sigma_g^2(1+\hat{\psi}^2)/T)$. The joint density function of $(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2)$ is given by

$$f(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2) = f(\hat{\psi}^2, \hat{\mu}_g) f(\hat{\sigma}_g^2) = f(\hat{\psi}^2) f(\hat{\mu}_g|\hat{\psi}^2) f(\hat{\sigma}_g^2). \tag{11}$$

The three sample efficiency set constants are given by $\hat{a} = \hat{\psi}^2 + \hat{\mu}_g^2/\hat{\sigma}_g^2$, $\hat{b} = \hat{\mu}_g/\hat{\sigma}_g^2$, and $\hat{c} = 1/\hat{\sigma}_g^2$.

Proposition 1 suggests that while $\hat{\mu}$ and $\hat{V}$ determine the sample efficiency set constants, one does not need to know μ and V to determine the distribution of the sample efficiency set constants. The joint distribution of $(\hat{a}, \hat{b}, \hat{c})$ is completely determined by the three population efficiency set constants (a, b, c) (together with N and T) or equivalently by $(\psi^2, \mu_g, \sigma_g^2)$. Proposition 1 also shows that $(\hat{\psi}^2, \hat{\mu}_g)$ are independent of $\hat{\sigma}_g^2$.

Besides knowing what are the determining parameters of the distribution of the sample efficiency set constants, Proposition 1 provides us a speedy approach of simulating $(\hat{a}, \hat{b}, \hat{c})$ that has its computation time independent of N and T. Traditionally, researchers need to simulate a $T \times N$ matrix of returns in order to simulate one realization of $(\hat{a}, \hat{b}, \hat{c})$. A faster method is to just simulate $\hat{\mu}$ and $\hat{V}$ using (7) and (8) to generate samples of $(\hat{a}, \hat{b}, \hat{c})$. This is still costly when N is large as the computational time is $O(N^2)$. However, relying on Proposition 1 all we need to do is simulate three independent random variables r, q, and x to obtain a sample of $(\hat{a}, \hat{b}, \hat{c})$. In addition, one does not need to specify μ and V for such a simulation.

Equipped with Proposition 1, we can easily obtain the joint moments of $(\hat{a}, \hat{b}, \hat{c})$. The following Lemma presents the mean and covariance matrix of $(\hat{a}, \hat{b}, \hat{c})$ as well as the unbiased estimators of (a, b, c). With appropriate adjustments of the scaling factor, some of the results in this lemma are available in Jobson and Korkie (1980).

Lemma 1 *The expected values of $\hat{a}$, $\hat{b}$, and $\hat{c}$ are given by*

$$E[\hat{a}] = \frac{N + Ta}{T - N - 2}, \qquad E[\hat{b}] = \frac{Tb}{T - N - 2}, \qquad E[\hat{c}] = \frac{Tc}{T - N - 2}. \tag{12}$$

The unbiased estimators of a, b, and c are given by

$$\hat{a}_u = \frac{(T-N-2)\hat{a}-N}{T}, \qquad \hat{b}_u = \frac{(T-N-2)\hat{b}}{T}, \qquad \hat{c}_u = \frac{(T-N-2)\hat{c}}{T}. \tag{13}$$

The covariance matrix of $\hat{a}$, $\hat{b}$, and $\hat{c}$ is given by

$$\mathrm{Var}[\hat{a}] = \frac{2T^2a^2 + 2(T-2)(N+2Ta)}{(T-N-2)^2(T-N-4)}, \tag{14}$$

$$\mathrm{Var}[\hat{b}] = \frac{T^2\left[ac + \left(\frac{T-2}{T}\right)c + \left(\frac{T-N}{T-N-2}\right)b^2\right]}{(T-N-1)(T-N-2)(T-N-4)}, \tag{15}$$

$$\mathrm{Var}[\hat{c}] = \frac{2T^2c^2}{(T-N-2)^2(T-N-4)}, \tag{16}$$

$$\mathrm{Cov}[\hat{a},\hat{b}] = \frac{2T^2ab + 2(T-2)Tb}{(T-N-2)^2(T-N-4)}, \tag{17}$$

$$\mathrm{Cov}[\hat{a},\hat{c}] = \frac{2T^2\left[ac + \left(\frac{T-2}{T}\right)c + (T-N-2)b^2\right]}{(T-N-1)(T-N-2)^2(T-N-4)}, \tag{18}$$

$$\mathrm{Cov}[\hat{b},\hat{c}] = \frac{2T^2bc}{(T-N-2)^2(T-N-4)}. \tag{19}$$

The joint moments of $\hat{\psi}^2$, $\hat{\mu}_g$ and $\hat{\sigma}_g^2$ are also easy to obtain. The following lemma presents the mean and covariance matrix of $(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2)$ as well as the unbiased estimators of $(\psi^2, \mu_g, \sigma_g^2)$. Some of these results are given in Jobson (1991).

Lemma 2 *The expected values of $\hat{\psi}^2$, $\hat{\mu}_g$, and $\hat{\sigma}_g^2$ are given by*

$$E[\hat{\psi}^2] = \frac{N-1+T\psi^2}{T-N-1}, \qquad E[\hat{\mu}_g] = \mu_g, \qquad E[\hat{\sigma}_g^2] = \frac{(T-N)\sigma_g^2}{T}. \tag{20}$$

The unbiased estimators of ψ^2, μ_g, and σ_g^2 are given by

$$\hat{\psi}_u^2 = \frac{(T-N-1)\hat{\psi}^2 - (N-1)}{T}, \qquad \hat{\mu}_{gu} = \hat{\mu}_g, \qquad \hat{\sigma}_{gu}^2 = \frac{T\hat{\sigma}_g^2}{T-N}. \tag{21}$$

The covariance matrix of $\hat{\psi}^2$, $\hat{\mu}_g$, and $\hat{\sigma}_g^2$ is given by

$$\mathrm{Var}[\hat{\psi}^2] = \frac{2T^2\psi^4 + 2(T-2)(N-1+2T\psi^2)}{(T-N-1)^2(T-N-3)}, \tag{22}$$

$$\mathrm{Var}[\hat{\mu}_g] = \frac{[T(1+\psi^2)-2]\sigma_g^2}{T(T-N-1)}, \tag{23}$$

$$\mathrm{Var}[\hat{\sigma}_g^2] = \frac{2(T-N)\sigma_g^4}{T^2}, \tag{24}$$

$$\mathrm{Cov}[\hat{\psi}^2,\hat{\mu}_g] = \mathrm{Cov}[\hat{\psi}^2,\hat{\sigma}_g^2] = \mathrm{Cov}[\hat{\mu}_g,\hat{\sigma}_g^2] = 0. \tag{25}$$

3. In-sample performance of sample minimum-variance portfolios

3.1 The four minimum-variance frontiers

If μ and V are known, then it is well known that the minimum-variance portfolio of the N risky assets with a targeted expected return of μ_p is given by

$$w = V^{-1}[\mu,\ 1_N]A^{-1}[\mu_p,\ 1]', \tag{26}$$

where

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}. \tag{27}$$

The variance of this optimal portfolio is given by $\sigma_p^2 = w'Vw$. It is easy to show that the relation between σ_p^2 and μ_p is

$$\sigma_p^2 = \frac{a - 2b\mu_p + c\mu_p^2}{ac - b^2}. \tag{28}$$

By varying μ_p in equation (28) we obtain the minimum-variance frontier. Instead of writing σ_p^2 using the efficiency set constants, it is more convenient for us to write σ_p^2 using ψ^2, μ_g and σ_g^2 as

$$\sigma_p^2 = \sigma_g^2 + \frac{(\mu_p - \mu_g)^2}{\psi^2}. \tag{29}$$

We call this frontier the *population minimum-variance frontier*.

When we do not know μ and V, we typically construct the minimum-variance portfolio using the sample moments. The weights of the sample minimum-variance portfolio are given by

$$\hat{w} = \hat{V}^{-1}[\hat{\mu},\ 1_N]\hat{A}^{-1}[\mu_p,\ 1]', \tag{30}$$

where

$$\hat{A} = \begin{bmatrix} \hat{a} & \hat{b} \\ \hat{b} & \hat{c} \end{bmatrix}. \tag{31}$$

For a given value of μ_p, the in-sample variance of the sample minimum-variance portfolio is given by

$$\hat{\sigma}_p^2 = \hat{w}'\hat{V}\hat{w} = \frac{\hat{a} - 2\hat{b}\mu_p + \hat{c}\mu_p^2}{\hat{a}\hat{c} - \hat{b}^2} = \hat{\sigma}_g^2 + \frac{(\mu_p - \hat{\mu}_g)^2}{\hat{\psi}^2}. \tag{32}$$

We call (32) the *sample minimum-variance frontier*. This is typically what researchers plot when they use historical data to estimate the population frontier. However, $\hat{\sigma}_p^2$ in (32) is a random variable as $\hat{\mu}$ and $\hat{V}$ are random. Although as $T \to \infty$, $\hat{\sigma}_p^2 \to \sigma_p^2$, the finite sample distribution of $\hat{\sigma}_p^2$ can differ significantly from σ_p^2 even when T is reasonably large.

Knowing that $\hat{\sigma}_p^2$ is random, we present its finite sample distribution in the next subsection. In addition, we present the expected value of $\hat{\sigma}_p^2$, which is denoted by $\bar{\sigma}_p^2 = E[\hat{\sigma}_p^2]$. The relation between $\bar{\sigma}_p^2$ and μ_p gives us a third frontier that we called the *expected value of the sample minimum-variance frontier*, or the *in-sample frontier* in short. The distribution and the expected value of the sample frontier give us an idea of how far on average the sample frontier is away from the population frontier. This information is useful for researchers to make inference concerning the population frontier. In many empirical tests of asset pricing models as well as investigation of international diversification benefits, making inference on the population frontier is the main task. An implicit assumption of this exercise is we assume that the economic agents know the population frontier and that it is only the econometrician that does not have this information.

In another line of research, we take the stand that the investors, just like the econometrician, do not know the true μ and V but instead only have the sample information $\hat{\mu}$ and $\hat{V}$. Under these circumstances, the population frontier is of little interest to investors as it is unattainable. Instead, they are interested in finding out what is the out-of-sample performance if they hold a sample minimum-variance portfolio $\hat{w}$. Recent studies like Mori (2004), Kan and Zhou (2007), Basak, Jagannathan, and Ma (2005), and Siegel and Woodgate (2007) start to address this issue. Note that for an investor who holds portfolio $\hat{w}$, the out-of-sample mean and variance are given by

$$\tilde{\mu}_p = \hat{w}'\mu, \tag{33}$$

$$\tilde{\sigma}_p^2 = \hat{w}'V\hat{w}. \tag{34}$$

Since $\hat{w}$ is random, the out-of-sample mean and variance of the sample minimum-variance portfolio are also random. An investor who holds $\hat{w}$ for the next period is more interested in $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ rather than their in-sample counterparts. In Section 4, we present the distribution of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$. Denote $\underline{\mu}_p = E[\tilde{\mu}_p]$ and $\underline{\sigma}_p^2 = E[\tilde{\sigma}_p^2]$ as the average out-of-sample mean and variance of the sample minimum-variance portfolio. Investors are interested in finding out the frontier based on $\underline{\mu}_p$ and $\underline{\sigma}_p^2$ because this represents a realistic expectation of the out-of-sample performance of the sample minimum-variance portfolios. We call this fourth frontier, which is constructed using $\underline{\mu}_p$ and $\underline{\sigma}_p^2$, the *out-of-sample frontier*. In Section 4, we provide analytical expressions for this out-of-sample frontier as well as delivering an unbiased forecast of it.

3.2 Distribution of the sample minimum-variance frontier

Using the results in Proposition 1, we can write the distribution of $\hat{\sigma}_p^2$ as function of the three random variables r, q, x. However, for convenience of presentation, we use an alternate characterization. Define

$$\delta = \frac{\mu_p - \mu_g}{\sigma_g}, \tag{35}$$

this allows us to write σ_p^2 as the following function of σ_p^2, ψ^2 and δ^2:

$$\sigma_p^2 = \sigma_g^2 \left(1 + \frac{\delta^2}{\psi^2}\right). \tag{36}$$

Using this characterization, the following Proposition presents the distribution of $\hat{\sigma}_p^2$.

Proposition 2 *Let $u \sim \chi^2_{N-1}(T\psi^2)$, $v \sim \chi^2_{T-N+1}$, and $\tilde{y} \sim N(\sqrt{T}\delta, 1)$, independent of each other. For a sample minimum-variance portfolio with target expected return of μ_p, the distribution of the sample variance $\hat{\sigma}_p^2$ is given by*[7]

$$\hat{\sigma}_p^2 = \frac{\breve{\sigma}_p^2 v}{T}, \tag{37}$$

where

$$\breve{\sigma}_p^2 = \sigma_g^2 \left(1 + \frac{\tilde{y}^2}{u}\right). \tag{38}$$

Proposition 2 and equation (33) suggest that the sample frontier, like the population frontier, is a quadratic function of μ_p. However, the coefficients are random and the distribution of

[7]The exact density function of $\hat{\sigma}_p^2$ can be written as an infinite series of confluent hypergeometric functions. Details are available upon request. Hillier and Satchell (2003) present a similar analysis as our Proposition 2. The only difference is that they analyze a minimum-variance portfolio that has the same sample mean as a target portfolio whereas we analyze a minimum-variance portfolio with a constant sample mean.

$\hat{\sigma}_p^2$ depends on three independent random variables. There are two sources of randomness in $\hat{\sigma}_p^2$. One is due to v, which is a function of $\hat{V}$ and the other one is due to u and $\tilde{y}$ in $\check{\sigma}_p^2$, which are functions of $\hat{\mu}$. With (37), simulating the distribution of $\hat{\sigma}_p^2$ is rather straightforward. It requires specifying (a, b, c), or equivalently $(\psi^2, \mu_g, \sigma_g^2)$ and then simulating u, v and $\tilde{y}$ to obtain the distribution of $\hat{\sigma}_p^2$. While one can always simulate the distribution of the sample minimum-variance frontier by simulating a $T \times N$ matrix of returns, our method is far more efficient. In addition, it is difficult to generalize the results from a Monte Carlo simulation as they depend on the choice of μ and V. Our analysis, however, makes it clear that the finite sample distribution only depends on the three efficiency set constants, and it is valid for all relevant combinations of μ and V.

Before we discuss the exact moments of $\hat{\sigma}_p^2$, we need to introduce an important integral that will be used repeatedly throughout the paper:

$$\phi = \frac{T\psi^2}{2}\int_0^1 e^{\frac{T\psi^2(y-1)}{2}} y^{\frac{N-3}{2}} \mathrm{d}y \quad \text{for } N \geq 2. \tag{39}$$

Note that we can also write ϕ using the confluent hypergeometric function as[8]

$$\phi = \frac{T\psi^2 e^{-\frac{T\psi^2}{2}}}{N-1} {}_1F_1\left(\frac{N-1}{2}, \frac{N+1}{2}; \frac{T\psi^2}{2}\right) = \frac{T\psi^2}{N-1} {}_1F_1\left(1, \frac{N+1}{2}; -\frac{T\psi^2}{2}\right). \tag{40}$$

Let P be an $N \times (N-1)$ orthonormal matrix such that its columns are orthogonal to $V^{-\frac{1}{2}}1_N$, i.e., $P'V^{-\frac{1}{2}}1_N = 0_{N-1}$. Define $z = \sqrt{T}P'V^{-\frac{1}{2}}\hat{\mu} \sim N(\mu_z, I_{N-1})$, where $\mu_z = \sqrt{T}P'V^{-\frac{1}{2}}\mu$. It is easy to verify that $\mu_z'\mu_z = T\psi^2$, so $u = z'z \sim \chi^2_{N-1}(T\psi^2)$. The following lemma expresses the expectation of various ratios of $z'\mu_z$ and $z'z$ in terms of ϕ.

Lemma 3 *Suppose $z \sim N(\mu_z, I_{N-1})$, and $\mu_z'\mu_z = T\psi^2$. We have*

$$E\left[\frac{1}{z'z}\right] = \frac{1-\phi}{N-3} \quad \text{for } N > 3, \tag{41}$$

$$E\left[\frac{1}{(z'z)^2}\right] = \frac{(N-5)\phi - T\psi^2(1-\phi) + 2}{2(N-3)(N-5)} \quad \text{for } N > 5, \tag{42}$$

$$E\left[\frac{z'\mu_z}{z'z}\right] = \phi \quad \text{for } N > 2, \tag{43}$$

$$E\left[\frac{z'\mu_z}{(z'z)^2}\right] = \frac{T\psi^2(1-\phi)}{2(N-3)} - \frac{\phi}{2} \quad \text{for } N > 3, \tag{44}$$

[8] A fast and numerically stable Matlab program for computing ϕ is available upon request.

$$E\left[\frac{(z'\mu_z)^2}{z'z}\right] = T\psi^2 - (N-2)\phi \quad \text{for } N > 1, \tag{45}$$

$$E\left[\frac{(z'\mu_z)^2}{(z'z)^2}\right] = \frac{(N-2)\phi}{2} - \frac{T\psi^2(N-4)(1-\phi)}{2(N-3)} \quad \text{for } N > 3. \tag{46}$$

With the help of this lemma, the following Proposition presents the expected value and variance of $\hat{\sigma}_p^2$. The expected value tells us on average where the sample frontier is located while the variance tells us how volatile the sample frontier is.

Proposition 3 *For a sample minimum-variance portfolio with target expected return of μ_p, its expected in-sample variance exists if and only if $N > 3$. When $N > 3$, the expected value of $\hat{\sigma}_p^2$ is given by*

$$\bar{\sigma}_p^2 = E[\hat{\sigma}_p^2] = \left(\frac{T-N+1}{T}\right) E[\check{\sigma}_p^2], \tag{47}$$

where $E[\check{\sigma}_p^2] = \sigma_g^2(1 + hE[u^{-1}])$ and $h = T\delta^2 + 1$. In addition, the variance of $\hat{\sigma}_p^2$ exists if and only if $N > 5$. When $N > 5$, the variance of $\hat{\sigma}_p^2$ is given by

$$\mathrm{Var}[\hat{\sigma}_p^2] = \frac{(T-N+1)[(T-N+3)\mathrm{Var}[\check{\sigma}_p^2] + 2E[\check{\sigma}_p^2]^2]}{T^2}, \tag{48}$$

where $\mathrm{Var}[\check{\sigma}_p^2] = \sigma_g^4\left[(h^2+4h-2)E[u^{-2}] - (hE[u^{-1}])^2\right]$, with $E[u^{-1}]$ and $E[u^{-2}]$ are given in (41) and (42).

To get a feeling for how well the sample frontier approximates the population frontier, consider the following example. We assume there are $N = 10$ risky assets, with population minimum-variance frontier determined by the following parameters: $\mu_g = 0.00745$, $\sigma_g = 0.04930$, and $\psi = 0.133$. These parameters are chosen based on the unbiased estimates of μ_g, σ_g^2 and ψ^2 using monthly returns on the ten size-ranked portfolios of the combined NYSE-AMEX-NASDAQ from 1926/1–2004/12.

In Figure 1, we provide information on the distribution of the sample minimum-variance frontier of these ten assets for different estimation period lengths (T). In each panel, the solid line is the population frontier, the dashed line is the expected value of the sample minimum-variance frontier ($\bar{\sigma}_p$) and the dotted lines are the 5th and 95th percentiles of $\hat{\sigma}_p$.[9]

[9]Note that $\bar{\sigma}_p$ is the square root of $E[\hat{\sigma}_p^2]$, but not the mean of $\hat{\sigma}_p$. The analytical expression for $E[\hat{\sigma}_p]$ is available upon request. The difference between $E[\hat{\sigma}_p]$ and $\bar{\sigma}_p$ is almost negligible even for T as small as 60.

The 5th and 95th percentiles of $\hat{\sigma}_p$ are obtained by simulating $\hat{\sigma}_p$ 100,000 times using (37). It is readily apparent that the sample minimum-variance frontier is hugely biased and the bias is most pronounced for portfolios that are far from the global minimum-variance portfolio. The bias is still quite pronounced even when using moments estimated with 50 years of monthly data.

Figure 1 about here

It is evident from this figure that the sample minimum-variance frontier tends to lie to the left of the population frontier, providing an overly optimistic assessment of the population frontier. To understand the reason for this, recall that the sample frontier is computed using the sample mean and covariance matrix that was also used to obtain the optimal portfolio weights of the sample minimum-variance portfolios. Due to the optimization, these portfolio weights are tilted toward assets with lower sample variances and lower sample covariances with other assets. However, low sample variance and covariances may be due to negative estimation error. As a result, the in-sample variance of these minimum-variance portfolios tends to be understated.

In the appendix, we present results for a second example based on $N = 25$ risky assets chosen to closely mimic the monthly value-weighted returns from the popular Fama and French (1993) 25 size and book-to-market portfolios over the period 1932/1–2004/12. The main difference that can be noted is that the bias increases with the number of assets.

3.3 Improved estimator of the minimum-variance frontier

Statistically, the main reason for the huge bias of the sample minimum-variance frontier is its use of $1/\hat{\psi}^2$ to estimate $1/\psi^2$. From Proposition 1 we know that $\hat{\psi}^2 \sim \chi^2_{N-1}(T\psi^2)/\chi^2_{T-N+1}$. Therefore, using the results in Lemma 3, the expected value of $1/\hat{\psi}^2$ is given by

$$E\left[\frac{1}{\hat{\psi}^2}\right] = \frac{(T-N+1)(1-\phi)}{N-3}. \tag{49}$$

The expectation of $1/\hat{\psi}^2$ is in general much less than $1/\psi^2$. In Figure 2, we plot the percentage bias of $1/\hat{\psi}^2$ as a function of $T\psi^2$ for the case that $N = 10$. Over a reasonable range of $T\psi^2$, the percentage bias of $1/\hat{\psi}^2$ is huge. Even for $T\psi^2$ as large as 20, the expectation of $1/\hat{\psi}^2$ is still less than 75% of the true value of $1/\psi^2$. This explains why the sample frontier has a much smaller variance than the true frontier, even for T as large as 600.

Although we know the functional form of the expectation of $1/\hat{\psi}^2$, correcting for the bias of $1/\hat{\psi}^2$ is not an easy task. This is because the bias is determined by N, T and ψ^2 through the function ϕ, which is a confluent hypergeometric function. The confluent hypergeometric function is a special function that is very difficult to approximate by polynomials.[10] As a result, when we attempt to use the standard bias reduction methods that rely on either Taylor series expansions or resampling techniques, the improvement is typically very small unless $T\psi^2$ is large. To illustrate, we plot the percentage bias of the jackknife version of $1/\hat{\psi}^2$ in Figure 2.[11] While the jackknife version of $1/\hat{\psi}^2$ provides an improvement over the sample estimator, its bias is still large.

Figure 2 about here

If we were able to find an adjusted estimator $\hat{\psi}_a^2$ that is a function of $\hat{\psi}^2$ such that $E[1/\hat{\psi}_a^2] = 1/\psi^2$, then we can construct an unbiased estimator of σ_p^2 as follows:

$$\hat{\sigma}_{pu}^2 = \frac{T\hat{\sigma}_g^2}{T-N} + \frac{1}{\hat{\psi}_a^2}\left[(\mu_p - \hat{\mu}_g)^2 - \frac{\hat{\sigma}_g^2(1+\hat{\psi}^2)}{T-N}\right]. \tag{50}$$

However, it does not appear to us that an unbiased estimator of $1/\psi^2$ exists, so we search for an improved estimator of $1/\psi^2$ instead.[12] After considering many estimators of $1/\psi^2$, we settled on the following adjusted estimator of $1/\psi^2$:

$$\frac{1}{\hat{\psi}_a^2} = \frac{T\int_0^z u^{\frac{T-N-1}{2}}(1-u)^{\frac{N-5}{2}}\mathrm{d}u}{2z^{\frac{T-N-1}{2}}(1-z)^{\frac{N-3}{2}}}, \tag{51}$$

[10]See Luh, Müller, and Vasundhra (2005) for a discussion of the difficulties in approximating confluent hypergeometric functions over a wide range of arguments.

[11]The jackknife version of an estimator $\hat{\theta}$ is $\hat{\theta}_{JK} = T\hat{\theta} - (T-1)\sum_{t=1}^{T}\hat{\theta}_{(t)}/T$, where $\hat{\theta}_{(t)}$ is the estimator $\hat{\theta}$ calculated without using the t-th observation of the data.

[12]There is a large literature of estimating noncentrality parameter of noncentral chi-squared and F-distributions using a single observation, see Johnson, Kotz, and Balakrishnan (1995, Chapters 29–30) for a review of this literature. However, none of the existing work appears to deal with the problem of estimating the inverse of the noncentrality parameter.

where $z = 1/(1+\hat{\psi}^2)$. While the expression of $1/\hat{\psi}_a^2$ looks complicated, the integral in the numerator is proportional to the incomplete beta function, so we can easily use the following Matlab command to compute $1/\hat{\psi}_a^2$:[13]

```
T*betainc(z,(T-N+1)/2,(N-3)/2)/(2*(1-z)*betapdf(z,(T-N+1)/2,(N-3)/2))
```

As it turns out, this new estimator of $1/\psi^2$ is not unbiased, but its bias is significantly less than that of $1/\hat{\psi}^2$. The following lemma presents the expectation of this adjusted estimator.

Lemma 4 *When $N > 3$, the expectation of $1/\hat{\psi}_a^2$ exists and it is given by*

$$E\left[\frac{1}{\hat{\psi}_a^2}\right] = \frac{1-e^{-\frac{T\psi^2}{2}}}{\psi^2}. \tag{52}$$

In addition, the relative bias of $1/\hat{\psi}_a^2$ is smaller than the relative bias of $1/\hat{\psi}^2$ when $N \geq 5$.

Note that the relative bias of this adjusted estimator is equal to $-e^{T\psi^2/2}$ which depends on $T\psi^2$ but not on N. In Figure 2, we plot the percentage bias of $1/\hat{\psi}_a^2$ as a function of $T\psi^2$ for the case that $N = 10$. When $T\psi^2$ is very small, there is still a large bias in our adjusted estimator. However, when $T\psi^2$ increases, the bias of $1/\hat{\psi}_a^2$ disappears significantly more quickly than $1/\hat{\psi}^2$. For example, when $T\psi^2 = 4$, the percentage bias of $1/\hat{\psi}_a^2$ is already at a low level of -13.5%, whereas $1/\hat{\psi}^2$ and its jackknife version still have percentage biases of -64.2% and -37.7%, respectively. Therefore, the improvement of $1/\hat{\psi}_a^2$ over $1/\hat{\psi}^2$ is remarkable. If one is still concerned with the bias of $1/\hat{\psi}_a^2$ when $T\psi^2$ is very small, one can use a jackknife version of $1/\hat{\psi}_a^2$. In Figure 2, we also plot the percentage bias of the jackknife version of $1/\hat{\psi}_a^2$. It shows that when $T\psi^2$ is small, the jackknife version of $1/\hat{\psi}_a^2$ greatly reduces the bias of $1/\hat{\psi}_a^2$, but when $T\psi^2$ becomes larger, it tends to overshoot and has a bigger bias than $1/\hat{\psi}_a^2$. Due to the computational time of the jackknife version and its additional volatility, we have decided to simply use $1/\hat{\psi}_a^2$ as our estimator of $1/\psi^2$.

[13]When N is odd, we can use integration by parts repeatedly to write

$$\frac{1}{\hat{\psi}_a^2} = \frac{T}{2}\sum_{s=0}^{\frac{N-5}{2}} \frac{\left(\frac{N-3}{2}-s\right)_s}{\left(\frac{T-N+1}{2}\right)_{s+1}} \frac{1}{\hat{\psi}^{2(s+1)}},$$

so $1/\hat{\psi}_a^2$ can be written as a polynomial in $1/\hat{\psi}^2$.

With our adjusted estimator $\hat{\psi}_a^2$, we can plug it in (50) to obtain an approximately unbiased estimator of σ_p^2. However, there is one additional problem with this estimator: $\hat{\sigma}_{pu}^2$ can be negative and hence is an inadmissible estimator of σ_p^2. To overcome this problem, we choose to use the following adjusted estimator of σ_p^2:

$$\hat{\sigma}_{pa}^2 = \frac{T\hat{\sigma}_g^2}{T-N} + \frac{1}{\hat{\psi}_a^2} \max\left[(\mu_p - \hat{\mu}_g)^2 - \frac{\hat{\sigma}_g^2(1+\hat{\psi}^2)}{T-N}, 0\right]. \tag{53}$$

To examine how well this adjusted estimator of σ_p^2 performs, we plot the expected value of our adjusted sample frontier (the dashed line) and the population frontier (the solid line) in Figure 3 for different lengths of estimation period (T). In addition, we plot the 5th and 95th percentiles of $\hat{\sigma}_{pa}$ using the dotted lines. By comparing the distribution of the adjusted sample frontier in Figure 3 with the distribution of the naïve sample frontier in Figure 1, we see that our adjusted estimator in equation (53) performs dramatically better. While there is still some bias in our adjusted sample frontier when T is small, the bias becomes almost negligible when $T > 120$ months.

Figure 3 about here

3.4 Distribution of the weights

Although this paper is primarily concerned with characterizing the exact behavior of the sample minimum-variance frontier, the distribution of the weights of the sample minimum-variance portfolio ($\hat{w}$) with a given target expected return is potentially interesting. In the following Proposition, we simplify the exact distribution of $\hat{w}$ and present the exact mean and covariance matrix of $\hat{w}$.

Proposition 4 *Let $\tilde{A} = [\hat{\mu},\ 1_N]'V^{-1}[\hat{\mu},\ 1_N]'$, $\breve{\sigma}_p^2 = [\mu_p,\ 1]\tilde{A}^{-1}[\mu_p,\ 1]'$, $x \sim N(0_{N-2}, I_{N-2})$, $\tilde{r} \sim \chi^2_{T-N+2}$, and Q be an $N \times (N-2)$ orthonormal matrix that is orthogonal to $V^{-\frac{1}{2}}[\hat{\mu},\ 1_N]$, with x, $\tilde{r}$ and $\hat{\mu}$ are independent of each other. The weights of the sample minimum-variance portfolio with a target expected return of μ_p can be written as*[14]

$$\hat{w} = V^{-1}[\hat{\mu},\ 1_N]\tilde{A}^{-1}[\mu_p,\ 1]' + V^{-\frac{1}{2}}Q\frac{\breve{\sigma}_p x}{\tilde{r}^{\frac{1}{2}}}. \tag{54}$$

[14]For the special case of $N = 2$, the second term drops out and it can be shown that $\hat{w}$ has the same distribution as $w_g + \text{sgn}(\mu_1 - \mu_2)\sigma_g\tilde{y}[1,\ -1]'/[([1,\ -1]V[1,\ -1]'z)^{\frac{1}{2}}z]$, where $\tilde{y} \sim N(\sqrt{T}\delta, 1)$, $z \sim N(\sqrt{T}\psi, 1)$,

When $N > 2$, the expected value of $\hat{w}$ exists and it is given by

$$E[\hat{w}] = w_g + \phi(w - w_g), \tag{55}$$

where w is the population counterpart of $\hat{w}$ and it is defined in (26), ϕ is defined in (39), and $w_g = V^{-1}1_N/(1_N'V^{-1}1_N)$ is the vector of the weights of the global minimum-variance portfolio. In addition, the covariance matrix of $\hat{w}$ exists when $N > 3$ and it is given by

$$\begin{aligned}\mathrm{Var}[\hat{w}] &= \left[\left(\frac{T-N-1}{T-N}\right)\sigma_g^2 h\left(\frac{1-\phi}{2(N-3)} - \frac{\phi}{2T\psi^2}\right) + \frac{E[\check{\sigma}_p^2]}{T-N} - \frac{\sigma_g^2\phi}{(T-N)T\psi^2}\right]V_a \\ &\quad + \left[\left(\frac{T-N-1}{T-N}\right)\sigma_g^2 h\left(\frac{(N-1)\phi}{2T\psi^2} - \frac{1-\phi}{2}\right) - \frac{\sigma_g^2}{T-N}\left(1 - \frac{(N-1)\phi}{T\psi^2}\right)\right]V_b \\ &\quad - \phi^2(w - w_g)(w - w_g)',\end{aligned} \tag{56}$$

where $h = T\delta^2 + 1$, $E[\check{\sigma}_p^2]$ is given in Proposition 3, and

$$V_a = V^{-1} - \frac{V^{-1}1_N 1_N' V^{-1}}{1_N' V^{-1} 1_N}, \tag{57}$$

$$V_b = \frac{V^{-1}(\mu - \mu_g 1_N)(\mu - \mu_g 1_N)' V^{-1}}{(\mu - \mu_g 1_N)' V^{-1}(\mu - \mu_g 1_N)}. \tag{58}$$

Proposition 4 extends the work of Britten-Jones (1999) who presents a test of linear restrictions on the weights of the tangency portfolio. It also extends the work of Okhrin and Schmid (2006) who present the exact distribution of the weights of the sample global minimum-variance portfolio and the exact mean and covariance matrix of the weights of the sample minimum-variance portfolio that maximizes an expected quadratic utility (as opposed to having a given target expected return as in our case).

3.5 Departures from normality

Our analytical results so far are derived under the assumption of multivariate normality. It is natural to question how well these results work when the returns are not normally distributed. To examine the robustness of our finite sample results to departures from

---

and they are independent of each other. An explicit expression of the probability density function of $\hat{w}$ for $N = 2$ is available upon request. For a slightly different problem, Hillier and Satchell (2003) also show that conditional on $\hat{\mu}$, the sample weights have a multivariate t-distribution.

normality, we consider the case in which the returns follow a multivariate t-distribution with five degrees of freedom. Under this t-distribution assumption, the returns have fat tails, which is what we often find in the data. As we cannot obtain the finite sample distribution of $\hat{\sigma}_p^2$ and $\hat{\sigma}_{pa}^2$ under the t-distribution assumption, we rely on simulation. In order to easily compare our results under the assumption of normality with results from the t-distribution, we simulate the returns of the assets using exactly the same μ and V as in the normality case. In Figure 4, we plot the distribution of $\hat{\sigma}_p$ and $\hat{\sigma}_{pa}$ under the normality and the t-distribution assumptions for the ten assets case using the same parameters as in Figures 1 and 3. To conserve space, we only provide the plots for $T = 60$ and $T = 240$ because there is a negligible difference between the results under the normality and t-distribution assumptions when $T = 600$. The left panels in Figure 4 are for the unadjusted sample frontiers ($\hat{\sigma}_p$) and the right panels are for the adjusted sample frontiers ($\hat{\sigma}_{pa}$). In each panel, the dashed lines represent the distribution of the sample frontiers under the multivariate t-distribution assumption and the dotted lines represent the distribution under the normality assumption. From the upper panels of Figure 4, we can see that when $T = 60$, $\hat{\sigma}_p$ and $\hat{\sigma}_{pa}$ under the t-distribution assumption is on average smaller than their distributions when returns are normally distributed. This result is expected because when returns are t-distributed, we get more extreme returns than in the normality case. As the mean-variance optimizer heavily overweighs stocks with extreme average returns, the sample variance of the optimal portfolio is likely to be smaller when extreme returns are more likely to occur. As T increases to 240, the finite sample distribution of $\hat{\sigma}_p$ under the t-distribution case is very close to the distribution under the normality case. This is because although the returns are not normally distributed, the distribution of $\hat{\mu}$ converges to the normal distribution when T is reasonably large. By comparing the distribution of the unadjusted sample frontier in the left panels with the adjusted sample frontier in the right panels, we can see that while our adjusted sample frontier is obtained under the normality assumption, it still does a very good job in reducing the bias of the unadjusted sample frontier even when returns exhibit very fat tails.

Figure 4 about here

Our robustness analysis suggests that our finite sample results can be sensitive to return distributions that exhibit fat tails, especially for $T \leq 60$ and when the number of assets is large. However, for $T \geq 240$, our finite sample results are very good approximations even when returns have a significant departure from the normality assumption. More importantly, our proposed adjusted estimator of the frontier still does a far better job than the unadjusted estimator even when returns are not normally distributed.

4. Out-of-sample performance of sample minimum-variance portfolios

4.1 Out-of-sample mean and variance

In this section, we are interested in the out-of-sample mean and variance of a sample minimum-variance portfolio. For the sample minimum-variance frontier, $\hat{w}$ is chosen such that $\hat{\mu}_p = \hat{w}'\hat{\mu} = \mu_p$, so there is no randomness in the in-sample mean. As a result, the only random variable is the in-sample variance $\hat{\sigma}_p^2 = \hat{w}'\hat{V}\hat{w}$. However, for the out-of-sample case, both the out-of-sample mean and variance, i.e., $\tilde{\mu}_p = \hat{w}'\mu$ and $\tilde{\sigma}_p^2 = \hat{w}'V\hat{w}$, are random variables. This means that there is one more random variable to study for the out-of-sample case.

Similar to the in-sample variance case, the joint distribution of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ only depends on $(\psi^2, \mu_g, \sigma_g^2)$. However, unlike $\hat{\sigma}_p^2$ whose distribution can be obtained by generating three random variables, our method of simulating the joint distribution of $(\tilde{\mu}_p, \tilde{\sigma}_p^2)$ requires generating six random variables. The additional random variables are needed because the expressions for $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ depend not only on $(\hat{\mu}, \hat{V})$ but also on (μ, V). The following Proposition presents the joint distribution of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$.

Proposition 5 *Let $\tilde{x} \sim N(0,1)$, $\tilde{q} \sim \chi^2_{N-3}$, $\tilde{r} \sim \chi^2_{T-N+2}$, $\tilde{u} \sim \chi^2_{N-2}$, $\tilde{y} \sim N(\sqrt{T}\delta, 1)$ and $\tilde{z} \sim N(\sqrt{T}\psi, 1)$, independent of each other. Let $u = \tilde{z}^2 + \tilde{u} \sim \chi^2_{N-1}(T\psi^2)$ and $\tilde{t} = \tilde{x}/\tilde{r}^{\frac{1}{2}}$. For a sample minimum-variance portfolio with target expected return of μ_p, its out-of-sample expected return and variance can be written as follows:*

$$\tilde{\mu}_p = \mu_g + \psi \left[\sigma_g \frac{\tilde{z}}{u} \tilde{y} + \check{\sigma}_p \left(\frac{\tilde{u}}{u} \right)^{\frac{1}{2}} \tilde{t} \right], \tag{59}$$

$$\tilde{\sigma}_p^2 = \breve{\sigma}_p^2 \left(1 + \frac{\tilde{x}^2 + \tilde{q}}{\tilde{r}}\right), \tag{60}$$

where $\breve{\sigma}_p^2$ is defined in Proposition 2.

There are two sources of randomness in the weights of the sample minimum-variance portfolio, one is due to $\hat{\mu}$ and the other is due to $\hat{V}$. From the proof of Proposition 5, we see that the random variables $\tilde{u}$, $\tilde{y}$ and $\tilde{z}$ are functions of $\hat{\mu}$, whereas $\tilde{x}$, $\tilde{q}$ and $\tilde{r}$ are functions of $\hat{V}$. If V is known but μ is not known, then the distribution of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ will be simplified to

$$\tilde{\mu}_p = \mu_g + \psi\sigma_g \frac{\tilde{z}}{u}\tilde{y}, \tag{61}$$

$$\tilde{\sigma}_p^2 = \breve{\sigma}_p^2. \tag{62}$$

Comparing these expressions with the ones in Proposition 5, we can see that the estimation risk of $\hat{V}$ adds to the uncertainty of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$. In particular, it will always increase $\tilde{\sigma}_p^2$ by a random factor of $1 + (\tilde{x}^2 + \tilde{q})/\tilde{r}$ and this further degrades the out-of-sample performance of the sample minimum-variance portfolio. On the other hand, if μ is known but V is not known, $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ will become

$$\tilde{\mu}_p = \mu_p, \tag{63}$$

$$\tilde{\sigma}_p^2 = \sigma_p^2 \left(1 + \frac{\tilde{x}^2 + \tilde{q}}{\tilde{r}}\right). \tag{64}$$

In this case, $\tilde{\mu}_p$ will always be equal to the target expected return μ_p and only $\tilde{\sigma}_p^2$ is random.

Since the out-of-sample mean and variance of a sample minimum-variance portfolio are random variables, we are interested in their expectations. These expectations give us an idea of what kind of average out-of-sample return and risk that an investor can obtain by holding a sample minimum-variance portfolio. The following Proposition presents the expected out-of-sample mean and variance of a sample minimum-variance portfolio.[15] In addition, it also presents the covariance matrix of $(\tilde{\mu}_p,\ \tilde{\sigma}_p^2,\ \hat{\sigma}_p^2)$.

Proposition 6 *For a sample minimum-variance portfolio with target expected return of μ_p, its expected out-of-sample mean exists if and only if $N > 2$. When $N > 2$, the expected value*

[15] Analytical results for expected out-of-sample standard deviation ($E[\tilde{\sigma}_p]$) are available upon request. The difference between $E[\tilde{\sigma}_p]$ and $\underline{\sigma}_p$ is almost negligible.

of $\tilde{\mu}_p$ *is given by*

$$\underline{\mu}_p = E[\tilde{\mu}_p] = \mu_p - (1-\phi)(\mu_p - \mu_g), \tag{65}$$

where ϕ *is defined in (39). In addition, the expected out-of-sample variance of the sample minimum-variance portfolio exists if and only if* $N > 3$*. When* $N > 3$*, the expected value of* $\tilde{\sigma}_p^2$ *is given by*

$$\underline{\sigma}_p^2 = E[\tilde{\sigma}_p^2] = \left(\frac{T-2}{T-N}\right) E[\check{\sigma}_p^2]. \tag{66}$$

For the covariance matrix of $\tilde{\mu}_p$ *and* $\tilde{\sigma}_p^2$*,* $\mathrm{Var}[\tilde{\mu}_p]$ *and* $\mathrm{Cov}[\tilde{\mu}_p, \tilde{\sigma}_p^2]$ *exist if and only if* $N > 3$*, and* $\mathrm{Var}[\tilde{\sigma}_p^2]$ *exists if and only if* $N > 5$*. When they exist, the variances and covariance are given by*

$$\begin{aligned}
\mathrm{Var}[\tilde{\mu}_p] &= \frac{\psi^2 E[\check{\sigma}_p^2] - \frac{\sigma_g^2}{T}\left(E\left[\frac{(z'\mu_z)^2}{z'z}\right] + hE\left[\frac{(z'\mu_z)^2}{(z'z)^2}\right]\right)}{T-N} + \frac{\sigma_g^2 h}{T} E\left[\frac{(z'\mu_z)^2}{(z'z)^2}\right] \\
&\quad - \phi^2(\mu_p - \mu_g)^2, && (67) \\
\mathrm{Var}[\tilde{\sigma}_p^2] &= \frac{(T-2)}{(T-N)(T-N-2)}\left[(T-4)\mathrm{Var}[\check{\sigma}_p^2] + \frac{2(N-2)E[\check{\sigma}_p^2]^2}{(T-N)}\right], && (68) \\
\mathrm{Cov}[\tilde{\mu}_p, \tilde{\sigma}_p^2] &= \frac{(T-2)}{(T-N)}(\mu_p - \mu_g)\sigma_g^2\left[(h+2)E\left[\frac{z'\mu_z}{(z'z)^2}\right] - \frac{h\phi(1-\phi)}{N-3}\right], && (69)
\end{aligned}$$

where $E[(z'\mu_z)^2/(z'z)]$*,* $E[(z'\mu_z)/(z'z)^2]$*, and* $E[(z'\mu_z)^2/(z'z)^2]$ *are given in Lemma 3, and* h*,* $E[\check{\sigma}_p^2]$ *and* $\mathrm{Var}[\check{\sigma}_p^2]$ *are defined in Proposition 3. The covariances of the in-sample variance* $\hat{\sigma}_p^2$ *with* $\tilde{\mu}_p$ *and* $\tilde{\sigma}_p^2$ *are given by*

$$\begin{aligned}
\mathrm{Cov}[\tilde{\mu}_p, \hat{\sigma}_p^2] &= \left[\frac{(T-N+1)(T-N)}{T(T-2)}\right]\mathrm{Cov}[\tilde{\mu}_p, \tilde{\sigma}_p^2], && (70) \\
\mathrm{Cov}[\tilde{\sigma}_p^2, \hat{\sigma}_p^2] &= \left[\frac{(T-N+1)(T-2)}{T(T-N)}\right]\mathrm{Var}[\check{\sigma}_p^2]. && (71)
\end{aligned}$$

In Figure 5 we plot three frontiers: the population minimum-variance frontier (the solid line), the expected value of the in-sample frontier (the outer dashed line), and the expected value of the out-of-sample frontier (the inner dashed line) for different estimation period lengths. From the figure it is evident that there is a huge difference between the three frontiers, even for T as large as 600 months. The in-sample frontier is in general a downward biased estimate of the population frontier, which in turn heavily dominates the out-of-sample frontier. If an investor chooses to hold a sample minimum-variance portfolio, he is interested in the out-of-sample frontier because that represents the average true performance of his portfolio. Nevertheless, what he observes from the data is the sample frontier, and it is on average much better than the out-of-sample frontier. If an investor relies on the sample frontier to indicate the out-of-sample performance of his portfolio, he will be grossly disappointed.

Figure 5 about here

There are two sources of disappointment for the investor. Using the fact that $0 < \phi < 1$ for $N \geq 3$, we can see from equation (65) that the average out-of-sample mean ($\underline{\mu}_p$) shrinks toward the expected return of the global minimum-variance portfolio (μ_g). As a result, $\underline{\mu}_p$ tends to be lower than the target expected return (μ_p) when $\mu_p > \mu_g$, so the investor will on average be disappointed with the out-of-sample mean of his portfolio. As for the out-of-sample variance of the portfolio, we use (47) and (66) to obtain

$$\underline{\sigma}_p^2 = \left[\frac{(T-2)T}{(T-N)(T-N+1)}\right]\bar{\sigma}_p^2, \tag{72}$$

so the average out-of-sample variance of the sample minimum-variance portfolio ($\underline{\sigma}_p^2$) tends to be much higher than its average in-sample variance ($\bar{\sigma}_p^2$), especially when N/T is large. Therefore, the investor will also be disappointed with the larger than expected risk of his portfolio. To understand the sources of these two disappointments, recall that the sample minimum-variance portfolio is obtained using the estimated moments of the returns. Suppose we form a portfolio with a high target expected return ($\mu_p > \mu_g$). Our portfolio will tend to be heavily invested in assets with high sample means and these sample means are high in part due to chance. Therefore, out-of-sample the returns will tend to be lower than they were in-sample. Similarly, our optimal portfolio tends to load heavily on assets with low sample variances and covariances which are in part low by chance. Out-of-sample, the covariances and variance will tend to be higher than they were historically. These two effects produce out-of-sample returns that have lower means and higher variances than the in-sample frontier implies.

Our results suggest that for a sample minimum-variance portfolio, the sample mean and variance are too optimistic as forecasts of its out-of-sample mean and variance. In order to come up with a reliable forecast of the true out-of-sample performance of a sample minimum-variance portfolio, we need to use a less optimistic forecast. The following Proposition provides the unbiased forecast of $\underline{\mu}_p$ and $\underline{\sigma}_p^2$.

Proposition 7 *The unbiased estimators of $\underline{\mu}_p$ and $\underline{\sigma}_p^2$ are given by*

$$\hat{\underline{\mu}}_{pu} = \mu_p - \frac{N-3}{(T-N+1)\hat{\psi}^2}(\mu_p - \hat{\mu}_g) \quad \text{for } N > 3, \tag{73}$$

$$\hat{\underline{\sigma}}_{pu}^2 = \left[\frac{(T-2)T}{(T-N)(T-N+1)}\right]\hat{\sigma}_p^2 \quad \text{for } N > 5. \tag{74}$$

Since $\underline{\mu}_p$ shrinks toward μ_g, our unbiased estimator also shrinks μ_p toward $\hat{\mu}_g$ to get an unbiased forecast of $\underline{\mu}_p$. In addition, the sample variance $\hat{\sigma}_p^2$ is too optimistic as a forecast of the out-of-sample variance. Proposition 7 suggests that one should multiply $\hat{\sigma}_p^2$ by an adjustment factor of $(T-2)T/[(T-N)(T-N+1)]$ to get an unbiased forecast of the out-of-sample variance of the sample minimum-variance portfolio. Note that similar formulas were presented by Siegel and Woodgate (2007), but their formulas are only approximately unbiased but not exact.

4.2 In-sample and out-of-sample Sharpe ratios

Many popular performance measures of a portfolio are functions of μ_p and σ_p^2. For example, one can measure the performance of a portfolio using its Sharpe ratio, defined as

$$\theta_p = \frac{\mu_p - r}{\sigma_p}, \tag{75}$$

where r is the risk-free rate. In order to estimate the Sharpe ratio of a portfolio, one often uses the sample Sharpe ratio, defined as

$$\hat{\theta}_p = \frac{\hat{\mu}_p - r}{\hat{\sigma}_p}. \tag{76}$$

When the portfolio weight vector w is fixed, $\hat{\theta}_p$ can be easily shown to have the following distribution (see Miller and Gehr (1978))

$$\hat{\theta}_p \sim \frac{t_{T-1}(\sqrt{T}\theta_p)}{\sqrt{T-1}}, \tag{77}$$

where $t_\nu(\delta)$ is a noncentral t-distribution with ν degrees of freedom and a noncentrality parameter of δ.[16]

However, for a sample minimum-variance portfolio, the weights of the portfolio are random, so the distribution result above does not apply and the exact distribution of its sample Sharpe ratio was previously unknown. With the results in our Proposition 2, we can now easily evaluate the distribution of $\hat{\theta}_p$. However, for an investor who holds a sample minimum-variance portfolio, he is not particularly concerned with the distribution of $\hat{\theta}_p$ but rather is more concerned about the out-of-sample Sharpe ratio of his portfolio, which is given by

$$\tilde{\theta}_p = \frac{\tilde{\mu}_p - r}{\tilde{\sigma}_p}. \tag{78}$$

With our results in Proposition 5, we can also easily evaluate the distribution of $\tilde{\theta}_p$.

In Figure 6, we provide information on the distribution of $\hat{\theta}_p$ and $\tilde{\theta}_p$ for $N = 10$ assets and various lengths of estimation window, using the same parameters as before and assuming $r = 0.5\%$/month. The solid line in Figure 6 represents the Sharpe ratios of portfolios on the population frontier for $\mu_g \le \mu_p \le \mu_T$, where μ_T is the expected return of the tangency portfolio. As expected, when we increase the target expected return from μ_g to μ_T, the population Sharpe ratio steadily increases and reaches the maximum at $\mu_p = \mu_T$. The dashed-dotted line and the surrounding dotted lines in Figure 6 represent the mean and the 5th and 95th percentiles of the in-sample Sharpe ratio ($\hat{\theta}_p$) for portfolios on the sample frontier. As $\hat{\sigma}_p^2$ tends to be smaller than σ_p^2, the in-sample Sharpe ratios are naturally much higher than the population ones. The high in-sample Sharpe ratios are of course unattainable out-of-sample. In Figure 6, we use the dashed line and the surrounding dotted lines to represent the mean and the 5th and 95th percentiles of the out-of-sample Sharpe ratio ($\tilde{\theta}_p$) of the sample minimum-variance portfolios. These represent the true performance of holding the sample minimum-variance portfolios. By comparing the distribution of $\hat{\theta}_p$ with $\tilde{\theta}_p$, we can see that $\hat{\theta}_p$ grossly overestimates $\tilde{\theta}_p$. For μ_p that is far away from μ_g, we often see that the 5th percentile of $\hat{\theta}_p$ is even higher than the 95th percentile of $\tilde{\theta}_p$. This suggests that when we choose an optimal portfolio based on sample estimates, we cannot

[16] For the fixed weights case, asymptotic distribution of $\hat{\theta}_p$ is provided by Jobson and Korkie (1981), Cadsby (1986), Lo (2002) and Memmel (2003).

trust the sample Sharpe ratio as an indicator of the future performance of the portfolio. It is also noteworthy that unlike the population and the in-sample Sharpe ratios, the distribution of the out-of-sample Sharpe ratio when $\mu_p = \mu_g$ is not all that different from the distribution when $\mu_p = \mu_T$, especially when we use a short estimation window. Therefore, the sample tangency portfolio is not all that superior to the sample global minimum-variance portfolio for investment purposes, unless the estimation window is very long.

Figure 6 about here

5. Conclusions

The minimum-variance frontier holds an important position in finance. However, relatively little is known about its sampling properties. In this paper we derive the finite-sample distribution of the efficient set constants and the sample minimum-variance frontier. We show that the sample minimum-variance frontier commonly graphed in the finance literature is significantly biased downward and suggest how a much improved estimate of the frontier can be obtained.

An important and relatively understudied problem is the distribution of the returns from holding a sample minimum-variance portfolio. We characterize the mean and variance of the out-of-sample return of a sample minimum-variance portfolio. The out-of-sample frontier is dramatically inferior to the in-sample frontier and failing to correct for the overoptimism leads to significant disappointment out-of-sample. In-sample optimism arises because the sample minimum-variance portfolios are tilted toward assets that experienced favorable in-sample returns, some of which is due to sampling variability. Stocks that experienced extreme returns, both good and bad, in-sample tend to perform less extreme out-of-sample because the random component of their sample moments is uncorrelated with future performance. We show how to construct an unbiased estimate of the mean and variance of the out-of-sample returns. Correcting for the in-sample optimism takes two forms. The out-of-sample expected return must be shrunk toward the average return of the sample global minimum-variance portfolio, and the out-of-sample variance must be increased by a constant factor.

Several extensions of the results in this paper are possible. For example, it is relatively easy to apply our distribution results to study the distributions of other out-of-sample performance measures that are functions of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$. It is also not too difficult to extend our distribution results on the sample minimum-variance frontier by incorporating equality constraints on the portfolio weights. However, there are some interesting problems that are technically challenging to us. One problem is how to construct a confidence interval for the population minimum-variance frontier.[17] Another problem is how to obtain the distribution of the sample frontier when we incorporate inequality constraints (like short-selling constraints) on the portfolio weights. Finally, a very interesting open-ended question is in the presence of estimation risk, how one should choose a portfolio to optimize its out-of-sample performance. We hope future research will provide solutions to these problems.

[17]Jobson (1991) provides various approaches of constructing confidence intervals for the frontier based on asymptotic results. Bodnar and Schmid (2006b) present an exact confidence region for the entire frontier. Exact confidence intervals on the minimum-variance frontier (for a given expected return or variance) are currently unavailable in the literature.

References

[1] Barry, C. B. 1974. Portfolio analysis under uncertain means, variances and covariance. *Journal of Finance* **29** 515–522.

[2] Basak, G. K., R. Jagannathan, R. Ma. 2005. Estimating the risk in sample efficient portfolios. Working paper, Northwestern University.

[3] Bawa, V. S., S. J. Brown, W. A. Klein. 1979. *Estimation Risk and Optimal Portfolio Choice*. Amsterdam, North-Holland.

[4] Bodnar, T. and W. Schmid. 2006a. Mean-variance analysis under parameter uncertainty. Working paper from the Postgraduate Research Programme "Capital Markets and Finance in the Enlarged Europe," No. 5/2006.

[5] Bodnar, T. and W. Schmid 2006b. Econometrical analysis of the sample efficient frontier. Working paper from the Postgraduate Research Programme "Capital Markets and Finance in the Enlarged Europe," No. 10/2006.

[6] Britten-Jones, M. 1999. The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance* **54** 655–671.

[7] Broadie, M. 1993. Computing efficient frontiers using estimated parameters. *Annals of Operations Research* **45** 21–58.

[8] Brown, S. J. 1976. *Optimal Portfolio Choice under Uncertainty.* Ph.D. dissertation, University of Chicago.

[9] Cadsby, C. B. 1986. Performance hypothesis testing with the Sharpe and Treynor measures: A Comment. *Journal of Finance* **41** 1175–1176.

[10] Clarke, R., H. de Silva, S. Thorley. 2006. Minimum-variance portfolios in the U.S. equity market. *Journal of Portfolio Management* **33** 10–24.

[11] Dickey, J. M. 1967. Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *Annals of Mathematical Statistics* **38** 511–518.

[12] Dickinson, J. P. 1974. The reliability of estimation procedures in portfolio analysis. *Journal of Financial and Quantitative Analysis* **9** 447–462.

[13] Fama, E. F., K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* **33** 3–56.

[14] Frankfurther, G. M., H. E. Phillips, J. P. Seagle. 1971. Portfolio selection: The effects of uncertain means, variances and covariances. *Journal of Financial and Quantitative Analysis* **6** 1251–1262.

[15] Garlappi, L., R. Uppal, T. Wang. 2007. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies* **20** 41–81.

[16] Haff, L. R. 1979. An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis* **9** 531–544.

[17] Haugen R., N. Baker. 1991. The efficient market inefficiency of capitalization-weighted stock portfolios. *Journal of Portfolio Management* **17** 35–40.

[18] Hillier, G. H., S. E. Satchell. 2003. Some exact results for efficient portfolios with given returns, in S. E. Satchell and A. Scowcroft (eds.) *Advances in portfolio construction and implementation* 310–325.

[19] Hillier, G., R. Kan, X. Xiaolu. 2008. Generating functions and short recursions, with applications to the moments of quadratic forms in noncentral normal vectors. Working Paper, University of Toronto, Ontario, Canada.

[20] Hoque, A. 1985. The exact moments of forecast error in the general dynamic model. *Sankhyā: The Indian Journal of Statistics* **47** (Series B, Part 1) 128–143.

[21] Jagannathan, R., T. Ma. 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance* **58** 1651–1683.

[22] Jobson, J. D. 1991. Confidence regions for the mean-variance efficient set: An alternative approach to estimation risk. *Review of Quantitative Finance and Accounting* **1** 235–257.

[23] Jobson, J. D., B. Korkie, V. Ratti. 1979. Improved estimation for Markowitz portfolios using James-Stein type estimators. *Proceedings of the American Statistical Association, Business and Economics Statistics Section* **41** 279–284.

[24] Jobson, J. D., B. Korkie. 1980. Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association* **75** 544–554.

[25] Jobson, J. D., B. Korkie. 1981. Putting Markowitz theory to work. *Journal of Portfolio Management* **7** 70–74.

[26] Jobson, J. D., B. Korkie. 1981. Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance* **36** 889–908.

[27] Johnson, N. L., S. Kotz, N. Balakrishnan. 1995. *Continuous Univariate Distributions, Volume 2.* Wiley, New York.

[28] Jorion, P. 1985. International portfolio diversification with estimation risk. *Journal of Business* **58** 259–278.

[29] Jorion, P. 1986. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* **21** 279–292.

[30] Kan, R., G. Zhou. 2007. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* **42** 621–656.

[31] Knight, J., S. E. Satchell. 2006. Exact properties of measures of optimal investment for institutional investors. Working paper, University of Western Ontario.

[32] Krishnan, M. 1967. The moments of a doubly noncentral t-distribution. *Journal of the American Statistical Association* **62** 278–287.

[33] Ledoit, O., M. Wolf. 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* **10** 603–621.

[34] Lo, A. 2002. The statistics of Sharpe ratios. *Financial Analyst Journal* **58** 36–52.

[35] Luh, W., J. Müller, P. Vasundhra. 2005. Polynomial series expansions for confluent and Guassian hypergeometric functions. *Mathematics of Computation* **74** 1937–1952.

[36] Magnus, J. R., 1986. The exact moments of a ratio of quadratic forms in normal variables. *Annales D'économie et de Statistique* **4** 95–109.

[37] Markowitz, H. M. 1952. Mean-variance analysis in portfolio choice and capital markets. *Journal of Finance* **7** 77–91.

[38] Memmel, C. 2003. Performance hypothesis testing with the Sharpe ratio. *Finance Letters* **1** 21–23.

[39] Merton, R. 1972. An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis* **7** 1851–1872.

[40] Miller, R. E., A. K. Gehr. 1978. Sample size bias and Sharpe's performance measure: A note. *Journal of Financial and Quantitative Analysis* **13** 943–946.

[41] Mori, H. 2004. Finite sample properties of estimators for the optimal portfolio weight. *Journal of Japan Statistical Society* **34** 27–46.

[42] Muirhead, R. J. 1982. *Aspects of Multivariate Statistical Theory.* Wiley, New York.

[43] Okhrin, Y., W. Schmid. 2006. Distributional properties of portfolio weights. *Journal of Econometrics* **134** 235–256.

[44] Sawa, T. 1972. Finite sample properties of the k-class estimators, *Econometrica* **40** 653–680.

[45] Siegel, A. F., A. Woodgate. 2007. Performance of portfolios optimized with estimation error. *Management Science* **53** 1005–1015.

[46] Sharpe, W. F. 1963. A simplified model for portfolio analysis. *Management Science* **9** 277–293.

[47] TerHorst, J. R., F. A. DeRoon, B. J. M. Werkerzx. 2002. Incorporating estimation risk in portfolio choice. Working paper, Tilburg University.

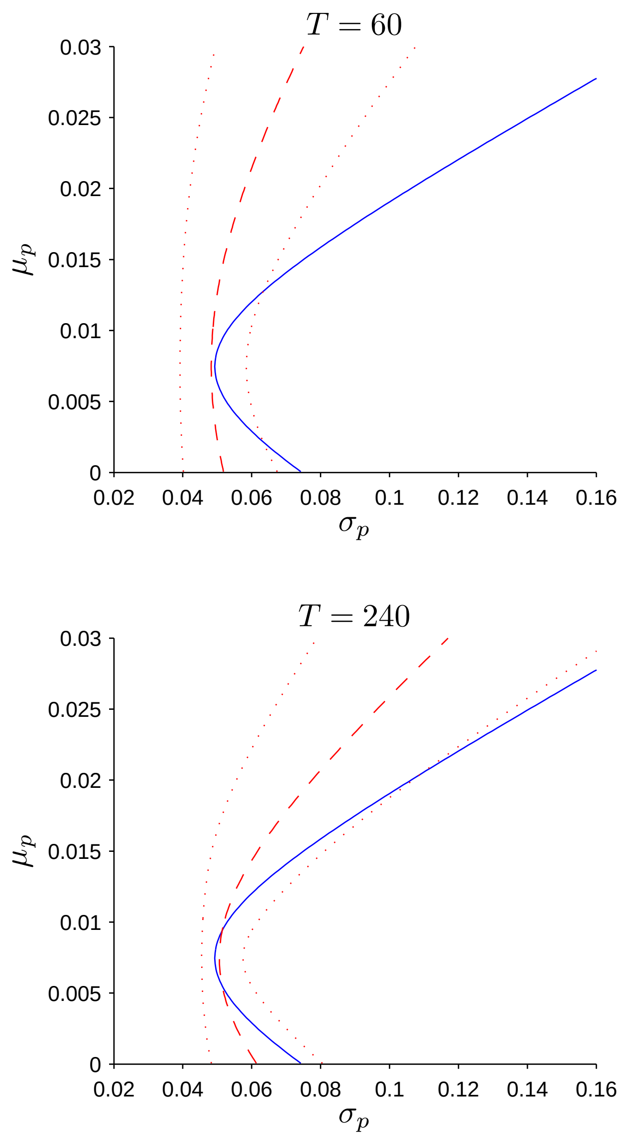

Figure 1. The figure provides information on the distribution of the sample minimum-variance frontier for different lengths of estimation period (T). The solid line represents the population frontier for $N = 10$ assets, with parameters $\mu_g = 0.00745$, $\sigma_g = 0.04930$, and $\psi = 0.133$. The dashed line represents $\bar{\sigma}_p$, where $\bar{\sigma}_p^2$ is the expected value of the sample minimum-variance frontier. The two dotted lines represent the 5th and the 95th percentiles of the sample minimum-variance frontier, based on 100,000 simulations.

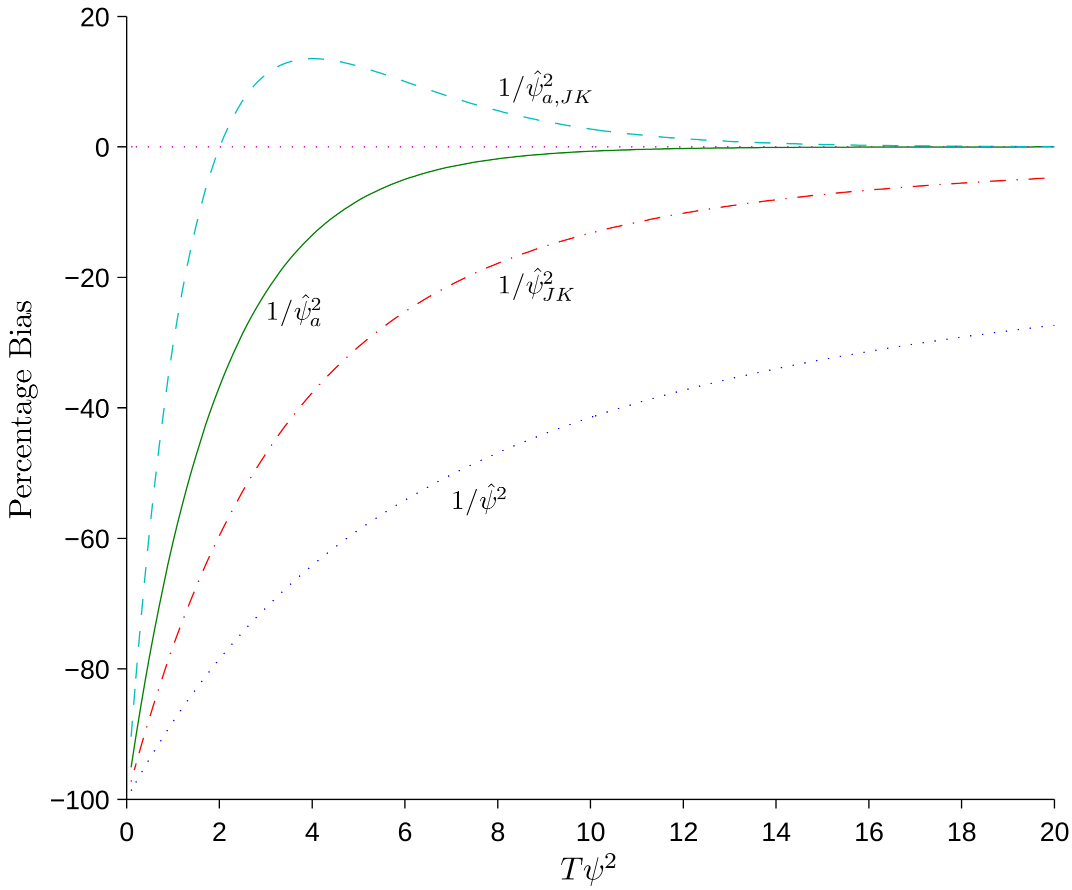

Figure 2. The figure plots the percentage biases of four estimators of the inverse of the squared slope of the asymptote to the minimum-variance frontier $(1/\psi^2)$ as a function of $T\psi^2$, where T is the number of time series observations. The dotted line represents the percentage bias of the sample estimator $1/\hat{\psi}^2$. The solid line represents the percentage bias of the adjusted estimator $1/\hat{\psi}^2_a$. The percentage biases of the jackknife version of the $\hat{\psi}^2$ and $\hat{\psi}^2_a$ are plotted using the dashed-dotted line and dashed line, respectively. The percentage biases of $1/\hat{\psi}^2$ and its jackknife version are computed under the assumption of $N = 10$ assets and $T = 120$, but the plot is quite insensitive to the assumption of T.

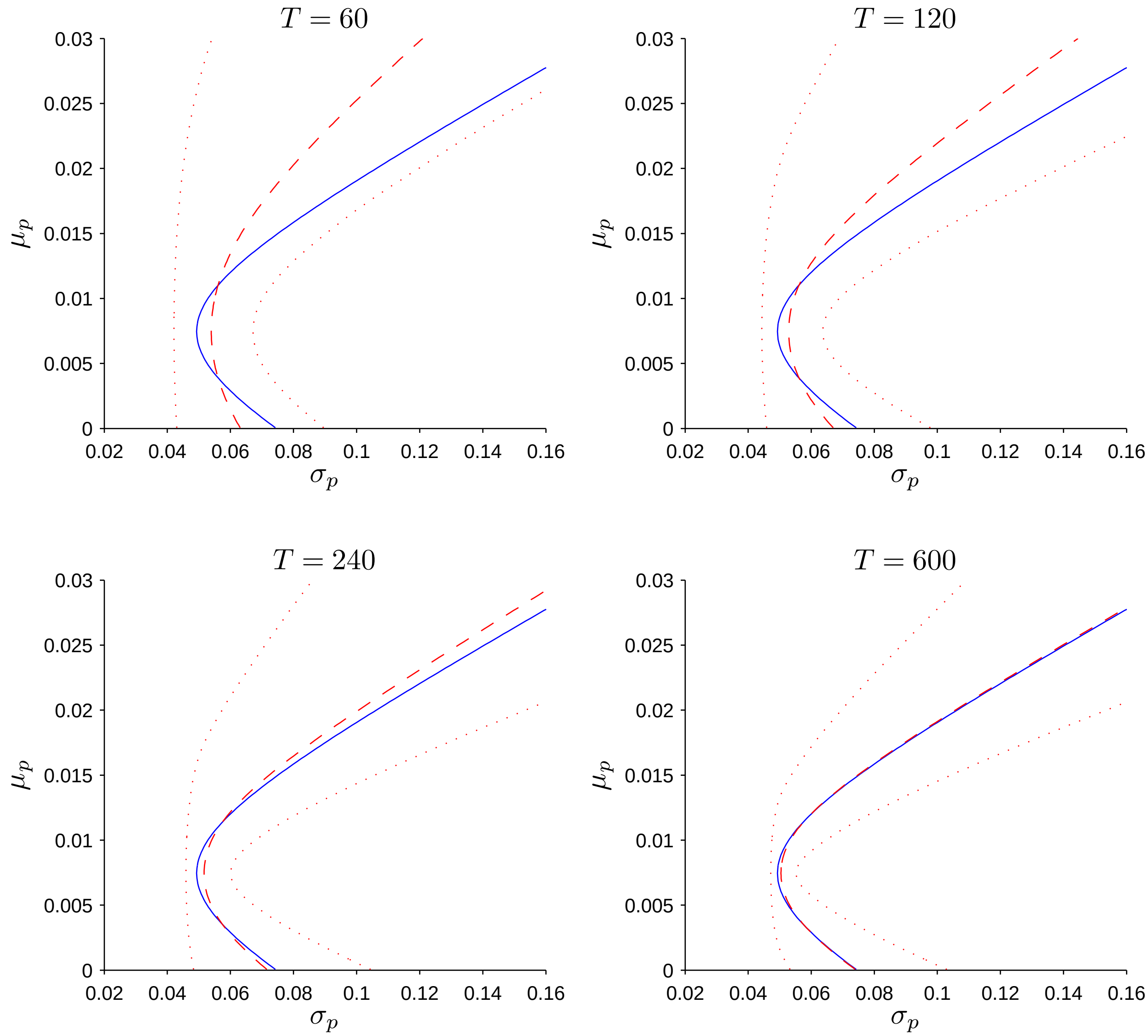

Figure 3. The figure provides information on the distribution of the adjusted sample minimum-variance frontier for different lengths of estimation period (T). The solid line represents the population frontier for $N = 10$ assets, with parameters $\mu_g = 0.00745$, $\sigma_g = 0.04930$, and $\psi = 0.133$. The dashed line represents the expected value of the adjusted sample minimum-variance frontier and the two dotted lines represent the 5th and the 95th percentiles of the adjusted sample minimum-variance frontier, all based on 100,000 simulations.

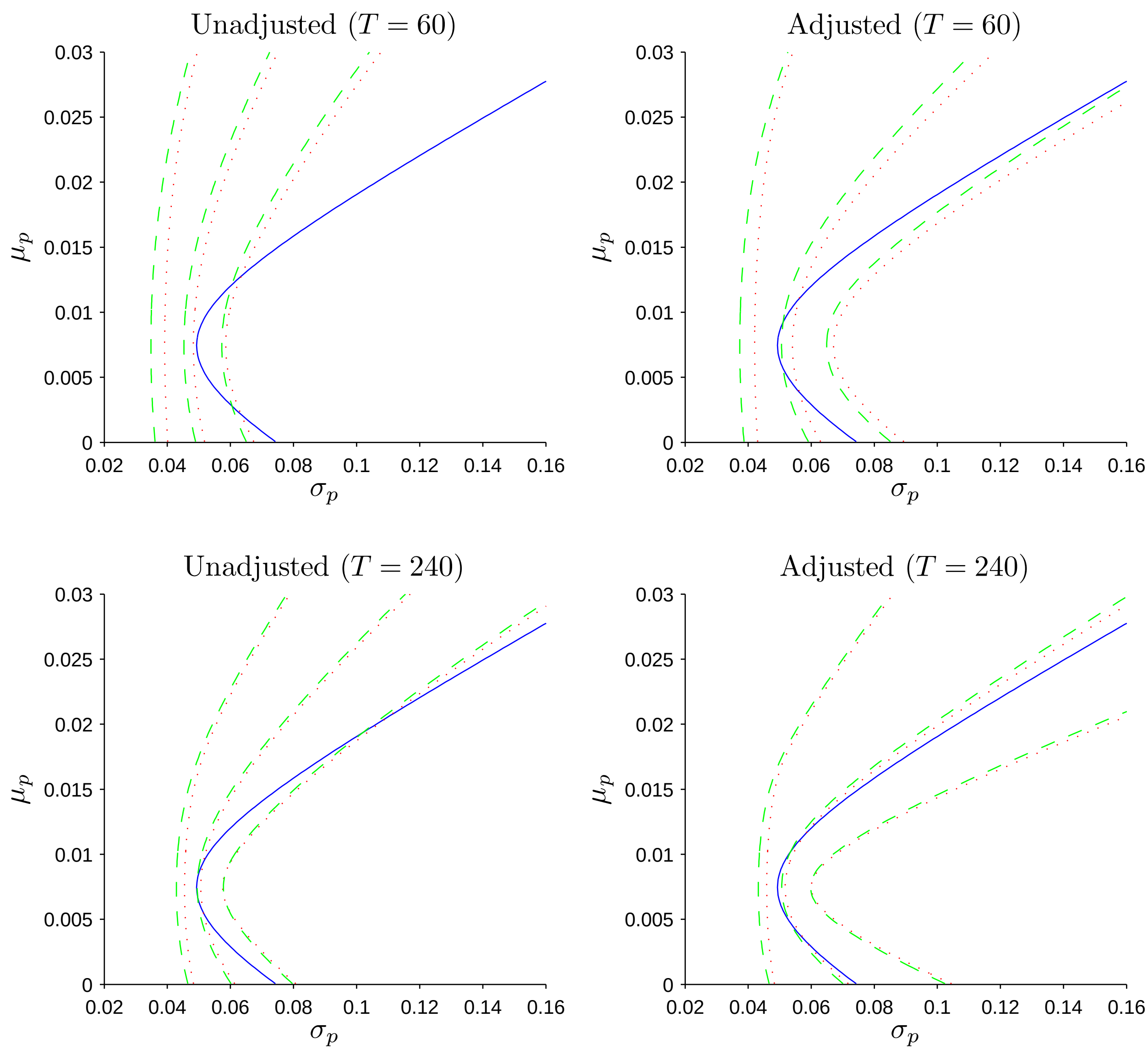

Figure 4. The figure provides information on the distribution of the unadjusted and adjusted sample minimum-variance frontier for different lengths of estimation period (T). The solid line represents the population frontier for $N = 10$ assets, with parameters $\mu_g = 0.00745$, $\sigma_g = 0.04930$, and $\psi = 0.133$. The three dashed lines represents the 5th percentile, expected value, and 95th percentile of the unadjusted and adjusted sample minimum-variance frontier under the assumption that returns are multivariate t-distributed with five degrees of freedom. The three dotted lines are the corresponding ones for the normality case. The distributions of the unadjusted and adjusted frontiers are based on 100,000 simulations.

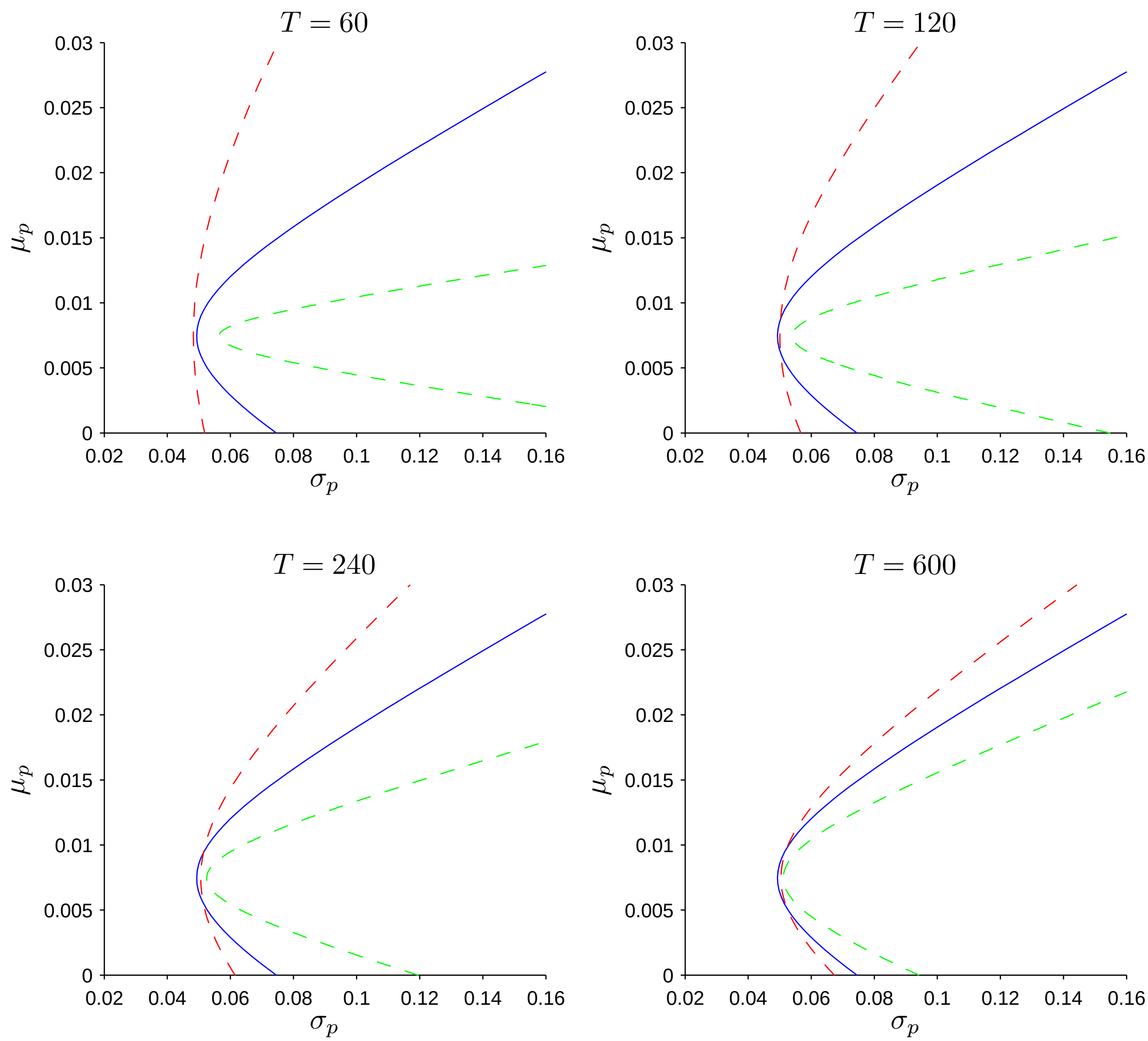

Figure 5. The figure provides information on the average out-of-sample performance of sample minimum-variance frontier for different lengths of estimation period (T). The solid line represents the population frontier for $N = 10$ assets, with parameters $\mu_g = 0.00745$, $\sigma_g = 0.04930$, and $\psi = 0.133$. The outer dashed line represents $\bar{\sigma}_p$, where $\bar{\sigma}_p^2$ is the expected value of the sample minimum-variance frontier. The inner dashed line represents $\underline{\sigma}_p$, where $\underline{\sigma}_p^2$ is the expected value of the out-of-sample variance of the sample minimum-variance frontier.

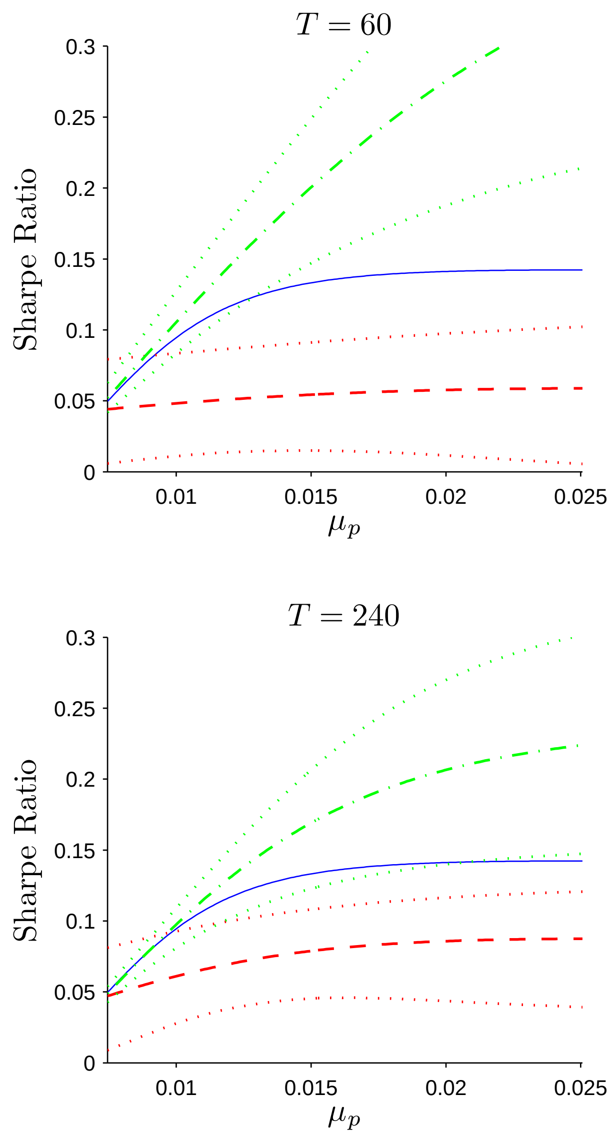

Figure 6. The figure provides information on the distribution of the in-sample and out-of-sample Sharpe ratio of sample minimum-variance portfolios with different target expected return for different lengths of estimation period (T). The solid line represents the Sharpe ratio of the population minimum-variance frontier for $N = 10$ assets, with parameters $\mu_g = 0.00745$, $\sigma_g = 0.04930$, $\psi = 0.133$ and a risk-free rate of 0.005. The dashed-dotted (dashed) line and surrounding dotted lines represent the average, 5th, and 95th percentiles of the in-sample (out-of-sample) Sharpe ratio of the sample minimum-variance portfolio with target expected return μ_p. The distributions of the in-sample and out-of-sample Sharpe ratio are based on 100,000 simulations.

Appendix: Proofs and extensions

A Proofs

Proof of Proposition 1. Let

$$\tilde{A} = [\hat{\mu},\ 1_N]'V^{-1}[\hat{\mu},\ 1_N], \tag{79}$$

$$\hat{A} = [\hat{\mu},\ 1_N]'\hat{V}^{-1}[\hat{\mu},\ 1_N] = \begin{bmatrix} \hat{a} & \hat{b} \\ \hat{b} & \hat{c} \end{bmatrix}. \tag{80}$$

From Theorem 3.2.11 of Muirhead (1982), conditional on $\hat{\mu}$, we have

$$\hat{A}^{-1} = \frac{1}{\hat{a}\hat{c} - \hat{b}^2} \begin{bmatrix} \hat{c} & -\hat{b} \\ -\hat{b} & \hat{a} \end{bmatrix} = \begin{bmatrix} \frac{1}{\hat{\psi}^2} & -\frac{\hat{\mu}_g}{\hat{\psi}^2} \\ -\frac{\hat{\mu}_g}{\hat{\psi}^2} & \hat{\sigma}_g^2 + \frac{\hat{\mu}_g^2}{\hat{\psi}^2} \end{bmatrix} \sim W_2(T - N + 1, \tilde{A}^{-1}/T). \tag{81}$$

Let P be an $N \times (N-1)$ orthonormal matrix such that its columns are orthogonal to $V^{-\frac{1}{2}}1_N$, i.e., $P'V^{-\frac{1}{2}}1_N = 0_{N-1}$. We define $y = \sqrt{T}1_N'V^{-1}\hat{\mu}/\sqrt{c} \sim N(\mu_y, 1)$ and $z = \sqrt{T}P'V^{-\frac{1}{2}}\hat{\mu} \sim N(\mu_z, I_{N-1})$, where $\mu_y = \sqrt{T}b/\sqrt{c} = \sqrt{T}\mu_g/\sigma_g$ and $\mu_z = \sqrt{T}P'V^{-\frac{1}{2}}\mu$. Note that y and z are independent of each other. With y and z defined, we can write

$$T\tilde{A} = \begin{bmatrix} y^2 + u & (Tc)^{\frac{1}{2}}y \\ (Tc)^{\frac{1}{2}}y & Tc \end{bmatrix}, \tag{82}$$

where $u = z'z = T\hat{\mu}[V^{-1} - V^{-1}1_N(1_N'V^{-1}1_N)^{-1}1_N'V^{-1}]\hat{\mu} \sim \chi^2_{N-1}(T\psi^2)$. Taking the inverse of $T\tilde{A}$, we have

$$\frac{\tilde{A}^{-1}}{T} = \begin{bmatrix} \frac{1}{u} & -\frac{y}{\sqrt{Tcu}} \\ -\frac{y}{\sqrt{Tcu}} & \frac{y^2+u}{Tcu} \end{bmatrix} = \begin{bmatrix} \frac{1}{u} & -\frac{\sigma_g y}{\sqrt{Tu}} \\ -\frac{\sigma_g y}{\sqrt{Tu}} & \frac{\sigma_g^2}{T}\left(1 + \frac{y^2}{u}\right) \end{bmatrix}. \tag{83}$$

Let $\hat{A}^{ij}$ be the (i,j)th element of $\hat{A}^{-1}$. Using Theorem 3.2.10 of Muirhead (1982), we have $1/\hat{\psi}^2 = \hat{A}^{11} \sim W_1(T-N+1, 1/u)$ and $\hat{\sigma}_g^2 = \hat{A}^{22} - \hat{A}^{21}(\hat{A}^{11})^{-1}\hat{A}^{12} \sim W_1(T-N, \sigma_g^2/T)$, and they are independent of each other. The first result suggests that $v = u/\hat{\psi}^2 \sim \chi^2_{T-N+1}$ and it is independent of u, so we have $\hat{\psi}^2 = u/v \equiv r$. The second result suggests $\hat{\sigma}_g^2 = \sigma_g^2 q/T$, where $q \sim \chi^2_{T-N}$. The theorem also suggests that when conditional on $\hat{A}^{11}$, or equivalently when conditional on $\hat{\psi}^2$, $\hat{\mu}_g = -\hat{A}^{21}(\hat{A}^{11})^{-1} \sim N(\sigma_g y/\sqrt{T}, \sigma_g^2\hat{\psi}^2/T)$. Therefore, conditional on $\hat{\psi}^2$, we can write

$$\hat{\mu}_g = \frac{\sigma_g(y + \hat{\psi}y_1)}{\sqrt{T}}, \tag{84}$$

where $y_1 \sim N(0,1)$ and it is independent of $\hat{\psi}^2$. Finally, by defining $x = (y + \hat{\psi}y_1 - \sqrt{T}\mu_g/\sigma_g)/(1+\hat{\psi}^2)^{\frac{1}{2}}$, we can write

$$\hat{\mu}_g = \mu_g + \left(\frac{1+\hat{\psi}^2}{T}\right)^{\frac{1}{2}} \sigma_g x. \tag{85}$$

Note that conditional on $\hat{\psi}^2$, $x \sim N(0,1)$ and its distribution does not depend on r and q, so x is independent of r and q. This completes the proof.

Proof of Lemma 1. For constant vectors e, and f, we use Theorem 3.2 of Haff (1979) to obtain the following identities.

$$E[e'\hat{V}^{-1}f] = \frac{T(e'V^{-1}f)}{T-N-2}, \tag{86}$$

$$E[(e'\hat{V}^{-1}e)(e'\hat{V}^{-1}f)] = \frac{T^2(e'V^{-1}e)(e'V^{-1}f)}{(T-N-2)(T-N-4)}, \tag{87}$$

$$E[(e'\hat{V}^{-1}f)^2] = \frac{T^2[(e'V^{-1}e)(f'V^{-1}f) + (T-N-2)(e'V^{-1}f)^2]}{(T-N-1)(T-N-2)(T-N-4)}, \tag{88}$$

$$E[(e'\hat{V}^{-1}e)(f'\hat{V}^{-1}f)] = \frac{T^2[(T-N-3)(e'V^{-1}e)(f'V^{-1}f) + 2(e'V^{-1}f)^2]}{(T-N-1)(T-N-2)(T-N-4)}. \tag{89}$$

For $E[\hat{a}]$, we put $e = f = \hat{\mu}$ in (86). Then using the independence between $\hat{\mu}$ and $\hat{V}$, we have

$$E[\hat{\mu}'\hat{V}^{-1}\hat{\mu}] = E[E[\hat{\mu}'\hat{V}^{-1}\hat{\mu}|\hat{\mu}]] = \frac{E[T\hat{\mu}'V^{-1}\hat{\mu}]}{T-N-2} = \frac{N+Ta}{T-N-2}. \tag{90}$$

The last equality follows because $T\hat{\mu}'V^{-1}\hat{\mu} \sim \chi_N^2(Ta)$ and its expected value is $N+Ta$. The proofs for $E[\hat{b}]$ and $E[\hat{c}]$ are similar. For the second moment of $\hat{a}$, we put $e = f = \hat{\mu}$ in (87) to obtain

$$E[(\hat{\mu}'\hat{V}^{-1}\hat{\mu})^2] = E[E[(\hat{\mu}'\hat{V}^{-1}\hat{\mu})^2|\hat{\mu}]] = \frac{E[(T\hat{\mu}'V^{-1}\hat{\mu})^2]}{(T-N-2)(T-N-4)} = \frac{(N+Ta)^2 + 2(N+2Ta)}{(T-N-2)(T-N-4)}. \tag{91}$$

The last equality follows because the second moment of a $\chi_N^2(Ta)$ random variable is given by $(N+Ta)^2 + 2(N+2Ta)$. Then using $\mathrm{Var}[\hat{a}] = E[\hat{a}^2] - E[\hat{a}]^2$ and with some simplification, we obtain (14). For $E[\hat{a}\hat{b}]$, we put $e = \hat{\mu}$ and $f = 1_N$ in (87) to obtain

$$\begin{aligned} E[(\hat{\mu}'\hat{V}^{-1}\hat{\mu})(\hat{\mu}'\hat{V}^{-1}1_N)] &= E[E[(\hat{\mu}'\hat{V}^{-1}\hat{\mu})(\hat{\mu}'\hat{V}^{-1}1_N)|\hat{\mu}]] \\ &= \frac{E[(T\hat{\mu}'V^{-1}\hat{\mu})(T\hat{\mu}'V^{-1}1_N)]}{(T-N-2)(T-N-4)} \\ &= \frac{(N+2+Ta)Tb}{(T-N-2)(T-N-4)}. \end{aligned} \tag{92}$$

The last equality follows because for $y \sim N(\mu_y, I_N)$, we have $E[(y'y)(y'h)] = (N+2+\mu_y'\mu_y)(\mu_y'h)$ for any constant vector h. Then using $\mathrm{Cov}[\hat{a},\hat{b}] = E[\hat{a}\hat{b}] - E[\hat{a}]E[\hat{b}]$ and with some simplification, we obtain (17). The derivations for the other elements of the covariance matrix are similar. This completes the proof.

Proof of Lemma 2. From the proof of Proposition 1, we have $\hat{\psi}^2 = u/v$ where $u \sim \chi^2_{N-1}(T\psi^2)$, $v \sim \chi^2_{T-N+1}$, and they are independent of each other. Using the fact $E[u] = N-1+T\psi^2$, $E[u^2] = (N-1+T\psi^2)^2 + 2(N-1+2T\psi^2)$, $E[v^{-1}] = 1/(T-N-1)$ and $E[v^{-2}] = 1/[(T-N-1)(T-N-3)]$, we can easily verify the expressions for $E[\hat{\psi}^2]$ and $\mathrm{Var}[\hat{\psi}^2]$. For $\hat{\mu}_g$, we have

$$E[\hat{\mu}_g] = E[E[\hat{\mu}_g|\hat{\psi}^2]] = \mu_g, \tag{93}$$

$$\mathrm{Var}[\hat{\mu}_g] = E[\mathrm{Var}[\hat{\mu}_g|\hat{\psi}^2]] + \mathrm{Var}[E[\hat{\mu}_g|\hat{\psi}^2]] = E\left[\frac{\sigma_g^2(1+\hat{\psi}^2)}{T}\right] = \frac{\sigma_g^2}{T}\left(1 + \frac{N-1+T\psi^2}{T-N-1}\right) \tag{94}$$

For $\hat{\sigma}_g^2$, Proposition 1 suggests that $\hat{\sigma}_g^2 = \sigma_g^2 q/T$, where $q \sim \chi^2_{T-N}$. Using the fact that $E[q] = T-N$ and $\mathrm{Var}[q] = 2(T-N)$, we can easily obtain the expressions for $E[\hat{\sigma}_g^2]$ and $\mathrm{Var}[\hat{\sigma}_g^2]$. For $\mathrm{Cov}[\hat{\mu}_g, \hat{\psi}^2]$, we use

$$\mathrm{Cov}[\hat{\mu}_g, \hat{\psi}^2] = E[\mathrm{Cov}[E[\hat{\mu}_g|\hat{\psi}^2], \hat{\psi}^2]] = \mathrm{Cov}[\mu_g, \hat{\psi}^2] = 0. \tag{95}$$

Finally, $\hat{\sigma}_g^2$ is independent of $\hat{\psi}^2$ and $\hat{\mu}_g$, which implies they are uncorrelated. This completes the proof.

Proof of Lemma 3. All six identities can be proved by using Lemma 1 of Sawa (1972). We only provide the proof of (46) here. The proof of the other identities are similar.[18] Let

[18](41)–(42) are the inverse first and second moment of a noncentral chi-squared distribution, and they were first presented in Krishnan (1967). (43) can be obtained using Lemma 4.2.1 of Muirhead (1982). (45) can be written as the expectation of a ratio of two quadratic forms of z, and its expression is available in Hoque (1985) and Magnus (1986).

$x_1 = z'z$, $x_2 = z'\mu_z$, the joint moment generating function of x_1 and x_2 is given by

$$\begin{aligned}
\phi(\theta_1, \theta_2) &= E[\exp(\theta_1 x_1 + \theta_2 x_2)] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{N-1}{2}}} e^{-\frac{1}{2}(z-\mu_z)'(z-\mu_z)+\theta_1 z'z+\theta_2 z'\mu_z} \mathrm{d}z_1 \ldots \mathrm{d}z_{N-1} \\
&= \frac{1}{(1-2\theta_1)^{\frac{N-1}{2}}} e^{\frac{1}{2}\left[\frac{(1+\theta_2)^2}{1-2\theta_1}-1\right]\mu_z'\mu_z}. \qquad (96)
\end{aligned}$$

From Lemma 1 of Sawa (1972), the expectation of $(x_2/x_1)^2$ is given by

$$E\left[\frac{x_2^2}{x_1^2}\right] = \int_{-\infty}^{0} -\theta_1 \left.\frac{\partial^2 \phi(\theta_1, \theta_2)}{\partial \theta_2^2}\right|_{\theta_2=0} \mathrm{d}\theta_1. \qquad (97)$$

Taking the partial derivative of $\phi(\theta_1, \theta_2)$ with respect to θ_2 twice and setting $\theta_2 = 0$, we have

$$\left.\frac{\partial^2 \phi(\theta_1, \theta_2)}{\partial \theta_2^2}\right|_{\theta_2=0} = \frac{2\theta e^{\theta\left(\frac{1}{1-2\theta_1}-1\right)}}{(1-2\theta_1)^{\frac{N-1}{2}+1}} \left(1 + \frac{2\theta}{1-2\theta_1}\right), \qquad (98)$$

where $\theta = (\mu_z'\mu_z)/2 = T\psi^2/2$. Using a transformation of $y = 1/(1-2\theta_1)$, we can write

$$\begin{aligned}
E\left[\frac{x_2^2}{x_1^2}\right] &= \frac{\theta}{2}\int_0^1 \left[y^{\frac{N-5}{2}} - (1-2\theta)y^{\frac{N-3}{2}} - 2\theta y^{\frac{N-1}{2}}\right] e^{\theta(y-1)}\mathrm{d}y \\
&= \frac{\theta}{2}\int_0^1 y^{\frac{N-5}{2}} e^{\theta(y-1)}\mathrm{d}y - \frac{(1-2\theta)\phi}{2} - \theta^2 \int_0^1 y^{\frac{N-1}{2}} e^{\theta(y-1)}\mathrm{d}y. \qquad (99)
\end{aligned}$$

Using integration by parts, the first term can be written as $\theta(1-\phi)/(N-3)$ and the third term can be written as $\theta - (N-1)\phi/2$. Simplifying the expression, we obtain (46). This completes the proof.

Proof of Propositions 2 and 5. Let $X = V^{-\frac{1}{2}}[\hat{\mu},\ 1_N]$ and $W = TV^{-\frac{1}{2}}\hat{V}V^{-\frac{1}{2}} \sim W_N(T-1, I_N)$. Using X and W, we can rewrite the following two matrices as

$$\hat{A}^{-1}[\hat{\mu},\ 1_N]\hat{V}^{-1}\mu = (X'W^{-1}X)^{-1}X'W^{-1}V^{-\frac{1}{2}}\mu, \qquad (100)$$

$$\hat{A}^{-1}[\hat{\mu},\ 1_N]\hat{V}^{-1}V\hat{V}^{-1}[\hat{\mu},\ 1_N]'\hat{A}^{-1} = (X'W^{-1}X)^{-1}(X'W^{-2}X)(X'W^{-1}X)^{-1}. \qquad (101)$$

Let Q be an $N \times (N-2)$ orthonormal matrix that is orthogonal to X, so $[X(X'X)^{-\frac{1}{2}},\ Q]$ is an orthonormal basis of $\mathbb{R}^N$. It follows that $X(X'X)^{-1}X' + QQ' = I_N$ and we can write

$$\begin{aligned}
(X'W^{-1}X)^{-1}X'W^{-1} &= (X'W^{-1}X)^{-1}X'W^{-1}[X(X'X)^{-1}X' + QQ'] \\
&= (X'X)^{-1}X' + (X'W^{-1}X)^{-1}(X'W^{-1}Q)Q'. \qquad (102)
\end{aligned}$$

Using this identity and the fact that $X'X = \tilde{A}$, where $\tilde{A}$ is defined in (79), we can write

$$\begin{aligned}
& [1,\ \mu_p](X'W^{-1}X)^{-1}X'W^{-1}V^{-\frac{1}{2}}\mu \\
= \ & \check{\mu}_p + [1,\ \mu_p](X'W^{-1}X)^{-1}(X'W^{-1}Q)Q'V^{-\frac{1}{2}}\mu, \qquad (103) \\
& [1,\ \mu_p](X'W^{-1}X)^{-1}(X'W^{-2}X)(X'W^{-1}X)^{-1}[1,\ \mu_p]' \\
= \ & \check{\sigma}_p^2 + [1,\ \mu_p](X'W^{-1}X)^{-1}(X'W^{-1}Q)(Q'W^{-1}X)(X'W^{-1}X)^{-1}[1,\ \mu_p]', \qquad (104)
\end{aligned}$$

where $\check{\mu}_p = [1,\ \mu_p]\tilde{A}^{-1}X'V^{-\frac{1}{2}}\mu$ and $\check{\sigma}_p^2 = [\mu_p,\ 1]\tilde{A}^{-1}[\mu_p,\ 1]'$. Since $\hat{\mu}$ is independent of $\hat{V}$, X and W are independent of each other. Conditional on X, we use Theorem 3.2.11 of Muirhead (1982) to show that

$$B = ([X(X'X)^{-\frac{1}{2}},\ Q]'W^{-1}[X(X'X)^{-\frac{1}{2}},\ Q])^{-1} \sim W_N(T-1, I_N). \qquad (105)$$

Partition B into two by two blocks with dimensions 2 and $N-2$, respectively. Using B_{ij} to denote the (i,j)th block of B, we can write

$$\begin{aligned}
B_{11} - B_{12}B_{22}^{-1}B_{21} &= (X'X)^{\frac{1}{2}}(X'W^{-1}X)^{-1}(X'X)^{\frac{1}{2}} = T\tilde{A}^{\frac{1}{2}}\hat{A}^{-1}\tilde{A}^{\frac{1}{2}}, \qquad (106) \\
-B_{22}^{-1}B_{21} &= Q'W^{-1}X(X'W^{-1}X)^{-1}(X'X)^{\frac{1}{2}} = Q'W^{-1}X(X'W^{-1}X)^{-1}\tilde{A}^{\frac{1}{2}}. \qquad (107)
\end{aligned}$$

From Theorem 3.2.10 of Muirhead (1982), we have $T\tilde{A}^{\frac{1}{2}}\hat{A}^{-1}\tilde{A}^{\frac{1}{2}} \sim W_2(T-N+1, I_2)$ and it is independent of $B_{22}^{-1}B_{21}$. It follows that

$$\hat{\sigma}_p^2 = [\mu_p,\ 1]\hat{A}^{-1}[\mu_p,\ 1]' = [\mu_p,\ 1]\tilde{A}^{-\frac{1}{2}}(\tilde{A}^{\frac{1}{2}}\hat{A}^{-1}\tilde{A}^{\frac{1}{2}})\tilde{A}^{-\frac{1}{2}}[\mu_p,\ 1]' \sim W_1(T-N+1, \check{\sigma}_p^2/T), \qquad (108)$$

so $T\hat{\sigma}_p^2/\check{\sigma}_p^2 \sim \chi^2_{T-N+1} \equiv v$ and v is independent of $\check{\sigma}_p^2$ and $B_{22}^{-1}B_{21}$. This proves Proposition 2.

Let $T_\nu(B,C)$ denotes the matric t-distribution with ν degrees of freedom and B and C are symmetric positive definite matrices.[19] Dickey (1967, p.514) shows that

$$B_{22}^{-1}B_{21} \sim T_{T-N+2}(I_{N-2}, I_2). \qquad (109)$$

It follows that

$$Y = Q'W^{-1}X(X'W^{-1}X)^{-1}[\mu_p,\ 1]' = -B_{22}^{-1}B_{21}\tilde{A}^{-\frac{1}{2}}[\mu_p,\ 1]' \sim T_{T-N+2}(I_{N-2}, \check{\sigma}_p^2), \qquad (110)$$

[19]The density function of $Y \sim T_\nu(B,C)$ is proportional to $|B|^{-\frac{q}{2}}|C|^{-\frac{p}{2}}|I_q + C^{-1}Y'B^{-1}Y|^{-\frac{\nu+p+q-1}{2}}$. Our notation of matric t-distribution differs slightly from the one used by Dickey (1967).

and $(T-N+2)^{\frac{1}{2}}Y/\breve{\sigma}_p$ has a standard multivariate t-distribution with $T-N+2$ degrees of freedom, and we can write

$$Y = \frac{\breve{\sigma}_p x}{\tilde{r}^{\frac{1}{2}}}, \tag{111}$$

where $x \sim N(0_{N-2}, I_{N-2})$ and $\tilde{r} \sim \chi^2_{T-N+2}$. Using (103) and (104), we express $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ as

$$\tilde{\mu}_p = \breve{\mu}_p + Y'Q'V^{-\frac{1}{2}}\mu = \breve{\mu}_p + \frac{x'Q'V^{-\frac{1}{2}}\mu}{\tilde{r}^{\frac{1}{2}}}\breve{\sigma}_p, \tag{112}$$

$$\tilde{\sigma}_p^2 = \breve{\sigma}_p^2 + Y'Y = \breve{\sigma}_p^2\left(1 + \frac{x'x}{\tilde{r}}\right). \tag{113}$$

Let $\tilde{d} = \mu'V^{-\frac{1}{2}}QQ'V^{-\frac{1}{2}}\mu = \mu'V^{-\frac{1}{2}}[I_N - X(X'X)^{-1}X']V^{-\frac{1}{2}}\mu$ and $\tilde{x} = \mu'V^{-\frac{1}{2}}Qx/\tilde{d}^{\frac{1}{2}}$. It is straightforward to show that $\tilde{x} \sim N(0,1)$, $\tilde{q} = x'x - \tilde{x}^2 \sim \chi^2_{N-3}$, and they are independent of each other. Therefore, we can write

$$\tilde{\mu}_p = \breve{\mu}_p + \breve{\sigma}_p\tilde{d}^{\frac{1}{2}}\frac{\tilde{x}}{\tilde{r}^{\frac{1}{2}}}, \tag{114}$$

$$\tilde{\sigma}_p^2 = \breve{\sigma}_p^2\left(1 + \frac{\tilde{x}^2 + \tilde{q}}{\tilde{r}}\right). \tag{115}$$

Let y and z as defined in the proof of Proposition 1, we have $\tilde{y} = \sqrt{T}\mu_p/\sigma_g - y \sim N(\sqrt{T}\delta, 1)$, $\tilde{z} = z'\mu_z/(\mu_z'\mu_z)^{\frac{1}{2}} = z'\mu_z/(\sqrt{T}\psi) \sim N(\sqrt{T}\psi, 1)$ and $\tilde{u} = z'[I_{N-1} - \mu_z(\mu_z'\mu_z)^{-1}\mu_z']z \sim \chi^2_{N-2}$, independent of each other, so $u = z'z = \tilde{z}^2 + \tilde{u}$. Using the representation of $\tilde{A}^{-1}$ in (83) and upon simplification, we have

$$\breve{\mu}_p = \mu_g + \left(\frac{z'\mu_z}{z'z}\right)\frac{\sigma_g\tilde{y}}{\sqrt{T}} = \mu_g + \frac{\psi\tilde{z}}{u}\sigma_g\tilde{y}, \tag{116}$$

$$\tilde{d} = \psi^2 - \frac{(z'\mu_z)^2}{T(z'z)} = \psi^2\left(1 - \frac{\tilde{z}^2}{u}\right) = \frac{\psi^2\tilde{u}}{u}, \tag{117}$$

$$\breve{\sigma}_p^2 = \sigma_g^2 + \frac{\sigma_g^2\tilde{y}^2}{z'z} = \sigma_g^2\left(1 + \frac{\tilde{y}^2}{u}\right). \tag{118}$$

Putting all the terms together, we obtain our expressions for $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$. This completes the proof.

Proof of Proposition 3. Using the independence of v and $\breve{\sigma}_p^2$, we have $\bar{\sigma}_p^2 = E[v/T]E[\breve{\sigma}_p^2] = (T-N+1)E[\breve{\sigma}_p^2]/T$. Using the fact that $E[\tilde{y}^2] = 1 + T\delta^2 = h$, $E[\breve{\sigma}_p^2]$ is given by

$$E[\breve{\sigma}_p^2] = \sigma_g^2(1 + E[u^{-1}]E[\tilde{y}^2]) = \sigma_g^2(1 + hE[u^{-1}]). \tag{119}$$

For $\text{Var}[\hat{\sigma}_p^2]$, we use the fact that $E[v^2] = (T-N+1)(T-N+3)$ and $\text{Var}[v] = 2(T-N+1)$ to obtain

$$\begin{aligned}\text{Var}[\hat{\sigma}_p^2] &= E[\text{Var}[\hat{\sigma}_p^2|v]] + \text{Var}[E[\hat{\sigma}_p^2|v]] \\ &= \frac{E[v^2]}{T^2}\text{Var}[\check{\sigma}_p^2] + \text{Var}\left[\frac{vE[\check{\sigma}_p^2]}{T}\right] \\ &= \frac{(T-N+1)(T-N+3)\text{Var}[\check{\sigma}_p^2]}{T^2} + \frac{2(T-N+1)E[\check{\sigma}_p^2]^2}{T^2}. \end{aligned} \tag{120}$$

For $\text{Var}[\check{\sigma}_p^2]$, we use the fact that $\text{Var}[\tilde{y}^2] = 2(2T\delta^2+1) = 2(2h-1)$ to obtain

$$\begin{aligned}\text{Var}[\check{\sigma}_p^2] &= E[\text{Var}[\check{\sigma}_p^2|u]] + \text{Var}[E[\check{\sigma}_p^2|u]] \\ &= \sigma_g^4 E[u^{-2}]\text{Var}[\tilde{y}^2] + \sigma_g^4 h^2 \text{Var}[u^{-1}] \\ &= \sigma_g^4\left[2(2h-1)E[u^{-2}] + h^2(E[u^{-2}] - E[u^{-1}]^2)\right]. \end{aligned} \tag{121}$$

Putting all the terms together, we obtain our expression of $\text{Var}[\hat{\sigma}_p^2]$. From Johnson, Kotz, and Balakrishnan (1995, Chapter 29), we know that $E[u^{-1}]$ exists if and only if $N > 3$ and $E[u^{-2}]$ exists if and only if $N > 5$. As $h > 0$, $E[\hat{\sigma}_p^2]$ also exists if and only if $N > 3$ and $\text{Var}[\hat{\sigma}_p^2]$ exists if and only if $N > 5$. This completes the proof.

Proof of Lemma 4. From the proof of Proposition 1, we know that $\hat{\psi}^2 = u/v$, where $u = \chi^2_{N-1}(T\psi^2)$ and $v \sim \chi^2_{T-N+1}$ are independent of each other. From Johnson, Kotz, and Balakrishnan (1995, p.484), the density function of $x = \hat{\psi}^2$ is given by

$$f(x) = e^{-\frac{T\psi^2}{2}} \sum_{j=0}^{\infty} q_j \frac{x^{\frac{N-3}{2}+j}}{(1+x)^{\frac{T}{2}+j}}, \tag{122}$$

where $q_j = (T\psi^2/2)^j / \left[B\left(\frac{N-1}{2}+j, \frac{T-N+1}{2}\right) j!\right]$. The expectation of $1/\hat{\psi}_a^2$ is then given by

$$\begin{aligned} E\left[\frac{1}{\hat{\psi}_a^2}\right] &= \frac{Te^{-\frac{T\psi^2}{2}}}{2}\sum_{j=0}^{\infty} q_j \int_0^\infty \frac{\int_0^{\frac{1}{1+x}} u^{\frac{T-N-1}{2}}(1-u)^{\frac{N-5}{2}}\mathrm{d}u}{\left(\frac{1}{1+x}\right)^{\frac{T-N-1}{2}}\left(\frac{x}{1+x}\right)^{\frac{N-3}{2}}} \frac{x^{\frac{N-1}{2}+j-1}}{(1+x)^{\frac{T}{2}+j}}\mathrm{d}x \\ &= \frac{Te^{-\frac{T\psi^2}{2}}}{2}\sum_{j=0}^{\infty} q_j \int_0^\infty \int_0^{\frac{1}{1+x}} \frac{x^j}{(1+x)^{j+2}} u^{\frac{T-N-1}{2}}(1-u)^{\frac{N-5}{2}}\mathrm{d}u\mathrm{d}x \\ &= \frac{Te^{-\frac{T\psi^2}{2}}}{2}\sum_{j=0}^{\infty} q_j \int_0^1 \left[\int_0^{\frac{1-u}{u}} \frac{x^j}{(1+x)^{j+2}}\mathrm{d}x\right] u^{\frac{T-N-1}{2}}(1-u)^{\frac{N-5}{2}}\mathrm{d}u \end{aligned}$$

$$= \frac{Te^{-\frac{T\psi^2}{2}}}{2}\sum_{j=0}^{\infty}\frac{q_j}{(j+1)}\int_0^1 u^{\frac{T-N-1}{2}}(1-u)^{\frac{N-3+j}{2}}\mathrm{d}u$$

$$= \frac{Te^{-\frac{T\psi^2}{2}}}{2}\sum_{j=0}^{\infty}\frac{\left(\frac{T\psi^2}{2}\right)^j}{(j+1)!}$$

$$= \frac{1-e^{-\frac{T\psi^2}{2}}}{\psi^2}. \tag{123}$$

Using Sawa's lemma, we can write the expected value of $1/\hat{\psi}^2$ as

$$E\left[\frac{1}{\hat{\psi}^2}\right] = (T-N+1)E\left[\frac{1}{u}\right] = \frac{T-N+1}{2}\int_0^1 e^{\frac{T\psi^2(y-1)}{2}}y^{\frac{N-5}{2}}\mathrm{d}y. \tag{124}$$

When $N \geq 5$, we have $0 < y^{\frac{N-5}{2}} \leq 1$ for $0 < y < 1$. It follows that

$$E\left[\frac{1}{\hat{\psi}^2}\right] < \frac{T}{2}\int_0^1 e^{\frac{T\psi^2(y-1)}{2}}y^{\frac{N-5}{2}}\mathrm{d}y \leq \frac{T}{2}\int_0^1 e^{\frac{T\psi^2(y-1)}{2}}\mathrm{d}y = \frac{1-e^{-\frac{T\psi^2}{2}}}{\psi^2} = E\left[\frac{1}{\hat{\psi}_a^2}\right]. \tag{125}$$

Therefore, $1/\hat{\psi}_a^2$ has a smaller relative bias than $1/\hat{\psi}^2$. This completes the proof.

Proof of Proposition 4. As in the proofs of Propositions 2 and 5, we define $X = V^{-\frac{1}{2}}[\hat{\mu},\ 1_N]$ and $W = TV^{-\frac{1}{2}}\hat{V}V^{-\frac{1}{2}} \sim W_N(T-1, I_N)$, which are independent of each other since $\hat{\mu}$ is independent of $\hat{V}$. Let Q be an $N\times(N-2)$ orthonormal matrix that is orthogonal to X, we use (102) to write $\hat{w}$ in (30) as

$$\hat{w} = V^{-\frac{1}{2}}W^{-1}X(X'W^{-1}X)^{-1}[\mu_p,\ 1]' = \tilde{w} + V^{-\frac{1}{2}}Q(Q'W^{-1}X)(X'W^{-1}X)^{-1}[\mu_p,\ 1]', \tag{126}$$

where

$$\tilde{w} = V^{-1}[\hat{\mu},\ 1_N]\tilde{A}^{-1}[\mu_p,\ 1]'. \tag{127}$$

Recall from the proof of Propositions 2 and 5 that conditional on $\hat{\mu}$, we have

$$Y = Q'W^{-1}X(X'W^{-1}X)^{-1}[\mu_p,\ 1]' = \frac{\check{\sigma}_p x}{\tilde{r}^{\frac{1}{2}}}, \tag{128}$$

where

$$\check{\sigma}_p^2 = [\mu_p,\ 1]\tilde{A}^{-1}[\mu_p,\ 1]' = \sigma_g^2\left(1+\frac{\tilde{y}^2}{u}\right), \tag{129}$$

with $\tilde{y} = \frac{\sqrt{T}\mu_p}{\sigma_g} - y \sim N(\sqrt{T}\delta, 1)$, $x \sim N(0_{N-2}, I_{N-2})$ and $\tilde{r} \sim \chi^2_{T-N+2}$ are independent of each other, and $\delta = (\mu_p - \mu_g)/\sigma_g$. Using these transformations, we obtain (54).

By using the expression of $\tilde{A}^{-1}$ in (83), we can further simplify $\tilde{w}$ to

$$\begin{aligned}
\tilde{w} &= V^{-1}[\hat{\mu},\ 1_N]\tilde{A}^{-1}[\mu_p,\ 1]' \\
&= TV^{-1}[\hat{\mu},\ 1_N]\begin{bmatrix} \frac{1}{u} & -\frac{\sigma_g y}{\sqrt{T}u} \\ -\frac{\sigma_g y}{\sqrt{T}u} & \frac{\sigma_g^2}{T}\left(1+\frac{y^2}{u}\right) \end{bmatrix}\begin{bmatrix} \mu_p \\ 1 \end{bmatrix} \\
&= TV^{-1}[\hat{\mu},\ 1_N]\begin{bmatrix} \frac{1}{u}\left(\mu_p - \frac{\sigma_g}{\sqrt{T}}y\right) \\ \frac{\sigma_g^2}{T} - \frac{\sigma_g y}{\sqrt{T}u}\left(\mu_p - \frac{\sigma_g}{\sqrt{T}}y\right) \end{bmatrix} \\
&= \frac{V^{-1}1_N}{1_N'V^{-1}1_N} + \left(\mu_p - \frac{\sigma_g}{\sqrt{T}}y\right)\frac{T}{u}V^{-1}[\hat{\mu},\ 1_N]\begin{bmatrix} 1 \\ -\frac{\sigma_g y}{\sqrt{T}} \end{bmatrix} \\
&= w_g + \left(\mu_p - \frac{\sigma_g}{\sqrt{T}}y\right)\frac{TV^{-\frac{1}{2}}[I_N - V^{-\frac{1}{2}}1_N(1_N'V^{-1}1_N)^{-1}1_N'V^{-\frac{1}{2}}]V^{-\frac{1}{2}}\hat{\mu}}{u} \\
&= w_g + \sigma_g\tilde{y}\frac{\sqrt{T}V^{-\frac{1}{2}}PP'V^{-\frac{1}{2}}\hat{\mu}}{u} \\
&= w_g + \sigma_g\tilde{y}\frac{V^{-\frac{1}{2}}Pz}{u},
\end{aligned} \tag{130}$$

where $\tilde{y} \sim N(\sqrt{T}\delta, 1)$ and $z \sim N(\sqrt{T}P'V^{-\frac{1}{2}}\mu, I_{N-1})$ are independent of each other and $u = z'z$.

In order to obtain the mean and the covariance matrix of $\hat{w}$, we need the following identities:

$$E\left[\frac{z}{z'z}\right] = \frac{\phi}{T\psi^2}\mu_z \quad \text{for } N > 2, \tag{131}$$

$$E\left[\frac{zz'}{z'z}\right] = \frac{\phi}{T\psi^2}I_{N-1} + \left(\frac{1}{T\psi^2} - \frac{(N-1)\phi}{T^2\psi^4}\right)\mu_z\mu_z' \quad \text{for } N > 1, \tag{132}$$

$$E\left[\frac{zz'}{(z'z)^2}\right] = \left(\frac{1-\phi}{2(N-3)} - \frac{\phi}{2T\psi^2}\right)I_{N-1} + \left(\frac{(N-1)\phi}{2T^2\psi^4} - \frac{1-\phi}{2T\psi^2}\right)\mu_z\mu_z' \quad \text{for } N > 3. \tag{133}$$

(131) can be proved using the Sawa's lemma as in the proof of Lemma 3. For (132) and (133), we can use Proposition 7 of Hillier, Kan, and Wang (2008) to obtain the following identity

$$\begin{aligned}
E\left[\frac{z'Az}{(z'z)^q}\right] &= \frac{\mathrm{tr}(A)}{2^q\left(\frac{N+1}{2}-q\right)_q}{}_1F_1\left(q; \frac{N+1}{2}; -\frac{T\psi^2}{2}\right) \\
&\quad + \frac{\mu_z'A\mu_z}{2^q\left(\frac{N+3}{2}-q\right)_q}{}_1F_1\left(1; \frac{N+3}{2}; -\frac{T\psi^2}{2}\right) \quad \text{for } N > 2q-1,
\end{aligned} \tag{134}$$

where q is a positive integer and A is a square matrix. In particular, if we choose A to be a zero matrix except its (i,j)th element is equal to one, we have

$$E\left[\frac{z_i z_j}{(z'z)^q}\right] = \frac{\delta_{ij}}{2^q\left(\frac{N+1}{2}-q\right)_q} {}_1F_1\left(q;\frac{N+1}{2};-\frac{T\psi^2}{2}\right) + \frac{E[z_i]E[z_j]}{2^q\left(\frac{N+3}{2}-q\right)_q} {}_1F_1\left(q;\frac{N+3}{2};-\frac{T\psi^2}{2}\right) \quad \text{for } N>2q-1, \quad (135)$$

where $\delta_{ij}=1$ when $i=j$ and zero otherwise. Putting together the elements, we obtain the following identity

$$E\left[\frac{zz'}{(z'z)^q}\right] = \frac{{}_1F_1\left(q;\frac{N+1}{2};-\frac{T\psi^2}{2}\right)}{2^q\left(\frac{N+1}{2}-q\right)_q} I_{N-1} + \frac{{}_1F_1\left(q;\frac{N+3}{2};-\frac{T\psi^2}{2}\right)}{2^q\left(\frac{N+3}{2}-q\right)_q}\mu_z\mu_z' \quad \text{for } N>2q-1. \quad (136)$$

By using the definition of ϕ as in (39) and the recurrence relation of the confluent hypergeometric function, we can write the following expressions in terms of ϕ

$$\frac{{}_1F_1\left(1;\frac{N+1}{2};-\frac{T\psi^2}{2}\right)}{N-1} = \frac{\phi}{T\psi^2}, \quad (137)$$

$$\frac{{}_1F_1\left(1;\frac{N+3}{2};-\frac{T\psi^2}{2}\right)}{N+1} = \frac{1}{T\psi^2} - \frac{(N-1)\phi}{T^2\psi^4}, \quad (138)$$

$$\frac{{}_1F_1\left(2;\frac{N+1}{2};-\frac{T\psi^2}{2}\right)}{(N-1)(N-3)} = \frac{1-\phi}{2(N-3)} - \frac{\phi}{2T\psi^2}, \quad (139)$$

$$\frac{{}_1F_1\left(2;\frac{N+3}{2};-\frac{T\psi^2}{2}\right)}{(N-1)(N+1)} = \frac{(N-1)\phi}{2T^2\psi^4} - \frac{1-\phi}{2T\psi^2}. \quad (140)$$

Finally, putting $q=1$ and $q=2$ in (136) and using (137)–(140), we prove (132) and (133).

With (131)–(133) proved, we proceed to derive the mean and the covariance matrix of $\hat{w}$. Since $E[x]=0_{N-2}$ and x is independent of $\hat{\mu}$ and $\tilde{r}$, we have

$$E[\hat{w}] = E[\tilde{w}] = w_g + (\mu_p-\mu_g)\sqrt{T}V^{-\frac{1}{2}}PE\left[\frac{z}{u}\right], \quad (141)$$

using the fact that $\tilde{y}$ is independent of z and $E[\tilde{y}]=\sqrt{T}\delta$. Then, using (131) and the fact that $\mu_z=\sqrt{T}P'V^{-\frac{1}{2}}\mu$, we obtain

$$\sqrt{T}V^{-\frac{1}{2}}PE\left[\frac{z}{u}\right] = \frac{\phi}{\psi^2}V^{-\frac{1}{2}}PP'V^{-\frac{1}{2}}\mu$$

$$= \frac{\phi}{\psi^2} V^{-\frac{1}{2}} \left(I_N - \frac{V^{-\frac{1}{2}} 1_N 1_N' V^{-\frac{1}{2}}}{1_N' V^{-1} 1_N} \right) V^{-\frac{1}{2}} \mu$$

$$= \frac{\phi}{\psi^2} V^{-1}(\mu - \mu_g 1_N) \tag{142}$$

for $N > 2$. It follows that the expected value of $\hat{w}$ when $N > 2$ is given by

$$E[\hat{w}] = w_g + \phi \frac{(\mu_p - \mu_g)}{\psi^2} V^{-1}(\mu - \mu_g 1_N) = w_g + \phi(w - w_g) \tag{143}$$

using the fact that

$$\begin{aligned} w &= V^{-1}[\mu,\ 1_N] A^{-1} [\mu_p,\ 1]' \\ &= \frac{V^{-1} 1_N}{1_N' V^{-1} 1_N} + \frac{(\mu_p - \mu_g)}{\psi^2} V^{-1}(\mu - \mu_g 1_N) \\ &= w_g + \frac{(\mu_p - \mu_g)}{\psi^2} V^{-1}(\mu - \mu_g 1_N). \end{aligned} \tag{144}$$

Since x has mean zero and is independent of $\hat{\mu}$, the two terms of $\hat{w}$ in (54) are uncorrelated and the second term has mean zero, so we have

$$\text{Var}[\hat{w}] = \text{Var}[\tilde{w}] + V^{-\frac{1}{2}} E\left[\frac{\check{\sigma}_p^2 Q x x' Q'}{\tilde{r}} \right] V^{-\frac{1}{2}}. \tag{145}$$

In order to determine the covariance matrix of $\tilde{w}$, we first obtain

$$\begin{aligned} E[\tilde{w}\tilde{w}'] &= w_g w_g' + (\mu_p - \mu_g) E\left[\frac{\sqrt{T} V^{-\frac{1}{2}} P z w_g'}{u} \right] + (\mu_p - \mu_g) E\left[\frac{\sqrt{T} w_g z' P' V^{-\frac{1}{2}}}{u} \right] \\ &\quad + E\left[\left(\mu_p - \frac{\sigma_g}{\sqrt{T}} y \right)^2 \right] E\left[\frac{T V^{-\frac{1}{2}} P z z' P' V^{-\frac{1}{2}}}{u^2} \right] \\ &= w_g w_g' + \phi(w - w_g) w_g' + \phi w_g (w - w_g)' \\ &\quad + \left[(\mu_p - \mu_g)^2 + \frac{\sigma_g^2}{T} \right] E\left[\frac{T V^{-\frac{1}{2}} P z z' P' V^{-\frac{1}{2}}}{u^2} \right]. \end{aligned} \tag{146}$$

Denoting $h = T\delta^2 + 1$, we obtain for $N > 3$

$$\text{Var}[\tilde{w}] = \sigma_g^2 h V^{-\frac{1}{2}} P E\left[\frac{z z'}{(z'z)^2} \right] P' V^{-\frac{1}{2}} - \phi^2 (w - w_g)(w - w_g)'. \tag{147}$$

Using the fact that $E[xx'] = I_{N-2}$, $E[1/\tilde{r}] = 1/(T - N)$ and x, $\tilde{r}$, and Q are independent of each other, we obtain

$$E\left[\frac{\check{\sigma}_p^2 Q x x' Q'}{\tilde{r}} \right] = \frac{1}{(T - N)} E\left[\check{\sigma}_p^2 V^{-\frac{1}{2}} Q Q' V^{-\frac{1}{2}} \right]. \tag{148}$$

Since Q is orthogonal to $V^{-\frac{1}{2}}1_N$ and $V^{-\frac{1}{2}}\hat{\mu}$, we can write QQ' as[20]

$$QQ' = PP' - \frac{PP'V^{-\frac{1}{2}}\hat{\mu}'\hat{\mu}V^{-\frac{1}{2}}PP'}{\hat{\mu}'V^{-\frac{1}{2}}PP'V^{-\frac{1}{2}}\hat{\mu}} = PP' - \frac{Pzz'P'}{z'z}. \tag{149}$$

As $\breve{\sigma}_p^2 = \sigma_g^2\left(1 + \frac{\tilde{y}^2}{u}\right)$ and $\tilde{y}$ is independent of u, we have

$$\begin{aligned} E\left[\breve{\sigma}_p^2 V^{-\frac{1}{2}}QQ'V^{-\frac{1}{2}}\right] &= E[\breve{\sigma}_p^2]V^{-\frac{1}{2}}PP'V^{-\frac{1}{2}} - \sigma_g^2 V^{-\frac{1}{2}}PE\left[\frac{zz'}{z'z}\right]P'V^{-\frac{1}{2}} \\ &\quad - \sigma_g^2 hV^{-\frac{1}{2}}PE\left[\frac{zz'}{(z'z)^2}\right]P'V^{-\frac{1}{2}} \end{aligned} \tag{150}$$

using the fact that $E[\tilde{y}^2] = h$. Combining these two terms, we have for $N > 3$,

$$\begin{aligned} \mathrm{Var}[\hat{w}] &= \mathrm{Var}[\tilde{w}] + \frac{E\left[\breve{\sigma}_p^2 V^{-\frac{1}{2}}QQ'V^{-\frac{1}{2}}\right]}{(T-N)} \\ &= \left(\frac{T-N-1}{T-N}\right)\sigma_g^2 hV^{-\frac{1}{2}}PE\left[\frac{zz'}{(z'z)^2}\right]P'V^{-\frac{1}{2}} + \frac{E[\breve{\sigma}_p^2]}{T-N}V^{-\frac{1}{2}}PP'V^{-\frac{1}{2}} \\ &\quad - \frac{\sigma_g^2}{T-N}V^{-\frac{1}{2}}PE\left[\frac{zz'}{z'z}\right]P'V^{-\frac{1}{2}} - \phi^2(w-w_g)(w-w_g)'. \end{aligned} \tag{151}$$

Substituting in (132) and (133) and using the definitions of V_a and V_b, we obtain (56). This completes the proof.

Proof of Proposition 6. We follow the notation in Proposition 5. For $\underline{\mu}_p$, the last term of $\tilde{\mu}_p$ has mean zero because $E[\tilde{t}] = 0$, so we have from (116)

$$\underline{\mu}_p = E\left[\mu_g + \left(\frac{z'\mu_z}{z'z}\right)\frac{\sigma_g\tilde{y}}{\sqrt{T}}\right] = \mu_g + E\left[\frac{z'\mu_z}{z'z}\right](\mu_p - \mu_g) = \mu_g + \phi(\mu_p - \mu_g). \tag{152}$$

The second equality relies on the fact that $\tilde{y}$ and z are independent, and the last equality uses (43) to obtain $E[z'\mu_z/(z'z)] = \phi$, which exists if and only if $N > 2$. For $\underline{\sigma}_p^2$, we have

$$\underline{\sigma}_p^2 = E\left[\left(1 + \frac{\tilde{x}^2 + \tilde{q}}{\tilde{r}}\right)\breve{\sigma}_p^2\right]\left(1 + \frac{N-2}{T-N}\right)E[\breve{\sigma}_p^2]. \tag{153}$$

The second equality relies on the fact that $\breve{\sigma}_p^2$ (which is only a function of $\tilde{z}$, $\tilde{u}$ and $\tilde{y}$) is independent of $\tilde{x}$, $\tilde{q}$ and $\tilde{r}$.

[20]This expression can also be directly obtained by using the fact that $QQ' = I_N - X(X'X)^{-1}X' = I_N - X\tilde{A}^{-1}X'$.

The proof of the variances and covariance requires repeated use of the two identities for unconditional variance and covariance $\text{Var}[X] = E[\text{Var}[X|Z]] + \text{Var}[E[X|Z]]$ and $\text{Cov}[X,Y] = E[\text{Cov}[X,Y|Z]] + \text{Cov}[E[X|Z], E[Y|Z]]$. We only provide the proof of $\text{Var}[\tilde{\mu}_p]$ here, the proof for the other elements are similar.

$$
\begin{aligned}
\text{Var}[\tilde{\mu}_p] &= \text{Var}\left[\psi\sigma_g \frac{\tilde{z}}{u}\tilde{y} + \psi\check{\sigma}_p\left(\frac{\tilde{u}}{u}\right)^{\frac{1}{2}}\tilde{t}\right] \\
&= \frac{1}{T-N}E\left[\psi^2\check{\sigma}_p^2\frac{\tilde{u}}{u}\right] + \text{Var}\left[\left(\frac{z'\mu_z}{z'z}\right)\frac{\sigma_g\tilde{y}}{\sqrt{T}}\right] \\
&= \frac{\psi^2 E[\check{\sigma}_p^2] - E\left[\frac{\check{\sigma}_p^2}{T}\frac{(z'\mu_z)^2}{z'z}\right]}{T-N} + E\left[\left(\frac{z'\mu_z}{z'z}\right)^2\frac{\sigma_g^2}{T}\right] + \text{Var}\left[\left(\frac{z'\mu_z}{z'z}\right)(\mu_p - \mu_g)\right] \\
&= \frac{\psi^2 E[\check{\sigma}_p^2] - \frac{\sigma_g^2}{T}\left(E\left[\frac{(z'\mu_z)^2}{z'z}\right] + hE\left[\frac{(z'\mu_z)^2}{(z'z)^2}\right]\right)}{T-N} + \frac{\sigma_g^2}{T}E\left[\frac{(z'\mu_z)^2}{(z'z)^2}\right] \\
&\quad + (\mu_p-\mu_g)^2\left(E\left[\frac{(z'\mu_z)^2}{(z'z)^2}\right] - E\left[\frac{z'\mu_z}{z'z}\right]^2\right) \\
&= \frac{\psi^2 E[\check{\sigma}_p^2] - \frac{\sigma_g^2}{T}\left(E\left[\frac{(z'\mu_z)^2}{z'z}\right] + hE\left[\frac{(z'\mu_z)^2}{(z'z)^2}\right]\right)}{T-N} + \frac{\sigma_g^2 h}{T}E\left[\frac{(z'\mu_z)^2}{(z'z)^2}\right] - \phi^2(\mu_p-\mu_g)^2. \qquad (154)
\end{aligned}
$$

The second equality is obtained by conditional on $\tilde{t}$ and using the fact that $\text{Var}[\tilde{t}] = E[\tilde{t}^2] = 1/(T-N)$. The third equality is obtained by conditional on $\tilde{y}$. Rearranging the terms, we obtain $\text{Var}[\tilde{\mu}_p]$. Note that $\text{Var}[\tilde{\mu}_p]$ exists if and only if $E[(z'\mu_z)^2/(z'z)^2]$ exists, i.e., $N > 3$. This completes the proof.

Proof of Proposition 7. From Proposition 1, we know $E[\hat{\mu}_g|\hat{\psi}^2] = \mu_g$. From (41), we have $E[\hat{\psi}^{-2}] = (T-N+1)(1-\phi)/(N-3)$, so we have

$$
E[\underline{\hat{\mu}}_{pu}] = \mu_p - \frac{N-3}{T-N+1}\left[\frac{T-N+1}{N-3}(1-\phi)(\mu_p-\mu_g)\right] = \mu_p - (1-\phi)(\mu_p-\mu_g) = \underline{\mu}_p. \quad (155)
$$

The unbiasedness of $\underline{\hat{\sigma}}_{pu}^2$ follows from (47) and (66). This completes the proof.

B Application using 25 portfolios

In this example, we assume there are $N = 25$ risky assets, with population minimum-variance frontier determined by the following parameters: $\mu_g = 0.01199$, $\sigma_g = 0.04108$, and $\psi = 0.209$. These parameters are chosen based on the unbiased estimates of μ_g, σ_g^2 and ψ^2 using monthly value-weighted returns from the popular Fama and French (1993) 25 portfolios over the period 1932/1–2004/12, formed based on size and book-to-market ratio.[21] Figure A.1 provides the distribution of the sample minimum-variance frontier of these 25 assets for different lengths of estimation period (T). As compared with Figure 1, we find that with more assets, the sample minimum-variance frontier exhibits a larger bias. In Figure A.1, we find that a significant part of the population frontier still falls outside of the 95th percentile of the sample frontier even when the estimation window is as long as $T = 600$ months. This suggests that with the typical length of time series that we use, the sample frontier is an extremely unreliable estimator of the population frontier.

In Figure A.2 we plot the percentage biases of different estimators of $1/\psi^2$. Compared with Figure 2, we observe that the percentage bias of $1/\hat{\psi}^2$ is larger when $N = 25$. For example, when $T\psi^2 = 4$, the percentage biases of $1/\hat{\psi}^2$ and its jackknife version have percentage biases of -87.6% and -73.9%, respectively. Even for $T\psi^2$ as large as 20, the expectation of $1/\hat{\psi}^2$ is still less than 40% of the true value of $1/\psi^2$.

Figure A.3 plots the adjusted frontier and we find it performs dramatically better than the sample frontier plotted in Figure A.1.

In Figure A.4, we provide evidence on the performance of the unadjusted and adjusted frontiers when returns are non-normal for the $N = 25$ assets case. The returns are assumed to follow a multivariate t-distribution with five degrees of freedom and have the same μ and V as in the normality case for Figures A.1 and A.3. The results are broadly consistent with the $N = 10$ assets case reported in Figure 4. However, with more assets, the fat tails in the t-distribution have a much stronger impact on the finite sample distribution of $\hat{\sigma}_p$ and $\hat{\sigma}_{pa}$, especially for $T = 60$. In general, the fat tails lower $\hat{\sigma}_p^2$ and $\hat{\sigma}_{pa}^2$ so that they have a bigger bias than the normality case. The effect of the fat tails is much smaller as T increases to 240 but there is still a noticeable difference between the finite sample distributions under the two distribution assumptions. In Figure A.4, we provide evidence on the performance of the unadjusted and adjusted frontiers when returns are non-normal for the $N = 25$ assets case using the same parameters for Figures A.1 and A.3. The results are broadly consistent with

[21] We are grateful to Ken French for making this data available on his website.

the $N = 10$ assets case reported in Figure 4. However, with more assets, the fat tails in the t-distribution have a much stronger impact on the finite sample distribution of $\hat{\sigma}_p$ and $\hat{\sigma}_{pa}$, especially for $T = 60$. In general, the fat tails lower $\hat{\sigma}_p^2$ and $\hat{\sigma}_{pa}^2$ so that they have a bigger bias than the normality case. The effect of the fat tails is much smaller as T increases to 240 but there is still a noticeable difference between the finite sample distributions under the two distribution assumptions.

Figure A.5 compares the population frontier with the in-sample and out-of-sample frontiers for the $N = 25$ assets case. We observe the downward bias of the in-sample frontier and the upward bias of the out-of-sample frontier are more severe when there are more assets.

Figure A.6 provides information about the distribution of $\hat{\theta}_p$ and $\tilde{\theta}_p$ for the $N = 25$ assets case. The patterns in Figure A.5 are largely similar to those in Figure 6 (for the $N = 10$ assets case), but with more assets, the overoptimism of the in-sample Sharpe ratio becomes even more pronounced.

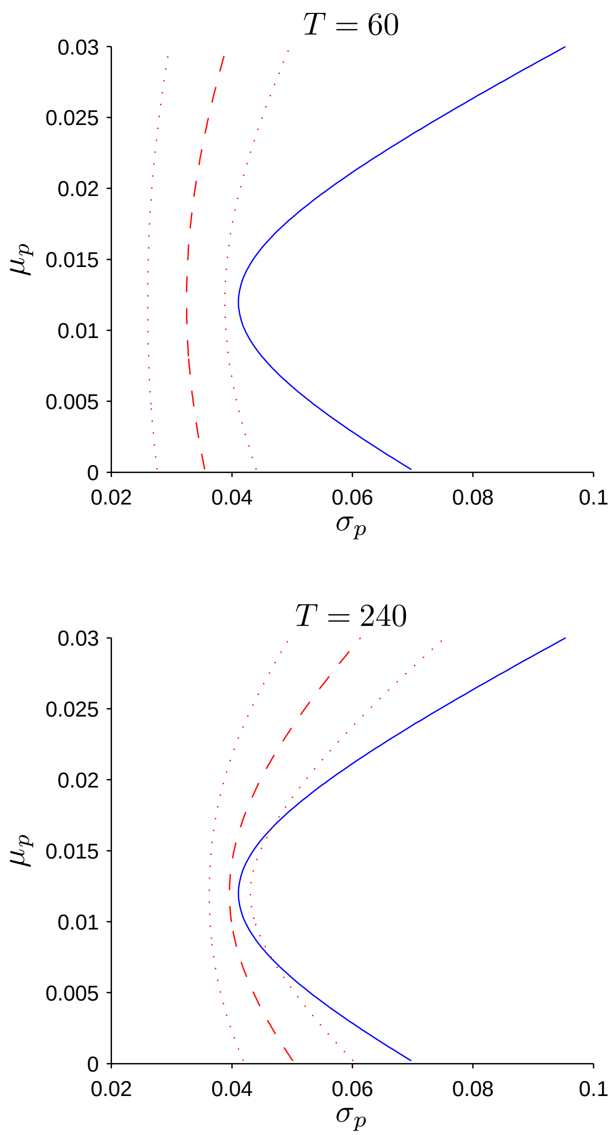

Figure A.1. The figure provides information on the distribution of the sample minimum-variance frontier for different lengths of estimation period (T). The solid line represents the population frontier for $N = 25$ assets, with parameters $\mu_g = 0.01199$, $\sigma_g = 0.04108$, and $\psi = 0.209$. The dashed line represents $\bar{\sigma}_p$, where $\bar{\sigma}_p^2$ is the expected value of the sample minimum-variance frontier. The two dotted lines represent the 5th and the 95th percentiles of the sample minimum-variance frontier, based on 100,000 simulations.

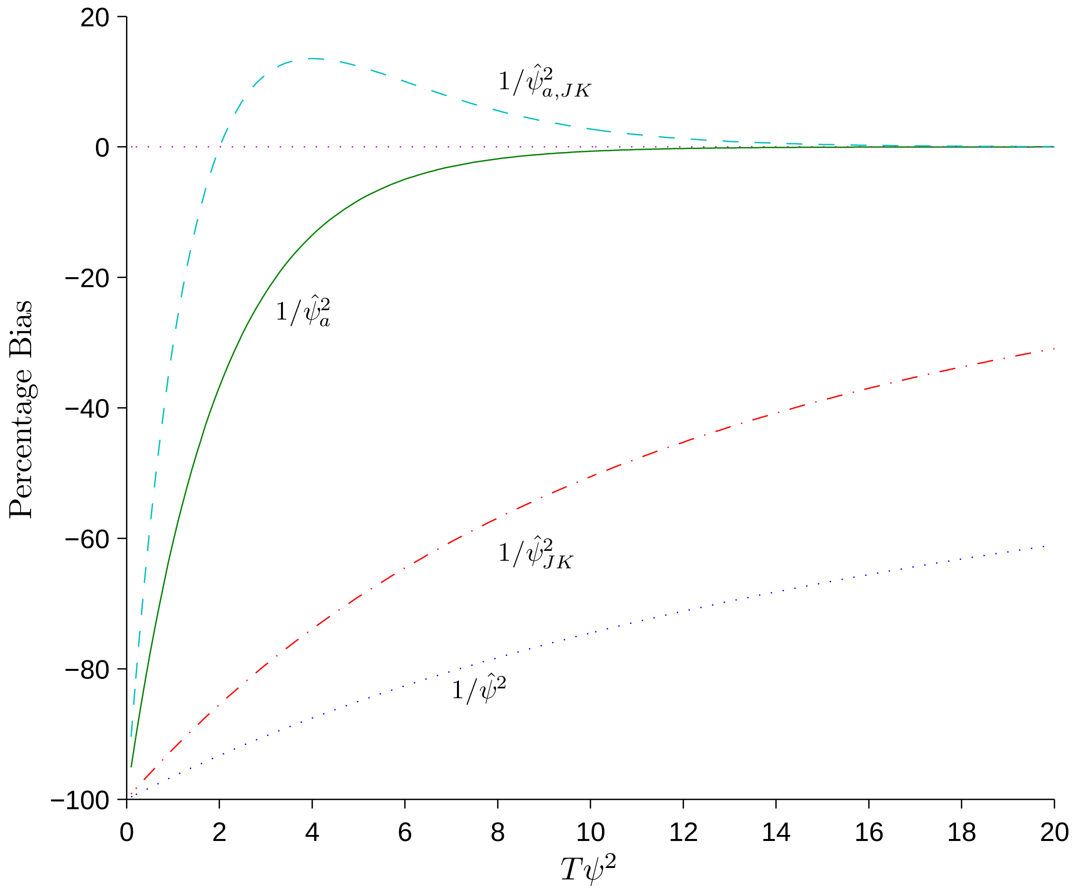

Figure A.2. The figure plots the percentage biases of four estimators of the inverse of the squared slope of the asymptote to the minimum-variance frontier $(1/\psi^2)$ as a function of $T\psi^2$, where T is the number of time series observations. The dotted line represents the percentage bias of the sample estimator $1/\hat{\psi}^2$. The solid line represents the percentage bias of the adjusted estimator $1/\hat{\psi}^2_a$. The percentage biases of the jackknife version of the $\hat{\psi}^2$ and $\hat{\psi}^2_a$ are plotted using the dashed-dotted line and dashed line, respectively. The percentage biases of $1/\hat{\psi}^2$ and its jackknife version are computed under the assumption of $N = 25$ assets and $T = 120$, but the plot is quite insensitive to the assumption of T.

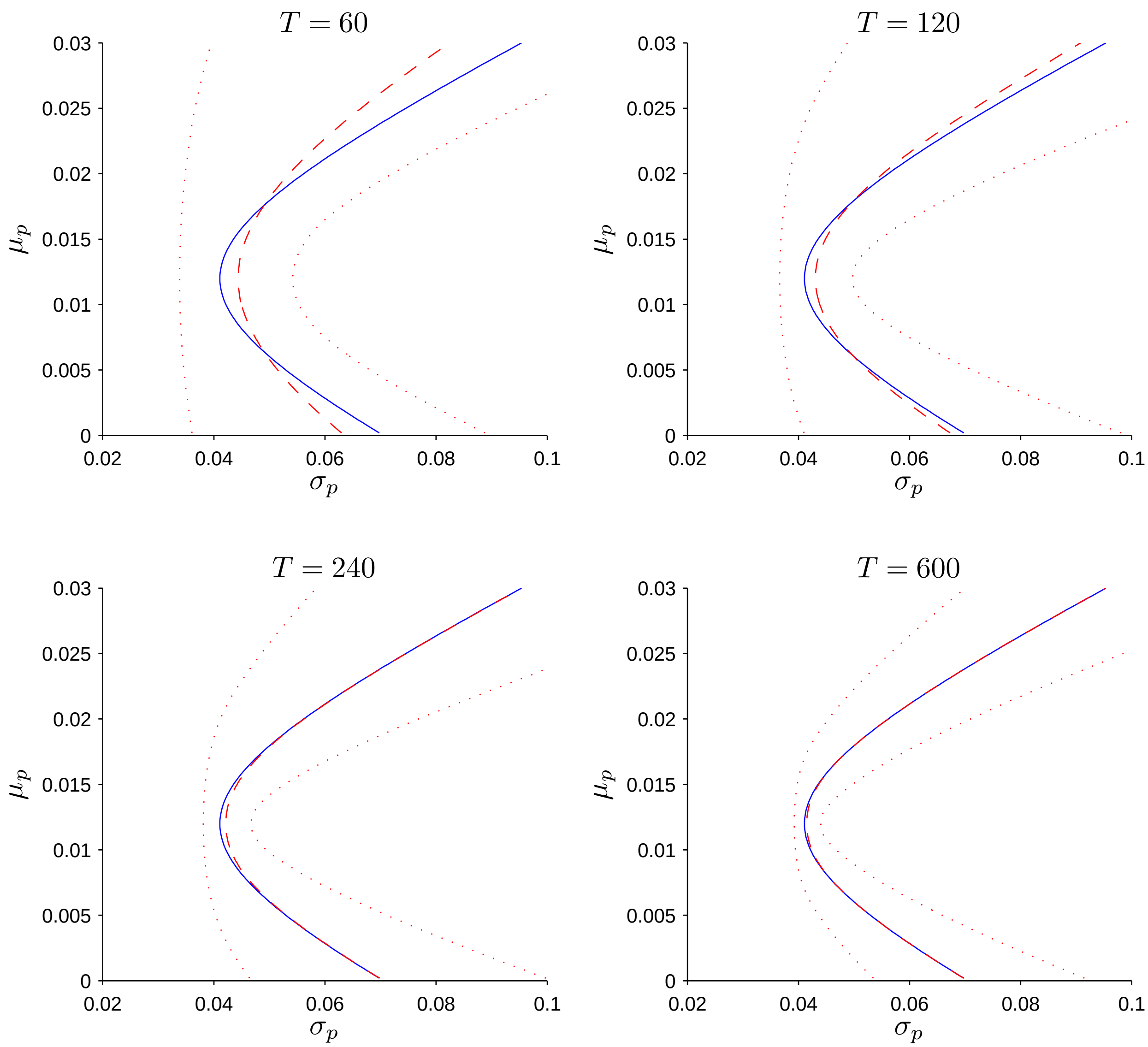

Figure A.3. The figure provides information on the distribution of the adjusted sample minimum-variance frontier for different lengths of estimation period (T). The solid line represents the population frontier for $N = 25$ assets, with parameters $\mu_g = 0.01199$, $\sigma_g = 0.04108$, and $\psi = 0.209$. The dashed line represents the expected value of the adjusted sample minimum-variance frontier and the two dotted lines represent the 5th and the 95th percentiles of the adjusted sample minimum-variance frontier, all based on 100,000 simulations.

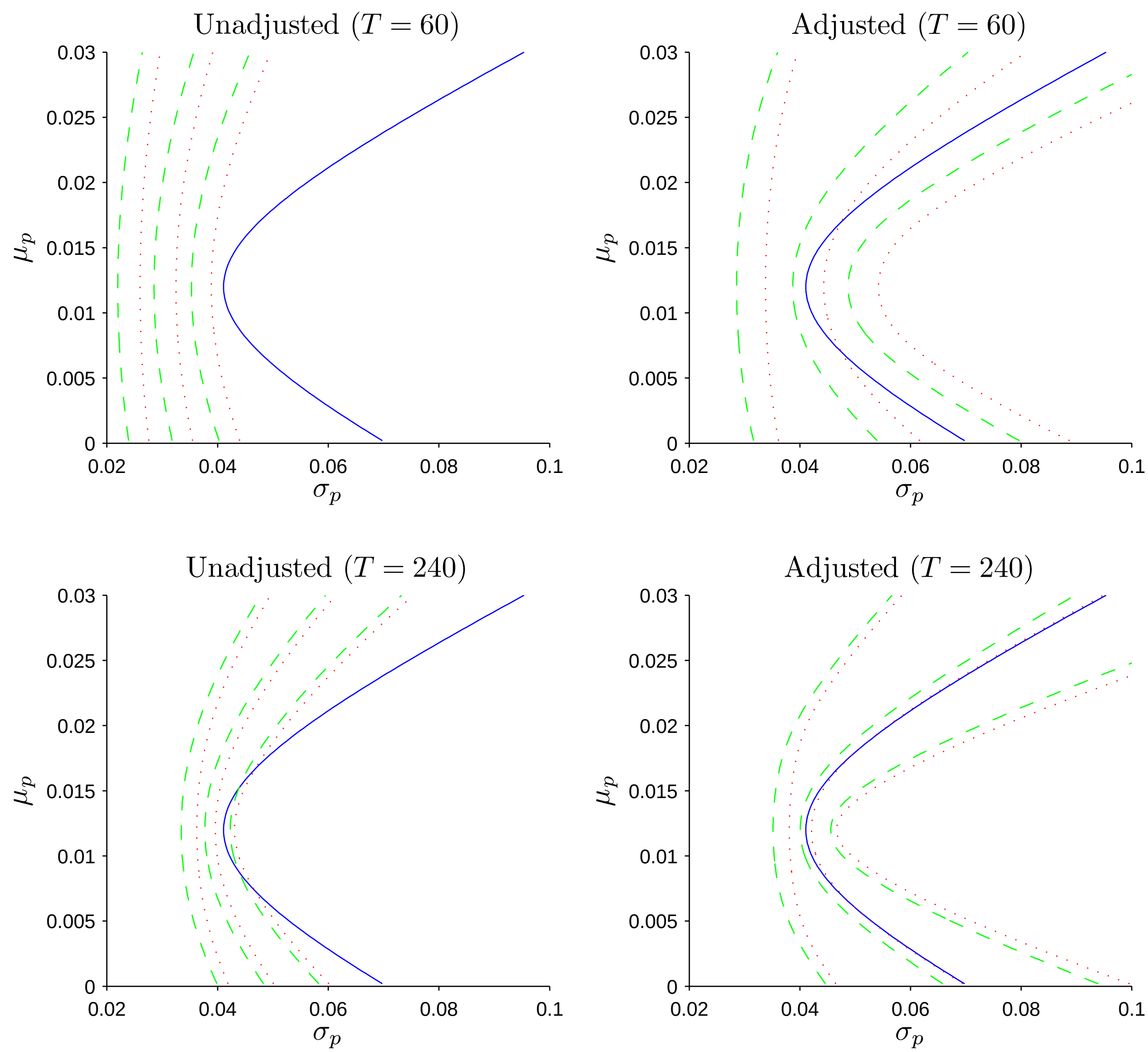

Figure A.4. The figure provides information on the distribution of the unadjusted and adjusted sample minimum-variance frontier for different lengths of estimation period (T). The solid line represents the population frontier for $N = 25$ assets, with parameters $\mu_g = 0.01199$, $\sigma_g = 0.04108$, and $\psi = 0.209$. The three dashed lines represents the 5th percentile, expected value, and 95th percentile of the unadjusted and adjusted sample minimum-variance frontier under the assumption that returns are multivariate t-distributed with five degrees of freedom. The three dotted lines are the corresponding ones for the normality case. The distributions of the unadjusted and adjusted frontiers are based on 100,000 simulations.

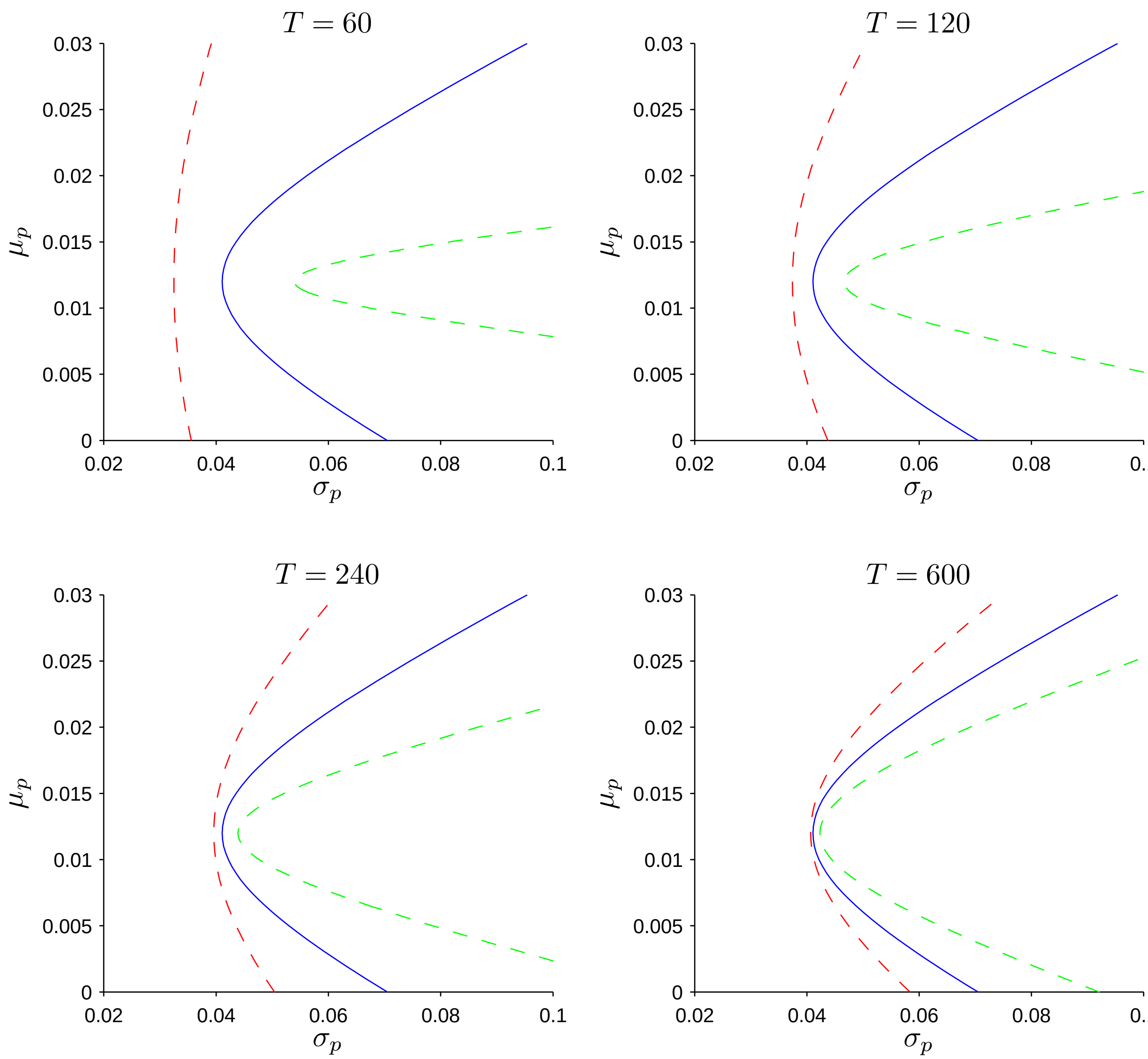

Figure A.5. The figure provides information on the average out-of-sample performance of sample minimum-variance frontier for different lengths of estimation period (T). The solid line represents the population frontier for $N = 25$ assets, with parameters $\mu_g = 0.01199$, $\sigma_g = 0.04108$, and $\psi = 0.209$. The outer dashed line represents $\bar{\sigma}_p$, where $\bar{\sigma}_p^2$ is the expected value of the sample minimum-variance frontier. The inner dashed line represents $\underline{\sigma}_p$, where $\underline{\sigma}_p^2$ is the expected value of the out-of-sample variance of the sample minimum-variance frontier.

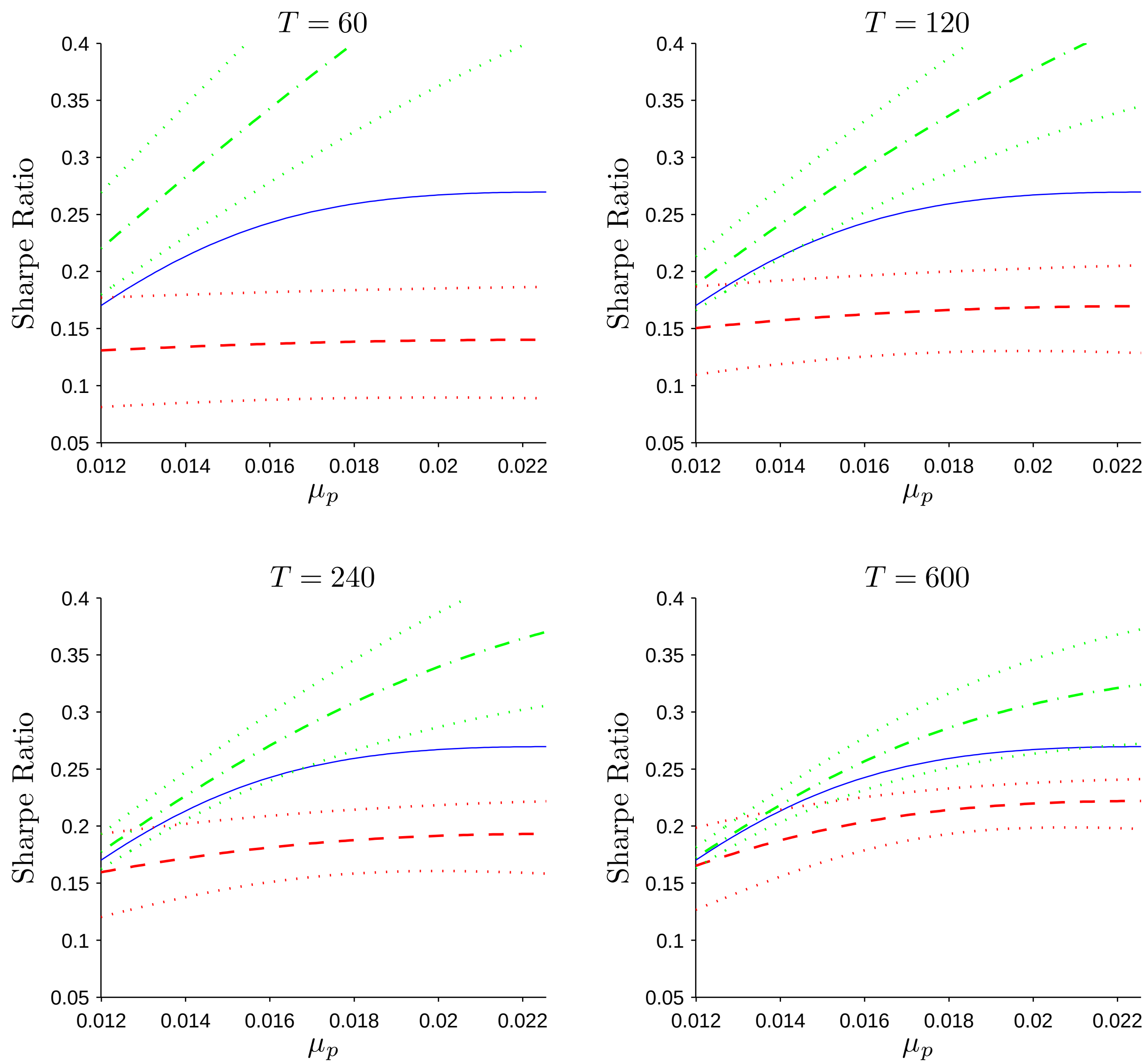

Figure A.6. The figure provides information on the distribution of the in-sample and out-of-sample Sharpe ratio of sample minimum-variance portfolios with different target expected return for different lengths of estimation period (T). The solid line represents the Sharpe ratio of the population minimum-variance frontier for $N = 10$ assets, with parameters $\mu_g = 0.01199$, $\sigma_g = 0.04108$, $\psi = 0.209$, and a risk-free rate of 0.005. The dashed-dotted (dashed) line and surrounding dotted lines represent the average, 5th, and 95th percentiles of the in-sample (out-of-sample) Sharpe ratio of the sample minimum-variance portfolio with target expected return μ_p. The distributions of the in-sample and out-of-sample Sharpe ratio are based on 100,000 simulations.