Chapter 10
Risk and Bias in Portfolio Optimization

Thomas Holgersson and Martin Singull

Abstract In this paper we derive weighted squared risk measures for a commonly used Stein-type estimator of the global minimum variance portfolio. The risk functions are conveniently split in terms of variance and squared bias over different weight matrices. It is argued that the common out-of-sample variance criteria should be used with care and that a simple unweighted risk function may be more appropriate.

10.1 Introduction

A financial portfolio is defined by a vector, $\mathbf{w} = (w_1, \ldots, w_p)$, say, whose element w_j describes the fraction of the portfolio investment that is allocated to some asset x_j. The random return of the portfolio is obtained as $r = x_1 w_1 + \cdots + x_p w_p = \mathbf{w}'\mathbf{x}$, where $\mathbf{x}$ is a vector of random returns. An investor is interested in holding a portfolio that maximizes the expected return $E[r]$ at a given level of risk ($var[r]$) or, equivalently, minimizes the risk $var[r]$ at a given level of the expected return. A portfolio satisfying such a maximization/minimization is called an optimal portfolio. Markowitz [20] developed a theory for such mean-variance portfolio optimization that still plays a fundamental role. The theory considers a fixed (non-random) portfolio $\mathbf{w}$, which could be determined qualitatively, and gives closed-form expressions for the optimal portfolio as a function of $\mathbf{w}$, which in turn depends on the mean vector and covariance matrix of $\mathbf{x}$.

While Markowitz's theory is concerned with analyses and conclusions drawn from a fixed portfolio vector, there has been a growing interest during the last fifty years in how one can use statistical methods to estimate $\mathbf{w}$. Because the

T. Holgersson (✉)
Department of Economics and Statistics, Linnaeus University, Växjö, Sweden
e-mail: thomas.holgersson@lnu.se

M. Singull
Department of Mathematics, Linköping University, Linköping, Sweden
e-mail: martin.singull@liu.se

optimal portfolio is a function of certain population parameters, one obtains the so-called standard estimator of **w** by substituting these unknown parameters by their sample counterparts. Some basic sampling properties such as low-order moments have long been known [6, 15], while the full normal-theory distribution of the standard estimator was derived by [22]. It has since been recognized, both from the theoretical distribution and from empirical findings, that the sampling variance of the standard estimator may be too large to be useful in investment strategies [4, 9, 17, 21].

As in most multivariate analyses, this applies in particular when the sample size n is close to the dimension p. A number of alternatives to the standard estimator have been derived recently. Jagannathan and Ma [13] noticed that imposing certain moment constraints leads to a reduction of sampling variance, while [4] proposed a generalized estimator that impose a threshold constraint on the portfolio vector. Bayesian methods have been proposed by [15] and [16], while inference solutions to the case of singular covariance have been derived by [1].

In this paper our interest lies in a particular family of estimators defined as a weighted mean between the standard estimator and a constant vector. This type of estimator, which is related to a family of estimators originally proposed by James and Stein [14], has been considered by [2, 8] and [12]. One of the main features of this weighted estimator is that it depends on a tuning coefficient, which also introduces a bias in the estimator. We derive some properties of this Stein-type estimator with respect to different kinds of weighted squared loss functions. Particular focus is set on the bias term since this has been given little attention in the literature previously. We will restrict ourselves to the specific case of the global minimum variance portfolio (GMVP), although many of the concerns in the paper also apply to more general portfolio optimizations. We will not attempt to derive any explicit estimators. Our primary interest lies in deriving and comparing different risk measures for a commonly used Stein-type estimator and discuss some of their differences.

10.2 Preliminaries

Consider a vector $\mathbf{x} : p \times 1$ of excess returns of p financial assets. A financial portfolio is defined as the weighted sum $\mathbf{w}'\mathbf{x}$, where $\mathbf{w} : p \times 1$ is called the portfolio weight. The vector **w** can be determined "qualitatively", i.e., based on expert knowledge about the market, or it could be estimated from historical data of returns, which is the objective of this article. An efficient portfolio is usually determined by minimizing the portfolio variance subject to a mean portfolio premium return and the additional constraint that investment proportions sum to one. According to [20], an efficient portfolio weight **w**, assuming absence of short sale constraints, is determined by

$$\min_{\mathbf{w} \in \mathbb{R}^p} \{\mathbf{w}' \Sigma \mathbf{w} \mid \mathbf{w}'\mathbf{1} = 1\}, \tag{10.1}$$

where Σ is the covariance matrix of $\mathbf{x}$ and $\mathbf{1} : p \times 1$ is a vector of ones. The well-known solution to (10.1) is given by

$$\mathbf{w} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}. \tag{10.2}$$

The vector $\mathbf{w}$ is known as the global minimum variance portfolio (GMVP). It is also possible to define portfolios under more general constraints than those of (10.1), see [15] and [4] for alternative formulations.

Since the quantity in (10.2) depends on an unknown parameter, it needs to be estimated from data. We are thus concerned with the problem of using a set of n observations on random returns, say $\mathbf{x}_1, \ldots, \mathbf{x}_n$, to develop an estimator of $\mathbf{w}$. We will assume a common setting where $\mathbf{x}_i \sim_{iid} N_p(\mu, \Sigma)$ under the assumptions $\max_j |\mu_j| \leq a_1 < \infty$, $\max_j \lambda_j(\Sigma) \leq a_2 < \infty$, and $0 < a_3 \leq \min_j \lambda_j(\Sigma)$, where $\mu = (\mu_1, \ldots, \mu_p)'$ and $\lambda_j(\Sigma)$ denote an eigenvalue of Σ. Although these assumptions are not strictly necessary for making inference of the GMPV weight, they simplify the technical treatments considerably.

An obvious estimator of $\mathbf{w}$ is obtained by replacing the unknown covariance matrix by its sample counterpart. The resulting estimate is commonly referred to as the *standard estimator*, henceforth denoted by $\widehat{\mathbf{w}}_0$. This estimator is central in the paper, and we state some basic properties for the sake of clarity.

Property 10.1 Assume $\mathbf{x}_i \sim_{iid} N_p(\mu, \Sigma)$, $i = 1, \ldots, n$, $p \geq 4$ and $n \geq p + 2$. Let $S = n^{-1} \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$, where $\bar{\mathbf{x}} = n^{-1} \sum_{i=1}^{n} \mathbf{x}_i$. Define

$$(i) \quad \mathbf{w} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}},$$

$$(ii) \quad \sigma^2 = \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}},$$

$$(iii) \quad \widehat{\mathbf{w}}_0 = (\widehat{w}_{0(1)}, \ldots, \widehat{w}_{0(p)})' = \frac{\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{S}^{-1}\mathbf{1}}.$$

Then

$$\widehat{\mathbf{w}}_0 \sim t_p \left(\mathbf{w}, \frac{\sigma^2 \Sigma^{-1} - \mathbf{w}\mathbf{w}'}{n - p + 1}, n - p + 1 \right),$$

where $t_p(\cdot)$ denotes a p-dimensional singular t-distribution with $n - p + 1$ degrees of freedom, location vector $\mathbf{w}$ and dispersion matrix $\sigma^2 \Sigma^{-1} - \mathbf{w}\mathbf{w}'$ with $rank(\sigma^2 \Sigma^{-1} - \mathbf{w}\mathbf{w}') = p - 1$.

Proof ([18, 22]) See also [19, Chapter 1], for a definition of the multivariate t-distribution. □

It is well known that the sampling variance of $\widehat{\mathbf{w}}_0$ can be substantial when n is small relative to the dimension p and hence of limited relevance to an investor. A considerable amount of research has been concerned with development of improved estimators of $\mathbf{w}$ [2, 4, 8, 12]. A common approach is to first decide a family of estimators, which usually depends on some tuning coefficient, and then use a risk function to identify the appropriate value of this coefficient. One concern with this approach, however, is that two different risk functions usually produce two different values of the tuning coefficient and hence that the distributional properties of our portfolio weight estimator may strongly depend on which specific risk function that is being used. The next section will discuss this matter further.

10.3   Risk Measures and Portfolio Estimation

The original view of portfolio optimization and risk as stated by [20] is that *The w_j's are not random variables, but are fixed by the investor* and that *"Risk" is described by the variance of the return*. The "risk" an investor is facing is accordingly determined by $var[\mathbf{w}'\mathbf{x}] = \mathbf{w}'\Sigma\mathbf{w}$, where $\Sigma = cov[\mathbf{x}_i]$. However, the definition of "risk" is less obvious when $\mathbf{w}$ is estimated from data because of the additional uncertainty of sampling variance and bias. From a perspective of statistical inference the term "risk" refers to the sampling variance and bias of a parameter estimator (say $\widehat{\mathbf{w}}$), while in portfolio theory "risk" primarily refers to variance of the return vector $\mathbf{x}$. Since the estimated portfolio return is defined by $r_i = \widehat{\mathbf{w}}'\mathbf{x}_i$, it involves risk in both senses.

Following Markowitz's view of a fixed (non-random) portfolio, the Lagrangian of the optimization problem in (10.1) may be formulated

$$L(\mathbf{w}, \Sigma, \lambda_0) = \frac{1}{2}\mathbf{w}'\Sigma\mathbf{w} - \lambda_0(\mathbf{w}'\mathbf{1} - 1), \qquad (10.3)$$

where λ_0 is a Lagrange multiplier. By taking derivatives of $L(\mathbf{w}, \Sigma, \lambda_0)$ w.r.t. $\mathbf{w}$ and equating at zero we get the condition

$$\mathbf{w} = \lambda_0 \Sigma^{-1}\mathbf{1}. \qquad (10.4)$$

Since $\mathbf{w}'\mathbf{1} = 1$ it follows that $\lambda_0 = (\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}$ and we obtain (10.2). Note that the identity (10.4) is completely determined by $\Sigma^{-1}\mathbf{1}$ in the sense that if we define $\theta = \Sigma^{-1}\mathbf{1}$, then the solution to the optimization problem depends on θ only. When it comes to random (estimated) portfolio weights, the optimization problem has been formulated in different ways in the literature. Jagannathan and Ma [13] specify a constrained portfolio variance minimization problem as

$$\min_{\mathbf{w}\in\mathbb{R}^p}\{\mathbf{w}'\widehat{\Sigma}\mathbf{w}|\mathbf{w}'\mathbf{1} = 1, 0 \leq w_j, w_j \leq \varpi, j = 1, \ldots, p\}, \qquad (10.5)$$

where $\widehat{\Sigma}$ is some estimator of Σ and ϖ is a finite constant such that $w_j \leq \varpi$ defines an upper bound constraint for the weights. Let $\lambda = \left(\lambda_1 \ldots \lambda_p\right)$ be the Lagrange multiplier for $0 \leq w_j$ and $\delta = \left(\delta_1 \ldots \delta_p\right)$ the Lagrange multiplier for $w_j \leq \varpi$ and define $\widetilde{\Sigma} = \widehat{\Sigma} + (\delta\mathbf{1}' + \mathbf{1}\delta') - (\lambda\mathbf{1}' + \mathbf{1}\lambda')$, where $\widehat{\Sigma}$ is the normal-theory unconstrained maximum likelihood (ML) estimate of Σ. Jagannathan and Ma [13] showed that $\widetilde{\Sigma}$ is positive semidefinite, and that constructing a constrained global minimum variance portfolio from $\widehat{\Sigma}$ (i.e., the solution to (10.5)) is equivalent to constructing the unconstrained minimum variance portfolio from $\widetilde{\Sigma}$. Thus the constraints can be imposed in the estimation stage instead of the optimization stage and the result would be the same. This is a fundamental result in portfolio theory because it connects Markowitz's theoretical portfolio theory with statistical inference theory, including both frequentistic and Bayesian treatments.

DeMiguel et al. [4] proposed a norm-constrained minimum variance portfolio as one that solves $\min_{\mathbf{w} \in \mathbb{R}^p} \{\mathbf{w}'\Sigma\mathbf{w} \mid \mathbf{w}'\mathbf{1} = 1, \|\mathbf{w}\| \leq \varpi\}$, where $\|\mathbf{w}\|$ denotes either the L_1 norm $\|\mathbf{w}\|_1 = \sum_{j=1}^{n} |w_j|$ or the L_2 norm $\|\mathbf{w}\|_2 = \sqrt{\mathbf{w}'Q\mathbf{w}}$, where Q is some positive definite matrix. They showed that the solution under the constraint $\|\mathbf{w}\|_1 = 1$ coincides with the shortsale constrained estimator.

[8] take a rather different view on the optimization problem and argue that, for some estimator $\widehat{\mathbf{w}}$ based on returns $\mathbf{x}_1, \ldots, \mathbf{x}_n$, the quantity $var[\mathbf{x}'_{n+1}\widehat{\mathbf{w}}|\mathscr{F}_n] = \widehat{\mathbf{w}}'\Sigma\widehat{\mathbf{w}}$, where $\mathscr{F}_n$ is the information set up to time n, represents the actual variance of the return belonging to the portfolio $\widehat{\mathbf{w}}$. References [2] and [23] take on a similar approach.

There currently seems to be no consensus, or unified theory, on how to formulate a statistical version of Markowitz's fixed-portfolio optimization problem. In what follows we will discuss some consequences of the view one takes on the optimization problem and the measure used to evaluate the properties of a weight estimator.

Portfolio Weight Estimators The GMVP weight vector $\mathbf{w} = (\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}\Sigma^{-1}$ only involves one unknown quantity, Σ^{-1}. One can therefore specify $\mathbf{w}$ as a function of Σ^{-1}, say $\mathbf{w} = f(\Sigma^{-1}) = (\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}\Sigma^{-1}$. A portfolio estimator may thus be obtained by substituting an estimator $\widehat{\Sigma}^{-1}$ into $\mathbf{w}$ and obtain $\widehat{\mathbf{w}} = f(\widehat{\Sigma}^{-1}) = (\mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1})^{-1}\widehat{\Sigma}^{-1}$. Any estimator $f(\widehat{\Sigma}^{-1})$ will be a sufficient statistic for $\mathbf{w}$ as long as $\widehat{\Sigma}^{-1}$ is a sufficient statistic for Σ^{-1}. Such "plug-in" estimators are therefore completely legitimate from an inferential point of view. The literature on estimation of Σ^{-1} is, in turn, extensive. Some important references include [5, 7, 10–12, 24, 26]. A comprehensive survey of improved estimators of Σ^{-1}, including Bayesian frameworks, is given in [25]. We will, however, not proceed on this path but instead consider a more restricted class of estimators.

Let $\widehat{\mathbf{w}}_0$ denote the standard estimator defined in Property 10.1 and $\mathbf{w}_{ref}$ denote some pre-determined reference portfolio, which could be fixed or random (but independent of $\widehat{\mathbf{w}}_0$). We denote a weighted mean between these two quantities as follows

$$\widehat{\mathbf{w}}_\alpha = (1 - \alpha)\widehat{\mathbf{w}}_0 + \alpha\mathbf{w}_{ref}, \quad 0 \leq \alpha \leq 1. \tag{10.6}$$

The estimator $\widehat{\mathbf{w}}_\alpha$, which is closely related to a family of estimators proposed by [14], has been used by [2, 8, 18]. In order to determine an appropriate value of the tuning coefficient α one usually applies a loss function. A common square loss is defined by $l = (\widehat{\mathbf{w}} - \mathbf{w})' Q (\widehat{\mathbf{w}} - \mathbf{w})$, where Q is a p.d. matrix. Taking the expected value of l we obtain a quadratic risk function for $\widehat{\mathbf{w}}_\alpha$ as

$$\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, Q) = E[(\widehat{\mathbf{w}}_\alpha - \mathbf{w})' Q (\widehat{\mathbf{w}}_\alpha - \mathbf{w})]. \tag{10.7}$$

The value of α that minimizes $\mathfrak{R}$ is then defined as the *optimal* α. By different choices of Q we obtain several common risk functions as special cases of (10.7). For example, $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, I)$ is the common mean squared error (MSE), which consists of the variance plus squared bias. The particular risk function $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, \Sigma)$ has been used by [8] and [18] while [2] used $l = (\widehat{\mathbf{w}}_\alpha - \mathbf{w})' \Sigma (\widehat{\mathbf{w}}_\alpha - \mathbf{w})$ directly, rather than its expected value, to derive an optimal portfolio estimator. Yet another way to evaluate the properties of a portfolio weight estimator is through its predictive, or out-of-sample properties. The predicted return of some estimator $\widehat{\mathbf{w}}$ is obtained by $\widehat{\mathbf{w}}' \mathbf{x}_t$, where $\mathbf{x}_t$ is a vector of returns not used in $\widehat{\mathbf{w}}$, i.e., $\mathbf{x}_t \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. The mean squared error of prediction (MSEP) of $\widehat{\mathbf{w}}$ may then be evaluated by $E[(\widehat{\mathbf{w}}' \mathbf{x}_t - \mathbf{w}' \mathbf{x}_t)^2]$. Note that MSEP is also a special case of (10.7), for

$$
\begin{aligned}
E[(\widehat{\mathbf{w}}' \mathbf{x}_t - \mathbf{w}' \mathbf{x}_t)^2] &= E[(\widehat{\mathbf{w}} - \mathbf{w})' \mathbf{x}_t \mathbf{x}_t' (\widehat{\mathbf{w}} - \mathbf{w})] \\
&= E[(\widehat{\mathbf{w}} - \mathbf{w})' (\Sigma + \mu\mu') (\widehat{\mathbf{w}} - \mathbf{w})] = \mathfrak{R}(\widehat{\mathbf{w}}, \Sigma + \mu\mu').
\end{aligned}
$$

In what follows we give explicit expressions of $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, Q)$ for some particular values of Q.

Proposition 10.1 *Let* $\widehat{\mathbf{w}}_\alpha = (1 - \alpha)\widehat{\mathbf{w}}_0 + \alpha p^{-1}\mathbf{1}$ *for some* $0 \leq \alpha \leq 1$, *and let* $E[\widehat{\mathbf{w}}_0] = \mathbf{w}$, $R = cov[\widehat{\mathbf{w}}_0] = \dfrac{\sigma^2 \Sigma^{-1} - \mathbf{w}\mathbf{w}'}{n - p - 1}$, *where* $\sigma^2 = (\mathbf{1}' \Sigma^{-1} \mathbf{1})^{-1}$. *Let* $\mathbf{x}_t$ *be an out-of-sample vector of returns, and define* $\mu = E[\mathbf{x}_t]$, $\mu_w = \mathbf{w}'\mu$ *and* $\bar{\mu} = p^{-1}\mathbf{1}'\mu$. *Then the following risk identities hold*

(a) $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, I) = (1 - \alpha)^2 \mathrm{tr}\{R\} + \alpha^2 (\mathbf{w}'\mathbf{w} - p^{-1})$,

(b) $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, \Sigma) = \dfrac{(1 - \alpha)^2}{n - p - 1}\sigma^2(p - 1) + \alpha^2(p^{-2}\mathbf{1}'\Sigma\mathbf{1} - \sigma^2)$,

(c) $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, \mu\mu') = (1 - \alpha)^2 \mu' R \mu + \alpha^2 (\mu_w - \bar{\mu})^2$,

(d) $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, \Sigma + \mu\mu') = \dfrac{(1 - \alpha)^2}{n - p - 1}(\sigma^2(p-1) + \sigma^2\mu'\Sigma^{-1}\mu + \mu_w^2) + \alpha^2((p^{-2}\mathbf{1}'\Sigma\mathbf{1} - \sigma^2) + (\mu_w - \bar{\mu})^2)$.

Before the proof of Proposition 10.1 we will state a useful identity in the following lemma.

Lemma 10.1 *Let* $\widehat{\mathbf{w}}_\alpha = (1-\alpha)\widehat{\mathbf{w}}_0 + \alpha p^{-1}\mathbf{1}$ *for some* $0 \le \alpha \le 1$, *and let* $E[\widehat{\mathbf{w}}_0] = \mathbf{w}$, $cov[\widehat{\mathbf{w}}_0] = R$. *Then it holds that*

$$E[(\widehat{\mathbf{w}}_\alpha - \mathbf{w})(\widehat{\mathbf{w}}_\alpha - \mathbf{w})'] = (1-\alpha)^2 R + \alpha^2 (\mathbf{w} - p^{-1}\mathbf{1})(\mathbf{w} - p^{-1}\mathbf{1})'.$$

Proof See Appendix.   □

We are now ready to give the proof of Proposition 10.1.

Proof Proof of Proposition 10.1. From Lemma 10.1 we find that

$$E[(\widehat{\mathbf{w}}_\alpha - \mathbf{w})(\widehat{\mathbf{w}}_\alpha - \mathbf{w})'] = \text{tr}\{E[(\widehat{\mathbf{w}}_\alpha - \mathbf{w})\widehat{\mathbf{w}}_\alpha - \mathbf{w}']\} = (1-\alpha)^2 \text{tr}\{R\} + \alpha^2(\mathbf{w}'\mathbf{w} - p^{-1}),$$

which establishes (a). Applying Lemma 10.1 again we obtain (b) since

$$\text{tr}\{E[(\widehat{\mathbf{w}}_\alpha - \mathbf{w})(\widehat{\mathbf{w}}_\alpha - \mathbf{w})'\Sigma]\} = (1-\alpha)^2 \text{tr}\{R\Sigma\} + \alpha^2(p^{-2}\mathbf{1}'\Sigma\mathbf{1} - \sigma^2)$$

and

$$\text{tr}\{R\Sigma\} = \frac{\text{tr}\{(\sigma^2\Sigma^{-1} - \mathbf{w}\mathbf{w}')\Sigma\}}{n-p-1} = \frac{\text{tr}\{\sigma^2 I - \mathbf{w}\mathbf{w}'\Sigma\}}{n-p-1} = \frac{p-1}{n-p-1}\sigma^2.$$

The identity (c) is similarly obtained by

$$\text{tr}\{E[(\widehat{\mathbf{w}}_\alpha - \mathbf{w})(\widehat{\mathbf{w}}_\alpha - \mathbf{w})'\mu\mu']\} = (1-\alpha)^2 \mu' R\mu + \alpha^2(\mathbf{w}'\mu - p^{-1}\mathbf{1}'\mu)^2,$$

while (d) is obtained by adding (b) and (c) and simplifying terms.   □

Remark 10.1 The MSEP measure in (d), i.e., $E[(\widehat{\mathbf{w}}'_\alpha\mathbf{x}_t - \mathbf{w}'\mathbf{x}_t)^2]$, describes the variability of $\widehat{\mathbf{w}}'_\alpha\mathbf{x}_t$ around the random term $\mathbf{w}'\mathbf{x}_t$. An alternative way of defining MSEP is by $E[(\widehat{\mathbf{w}}'_\alpha\mathbf{x}_t - \mathbf{w}'\mu)^2]$ which describes the variability of $\widehat{\mathbf{w}}'_\alpha\mathbf{x}_t$ around the non-random point $\mathbf{w}'E[\mathbf{x}_t]$. These two differ in that

$$E[(\widehat{\mathbf{w}}'_\alpha\mathbf{x}_t - \mathbf{w}'\mu)^2] = E[(\widehat{\mathbf{w}}'_\alpha\mathbf{x}_t - \mathbf{w}'\mathbf{x}_t)^2] + (\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}.$$

See Appendix for a proof of this identity. Note that the smallest possible return prediction variance for any of the risks (a)–(d) of Proposition 10.1 is 0, which is attained at $\alpha = 1$, whereas for $E[(\widehat{\mathbf{w}}'_\alpha\mathbf{x}_t - \mathbf{w}'\mu)^2]$ the minimum prediction variance is determined by $(\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}$ regardless of the value of α.

For purposes of deriving an estimator of $\mathbf{w}$, the choice between $E[(\widehat{\mathbf{w}}'_\alpha\mathbf{x}_t - \mathbf{w}'\mathbf{x}_t)^2]$ and $E[(\widehat{\mathbf{w}}'_\alpha\mathbf{x}_t - \mathbf{w}'\mu)^2]$ makes no difference, but for calculation of prediction intervals our view of the centre point does matter.

Remark 10.2 The bias terms vanish trivially if $\alpha = 0$ and/or $\Sigma = I$. More generally, the bias terms for $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, I)$, $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, \Sigma)$, and $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, \mu\mu')$ vanish when $p\mathbf{1}'\Sigma^{-2}\mathbf{1} = (\mathbf{1}'\Sigma^{-1}\mathbf{1})^2$, $p^{-2}\mathbf{1}'\Sigma\mathbf{1} = \sigma^2$ and $\mu_w - \bar{\mu} = 0$, respectively. While

the first two conditions are almost identical to the condition $\Sigma = I$, the identity $\mu_w = \bar{\mu}$, holds when the portfolio return equals the average asset mean.

Remark 10.3 The optimal value of the tuning coefficient α may be derived by equating $\dfrac{\partial \mathfrak{R}(\widehat{\mathbf{w}}_\alpha, Q)}{\partial \alpha}$ at zero and solving for α. The resulting value, say α_{opt}, then yields an adaptive estimator $\widehat{\mathbf{w}}_{\alpha_{opt}} = (1 - \alpha_{opt})\widehat{\mathbf{w}}_0 + \alpha_{opt} p^{-1}\mathbf{1}$. However, α_{opt} is by necessity a function of unknown population parameters. Operational, or "bona-fide", estimators are usually obtained by substituting these unknown parameters by some estimators to obtain an approximation $\widehat{\alpha}$. The adapted portfolio estimator is then defined by $\widehat{\mathbf{w}}_{\widehat{\alpha}} = (1 - \widehat{\alpha})\widehat{\mathbf{w}}_0 + \widehat{\alpha} p^{-1}\mathbf{1}$. The coefficient $\widehat{\alpha}$ is hence random, which in turn distorts the otherwise known distribution of $\widehat{\mathbf{w}}_\alpha$. It is still possible to conduct inferential analysis (interval estimation, etc.) through the theory of oracle inequalities but at the price of becoming more technically involved (see [3] for a comprehensive survey of the topic). Another option is to determine α qualitatively, for example based on expert knowledge of the financial market. This method has the appealing property of retaining the sampling distribution of $\widehat{\mathbf{w}}_\alpha$.

Remark 10.4 The original portfolio problem as posed by [20] considers the return of a non-random portfolio at data point t (e.g., a time point), determined by $\mathbf{w}'\mathbf{x}_t$ with variance $var[\mathbf{w}'\mathbf{x}_t] = \text{tr}\{cov[\mathbf{x}_t]\mathbf{w}\mathbf{w}'\} = \mathbf{w}'\Sigma\mathbf{w}$, which is void of concerns about consistency and bias, etc. The fixed-portfolio theory uses Lagrange multipliers to derive the global minimum weight portfolio $\mathbf{w} = \dfrac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$, which is the minimum of $var[\mathbf{w}'\mathbf{x}_t]$ subject to the constraint $\sum_{j=1}^{p} w_j = 1$. How to proceed from there to statistical estimation is less obvious. It is customary to apply the criterion $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, \Sigma) = E[(\widehat{\mathbf{w}} - \mathbf{w})'\Sigma(\widehat{\mathbf{w}} - \mathbf{w})]$ to derive estimators of $\mathbf{w}$, but it is not clear what is gained by minimizing this risk instead of the unweighted risk $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, I) = E[(\widehat{\mathbf{w}} - \mathbf{w})'(\widehat{\mathbf{w}} - \mathbf{w})]$. For any consistent estimator, say $\widehat{\mathbf{w}} = \dfrac{\widehat{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1}}$, where $\widehat{\Sigma}^{-1}$ some consistent estimator of Σ^{-1}, the asymptotic return variance is $\lim_{n\to\infty} var[\widehat{\mathbf{w}}'\mathbf{x}_t] = \mathbf{w}'\Sigma\mathbf{w}$ regardless of which risk function was used to obtain $\widehat{\mathbf{w}}$. The weight $Q = \Sigma$ has already been used in the minimization of $\mathbf{w}'\Sigma\mathbf{w}$ to obtain the functional form $\dfrac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$, and there is no apparent gain in using the weighting $(\widehat{\mathbf{w}} - \mathbf{w})'\Sigma(\widehat{\mathbf{w}} - \mathbf{w})$ once again in the estimation stage. In fact, Σ^{-1} is the only unknown parameter in $\dfrac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$. Knowing Σ (up to a scalar multiplication) is therefore equivalent to knowing $\mathbf{w}$. Since Σ is a one-to-one transformation of Σ^{-1}, the weighted loss $(\widehat{\mathbf{w}} - \mathbf{w})'\Sigma(\widehat{\mathbf{w}} - \mathbf{w})$ in a sense uses the true parameter to derive its own estimator. In this view the unweighted risk $\mathfrak{R}(\widehat{\mathbf{w}}_\alpha, I) = E[(\widehat{\mathbf{w}} - \mathbf{w})'(\widehat{\mathbf{w}} - \mathbf{w})]$ may better reflect the actual risk of an estimator $\widehat{\mathbf{w}}$.

Appendix

First we give the proof of Lemma 10.1.

Proof Note that

$$\widehat{\mathbf{w}}_\alpha - E[\widehat{\mathbf{w}}_\alpha] = \widehat{\mathbf{w}}_0 - \alpha(\widehat{\mathbf{w}}_0 - p^{-1}\mathbf{1}) - \mathbf{w} - \alpha(\mathbf{w} - p^{-1}\mathbf{1}) = (\alpha - 1)(\widehat{\mathbf{w}}_0 - \mathbf{w}).$$

Hence,

$$cov[\widehat{\mathbf{w}}_\alpha] = (\alpha - 1)^2 E[(\widehat{\mathbf{w}}_0 - \mathbf{w})(\widehat{\mathbf{w}}_0 - \mathbf{w})'] = (\alpha - 1)^2 cov[\widehat{\mathbf{w}}_0] = (\alpha - 1)^2 R.$$

Similarly, define the bias term $\mathbf{b} = E[\widehat{\mathbf{w}}_\alpha] - \mathbf{w}$. Then

$$E[\widehat{\mathbf{w}}_0] - \alpha(E[\widehat{\mathbf{w}}_0] - p^{-1}\mathbf{1}) - \mathbf{w} = -\alpha(\mathbf{w} - p^{-1}\mathbf{1}).$$

Hence it follows that $\mathbf{b}\mathbf{b}' = \alpha^2(\mathbf{w} - p^{-1}\mathbf{1})(\mathbf{w} - p^{-1}\mathbf{1})'$ and

$$E[(\widehat{\mathbf{w}}_\alpha - \mathbf{w})(\widehat{\mathbf{w}}_\alpha - \mathbf{w})'] = cov[\widehat{\mathbf{w}}_\alpha] + \mathbf{b}\mathbf{b}' = (1 - \alpha)^2 R + \alpha^2(\mathbf{w} - p^{-1}\mathbf{1})(\mathbf{w} - p^{-1}\mathbf{1})'.$$

$$\square$$

Next we give the proof of the identity in Remark 10.1.

Proof We have

$$\begin{aligned}
E[(\widehat{\mathbf{w}}_\alpha'\mathbf{x}_t - \mathbf{w}'\mu)^2] &= E[((\widehat{\mathbf{w}}_\alpha'\mathbf{x}_t - \mathbf{w}'\mathbf{x}_t) + (\mathbf{w}'\mathbf{x}_t - \mathbf{w}'\mu))^2] \\
&= \mathrm{tr}\{E[(\widehat{\mathbf{w}}_\alpha - \mathbf{w})\mathbf{x}_t'\mathbf{x}_t(\widehat{\mathbf{w}}_\alpha - \mathbf{w})]\} \\
&\quad + 2\mathrm{tr}\{E[(\widehat{\mathbf{w}}_\alpha'\mathbf{x}_t - \mathbf{w}'\mathbf{x}_t)(\mathbf{w}'\mathbf{x}_t - \mathbf{w}'\mu)]\} \\
&\quad + E[\mathbf{w}'(\mathbf{x}_t - \mu)(\mathbf{x}_t - \mu)'\mathbf{w}] \\
&= A + 2B + D,
\end{aligned}$$

say. The identities $A = \mathrm{tr}\{(\Sigma + \mu\mu')E[(\widehat{\mathbf{w}}_\alpha - \mathbf{w})(\widehat{\mathbf{w}}_\alpha - \mathbf{w})']\}$ and $D = \mathrm{tr}\{E[(\mathbf{x}_t - \mu)(\mathbf{x}_t - \mu)'\mathbf{w}\mathbf{w}']\} = \mathrm{tr}\{\Sigma\mathbf{w}\mathbf{w}'\} = \mathbf{w}'\Sigma\mathbf{w}$ are immediate. For the middle term B we have

$$\begin{aligned}
B &= \mathrm{tr}\{E[\mathbf{x}_t\mathbf{x}_t'\mathbf{w}\widehat{\mathbf{w}}_\alpha - \mathbf{x}_t\mathbf{x}_t'\mathbf{w}\mathbf{w}' - \mu\mu'\mathbf{w}\widehat{\mathbf{w}}_\alpha' + \mu\mu'\mathbf{w}\mathbf{w}']\} \\
&= \mathrm{tr}\{E[\mathbf{x}_t\mathbf{x}_t'(\mathbf{w}(\mathbf{w} + \alpha(\mathbf{w} - p^{-1}\mathbf{1}))' - \mathbf{w}\mathbf{w}') - \mu\mu'(\mathbf{w}(\mathbf{w} + \alpha(\mathbf{w} - p^{-1}\mathbf{1}))' - \mathbf{w}\mathbf{w}')]\} \\
&= \mathrm{tr}\{\alpha E[\mathbf{x}_t\mathbf{x}_t' - \mu\mu']\mathbf{w}(\mathbf{w} - p^{-1}\mathbf{1})'\} = \alpha(\mathbf{w} - p^{-1}\mathbf{1})'\Sigma\mathbf{w} = 0. \qquad \square
\end{aligned}$$

References

1. Bodnar, T., Mazur, S., Podgórski, K.: Singular inverse Wishart distribution and its application to portfolio theory. J. Multivar. Anal. **143**, 314–326 (2016)
2. Bodnar, T., Parolya, N., Schmid, W.: Estimation of the global minimum variance portfolio in high dimensions. Eur. J. Oper. Res. **266**(1), 371–390 (2018)
3. CandŔes, E.J.: Modern statistical estimation via oracle inequalities. Acta Numer. **15**, 257–325 (2006)
4. DeMiguel, V., Garlappi, L., Nogales, F.J., Uppal, R.: A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. Manage. Sci. **55**(5), 798–812 (2009)
5. Dey, D.K., Ghosh, M., Srinivasan, C.: A new class of improved estimators of a multinormal precision matrix. Stat. Decis. **8**(2), 141–152 (1990)
6. Dickinson, J.P.: The reliability of estimation procedures in portfolio analysis. J. Financ. Quant. Anal. **9**(3), 447–462 (1974)
7. Efron, B., Morris, C.: Multivariate empirical Bayes and estimation of covariance matrices. Ann. Stat. **4**(1), 22–32 (1976)
8. Frahm, G., Memmel, C.: Dominating estimators for minimum-variance portfolios. J. Econom. **159**(2), 289–302 (2010)
9. Frankfurter, G.M., Phillips H.E., Seagle, J.P.: Portfolio selection: the effects of uncertain means, variances, and covariances. J. Financ. Quant. Anal. **6**(5), 1251–1262 (1971)
10. Haff, L.R.: Estimation of the inverse covariance matrix: random mixtures of the inverse Wishart matrix and the identity. Ann. Stat. **7**(6), 1264–1276 (1979)
11. Haff, L.R.: The variational form of certain bayes estimators. Ann. Stat. **19**(3), 1163–1190 (1991)
12. Holgersson, T., Karlsson, P., Stephan, A.: A risk perspective of estimating portfolio weights of the global minimum variance portfolio. Adv. Stat. Anal. **104**(1), 59–80 (2020)
13. Jagannathan, R., Ma, T.: Risk reduction in large portfolios: why imposing the wrong constraints helps. J. Finance **58**(4), 1651–1684 (2003)
14. James, W., Stein, C.: Estimation with quadratic loss. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 361–379 (1961)
15. Jobson, J.D., Korkie, B.: Estimation for Markowitz efficient portfolios. J. Am. Stat. Assoc. **75**(371), 544–554 (1980)
16. Jorion, P.: International portfolio diversification with estimation risk. J. Bus. **58**(3), 259–278 (1985)
17. Kan, R., Zhou, G.: Optimal portfolio choice with parameter uncertainty. J. Financ. Quant. Anal. **42**(3), 621–656 (1985)
18. Kempf, A., Memmel, C.: Estimating the global minimum variance portfolio. Schmalenbach Bus. Rev. **58**, 332–348 (2006)
19. Kotz, S., Nadarajah, S.: Multivariate T-Distributions and Their Applications, 1st edn. Cambridge University Press, Cambridge (2004)
20. Markowitz, H.: Portfolio selection. J. Financ. **7**(1), 77–91 (1952)
21. Michaud, R.O.: The Markowitz optimization enigma: is optimized optimal? ICFA Continuing Educ. Ser. **4**, 43–54 (1989)
22. Okhrin, Y., Schmid, W.: Distributional properties of portfolio weights. J. Econom. **134**(1), 235–256 (2006)
23. Rubio, F., Mestre, X., Palomar, D.P.: Performance analysis and optimal selection of large minimum variance portfolios under estimation risk. IEEE J. Select. Top. Signal Process. **6**(4), 337–350 (2012)
24. Serdobolskii, V.I.: Estimation of high-dimensional inverse covariance matrices. In: Multivariate Statistical Analysis: A High-Dimensional Approach, pp. 87–101. Springer, New York, NY (2000)

25. Tsukuma, H., Konno, Y.: On improved estimation of normal precision matrix and discriminant coefficients. J. Multivar. Anal. **97**, 1477–1500 (2006)
26. Zhou, X., Sun, X., Wang, J.: Estimation of the multivariate normal precision matrix under the entropy loss. Ann. Inst. Stat. Math. **53**(4), 760–768 (2001)