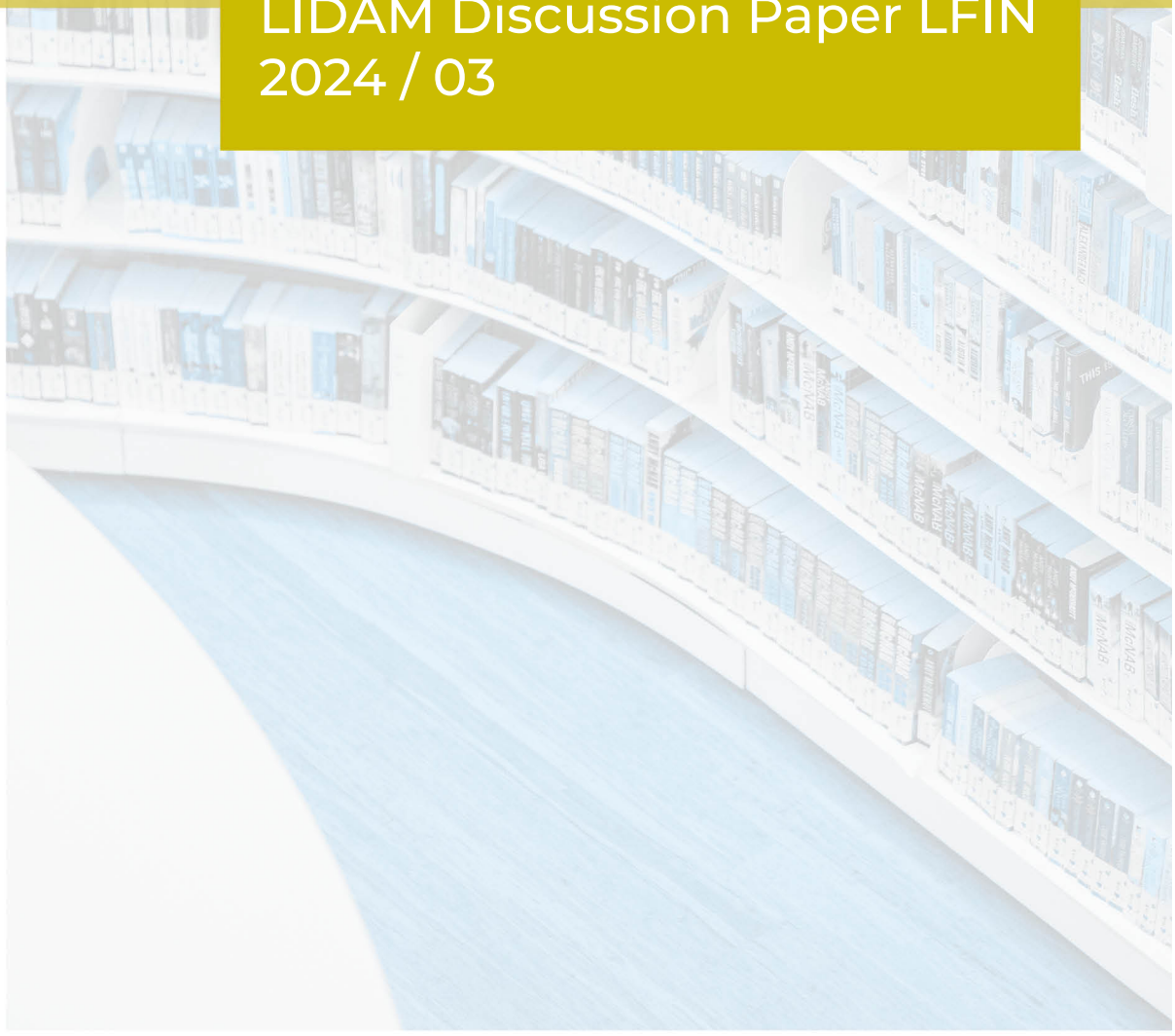
THE ECONOMIC VALUE OF MEAN SQUARED ERROR: EVIDENCE FROM PORTFOLIO SELECTION

Zhaokun Cai, Zhenyu Cui, Nathan Lassance, Majeed Simaan

LIDAM Discussion Paper LFIN
2024 / 03



UCLouvain

LFIN | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

LFIN

Voie du Roman Pays 34, L1.03.01 B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/lfin/publications.html

The Economic Value of Mean Squared Error: Evidence from Portfolio Selection*

Zhaokun Cai[1] Zhenyu Cui[1] Nathan Lassance[2] Majeed Simaan[1]

June 6, 2024

Abstract

When designing and evaluating estimators, the mean squared error (MSE) is the most commonly used generic statistical loss function because it captures the bias-variance tradeoff and allows easy analytical and numerical treatment. However, MSE estimators are often applied to decision problems for which the loss function is different, raising questions about how much value there is in using a generic statistical loss function like the MSE rather than a decision loss function. We elucidate this question through the lens of the portfolio selection problem by showing that for several important portfolio rules, there is a positive linear relation between the MSE and a portfolio-decision loss function. Moreover, shrinkage portfolio estimators derived under these two loss functions are typically close to each other. Our findings highlight the economic value of MSE to serve as a general-purpose statistical loss function in portfolio selection.

Keywords: loss functions, decision theory, out-of-sample risk, investment.

*Majeed Simaan is corresponding author. (1) School of Business, Stevens Institute of Technology, Hoboken, NJ, USA. (2) LFIN/LIDAM, UCLouvain, Mons, Belgium.

1 Introduction

When evaluating and designing estimators, the decision-maker needs to select a loss function. For that purpose, one can boil down the set of loss functions into two possible choices. The first option, the most common and a standard in statistics, econometrics, and machine learning, is to select a generic statistical loss function like the mean squared error (MSE), which we choose as a representative benchmark. The MSE is appealing because it not only captures the bias-variance tradeoff but also allows easy analytical and numerical treatment. Thus, by using a generic loss function like the MSE, one expects to obtain estimators that are valuable for a wide range of applications. The second option, as advocated by statistical decision theory (Berger, 1980), is to select a loss function consistent with the specific decision problem to which the estimator will be applied, i.e., a decision loss function. Decision loss functions have the benefit of delivering optimal estimators for the problem of interest. However, they may be more challenging both analytically and numerically than the MSE, formulating a specific decision loss function is not always trivial for multifaceted problems, and the resulting estimators may not pertain to other applications.

In this context, the more generic MSE loss function actually offers greater flexibility for dealing with decision-related tasks, even though it may be deemed as not directly related to the downstream applications. In this paper, our objective is to investigate analytically and empirically how much economic value MSE estimators carry relative to decision loss functions.[1]

We can expect that there is no unequivocal answer to this research question and that it depends on the specific setting considered. Indeed, the current state of the question in the literature is mixed. For example, the linear shrinkage method for estimating covariance matrices introduced by Ledoit and Wolf (2003a,b, 2004) is based on the MSE loss function and is successful in a variety of applications. At the same time, their recent nonlinear shrinkage method in Ledoit and Wolf (2017) builds on a decision loss function for solving the portfolio selection problem and outperforms linear shrinkage. Similarly, Butler and Kwon (2023) find that optimizing mean-variance portfolio weights and predicting the necessary inputs in a single step using the predict-and-optimize framework (Elmachtoub and Grigas, 2022) outperforms the standard decoupled approach. As another example,

[1] In the operations research and optimization literature, this objective relates to the predict-then-optimize framework versus the predict-and-optimize one (Elmachtoub and Grigas, 2022). As Vanderschueren et al. (2022) clearly explain in their abstract, "Predictive models are increasingly being used to optimize decision-making and minimize costs. A conventional approach is predict-then-optimize: first, a predictive model is built; then, this model is used to optimize decision-making. A drawback of this approach, however, is that it only incorporates costs in the second stage. Conversely, the predict-and-optimize approach proposes learning a predictive model by directly minimizing the cost of the downstream decision-making task."

the calibration of option pricing models is historically based on the MSE (Bakshi et al., 1997), but the models are then often evaluated using a loss function that reflects the model application, such as pricing, hedging, or speculating. In this context, Christoffersen and Jacobs (2004) warn against the inconsistency of using a loss function for estimating option pricing parameters that is different from the one used for their evaluations. A further illustration is that at least since Fama and French (1993), the asset pricing literature predominantly uses the MSE (or, equivalently, the regression R^2) for evaluating the accuracy of expected return estimates implied by asset pricing models. However, Cenesizoglu and Timmermann (2012), Kelly et al. (2024), and Liu et al. (2024) suggest that the MSE of expected return estimates is not always a good indicator of their economic value, and using a decision loss function for estimating expected returns yields portfolios with larger Sharpe ratios.[2] In general, the MSE and other generic statistical loss functions are at the foundation of regression and machine learning methods, which are applied to countless problems.[3]

In this paper, we provide a formal analytical evaluation of the economic value of the MSE loss function in the context of the important portfolio selection problem. We make three main contributions to the literature. First, we link the MSE and decision loss functions for two fixed portfolio rules: the sample mean-variance (SMV) portfolio with a risk-free asset and the fully invested SMV portfolio. The sample global minimum-variance (GMV) portfolio, another important portfolio rule, is a special case of the latter. The decision loss function we use for these portfolios is the expected out-of-sample utility loss (Kan and Zhou, 2007; Kan et al., 2022), which nests the expected out-of-sample variance loss (Frahm and Memmel, 2010; Bodnar et al., 2018) as a special case. Although the MSE and decision loss functions are fundamentally different, they are both minimized and equal to zero when the portfolio weights are estimated without error, so we can expect some relation between the two. We formalize this intuition and demonstrate theoretically that there is indeed a positive linear relation between the MSE and decision loss functions for the two portfolio rules. That is, the MSE carries economic value. Moreover, the intercept in this linear relation has a minor contribution, and thus, to a very good approximation, the MSE and decision loss functions are simply proportional to each other.

[2] Cenesizoglu and Timmermann (2012) write the following in their abstract: "While there is generally a positive correlation between a return prediction model's out-of-sample statistical performance and its ability to add economic value, the relation tends to be weak [...]." Similarly, Kelly et al. (2024, p.4) note that "the out-of-sample R^2 from a prediction model is an incomplete measure of its economic value. A market timer can generate significant economic profits even when the predictive R^2 is negative."

[3] As Elmachtoub and Grigas (2022) write in their abstract, "By and large, machine learning tools are intended to minimize prediction error and do not account for how the predictions will be used in the downstream optimization problem."

Second, we compare optimal shrinkage portfolio estimators derived under the MSE and decision loss functions. Specifically, we consider the shrinkage of the SMV portfolio toward the risk-free asset as in Kan and Zhou (2007), and the shrinkage of the fully invested SMV portfolio toward the sample GMV portfolio as in Kan et al. (2022) and Lassance et al. (2024).[4] In either case, we calibrate the shrinkage intensities based on the MSE and decision loss functions. We show that although the two loss functions do not deliver the same shrinkage intensities, these intensities are typically close to each other. Therefore, although the MSE is unrelated to portfolio selection,[5] the theoretical performance loss incurred by relying on the MSE to calibrate shrinkage portfolios is very small. Interestingly, we also demonstrate that the MSE typically delivers a more intensive shrinkage than the decision loss function and, thus, more conservative portfolios.

Third, we back our theoretical results with empirical evidence across several datasets. We introduce empirical counterparts to the MSE and decision loss functions and show that the positive linear relation between the two also holds in the data, in which our theoretical assumptions are typically not fulfilled. Moreover, we compare the MSE-optimized and decision-optimized shrinkage portfolios and find that they deliver similar out-of-sample performance, further confirming the economic value of the generic MSE statistical loss function.

Our conclusions mirror similar ones in the risk-neutral world, i.e., under the $\mathbb{Q}$ probability measure. Specifically, Christoffersen and Jacobs (2004) argue for MSE as a general-purpose loss function for option valuation, and later their conjecture is confirmed by Bams et al. (2009). What we do is the counterpart of that in the real world, i.e., under the $\mathbb{P}$ probability measure, using portfolio selection as the application rather than option valuation. As the authors show, MSE is a general-purpose loss function in the $\mathbb{Q}$ world, and we demonstrate MSE is general-purpose in the $\mathbb{P}$ world.

The rest of the paper is organized as follows. In Section 2, we present our general definition of the MSE and decision loss functions and how to establish the link between the two. In Section 3, we define the portfolio selection problem with parameter uncertainty. In Section 4, we derive the relation between the MSE and decision loss functions for several fixed portfolio rules. In Section 5, we compare shrinkage portfolio estimators calibrated to the MSE and to the decision loss function,

[4] Shrinking portfolio weights is one among other existing methods to alleviate estimation risk such as shrinkage of inputs (Ledoit and Wolf, 2017; Barroso and Saxena, 2021), robust optimization (El Ghaoui et al., 2003), robust estimation (DeMiguel and Nogales, 2009), weight constraints (DeMiguel et al., 2009a; Yen, 2016), and principal and independent component analysis (Zhao et al., 2019; Lassance et al., 2022).

[5] We demonstrate that the decision loss function corresponds to a more general MSE that uses the return covariance matrix as a scaling matrix. Therefore, both the MSE and decision loss functions can be interpreted as a measure of distance between the estimated and population portfolios. The difference is that the distance measure used by the MSE is unrelated to the portfolio problem while that used by the decision loss is.

respectively. In Section 6, we derive unbiased estimators of the quantities involved in our optimal shrinkage intensities. In Section 7, we provide empirical evidence supporting our theoretical findings. Finally, Section 8 concludes. The proofs of all results are available in the appendix.

2 Definition of MSE and decision loss functions

Let $\hat{\theta}$ be an estimator of $\theta \in \mathbb{R}^N$ obtained from an i.i.d. data sample drawn from a distribution $\mathcal{D}$. We define the MSE of $\hat{\theta}$ as follows.

Definition 1 *The MSE loss function for the estimator $\hat{\theta}$ of θ is*

$$M_A(\hat{\theta}) = \mathbb{E}\Big[L_A(\hat{\theta} - \theta)\Big] = \mathbb{E}\Big[(\hat{\theta} - \theta)^\top A(\hat{\theta} - \theta)\Big] \geq 0, \tag{2.1}$$

where $A \in \mathbb{R}^{N \times N}$ is a constant positive-definite scaling matrix, $L_A(x) = x^\top Ax$ is the A-norm of the vector x, and the expectation is evaluated under the distribution $\mathcal{D}$.

This definition extends the standard MSE that uses $A = I$, the identity matrix. The MSE is a popular choice in statistics because it captures the bias-variance tradeoff. Specifically, we have

$$M_A(\hat{\theta}) = L_A(\mathbb{E}[\hat{\theta}] - \theta) + \operatorname{tr}(A\mathbb{V}[\hat{\theta}]), \tag{2.2}$$

where the first term is zero when the estimator is unbiased, and the second term is zero when the estimator has zero variance.

Next, consider the case in which θ solves a particular decision problem, i.e.,

$$\theta = \operatorname*{argmax}_{x \in \mathcal{C}} P(x) \tag{2.3}$$

for some performance measure P and constraint set $\mathcal{C}$.[6] We define the decision loss function for $\hat{\theta}$ as follows.

Definition 2 *The decision loss function for the estimator $\hat{\theta} \in \mathcal{C}$ of $\theta = \operatorname*{argmax}_{x \in \mathcal{C}} P(x)$ is*

$$D_P(\hat{\theta}) = P(\theta) - \mathbb{E}[P(\hat{\theta})] \geq 0, \tag{2.4}$$

where the expectation is evaluated under the distribution $\mathcal{D}$.

[6] For instance, if θ denotes the portfolio weights and $P(\cdot)$ is a mean-variance utility function, then the optimal θ in Equation (2.3) is the mean-variance efficient portfolio in the set $\mathcal{C}$.

Our objective in the specific portfolio selection problem is twofold. First, for a fixed estimator $\hat{\theta}$ and a chosen matrix A, find a relation between $M_A(\hat{\theta})$ and $D_P(\hat{\theta})$, i.e., some function f such that

$$M_A(\hat{\theta}) = f(D_P(\hat{\theta})). \tag{2.5}$$

In other words, how proportional is the MSE to the decision loss function? Second, for a parameterized estimator $\hat{\theta}(\delta)$, compare $\delta_M = \operatorname{argmin}_\delta M_A(\hat{\theta}(\delta))$ with $\delta_D = \operatorname{argmin}_\delta D_P(\hat{\theta}(\delta))$ and evaluate the increase in decision loss:

$$\Delta_P(\hat{\theta}(\delta)) = D_P(\hat{\theta}(\delta_M)) - D_P(\hat{\theta}(\delta_D)) = \mathbb{E}[P(\hat{\theta}(\delta_D))] - \mathbb{E}[P(\hat{\theta}(\delta_M))] \geq 0. \tag{2.6}$$

Put differently, what is the cost of using a generic loss function like the MSE? If f in (2.5) is an increasing function, and if Δ_P in (2.6) is small, then the MSE is a valuable loss function for the decision problem of interest. Note that f being an increasing function and Δ_P being small do not necessarily imply each other.

3 Portfolio selection problem

We now present the portfolio selection problem with parameter uncertainty, an important problem in the finance literature (Jagannathan and Ma, 2003; Kan and Zhou, 2007; DeMiguel et al., 2009b; Tu and Zhou, 2011; Ao et al., 2019; Kan et al., 2022; Barroso and Saxena, 2021). We start with the simpler case in which the parameters are known.

3.1 Portfolio selection without parameter uncertainty

Let $r \in \mathbb{R}^N$ be a vector of excess returns on $N \geq 2$ assets with mean μ and positive-definite covariance matrix Σ. The portfolio selection problem consists in finding an optimal portfolio $w \in \mathbb{R}^N$, where w_i is the percentage of capital allocated to asset i, according to a performance measure P capturing some aspects of the distribution of the portfolio return $w^\top r$. The performance measure P we consider is the mean-variance utility,

$$P(w) = w^\top \mu - \frac{\gamma}{2} w^\top \Sigma w, \tag{3.1}$$

where $\gamma > 0$ is the risk-aversion coefficient, which also comprises the portfolio-return variance as a special case when $\gamma \to \infty$.

The mean-variance utility performance measure yields two portfolio rules of interest. The first

one is the mean-variance (MV) portfolio[7] that maximizes $P(w)$ without any constraint, thus allowing a risk-free asset,

$$w_m = \underset{w}{\operatorname{argmax}}\, P(w) = \frac{1}{\gamma}\Sigma^{-1}\mu. \tag{3.2}$$

The performance delivered by w_m is

$$P(w_m) = \frac{\mu^\top \Sigma^{-1}\mu}{2\gamma}. \tag{3.3}$$

The second portfolio rule is the fully invested MV portfolio that maximizes $P(w)$ under the constraint $w^\top \iota = 1$ with ι a vector of ones,

$$w_{mf} = \underset{w:w^\top\iota=1}{\operatorname{argmax}}\, P(w) = w_g + w_a, \tag{3.4}$$

where w_g is the global minimum-variance (GMV) portfolio,

$$w_g = \frac{\Sigma^{-1}\iota}{\iota^\top \Sigma^{-1}\iota}, \tag{3.5}$$

and w_a is the arbitrage portfolio that maximizes $P(w)$ under the constraint $w^\top \iota = 0$,

$$w_a = \frac{1}{\gamma}Q\mu \quad \text{with} \quad Q = \Sigma^{-1}\Big(I - \iota w_g^\top\Big). \tag{3.6}$$

The performance delivered by w_{mf} is

$$P(w_{mf}) = P(w_g) + P(w_a) = \left(\mu_g - \frac{\gamma}{2}\sigma_g^2\right) + \frac{\mu^\top Q\mu}{2\gamma}, \tag{3.7}$$

where $\mu_g = w_g^\top \mu$ and $\sigma_g^2 = w_g^\top \Sigma w_g = 1/(\iota^\top \Sigma^{-1}\iota)$. Note that the GMV portfolio w_g is a special case of the fully invested MV portfolio w_{mf} as $\gamma \to \infty$.

3.2 Portfolio selection with parameter uncertainty

Now, in practice, μ and Σ are unknown, and thus, so are the optimal portfolios w_m and w_{mf}. Therefore, we assume that we have an i.i.d. sample of returns of size T, $(r_1, \ldots, r_T)$, which gives us the following sample estimators of μ and Σ,

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t, \tag{3.8}$$

[7] Even when investors have preferences for higher-order moments, holding a mean-variance portfolio is often preferable to holding the global-optimal portfolio due to estimation risk; see Simaan (2014), Khashanah et al. (2022), and Lassance (2022).

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(r_t - \hat{\mu})(r_t - \hat{\mu})^\top, \tag{3.9}$$

where we assume $T > N$ so that $\hat{\Sigma}$ is invertible. We make the following assumption about the distribution $\mathcal{D}$ from which the sample returns are drawn.

Assumption 1 *The i.i.d. sample of returns,* $(r_1, \ldots, r_T)$*, is drawn from a multivariate normal distribution,* $\mathcal{D} \sim \mathcal{N}(\mu, \Sigma)$*.*

Assumption 1 is the common assumption made in the literature to study the out-of-sample performance of estimated portfolios under estimation risk and in finite samples (Okhrin and Schmid, 2006; Kan and Zhou, 2007; Tu and Zhou, 2011; DeMiguel et al., 2015; Clark et al., 2022; Kan et al., 2022; Lassance et al., 2023, 2024).

Under Assumption 1, we can show from Okhrin and Schmid (2006) that the sample unbiased estimators of the optimal portfolios w_m and w_{mf} in (3.2) and (3.4) are

$$\hat{w}_m = \frac{T - N - 2}{\gamma T}\hat{\Sigma}^{-1}\hat{\mu}, \tag{3.10}$$

$$\hat{w}_{mf} = \hat{w}_g + \hat{w}_a = \frac{\hat{\Sigma}^{-1}\iota}{\iota^\top\hat{\Sigma}^{-1}\iota} + \frac{T - N - 1}{\gamma T}\hat{Q}\hat{\mu} \quad \text{with} \quad \hat{Q} = \hat{\Sigma}^{-1}\Big(I - \iota\hat{w}_g^\top\Big). \tag{3.11}$$

We refer to $\hat{w}_m$ and w_{mf} as the sample mean-variance (SMV) portfolio and the fully invested SMV portfolio, respectively. In Section 4, we measure the degree of estimation errors in these estimates using MSE and decision loss functions, and we relate the two loss functions together. Then, in Section 5, we introduce shrinkage estimators of these portfolios and compare the optimal shrinkage intensities under the MSE and decision loss functions, respectively.

4 Relating the MSE and decision loss functions

In the following proposition, we address the first objective outlined in Section 2 by relating the MSE and decision loss functions for $\hat{w}_m$ and $\hat{w}_{mf}$ in (3.10)–(3.11). We consider two scaling matrices for defining the MSE, $A = \Sigma$ and $A = I$. The decision loss functions $D_P(\hat{w}_m)$ and $D_P(\hat{w}_{mf})$ are known as the expected out-of-sample utility loss in the portfolio selection literature (Kan and Zhou, 2007; Tu and Zhou, 2011), which nests the expected out-of-sample variance loss (Frahm and Memmel, 2010; Bodnar et al., 2018) as a special case when $\gamma \to \infty$.

Proposition 1 *Let Assumption 1 hold and $T > N + 4$. Then,*

1. *The decision loss function for the portfolio estimators $\hat{w}_m$ and $\hat{w}_{mf}$ in (3.10)–(3.11) is*

$$D_P(\hat{w}_m) = (c_1 - 1)P(w_m) + \frac{c_1}{2\gamma}\frac{N}{T}, \tag{4.1}$$

$$D_P(\hat{w}_{mf}) = D_P(\hat{w}_g) + D_P(\hat{w}_a) = \frac{\gamma}{2}(c_0 - 1)\sigma_g^2 + (c_2 - 1)P(w_a) + \frac{c_2}{2\gamma}\frac{N-1}{T}, \tag{4.2}$$

where $c_0 = \frac{T-2}{T-N-1}$, $c_1 = \frac{(T-2)(T-N-2)}{(T-N-1)(T-N-4)}$, and $c_2 = \frac{(T-2)(T-N-1)}{(T-N)(T-N-3)}$ are larger than one.

2. *When the scaling matrix $A = \Sigma$, the MSE and decision loss functions are equivalent. Specifically,*

$$M_\Sigma(\hat{w}_m) = \frac{2}{\gamma}D_P(\hat{w}_m), \tag{4.3}$$

$$M_\Sigma(\hat{w}_{mf}) = \frac{2}{\gamma}D_P(\hat{w}_{mf}). \tag{4.4}$$

3. *When the scaling matrix $A = I$, the MSE and decision loss functions have a positive linear relation. Specifically,*

$$M_I(\hat{w}_m) = \frac{2\bar{\tau}}{\gamma}D_P(\hat{w}_m) + \left(\frac{T-N-2}{T-2}c_1 - 1\right)\frac{\kappa_1}{\gamma^2}, \tag{4.5}$$

$$M_I(\hat{w}_{mf}) = M_I(\hat{w}_g) + M_I(\hat{w}_a) = \frac{2\bar{q}}{\gamma}D_P(\hat{w}_{mf}) + \left(\frac{T-N-1}{T-2}c_2 - 1\right)\frac{\kappa_2}{\gamma^2}, \tag{4.6}$$

where $M_I(\hat{w}_g) = \bar{q}(c_0 - 1)\sigma_g^2$, $\bar{\tau} = \frac{1}{N}\mathrm{tr}(\Sigma^{-1})$, $\bar{q} = \frac{1}{N-1}\mathrm{tr}(Q)$, and

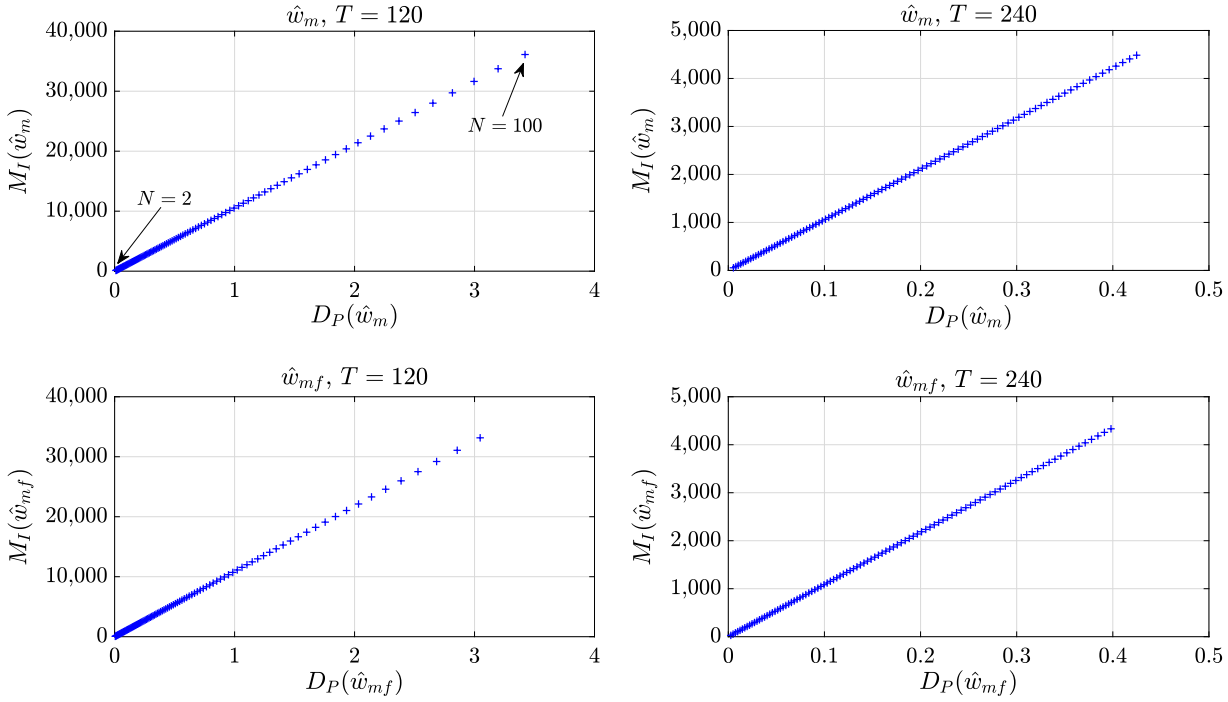
$$\kappa_1 = \mu^\top \Sigma^{-1}(\Sigma^{-1} - \bar{\tau}I)\mu, \tag{4.7}$$

$$\kappa_2 = \mu^\top Q(Q - \bar{q}I)\mu. \tag{4.8}$$

Several comments are in order. First, Part 1 of Proposition 1 shows that the decision loss functions increase with N and decrease with T, as one would expect. Second, Part 2 of Proposition 1 shows that provided we define the MSE with a scaling matrix $A = \Sigma$, the MSE and decision loss functions are equivalent, though with a scaling adjustment attributable to γ.

Third, Part 3 of Proposition 1 shows that in the more common definition of the MSE with a scaling matrix $A = I$, there is a positive linear relation between the MSE and decision loss functions, meaning that the MSE carries valuable economic content for the portfolio problem. Figure 1 illustrates this positive relation, where we consider a sample size of either $T = 120$ or 240 months and a risk-aversion coefficient $\gamma = 1$. We calibrate (μ, Σ) to a dataset of monthly returns on

Figure 1: Positive linear relation between the MSE and decision loss functions



Notes. This figure depicts the relation between the mean squared error (MSE) and decision loss functions for the sample mean-variance portfolio ($\hat{w}_m$ in (3.10)) and the fully invested sample mean-variance portfolio ($\hat{w}_{mf}$ in (3.11)). The decision loss function is the expected out-of-sample utility loss and is given by $D_P(\hat{w}_m)$ and $D_P(\hat{w}_{mf})$ in (4.1)–(4.2). The positive linear relation between the MSE, $M_I(\hat{w}_m)$ and $M_I(\hat{w}_{mf})$, and decision loss functions is given by (4.5)–(4.6). The left and right plots consider a sample size of $T = 120$ and 240 months, respectively. In each plot, the blue crosses depict the relation for N going from 2 to 100. We consider a risk-aversion coefficient $\gamma = 1$ and we calibrate the mean and covariance matrix (μ, Σ) to a dataset of monthly returns on 25 portfolios of firms sorted on size and book-to-market spanning July 1963 to December 2023.

25 portfolios of firms sorted on size and book-to-market (25SBM) spanning July 1963 to December 2023. Then, we depict the relation between the MSE and decision loss functions for N increasing from 2 to 100. We can clearly observe the positive linear relation and that both loss functions increase with N and decrease with T.

Fourth, the signs of the intercepts in the linear relations between the MSE and decision loss functions in (4.5)–(4.6) depend on the signs of κ_1 and κ_2 in (4.7)–(4.8), respectively.[8] It is important to understand whether κ_1 and κ_2 tend to be positive or negative because, as we will show in Section 5, they also control which of the MSE or decision loss functions delivers a more intensive shrinkage of the SMV and fully invested SMV portfolios. In the next proposition, we show that these two parameters are negative if the squared Sharpe ratios of the principal components (PCs) form a decreasing sequence.

[8]Indeed, $c_1(T - N - 2)/(T - 2) - 1$ and $c_2(T - N - 1)/(T - 2) - 1$ are both positive.

Proposition 2 *Let $s_i^2 = (v_i^\top \mu)^2/(v_i^\top \Sigma v_i)$ be the squared Sharpe ratio of the ith PC of asset returns, where v_i is the ith eigenvector of the covariance matrix Σ. Then, a sufficient condition under which $\kappa_1 \leq 0$ is that $s_1^2 \geq s_2^2 \geq \cdots \geq s_N^2$. The same condition also implies $\kappa_2 \leq 0$ if, in addition, the first eigenvector is constant, i.e., $v_1 = \iota/\sqrt{N}$.*

Kozak et al. (2020) explain that the non-existence of near-arbitrage opportunities in the market implies that PCs should have decreasing Sharpe ratios. Discussing two classes of asset pricing models, they write "To the extent it exists, mispricing then appears in the SDF mainly through the risk prices of high-eigenvalue PCs. Thus, within both classes of asset pricing models, we would expect Sharpe ratios to be increasing in the eigenvalue." Regarding the additional condition that v_1 is constant, Kozak et al. (2018) make a similar assumption, which we can interpret as the market explaining most of the variations in asset returns.

However, note that the squared Sharpe ratios of PCs forming a decreasing sequence is only a sufficient condition for $\kappa_1 \leq 0$, and $\kappa_2 \leq 0$ if, in addition, v_1 is a constant. In all four equity datasets we use in our empirical analysis, we observe that there is a strong enough factor structure for both parameters to be mostly negative over time even if the s_i^2 are not exactly monotonically decreasing and even if v_1 is not exactly a constant vector; see Figure 5.

Finally, we find numerically that in the decomposition of the MSE of $\hat{w}_m$ and $\hat{w}_{mf}$ in (4.5)–(4.6), the first term related to the decision loss function dominates the second term by several orders of magnitude. For example, let $T = 120$, $N = 25$, $\gamma = 1$, and (μ, Σ) be calibrated to the 25SBM dataset as above. Then, we have

$$M_I(\hat{w}_m) = \frac{2\bar{\tau}}{\gamma} D_P(\hat{w}_m) + \left(\frac{T-N-2}{T-2} c_1 - 1\right)\frac{\kappa_1}{\gamma^2} = 1{,}666 - 3.235, \tag{4.9}$$

$$M_I(\hat{w}_{mf}) = \frac{2\bar{q}}{\gamma} D_P(\hat{w}_{mf}) + \left(\frac{T-N-1}{T-2} c_2 - 1\right)\frac{\kappa_2}{\gamma^2} = 1{,}551 - 1.980. \tag{4.10}$$

Therefore, to a very good approximation, the MSE and decision loss functions are proportional to each other. In unreported numerical results, we find that this result is highly robust to different values of N, T, and (μ, Σ).

5 MSE versus decision loss for shrinkage portfolios

We now address the second objective outlined in Section 2, which is to consider parametrized estimators and to compare the optimal parameters found under the MSE and decision loss functions

as well as the increase in decision loss caused by the reliance on MSE. The two parametrized estimators we consider are shrinkage portfolio estimators that shrink $\hat{w}_m$ and $\hat{w}_{mf}$ in (3.10)–(3.11) to alleviate the estimation risk they are facing.

The first shrinkage estimator, studied by Kan and Zhou (2007), shrinks the SMV portfolio $\hat{w}_m$ toward the risk-free asset,

$$\hat{w}_m(\delta) = (1-\delta)\hat{w}_m. \tag{5.1}$$

The second shrinkage estimator, studied by Bodnar et al. (2022), Kan et al. (2022), and Lassance et al. (2024), shrinks the fully invested SMV portfolio toward the sample GMV portfolio,

$$\hat{w}_{mf}(\delta) = (1-\delta)\hat{w}_{mf} + \delta\hat{w}_g = \hat{w}_g + (1-\delta)\hat{w}_a. \tag{5.2}$$

In the next proposition, we derive the optimal shrinkage intensities δ under the MSE and decision loss functions, and we evaluate Δ_P in (2.6), which tells us how much portfolio performance we lose by calibrating δ based on the MSE. Given Part 2 of Proposition 1, it is not difficult to show that when the scaling matrix $A = \Sigma$ in the MSE, the optimal shrinkage intensities under the MSE and decision loss functions are the same. Therefore, we focus on the case $A = I$.

Proposition 3 *Let Assumption 1 hold and $T > N + 4$. Then,*

1. *The optimal shrinkage intensities under the decision loss function for the shrinkage estimators $\hat{w}_m(\delta)$ and $\hat{w}_{mf}(\delta)$ in (5.1)–(5.2) are*

$$\delta_{m,D} = \operatorname*{argmin}_{\delta} D_P(\hat{w}_m(\delta)) = \frac{D_P(\hat{w}_m)}{D_P(\hat{w}_m) + P(w_m)}, \tag{5.3}$$

$$\delta_{mf,D} = \operatorname*{argmin}_{\delta} D_P(\hat{w}_{mf}(\delta)) = \frac{D_P(\hat{w}_a)}{D_P(\hat{w}_a) + P(w_a)}, \tag{5.4}$$

   which are both between zero and one and independent of γ. Moreover, the decision loss delivered by these optimal shrinkage intensities is

$$D_P(\hat{w}_m(\delta_{m,D})) = \delta_{m,D} P(w_m), \tag{5.5}$$

$$D_P(\hat{w}_{mf}(\delta_{mf,D})) = D_P(\hat{w}_g) + \delta_{mf,D} P(w_a). \tag{5.6}$$

2. *The optimal shrinkage intensities under the MSE loss function with a scaling matrix $A = I$*

are

$$\delta_{m,M} = \underset{\delta}{\operatorname{argmin}}\, M_I(\hat{w}_m(\delta)) = \frac{M_I(\hat{w}_m)}{M_I(\hat{w}_m) + L_I(w_m)}, \tag{5.7}$$

$$\delta_{mf,M} = \underset{\delta}{\operatorname{argmin}}\, M_I(\hat{w}_{mf}(\delta)) = \frac{M_I(\hat{w}_a)}{M_I(\hat{w}_a) + L_I(w_a)}, \tag{5.8}$$

which are both between zero and one and independent of γ*. Moreover, the MSE delivered by these optimal shrinkage intensities is*

$$M_I(\hat{w}_m(\delta_{m,M})) = \delta_{m,M} L_I(w_m), \tag{5.9}$$

$$M_I(\hat{w}_{mf}(\delta_{mf,M})) = M_I(\hat{w}_g) + \delta_{mf,M} L_I(w_a). \tag{5.10}$$

3. *The signs of* $\delta_{m,D} - \delta_{m,M}$ *and* $\delta_{mf,D} - \delta_{mf,M}$ *are equal to those of* κ_1 *and* κ_2 *in* (4.7)–(4.8)*, respectively.*

4. *Under the approximation* $M_I(\hat{w}_m) \approx \frac{2\bar{\tau}}{\gamma} D_P(\hat{w}_m)$ *and* $M_I(\hat{w}_{mf}) \approx \frac{2\bar{q}}{\gamma} D_P(\hat{w}_{mf})$ *discussed in Section 4, the shrinkage intensities minimizing the MSE are approximately*

$$\delta_{m,M} \approx \frac{2\gamma D_P(\hat{w}_m)}{2\gamma D_P(\hat{w}_m) + (\gamma^2/\bar{\tau}) L_I(w_m)} = \delta_{m,D}\left(1 - \frac{\kappa_1/\bar{\tau}}{2\gamma D_P(\hat{w}_m) + (\gamma^2/\bar{\tau}) L_I(w_m)}\right), \tag{5.11}$$

$$\delta_{mf,M} \approx \frac{2\gamma D_P(\hat{w}_a)}{2\gamma D_P(\hat{w}_a) + (\gamma^2/\bar{q}) L_I(w_a)} = \delta_{mf,D}\left(1 - \frac{\kappa_2/\bar{q}}{2\gamma D_P(\hat{w}_a) + (\gamma^2/\bar{q}) L_I(w_a)}\right). \tag{5.12}$$

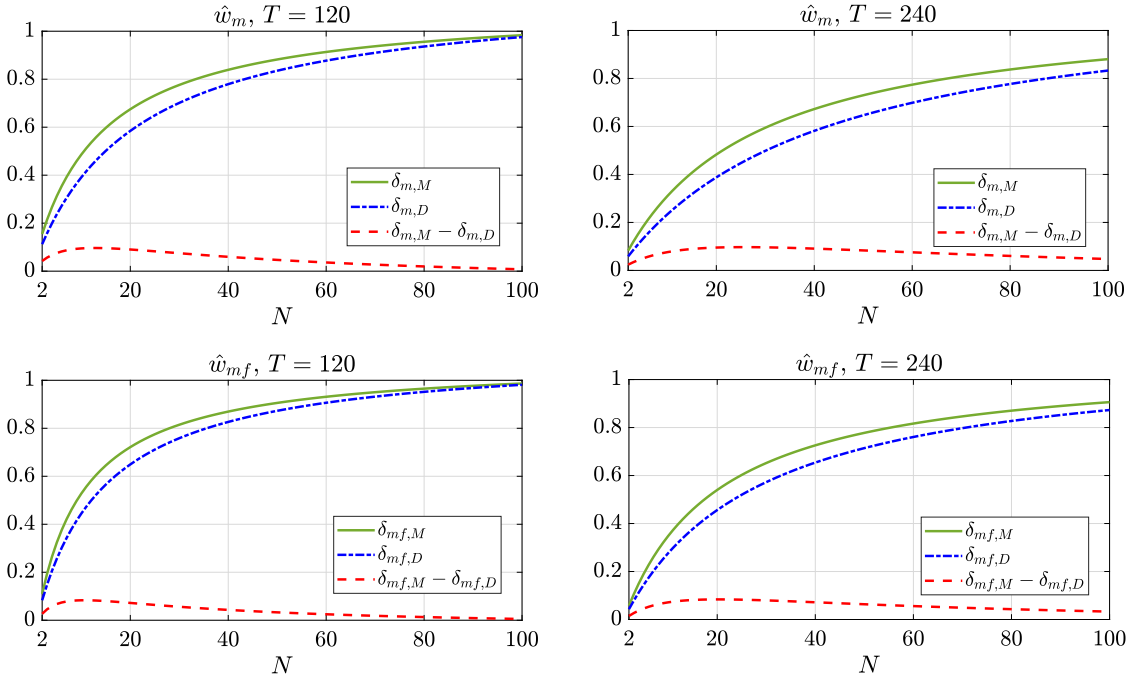
5. *The increase in decision loss due to relying on the MSE loss function instead of the decision loss function to optimize the shrinkage intensities is*

$$\begin{aligned} \Delta_P(\hat{w}_m(\delta)) &= D_P(\hat{w}_m(\delta_{m,M})) - D_P(\hat{w}_m(\delta_{m,D})) \\ &= (\delta_{m,D} - \delta_{m,M})^2 [D_P(\hat{w}_m) + P(w_m)], \end{aligned} \tag{5.13}$$

$$\begin{aligned} \Delta_P(\hat{w}_{mf}(\delta)) &= D_P(\hat{w}_{mf}(\delta_{mf,M})) - D_P(\hat{w}_{mf}(\delta_{mf,D})) \\ &= (\delta_{mf,D} - \delta_{mf,M})^2 [D_P(\hat{w}_a) + P(w_a)]. \end{aligned} \tag{5.14}$$

A first observation is that the optimal shrinkage intensities in Parts 1 and 2 of Proposition 3 take a natural expression. When they minimize the decision loss in (5.3)–(5.4), they increase with the decision loss of the portfolio that is shrunk, and they decrease with the performance inefficiency of the portfolio we shrink toward. A similar intuition holds for the shrinkage intensities that minimize the MSE in (5.7)–(5.8), which increase with the MSE of the portfolio that is shrunk and decrease with the MSE of the portfolio we shrink toward.

Figure 2: Shrinkage intensities under the MSE and decision loss functions

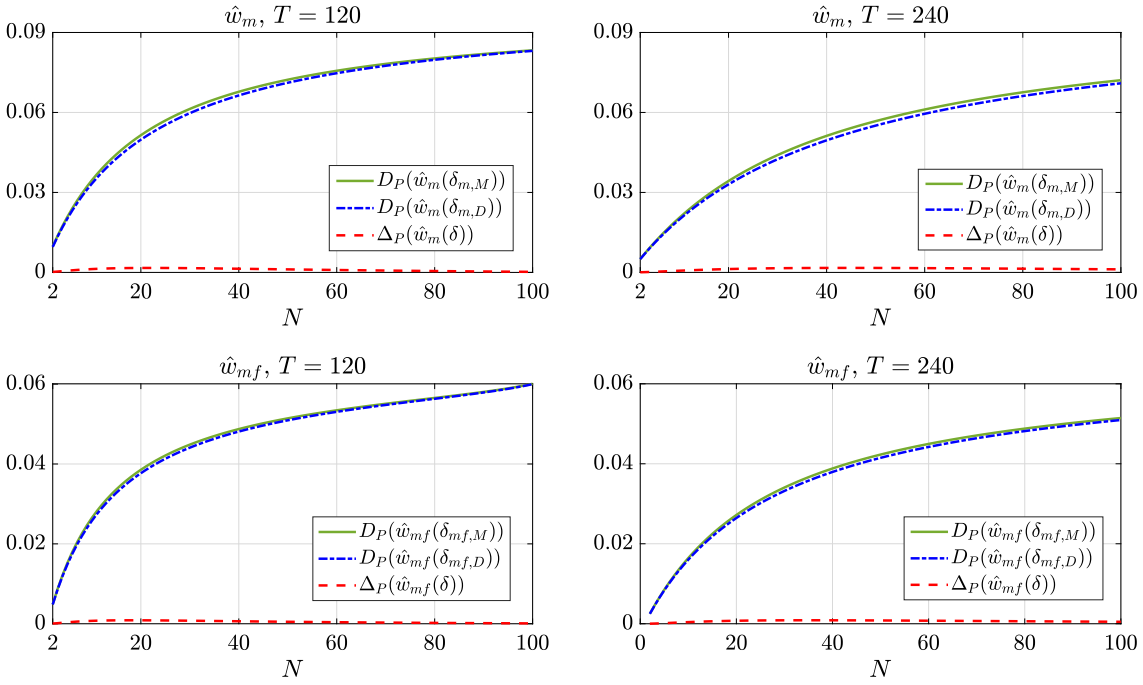


Notes. This figure depicts the shrinkage intensities that minimize the mean squared error (MSE) and decision loss functions for the shrinkage estimators of the sample mean-variance portfolio ($\hat{w}_m(\delta)$ in (5.1)) and the fully invested sample mean-variance portfolio ($\hat{w}_{mf}(\delta)$ in (5.2)). The shrinkage intensities that minimize the MSE are $(\delta_{m,M}, \delta_{mf,M})$ in (5.7)–(5.8) and those that minimize the decision loss are $(\delta_{m,D}, \delta_{mf,D})$ in (5.3)–(5.4). The left and right plots consider a sample size of $T = 120$ and 240 months, respectively. The number of assets N varies from 2 to 100. We calibrate the mean and covariance matrix (μ, Σ) to a dataset of monthly returns on 25 portfolios of firms sorted on size and book-to-market spanning July 1963 to December 2023.

Another result, in Part 3 of Proposition 3, is that although the MSE and decision loss functions for the SMV portfolio and the fully invested SMV portfolio are linearly related to one another, as shown in Part 3 of Proposition 1, the optimal shrinkage intensity to apply to these portfolios is not the same under the two loss functions. In particular, the sign of $\delta_{m,D} - \delta_{m,M}$ is equal to that of κ_1, and the sign of $\delta_{mf,D} - \delta_{mf,M}$ is equal to that of κ_2, respectively. As discussed in Section 4, κ_1 and κ_2 are typically negative, which implies that $\delta_{m,M} \geq \delta_{m,D}$ and $\delta_{mf,M} \geq \delta_{mf,D}$, i.e., we typically shrink more intensively under the MSE loss function than the decision loss objective. In that sense, the MSE leads to more conservative portfolio decisions.

Then, to address the second objective outlined in Section 2, we need to assess how close are the shrinkage intensities minimizing the decision loss and the MSE and what is the increase in decision loss due to relying on the MSE. These two points are covered in Parts 4 and 5 of Proposition 3. Concerning the comparison of shrinkage intensities, how close they are under the MSE and decision loss functions depends respectively on how small κ_1 is relative to $\bar{\tau}$ and how small κ_2 is relative to $\bar{q}$, as shown by (5.11)–(5.12). The answer to this question is an empirical exercise because it

Figure 3: Increase in decision loss due to calibrating shrinkage estimators with the MSE



Notes. This figure depicts the decision loss function for the shrinkage estimators of the sample mean-variance portfolio ($\hat{w}_m(\delta)$ in (5.1)) and the fully invested sample mean-variance portfolio ($\hat{w}_{mf}(\delta)$ in (5.2)). The shrinkage intensities minimize either the MSE, $(\delta_{m,M}, \delta_{mf,M})$ in (5.7)–(5.8), or the decision loss, $(\delta_{m,D}, \delta_{mf,D})$ in (5.3)–(5.4). The increase in decision loss due to using $(\delta_{m,M}, \delta_{mf,M})$ instead of $(\delta_{m,D}, \delta_{mf,D})$ is given by $\Delta_P(\hat{w}_m(\delta))$ and $\Delta_P(\hat{w}_{mf}(\delta))$ in (5.13)–(5.14). The left and right plots consider a sample size of $T = 120$ and 240 months, respectively. The number of assets N varies from 2 to 100. We calibrate the mean and covariance matrix (μ, Σ) to a dataset of monthly returns on 25 portfolios of firms sorted on size and book-to-market spanning July 1963 to December 2023.

depends on (μ, Σ). In the empirical analysis of Section 7, we consistently observe across datasets that the differences between $(\delta_{m,D}, \delta_{mf,D})$ in (5.3)–(5.4) and $(\delta_{m,M}, \delta_{mf,M})$ in (5.7)–(5.8) are small. We illustrate this in Figure 2 using a similar setting to that in Figure 1. When $T = 120$ and 240 while N varies from 2 to 100, the differences $\delta_{m,M} - \delta_{m,D}$ and $\delta_{mf,M} - \delta_{mf,D}$ do not go beyond 0.1, and thus, remain small.

Finally, the increase in decision loss due to relying on the MSE loss function is given by $\Delta_P(\hat{w}_m(\delta))$ in (5.13) and $\Delta_P(\hat{w}_{mf}(\delta))$ in (5.14), which we expect to be small because the differences $\delta_{m,M} - \delta_{m,D}$ and $\delta_{mf,M} - \delta_{mf,D}$ are generally small, as discussed above. Figure 3 depicts all three terms in $\Delta_P(\hat{w}_m(\delta)) = D_P(\hat{w}_m(\delta_{m,M})) - D_P(\hat{w}_m(\delta_{m,D}))$ and in $\Delta_P(\hat{w}_{mf}(\delta)) = D_P(\hat{w}_{mf}(\delta_{mf,M})) - D_P(\hat{w}_{mf}(\delta_{mf,D}))$ using a similar setting to that in Figures 1 and 2. We observe that the Δ_P's are very close to zero. That is, calibrating shrinkage estimators with the MSE instead of the decision loss function barely makes any difference in portfolio performance, and thus, the MSE contains economic value for the portfolio problem.

6 Estimation of optimal shrinkage intensities

The shrinkage intensities proposed in the previous section are oracle, i.e., infeasible to implement in practice, because they depend on the population parameters (μ, Σ). As a result, we propose feasible estimators of the shrinkage intensities in this section. In particular, we now show how to estimate the shrinkage intensities $(\delta_{m,D}, \delta_{mf,D})$ that minimize the decision loss, given by Part 1 of Proposition 3, and the shrinkage intensities $(\delta_{m,M}, \delta_{mf,M})$ that minimize the MSE, given by Part 2 of Proposition 3. Note that $\delta_{m,D}$ requires estimating $(D_P(\hat{w}_m), P(w_m))$, $\delta_{mf,D}$ requires $(D_P(\hat{w}_a), P(w_a))$, $\delta_{m,M}$ requires $(M_I(\hat{w}_m), L_I(w_m))$, and $\delta_{mf,M}$ requires $(M_I(\hat{w}_a), L_I(w_a))$. In this regard, we derive unbiased estimators of these eight quantities in the next proposition.

Proposition 4 *Let Assumption 1 hold and $T > N+4$. Then, the unbiased estimators of $D_P(\hat{w}_m)$, $D_P(\hat{w}_a), P(w_m), P(w_a), M_I(\hat{w}_m), M_I(\hat{w}_a), L_I(w_m), L_I(w_a)$ are*

$$\widehat{D}_P(\hat{w}_m) = \frac{(c_1 - 1)(T - N - 2)\hat{\mu}^\top \hat{\Sigma}^{-1}\hat{\mu} + N}{2\gamma T}, \tag{6.1}$$

$$\widehat{D}_P(\hat{w}_a) = \frac{(c_2 - 1)(T - N - 1)\hat{\mu}^\top \hat{Q}\hat{\mu} + (N - 1)}{2\gamma T}, \tag{6.2}$$

$$\widehat{P}(w_m) = \frac{(T - N - 2)\hat{\mu}^\top \hat{\Sigma}^{-1}\hat{\mu} - N}{2\gamma T}, \tag{6.3}$$

$$\widehat{P}(w_a) = \frac{(T - N - 1)\hat{\mu}^\top \hat{Q}\hat{\mu} - (N - 1)}{2\gamma T}, \tag{6.4}$$

$$\widehat{M_I}(\hat{w}_m) = \frac{(T - N)(T - N - 4) + 2}{\gamma^2 T^2 (T - N - 1)}\hat{\mu}^\top \hat{\Sigma}^{-2}\hat{\mu} + \frac{T - N - 2}{\gamma^2 T^2}\mathrm{tr}(\hat{\Sigma}^{-1})(1 + \hat{\mu}^\top \hat{\Sigma}^{-1}\hat{\mu}), \tag{6.5}$$

$$\widehat{M_I}(\hat{w}_a) = \frac{(T - N + 1)(T - N - 3) + 2}{\gamma^2 T^2 (T - N)}\hat{\mu}^\top \hat{Q}^2\hat{\mu} + \frac{T - N - 1}{\gamma^2 T^2}\mathrm{tr}(\hat{Q})(1 + \hat{\mu}^\top \hat{Q}\hat{\mu}), \tag{6.6}$$

$$\widehat{L}_I(w_m) = \frac{(T - N - 2)(T - N - 4) + 2}{\gamma^2 T^2}\hat{\mu}^\top \hat{\Sigma}^{-2}\hat{\mu} - \frac{T - N - 2}{\gamma^2 T^2}\mathrm{tr}(\hat{\Sigma}^{-1})(1 + \hat{\mu}^\top \hat{\Sigma}^{-1}\hat{\mu}), \tag{6.7}$$

$$\widehat{L}_I(w_a) = \frac{(T - N - 1)(T - N - 3) + 2}{\gamma^2 T^2}\hat{\mu}^\top \hat{Q}^2\hat{\mu} - \frac{T - N - 1}{\gamma^2 T^2}\mathrm{tr}(\hat{Q})(1 + \hat{\mu}^\top \hat{Q}\hat{\mu}). \tag{6.8}$$

Note that the unbiased estimators of $P(w_m), P(w_a), L_I(w_m)$, and $L_I(w_a)$ can turn negative even though the population value is always positive. Therefore, we estimate the optimal shrinkage intensities in Proposition 3 by plugging the unbiased estimators in Proposition 4 and taking the positive part of the unbiased estimators of $P(w_m), P(w_a), L_I(w_m)$, and $L_I(w_a)$. This gives

$$\hat{\delta}_{m,D} = \frac{\widehat{D}_P(\hat{w}_m)}{\widehat{D}_P(\hat{w}_m) + \max(\widehat{P}(w_m), 0)}, \tag{6.9}$$

$$\hat{\delta}_{mf,D} = \frac{\widehat{D}_P(\hat{w}_a)}{\widehat{D}_P(\hat{w}_a) + \max(\widehat{P}(w_a), 0)}, \tag{6.10}$$

$$\hat{\delta}_{m,M} = \frac{\widehat{M}_I(\hat{w}_m)}{\widehat{M}_I(\hat{w}_m) + \max(\widehat{L}_I(w_m), 0)}, \tag{6.11}$$

$$\hat{\delta}_{mf,M} = \frac{\widehat{M}_I(\hat{w}_a)}{\widehat{M}_I(\hat{w}_a) + \max(\widehat{L}_I(w_a), 0)}. \tag{6.12}$$

We conclude this section with the following corollary of Proposition 4, in which we derive unbiased estimators of the parameters κ_1 and κ_2 in (4.7)–(4.8). This result will be useful for the empirical analysis of Section 7.

Corollary 1 *Let Assumption 1 hold and $T > N + 4$. Then, the unbiased estimators of κ_1 and κ_2 in* (4.7)–(4.8) *are*

$$\hat{\kappa}_1 = \frac{N(T-N-2)(T-N-4)+2(T-2)}{NT^2}\hat{\mu}^\top\hat{\Sigma}^{-2}\hat{\mu} - \frac{(T-2)(T-N-2)}{NT^2}\mathrm{tr}(\hat{\Sigma}^{-1})\hat{\mu}^\top\hat{\Sigma}^{-1}\hat{\mu}, \tag{6.13}$$

$$\hat{\kappa}_2 = \frac{(N-1)(T-N-1)(T-N-3)+2(T-2)}{(N-1)T^2}\hat{\mu}^\top\hat{Q}^2\hat{\mu} - \frac{(T-2)(T-N-1)}{(N-1)T^2}\mathrm{tr}(\hat{Q})\hat{\mu}^\top\hat{Q}\hat{\mu}. \tag{6.14}$$

7 Empirical evidence

We now study the relation between the MSE and decision loss functions for the SMV portfolio, the fully invested SMV portfolio, and their shrinkage estimators.

7.1 Data

We consider four datasets of monthly stock returns grouped according to firm characteristics and industries, which we download from Kenneth French's website. The first dataset is composed of 25 portfolios of stocks sorted on size and book-to-market (25SBM). The second dataset is composed of 25 portfolios of stocks sorted on book-to-market and operating profitability (25BMOP). The third dataset is composed of 25 portfolios of stocks sorted on operating profitability and investment (25OPINV). The fourth and last dataset is composed of 30 portfolios of stocks grouped into industries (30IND). The time period for each dataset spans July 1963 to December 2023. These datasets use a rather small N of 25 or 30, but we can easily vary the magnitude of estimation risk by changing the sample size T.

7.2 Empirical MSE and decision loss functions for SMV portfolios

Our first objective is to relate empirical counterparts to the MSE and decision loss functions for the SMV portfolio $\hat{w}_m$ and the fully invested SMV portfolio $\hat{w}_{mf}$. We construct the empirical counterparts as follows. Let τ be the total number of months in the dataset, i.e., $\tau = 726$. We take as the population mean and covariance matrix of asset returns, μ and Σ, the sample mean and covariance matrix computed on the whole sample, i.e.,

$$\mu = \frac{1}{\tau}\sum_{m=1}^{\tau} r_m, \tag{7.1}$$

$$\Sigma = \frac{1}{\tau}\sum_{m=1}^{\tau} (r_m - \mu)(r_m - \mu)^\top. \tag{7.2}$$

Then, we fix a sample size T and construct $\tau - T + 1$ portfolio weight vectors using rolling windows. Specifically, let $\hat{\mu}_t$ and $\hat{\Sigma}_t$ be the sample mean and covariance matrix computed from months $t-T+1$ to t,

$$\hat{\mu}_t = \frac{1}{T}\sum_{m=t-T+1}^{t} r_m, \tag{7.3}$$

$$\hat{\Sigma}_t = \frac{1}{T}\sum_{m=t-T+1}^{t} (r_m - \hat{\mu})(r_m - \hat{\mu})^\top, \tag{7.4}$$

such that the $\tau - T + 1$ SMV and fully invested SMV portfolios are

$$\hat{w}_{m,t} = \frac{T-N-2}{\gamma T}\hat{\Sigma}_t^{-1}\hat{\mu}_t, \quad t = T, \ldots, \tau, \tag{7.5}$$

$$\hat{w}_{mf,t} = \frac{\hat{\Sigma}_t^{-1}\iota}{\iota^\top\hat{\Sigma}_t^{-1}\iota} + \frac{T-N-1}{\gamma T}\hat{Q}_t\hat{\mu}_t, \quad t = T, \ldots, \tau, \tag{7.6}$$

where $\hat{Q}_t = \hat{\Sigma}_t^{-1} - \hat{\Sigma}_t^{-1}\iota\iota^\top\hat{\Sigma}_t^{-1}/(\iota^\top\hat{\Sigma}_t^{-1}\iota)$. Finally, the empirical decision loss function for $\hat{w}_m$ and $\hat{w}_{mf}$ is

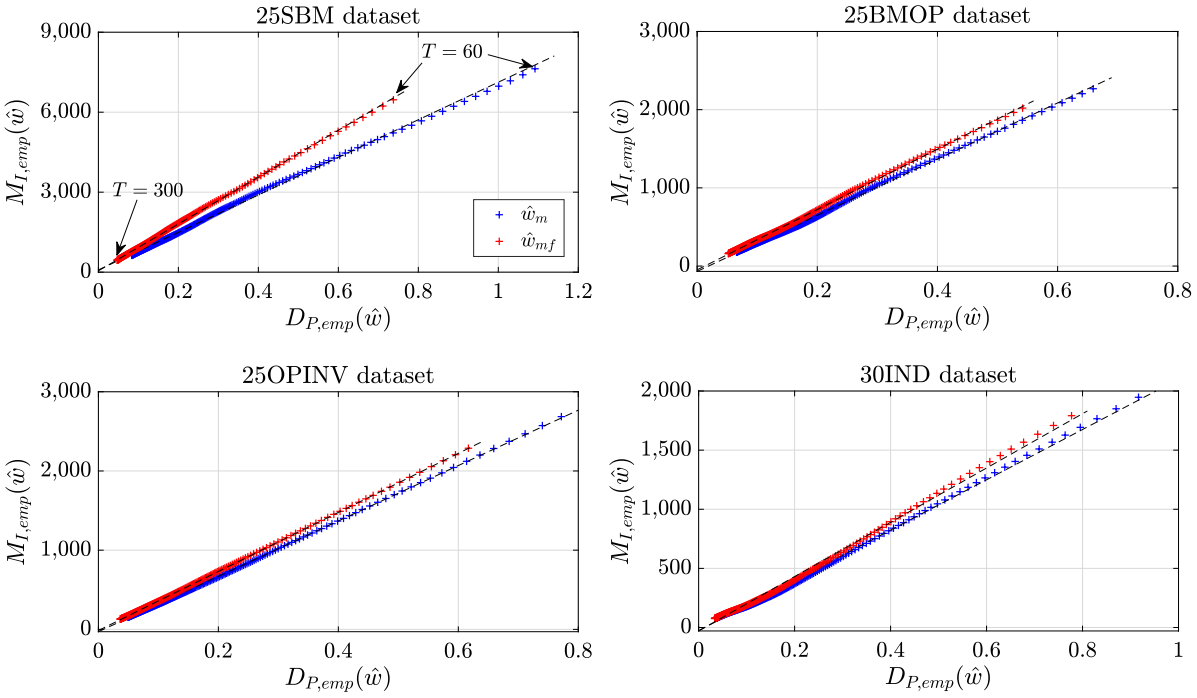
$$D_{P,emp}(\hat{w}_m) = P(w_m) - \frac{1}{\tau - T + 1}\sum_{t=T}^{\tau} P(\hat{w}_{m,t}), \tag{7.7}$$

$$D_{P,emp}(\hat{w}_{mf}) = P(w_{mf}) - \frac{1}{\tau - T + 1}\sum_{t=T}^{\tau} P(\hat{w}_{mf,t}), \tag{7.8}$$

where $P(w_m)$ and $P(w_{mf})$ are given by (3.3) and (3.7), and the MSE loss function is

$$M_{I,emp}(\hat{w}_m) = \frac{1}{\tau - T + 1}\sum_{t=T}^{\tau} L_I(\hat{w}_{m,t} - w_m), \tag{7.9}$$

Figure 4: Empirical relation between the MSE and decision loss functions



Notes. This figure depicts the relation between the empirical mean squared error (MSE) and decision loss functions for the sample mean-variance portfolio, $\hat{w}_m$ in (3.10), and the fully invested sample mean-variance portfolio, $\hat{w}_{mf}$ in (3.11). The empirical decision loss function is the average out-of-sample utility loss and is given by $D_{P,emp}(\hat{w}_m)$ and $D_{P,emp}(\hat{w}_{mf})$ in (7.7)–(7.8). The empirical MSE, $M_{I,emp}(\hat{w}_m)$ and $M_{I,emp}(\hat{w}_{mf})$, is given by (7.9)–(7.10). In each plot, the blue crosses depict the relation for $\hat{w}_m$, and the red crosses for $\hat{w}_{mf}$, with the sample size T varying between 60 and 300 months. We consider a risk-aversion coefficient $\gamma = 1$. The four datasets span July 1963 to December 2023 and are described in Section 7.1.

$$M_{I,emp}(\hat{w}_{mf}) = \frac{1}{\tau - T + 1} \sum_{t=T}^{\tau} L_I(\hat{w}_{mf,t} - w_{mf}). \tag{7.10}$$

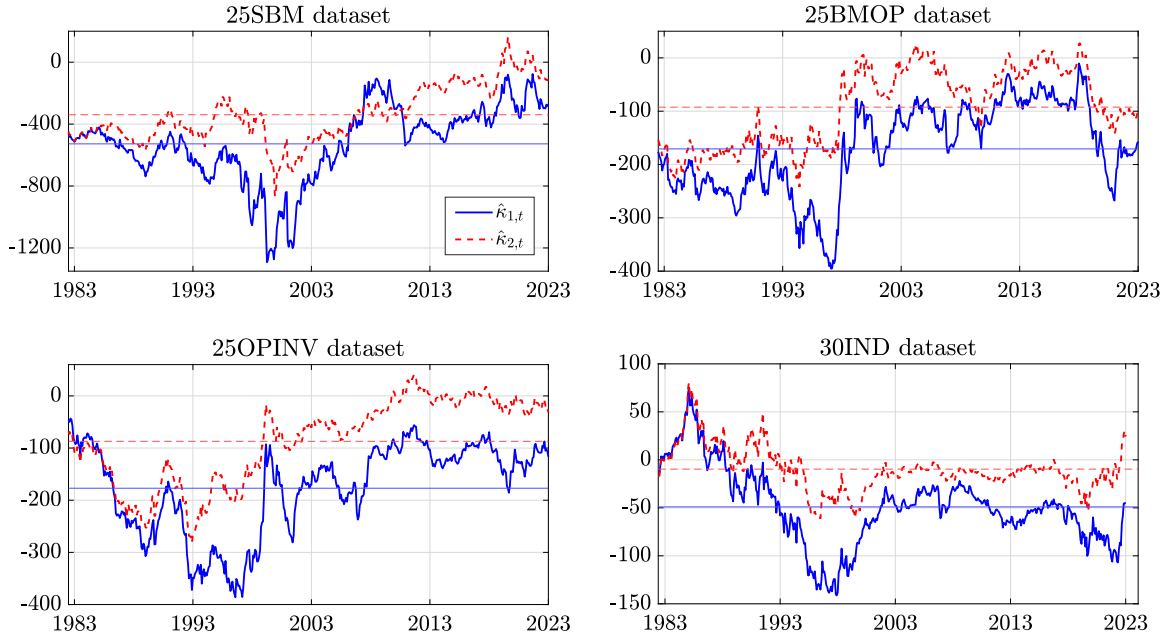
In Figure 4, we depict the relation between the empirical MSE and decision loss functions for each of the four datasets, a risk-aversion coefficient $\gamma = 1$, and a sample size that varies from $T = 60$ to 300. In line with the theoretical predictions from Proposition 1, we observe that the relation between the two loss functions is positive and essentially linear.

7.3 Comparison of empirical shrinkage portfolios

Our second objective in this empirical analysis is to compare the empirical decision loss of the shrinkage portfolios $\hat{w}_m(\delta)$ in (5.1) and $\hat{w}_{mf}(\delta)$ in (5.1) for which the shrinkage intensity δ minimizes either the decision loss or the MSE.

In Proposition 3, we show that the signs of $\delta_{m,D} - \delta_{m,M}$ and $\delta_{mf,D} - \delta_{mf,M}$ are equal to those of κ_1 and κ_2 in (4.7)–(4.8), respectively. To evaluate these signs, we use the unbiased estimators of

Figure 5: Empirical unbiased estimates of κ_1 and κ_2 over time

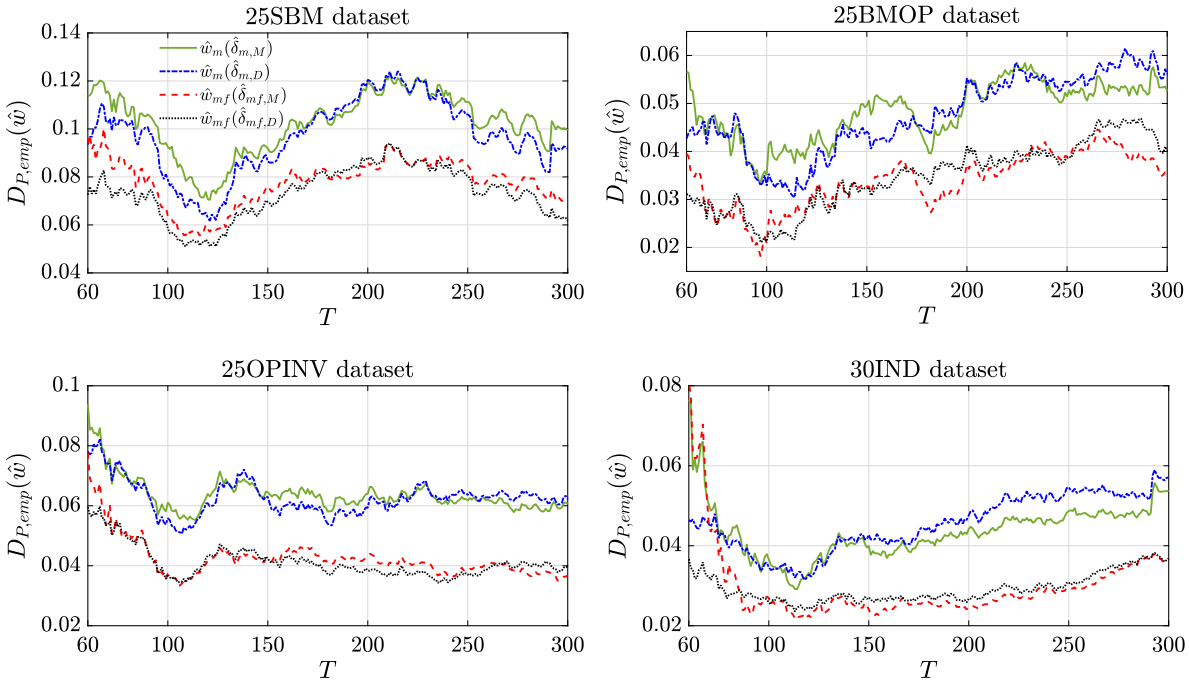


Notes. This figure depicts the unbiased empirical estimates of κ_1 and κ_2 in (6.13)–(6.14) using rolling windows of size $T = 240$ months. The horizontal lines depict the average value across all rolling windows. We consider a risk-aversion coefficient $\gamma = 1$. The four datasets span July 1963 to December 2023 and are described in Section 7.1. As shown in Part 3 of Proposition 1, a negative sign of κ_1 (κ_2) means that the mean squared error loss function delivers a larger shrinkage intensity δ in the shrinkage portfolio $\hat{w}_m(\delta)$ in (5.1) ($\hat{w}_{mf}(\delta)$ in (5.2)) than that under the decision loss function.

κ_1 and κ_2 in Corollary 1. Specifically, using rolling windows of size T as in Section 7.2, we obtain $\tau - T + 1$ estimates $\hat{\kappa}_{1,t}$ and $\hat{\kappa}_{2,t}$ by plugging $(\hat{\mu}_t, \hat{\Sigma}_t)$ from (7.3)–(7.4) into (6.13)–(6.14). In Figure 5, we depict $\hat{\kappa}_{1,t}$ and $\hat{\kappa}_{2,t}$ across the four datasets for a sample size of $T = 240$ months. We find that both $\hat{\kappa}_{1,t}$ and $\hat{\kappa}_{2,t}$ are negative in most rolling windows, which implies that the MSE loss function generally delivers a more intensive degree of shrinkage than the decision loss function.

Then, we evaluate the empirical decision loss of the estimated shrinkage portfolios $\hat{w}_m(\hat{\delta}_m)$ and $\hat{w}_{mf}(\hat{\delta}_{mf})$, where $\hat{\delta}_m$ and $\hat{\delta}_{mf}$ are designed to minimize either the MSE ($\hat{\delta}_{m,M}$ and $\hat{\delta}_{mf,M}$) or the decision loss ($\hat{\delta}_{m,D}$ and $\hat{\delta}_{mf,D}$). We follow the standard approach in the literature (e.g., DeMiguel et al. (2009b), Kan et al. (2022)) by constructing a series of out-of-sample returns using rolling windows of past data. Specifically, let τ be the total number of months in the dataset, i.e., $\tau = 726$, and T be a chosen sample size. Then, we estimate the portfolio and shrinkage intensity at time t, $\hat{w}_{m,t}(\hat{\delta}_{m,t})$ and $\hat{w}_{mf,t}(\hat{\delta}_{mf,t})$, using returns from month $t - T + 1$ to t. The estimated shrinkage intensities $\hat{\delta}_{m,t}$ and $\hat{\delta}_{mf,t}$ are obtained by plugging $(\hat{\mu}_t, \hat{\Sigma}_t)$ in (7.3)–(7.4) into (6.9)–(6.12). Then, the out-of-sample portfolio return at time $t + 1$ is $r_{m,t+1} = \hat{w}_{m,t}(\hat{\delta}_{m,t})^\top r_{t+1}$ and

Figure 6: Empirical decision loss of shrinkage portfolios



Notes. This figure depicts the empirical decision loss of estimated shrinkage portfolios that are designed to minimize either the mean squared error (MSE), $\hat{w}_m(\hat{\delta}_{m,M})$ and $\hat{w}_{mf}(\hat{\delta}_{mf,M})$, or the decision loss function, $\hat{w}_m(\hat{\delta}_{m,D})$ and $\hat{w}_{mf}(\hat{\delta}_{mf,D})$. The shrinkage intensities are estimated using (6.9)–(6.12). The empirical decision loss function is the average out-of-sample utility loss and is given by $D_{P,emp}(\hat{w}_m(\hat{\delta}_m))$ and $D_{P,emp}(\hat{w}_{mf}(\hat{\delta}_{mf}))$ in (7.11)–(7.12). In each plot, the sample size T varies between 60 and 300 months. We consider a risk-aversion coefficient $\gamma = 1$. The four datasets span July 1963 to December 2023 and are described in Section 7.1.

$r_{mf,t+1} = \hat{w}_{mf,t}(\hat{\delta}_{mf,t})^\top r_{t+1}$. Finally, the empirical decision loss function is given by

$$D_{P,emp}(\hat{w}_m(\hat{\delta}_m)) = P(w_m) - \left(\bar{r}_m - \frac{\gamma}{2}\sigma_m^2\right), \tag{7.11}$$

$$D_{P,emp}(\hat{w}_{mf}(\hat{\delta}_{mf})) = P(w_{mf}) - \left(\bar{r}_{mf} - \frac{\gamma}{2}\sigma_{mf}^2\right), \tag{7.12}$$

where $\bar{r}_m = \frac{1}{\tau - T}\sum_{t=T+1}^{\tau} r_{m,t}$, $\sigma_m^2 = \frac{1}{\tau - T}\sum_{t=T+1}^{\tau}(r_{m,t} - \bar{r}_m)^2$, and similarly for $\bar{r}_{mf}$ and σ_{mf}^2, and $P(w_m)$ and $P(w_{mf})$ are obtained by letting the population mean μ and covariance matrix Σ be computed on the out-of-sample period from month $T+1$ to τ. This way of evaluating the empirical decision loss function is not the same as that for the SMV portfolios in Section 7.2. In particular, μ and Σ in Section 7.2 are fixed and obtained from the whole sample of returns to obtain empirical MSE and decision loss functions that mirror their theoretical counterparts for which μ and Σ are fixed. In contrast, in this section, μ and Σ vary with the out-of-sample period considered and thus also with the sample size T, so as to mirror what investors can implement in real time and thus can gain with the shrinkage portfolios.

In Figure 6, we depict the empirical decision loss of the shrinkage portfolios that minimize the MSE, $D_{P,emp}(\hat{w}_m(\hat{\delta}_{m,M}))$ and $D_{P,emp}(\hat{w}_m(\hat{\delta}_{mf,M}))$, and those that minimize the decision loss, $D_{P,emp}(\hat{w}_m(\hat{\delta}_{m,D}))$ and $D_{P,emp}(\hat{w}_m(\hat{\delta}_{mf,D}))$, for each of the four datasets, a risk-aversion coefficient $\gamma = 1$, and a sample size that varies from $T = 60$ to 300. We first observe that the shrinkage portfolios deliver decision losses significantly smaller than those delivered when we do not apply shrinkage in Figure 4. Then, concerning the comparison of different shrinkage portfolios, the results are overall consistent with those in Figure 3 in which the shrinkage intensities are known without error and returns follow an i.i.d. multivariate normal distribution. In particular, the shrinkage intensities that minimize the MSE, $\hat{\delta}_{m,M}$ and $\hat{\delta}_{mf,M}$, deliver a similar decision loss than that delivered by the shrinkage intensities that minimize the decision loss, $\hat{\delta}_{m,D}$ and $\hat{\delta}_{mf,D}$.[9] This gives empirical evidence that the MSE contains valuable economic content for the purpose of designing shrinkage portfolio estimators.

8 Conclusion

Which loss function should we use when designing estimators? Ideally, one would identify a decision loss function corresponding to the economic decision problem the estimator is meant to solve. However, this identification may not be an easy task, the resulting loss function may be challenging to handle numerically and analytically, and the estimator may not be applicable to other problems. This renders generic statistical loss functions, like the MSE, an appealing alternative because they can be applied widely and are easy to handle. And indeed, the MSE is at the foundation of regression and machine learning methods. The pending question is whether the MSE carries economic value, i.e., delivers estimators that perform well for the decision problem of interest.

We analyze this question for the portfolio selection problem in a setting where both the MSE and a decision loss function—the expected out-of-sample utility loss—can be derived analytically and thus be compared to one another. We begin by comparing the MSE and decision loss functions for fixed mean-variance portfolio rules and find a positive linear relation between the two loss functions. We then compare the MSE and decision loss functions for designing optimal shrinkage portfolio estimators. We find that the shrinkage intensities implied by the two loss functions are close to one another and verify empirically that the shrinkage portfolios under both loss functions offer a similar decision loss.

[9] The decision losses in Figure 6 do not uniformly decrease with the sample size T because returns are not i.i.d. and because the out-of-sample period is different as we vary T.

Remarkably, the comparison between the two loss functions is driven by a single parameter: either κ_1 in (4.7) for the case with a risk-free asset or κ_2 in (4.8) for the fully invested case. If these parameters equal zero, then the MSE and decision loss functions are proportional to each other and deliver the same optimal shrinkage intensities. We derive a sufficient condition under which κ_1 and κ_2 are negative, which is that the Sharpe ratios of principal components form a decreasing sequence. We show empirically that the unbiased estimators of κ_1 and κ_2 that we derive are generally negative over time. This negative sign indicates that the MSE loss function delivers a more intensive degree of shrinkage, i.e., more conservative shrinkage portfolios.

Overall, our results suggest that the MSE carries economic value for portfolio selection. More generally, we aim to motivate researchers to pay more attention to the implications of their choice of loss functions in their estimation tasks and to show how these can be studied systematically.

We envision several avenues for future research. One can investigate other types of portfolio strategies, such as machine learning algorithms using decision loss functions as in Liu et al. (2024). It is also of interest to compare loss functions applied to the inputs of portfolio strategies instead of the portfolio weights themselves. For example, when shrinking the sample covariance matrix toward the identity matrix, what is the difference between calibrating the shrinkage intensity based on the MSE as in Ledoit and Wolf (2004) or based on the expected out-of-sample performance loss of the portfolio that will use this shrinkage covariance matrix estimator? Finally, we find a linear relation between the MSE and decision loss functions because they are both quadratic functions of portfolio weights. However, DeMiguel and Nogales (2009) recommend using robust alternatives to the MSE loss function in portfolio selection that do not grow as fast as the square function, such as the Huber's loss function. Thus, future research can investigate the relation between such robust loss functions and decision loss functions, which will imply a nonlinear relation. Exploring this link with robust statistical loss functions would also help inspire more robust decision loss functions than the standard expected out-of-sample utility loss.

A Proofs of results

In this appendix, we provide the proofs of Propositions 1 to (4) as well as Corollary 1. Throughout the proofs, we use the covariance matrix of the unbiased SMV portfolio $\hat{w}_m$ in (3.10) and fully invested SMV portfolio $\hat{w}_{mf}$ in (3.10). From Javed et al. (2021, Corollary 2.2), the covariance

matrix of $\hat{w}_m$ under Assumption 1 exists if $T > N + 4$ and is given by

$$\mathbb{V}[\hat{w}_m] = \frac{c_1}{\gamma^2}\left(\frac{2\gamma P(w_m)}{T-2} + \frac{1}{T}\right)\Sigma^{-1} + \left(\frac{T-N-2}{T-2}c_1 - 1\right)w_m w_m^\top. \tag{A.1}$$

From Okhrin and Schmid (2006, Theorem 1), the covariance matrix of $\hat{w}_{mf}$ under Assumption 1 exists if $T > N + 3$ and is given by

$$\mathbb{V}[\hat{w}_{mf}] = \mathbb{V}[\hat{w}_g] + \mathbb{V}[\hat{w}_a], \tag{A.2}$$

where

$$\mathbb{V}[\hat{w}_g] = \frac{\sigma_g^2}{T-N-1}Q, \tag{A.3}$$

$$\mathbb{V}[\hat{w}_a] = \frac{c_2}{\gamma^2}\left(\frac{2\gamma P(w_a)}{T-2} + \frac{1}{T}\right)Q + \left(\frac{T-N-1}{T-2}c_2 - 1\right)w_a w_a^\top. \tag{A.4}$$

A.1 Proof of Proposition 1

Part 1. The decision loss function for $\hat{w}_m$ is

$$D_P(\hat{w}_m) = P(w_m) - \mathbb{E}[P(\hat{w}_m)] = P(w_m) - \mathbb{E}[\hat{w}_m^\top \mu] + \frac{\gamma}{2}\mathbb{E}[\hat{w}_m^\top \Sigma \hat{w}_m], \tag{A.5}$$

where

$$\mathbb{E}[\hat{w}_m^\top \mu] = w_m^\top \mu = 2P(w_m), \tag{A.6}$$

$$\mathbb{E}[\hat{w}_m^\top \Sigma \hat{w}_m] = w_m^\top \Sigma w_m + \mathrm{tr}(\mathbb{V}[\hat{w}_m]\Sigma) = c_1\left(\frac{2}{\gamma}P(w_m) + \frac{1}{\gamma^2}\frac{N}{T}\right), \tag{A.7}$$

which yields Equation (4.1). Then, the decision loss function for $\hat{w}_{mf}$ is

$$D_P(\hat{w}_m) = P(w_{mf}) - \mathbb{E}[P(\hat{w}_{mf})] = \mu_g - \frac{\gamma}{2}\sigma_g^2 + P(w_a) - \mathbb{E}[\hat{w}_{mf}^\top \mu] + \frac{\gamma}{2}\mathbb{E}[\hat{w}_{mf}^\top \Sigma \hat{w}_{mf}], \tag{A.8}$$

where

$$\mathbb{E}[\hat{w}_{mf}^\top \mu] = w_{mf}^\top \mu = \mu_g + 2P(w_a), \tag{A.9}$$

$$\mathbb{E}[\hat{w}_{mf}^\top \Sigma \hat{w}_{mf}] = w_{mf}^\top \Sigma w_{mf} + \mathrm{tr}(\mathbb{V}[\hat{w}_{mf}]\Sigma) = c_0\sigma_g^2 + c_2\left(\frac{2}{\gamma}P(w_a) + \frac{1}{\gamma^2}\frac{N-1}{T}\right), \tag{A.10}$$

where we use $\mathrm{tr}(Q\Sigma) = N - 1$, which yields Equation (4.2).

Part 2. The MSE of $\hat{w}_m$ with a scaling matrix $A = \Sigma$ is

$$M_\Sigma(\hat{w}_m) = \mathrm{tr}(\mathbb{V}[\hat{w}_m]\Sigma) = (c_1 - 1)\frac{2}{\gamma}P(w_m) + \frac{c_1}{\gamma^2}\frac{N}{T}, \tag{A.11}$$

which is equal to $\frac{2}{\gamma}D_P(\hat{w}_m)$. Similarly, the MSE of $\hat{w}_{mf}$ with $A = \Sigma$ is

$$M_\Sigma(\hat{w}_{mf}) = \mathrm{tr}(\mathbb{V}[\hat{w}_{mf}]\Sigma) = (c_0 - 1)\sigma_g^2 + (c_2 - 1)\frac{2}{\gamma}P(w_a) + \frac{c_2}{\gamma^2}\frac{N-1}{T}, \tag{A.12}$$

which is equal to $\frac{2}{\gamma}D_P(\hat{w}_{mf})$.

Part 3. The MSE of $\hat{w}_m$ with a scaling matrix $A = I$ is

$$M_I(\hat{w}_m) = \mathrm{tr}(\mathbb{V}[\hat{w}_m]) = \frac{\bar{\tau}c_1 N}{\gamma^2}\Big(\frac{2\gamma P(w_m)}{T-2} + \frac{1}{T}\Big) + \Big(\frac{T-N-2}{T-2}c_1 - 1\Big)\frac{\mu^\top \Sigma^{-2}\mu}{\gamma^2}, \tag{A.13}$$

which is equal to Equation (4.5) with $D_P(\hat{w}_m)$ given by (4.1). Similarly, the MSE of $\hat{w}_{mf}$ with $A = I$ is

$$\begin{aligned} M_I(\hat{w}_{mf}) &= \mathrm{tr}(\mathbb{V}[\hat{w}_{mf}]) \\ &= \bar{q}(c_0 - 1)\sigma_g^2 + \frac{\bar{q}c_2(N-1)}{\gamma^2}\Big(\frac{2\gamma P(w_a)}{T-2} + \frac{1}{T}\Big) + \Big(\frac{T-N-1}{T-2}c_2 - 1\Big)\frac{\mu^\top Q^2\mu}{\gamma^2}, \end{aligned} \tag{A.14}$$

which is equal to Equation (4.6) with $D_P(\hat{w}_{mf})$ given by (4.2). This completes the proof.

A.2 Proof of Proposition 2

Consider the eigenvalue decomposition of the inverse covariance matrix, $\Sigma^{-1} = V\Lambda^{-1}V^\top$, where $V = [v_1,\ v_2,\ \ldots,\ v_N]$ is the matrix of eigenvectors and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$ is the diagonal matrix of eigenvalues, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ and $\lambda_N > 0$ because Σ is assumed positive definite. Then, we can show that $\kappa_1/\bar{\tau}$, which has the same sign as that of κ_1, is given by

$$\frac{\kappa_1}{\bar{\tau}} = N\frac{\sum_{i=1}^N s_i^2/\lambda_i}{\sum_{i=1}^N 1/\lambda_i} - \sum_{i=1}^N s_i^2. \tag{A.15}$$

Similarly, under the condition $v_1 = \iota/\sqrt{N}$, we can show that

$$Q = \Sigma^{-1}\Big(I - v_1 v_1^\top\Big) = V\mathrm{diag}(0, 1/\lambda_2, 1/\lambda_3, \ldots, 1/\lambda_N)V^\top, \tag{A.16}$$

and thus, $\kappa_2/\bar{q}$, which has the same sign as that of κ_2, is given by

$$\frac{\kappa_2}{\bar{q}} = (N-1)\frac{\sum_{i=2}^N s_i^2/\lambda_i}{\sum_{i=2}^N 1/\lambda_i} - \sum_{i=2}^N s_i^2. \tag{A.17}$$

We can see that $\kappa_2/\bar{q}$ has a similar expression to that for $\kappa_1/\bar{\tau}$. In particular, the sufficient condition we derive under which $\kappa_1 \leq 0$ is also sufficient for $\kappa_2 \leq 0$, provided that $v_1 = \iota/\sqrt{N}$.

Therefore, we now prove that $\kappa_1/\bar{\tau} \leq 0$ if $s_1^2 \geq s_2^2 \geq \cdots \geq s_N^2$. We have

$$\begin{aligned}
\frac{\kappa_1}{\bar{\tau}} &= N\frac{\sum_{i=1}^{N-1}(s_i^2 - s_N^2)/\lambda_i}{\sum_{i=1}^{N} 1/\lambda_i} - \left(\sum_{i=1}^{N-1} s_i^2 - (N-1)s_N^2\right) \\
&= \left(\sum_{i=1}^{N-1} s_i^2 - (N-1)s_N^2\right)\left[-1 + \frac{N}{\sum_{i=1}^{N} 1/\lambda_i}\sum_{i=1}^{N-1}\frac{1}{\lambda_i}\left(\frac{s_i^2 - s_N^2}{\sum_{j=1}^{N-1} s_j^2 - (N-1)s_N^2}\right)\right] \\
&= \left(\sum_{j=1}^{N-1} s_j^2 - (N-1)s_N^2\right)\left[-1 + \frac{N/(N-1)}{\sum_{i=1}^{N} 1/\lambda_i}\times\right. \\
&\quad \left.\left(\sum_{i=1}^{N-1}\frac{1}{\lambda_i} + \frac{1}{\sum_{i=1}^{N-1} s_i^2 - (N-1)s_N^2}\sum_{i=1}^{N-1}\frac{1}{\lambda_i}\left\{(N-1)s_i^2 - \sum_{j=1}^{N-1} s_j^2\right\}\right)\right]. \qquad \text{(A.18)}
\end{aligned}$$

We clearly have $\sum_{j=1}^{N-1} s_j^2 - (N-1)s_N^2 \geq 0$ if $s_1^2 \geq s_2^2 \geq \cdots \geq s_N^2$. Moreover,

$$-1 + \frac{N}{N-1}\frac{\sum_{i=1}^{N-1} 1/\lambda_i}{\sum_{i=1}^{N} 1/\lambda_i} = \frac{\sum_{i=1}^{N-1} 1/\lambda_i - (N-1)/\lambda_N}{(N-1)\sum_{i=1}^{N} 1/\lambda_i} \leq 0, \qquad \text{(A.19)}$$

where the inequality holds because $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. Therefore, $\kappa_1/\bar{\tau}$ in (A.18) can be bounded above as

$$\begin{aligned}
\frac{\kappa_1}{\bar{\tau}} &\leq \frac{N/(N-1)}{\sum_{i=1}^{N} 1/\lambda_i}\sum_{i=1}^{N-1}\frac{1}{\lambda_i}\left((N-1)s_i^2 - \sum_{j=1}^{N-1} s_j^2\right) \\
&= \frac{N/(N-1)}{\sum_{i=1}^{N} 1/\lambda_i}\left(\sum_{i=1}^{N-2} s_i^2 - (N-2)s_{N-1}^2\right)\left[\sum_{i=1}^{N-1}\frac{1}{\lambda_i}\left(\frac{(N-1)s_i^2 - \sum_{j=1}^{N-1} s_j^2}{\sum_{j=1}^{N-2} s_j^2 - (N-2)s_{N-1}^2}\right)\right]. \qquad \text{(A.20)}
\end{aligned}$$

Now, we can show that

$$\frac{(N-1)s_i^2 - \sum_{j=1}^{N-1} s_j^2}{\sum_{j=1}^{N-2} s_j^2 - (N-2)s_{N-1}^2} = \frac{1}{N-2} + \frac{N-1}{N-2}\frac{(N-2)s_i^2 - \sum_{j=1}^{N-2} s_j^2}{\sum_{j=1}^{N-2} s_j^2 - (N-2)s_{N-1}^2}, \qquad \text{(A.21)}$$

which is equal to -1 when $i = N-1$. Therefore, inequality (A.20) becomes

$$\begin{aligned}
\frac{\kappa_1}{\bar{\tau}} &\leq \frac{N/(N-1)}{\sum_{i=1}^{N} 1/\lambda_i}\left(\sum_{i=1}^{N-2} s_i^2 - (N-2)s_{N-1}^2\right)\times \\
&\quad \left[\frac{1}{N-2}\sum_{i=1}^{N-2}\frac{1}{\lambda_i} - \frac{1}{\lambda_{N-1}} + \frac{N-1}{N-2}\sum_{i=1}^{N-2}\frac{1}{\lambda_i}\left\{\frac{(N-2)s_i^2 - \sum_{j=1}^{N-2} s_j^2}{\sum_{j=1}^{N-2} s_j^2 - (N-2)s_{N-1}^2}\right\}\right]. \qquad \text{(A.22)}
\end{aligned}$$

We clearly have $\sum_{j=1}^{N-2} s_j^2 - (N-2)s_{N-1}^2 \geq 0$ if $s_1^2 \geq s_2^2 \geq \cdots \geq s_{N-1}^2$, and moreover, $\frac{1}{N-2}\sum_{i=1}^{N-2}\frac{1}{\lambda_i} - \frac{1}{\lambda_{N-1}} \leq 0$. Thus, we obtain the inequality

$$\frac{\kappa_1}{\bar{\tau}} \leq \frac{N/(N-2)}{\sum_{i=1}^{N} 1/\lambda_i}\left[\sum_{i=1}^{N-2}\frac{1}{\lambda_i}\left((N-2)s_i^2 - \sum_{j=1}^{N-2} s_j^2\right)\right]. \qquad \text{(A.23)}$$

We have that $(N-2)s_i^2 - \sum_{j=1}^{N-2} s_j^2$, $i = 1, \ldots, N-2$, is decreasing with the index i if $s_1^2 \geq s_2^2 \geq \cdots \geq s_{N-2}^2$. Therefore, the sum between brackets in (A.23) is lower than the case when all eigenvalues are equal to the largest one, λ_1. That is, we have

$$\frac{\kappa_1}{\bar{\tau}} \leq \frac{N/(N-2)}{\sum_{i=1}^{N} \lambda_1/\lambda_i} \sum_{i=1}^{N-2} \left((N-2)s_i^2 - \sum_{j=1}^{N-2} s_j^2 \right) \leq 0, \tag{A.24}$$

where the last inequality holds because $\sum_{i=1}^{N-2} \sum_{j=1}^{N-2} s_j^2 = (N-2) \sum_{i=1}^{N-2} s_i^2$. This completes the proof.

A.3 Proof of Proposition 3

Part 1. Given the expression for $\mathbb{E}[\hat{w}_m^\top \mu]$ in (A.6) and $\mathbb{E}[\hat{w}_m^\top \Sigma \hat{w}_m]$ in (A.7), we have

$$\begin{aligned}
D_P(\hat{w}_m(\delta)) &= P(w_m) - \mathbb{E}[\hat{w}_m(\delta)] \\
&= P(w_m) - (1-\delta)\mathbb{E}[\hat{w}_m^\top \mu] + \frac{\gamma}{2}(1-\delta)^2 \mathbb{E}[\hat{w}_m^\top \Sigma \hat{w}_m] \\
&= P(w_m) - 2(1-\delta)P(w_m) + (1-\delta)^2 c_1 \left(P(w_m) + \frac{1}{2\gamma}\frac{N}{T} \right) \\
&= \delta^2 P(w_m) + (1-\delta)^2 D_P(\hat{w}_m),
\end{aligned} \tag{A.25}$$

where $D_P(\hat{w}_m)$ is given by (4.1). Given (A.25), it is easy to show that $\delta_{m,D}$ minimizing $D_P(\hat{w}_m(\delta))$ is given by (5.3), and that plugging $\delta_{m,D}$ into (A.25) gives (5.5). Turning to $\hat{w}_{mf}$, we have similarly that given the expression for $\mathbb{E}[\hat{w}_{mf}^\top \mu]$ in (A.9) and $\mathbb{E}[\hat{w}_{mf}^\top \Sigma \hat{w}_{mf}]$ in (A.10),

$$\begin{aligned}
D_P(\hat{w}_{mf}(\delta)) &= P(w_{mf}) - \mathbb{E}[\hat{w}_{mf}(\delta)] \\
&= \mu_g - \frac{\gamma}{2}\sigma_g^2 + P(w_a) - \mathbb{E}\left[\hat{w}_g^\top \mu + (1-\delta)\hat{w}_a^\top \mu \right] + \frac{\gamma}{2}\mathbb{E}\left[\hat{w}_g^\top \Sigma \hat{w}_g + (1-\delta)^2 \hat{w}_a^\top \Sigma \hat{w}_a \right] \\
&= \mu_g - \frac{\gamma}{2}\sigma_g^2 + P(w_a) - [\mu_g + 2(1-\delta)P(w_a)] \\
&\quad + \left[\frac{\gamma}{2} c_0 \sigma_g^2 + (1-\delta)^2 c_2 \left(P(w_a) + \frac{1}{2\gamma}\frac{N-1}{T} \right) \right] \\
&= \frac{\gamma}{2}(c_0 - 1)\sigma_g^2 + \delta^2 P(w_a) + (1-\delta)^2 D_P(\hat{w}_a),
\end{aligned} \tag{A.26}$$

where $D_P(\hat{w}_a)$ is given by (4.2). Given (A.26), it is easy to show that $\delta_{mf,D}$ minimizing $D_P(\hat{w}_{mf}(\delta))$ is given by (5.4), and that plugging $\delta_{mf,D}$ into (A.26) gives (5.6). Finally, $\delta_{m,D}$ and $\delta_{mf,D}$ are both between zero and one and independent of γ because $P(w_m)$, $P(w_a)$, $D_P(\hat{w}_m)$, and $D_P(\hat{w}_a)$ are all positive and proportional to $1/(2\gamma)$.

Part 2. Given that $\hat{w}_m$ is unbiased, we have

$$M_I(\hat{w}_m(\delta)) = \delta^2 L_I(w_m) + (1-\delta)^2 M_I(\hat{w}_m), \tag{A.27}$$

where $M_I(\hat{w}_m)$ is given by (4.5). Given (A.27), it is easy to show that $\delta_{m,M}$ minimizing $M_I(\hat{w}_m(\delta))$ is given by (5.7), and that plugging $\delta_{m,M}$ into (A.27) gives (5.9). Turning to $\hat{w}_{mf}$, we have similarly that because $\hat{w}_{mf}$ is unbiased,

$$M_I(\hat{w}_{mf}(\delta)) = \delta^2 L_I(w_a) + M_I(\hat{w}_g) + (1-\delta)^2 M_I(\hat{w}_a), \tag{A.28}$$

where $M_I(\hat{w}_g)$ and $M_I(\hat{w}_a)$ are given by (4.6). Given (A.28), it is easy to show that $\delta_{mf,M}$ minimizing $M_I(\hat{w}_{mf}(\delta))$ is given by (5.8), and that plugging $\delta_{mf,M}$ into (A.28) gives (5.10). Finally, $\delta_{m,M}$ and $\delta_{mf,M}$ are both between zero and one and independent of γ because $L_I(w_m)$, $L_I(w_a)$, $M_I(\hat{w}_m)$, and $M_I(\hat{w}_a)$ are all positive and proportional to $1/\gamma^2$.

Part 3. Given the relation between the MSE and decision loss function for $\hat{w}_m$ in (4.5), we have

$$\begin{aligned}
\delta_{m,D} - \delta_{m,M} &= \frac{D_P(\hat{w}_m)}{D_P(\hat{w}_m) + P(w_m)} - \frac{\frac{2\bar{\tau}}{\gamma} D_P(\hat{w}_m) + \left(\frac{T-N-2}{T-2} c_1 - 1\right)\frac{\kappa_1}{\gamma^2}}{\frac{2\bar{\tau}}{\gamma} D_P(\hat{w}_m) + \left(\frac{T-N-2}{T-2} c_1 - 1\right)\frac{\kappa_1}{\gamma^2} + L_I(w_m)} \\
&\propto D_P(\hat{w}_m)\left[\frac{2\bar{\tau}}{\gamma} D_P(\hat{w}_m) + \left(\frac{T-N-2}{T-2} c_1 - 1\right)\frac{\kappa_1}{\gamma^2} + L_I(w_m)\right] \\
&\quad - (D_P(\hat{w}_m) + P(w_m))\left[\frac{2\bar{\tau}}{\gamma} D_P(\hat{w}_m) + \left(\frac{T-N-2}{T-2} c_1 - 1\right)\frac{\kappa_1}{\gamma^2}\right] \\
&= D_P(\hat{w}_m)\left(L_I(w_m) - \frac{2\bar{\tau}}{\gamma} P(w_m)\right) - \left(\frac{T-N-2}{T-2} c_1 - 1\right)\frac{\kappa_1}{\gamma^2} P(w_m).
\end{aligned} \tag{A.29}$$

Given that $L_I(w_m) - \frac{2\bar{\tau}}{\gamma} P(w_m) = \kappa_1/\gamma^2$, we have

$$\delta_{m,D} - \delta_{m,M} \propto \frac{\kappa_1}{\gamma^2}\left[D_P(\hat{w}_m) - \left(\frac{T-N-2}{T-2} c_1 - 1\right) P(w_m)\right] = \frac{\kappa_1 c_1 N}{\gamma^2}\left(\frac{P(w_m)}{T-2} + \frac{1}{2\gamma T}\right), \tag{A.30}$$

whose sign is equal to that of κ_1. The proof is similar for the case $\delta_{mf,D} - \delta_{mf,M}$.

Part 4. When $M_I(\hat{w}_m) \approx \frac{2\bar{\tau}}{\gamma} D_P(\hat{w}_m)$, the shrinkage intensity $\delta_{m,M}$ in (5.7) is approximately given by

$$\delta_{m,M} \approx \tilde{\delta}_{m,M} = \frac{\frac{2\bar{\tau}}{\gamma} D_P(\hat{w}_m)}{\frac{2\bar{\tau}}{\gamma} D_P(\hat{w}_m) + L_I(w_m)} = \frac{2\gamma D_P(\hat{w}_m)}{2\gamma D_P(\hat{w}_m) + (\gamma^2/\bar{\tau}) L_I(w_m)}. \tag{A.31}$$

The ratio $\tilde{\delta}_{m,M}/\delta_{m,D}$ is

$$\frac{\tilde{\delta}_{m,M}}{\delta_{m,D}} = \frac{2\gamma[D_P(\hat{w}_m) + P(w_m)]}{2\gamma D_P(\hat{w}_m) + (\gamma^2/\bar{\tau}) L_I(w_m)}$$

$$= 1 - \left(1 - \frac{2\gamma[D_P(\hat{w}_m) + P(w_m)]}{2\gamma D_P(\hat{w}_m) + (\gamma^2/\bar{\tau})L_I(w_m)}\right)$$
$$= 1 - \frac{\kappa_1/\bar{\tau}}{2\gamma D_P(\hat{w}_m) + (\gamma^2/\bar{\tau})L_I(w_m)}, \tag{A.32}$$

which corresponds to Equation (5.11). The proof is similar for Equation (5.12).

Part 5. Given the expression for $D_P(\hat{w}_m(\delta))$ in (A.25), we have

$$\begin{aligned}\Delta_P(\hat{w}_m(\delta)) &= (\delta_{m,M}^2 - \delta_{m,D}^2)P(w_m) + [(1-\delta_{m,M})^2 - (1-\delta_{m,D})^2]D_P(\hat{w}_m) \\ &= (\delta_{m,M}^2 - \delta_{m,D}^2)[D_P(\hat{w}_m) + P(w_m)] + 2(\delta_{m,D} - \delta_{m,M})D_P(\hat{w}_m). \end{aligned} \tag{A.33}$$

Now, Equation (5.13) follows because

$$\begin{aligned}&\Delta_P(\hat{w}_m(\delta)) - (\delta_{m,D} - \delta_{m,M})^2[D_P(\hat{w}_m) + P(w_m)] \\ &= 2(\delta_{m,M} - \delta_{m,D})(\delta_{m,D}[D_P(\hat{w}_m) + P(w_m)] - D_P(\hat{w}_m)) \\ &= 0, \end{aligned} \tag{A.34}$$

where the last equality holds because $\delta_{m,D} = D_P(\hat{w}_m)/[D_P(\hat{w}_m) + P(w_m)]$. The proof is similar for Equation (5.13), and this completes the proof.

A.4 Proof of Proposition 4

Here we give the proof for the unbiased estimators of $D_P(\hat{w}_m), P(w_m), M_I(\hat{w}_m)$, and $L_I(w_m)$. The proof for the unbiased estimators of $D_P(\hat{w}_a), P(w_a), M_I(\hat{w}_a)$, and $L_I(w_a)$ follows from a similar reasoning, and these estimators admit similar formulas as those for the unbiased estimators of $D_P(\hat{w}_m), P(w_m), M_I(\hat{w}_m)$, and $L_I(w_m)$ with $\hat{\Sigma}^{-1}$ replaced by $\hat{Q}$ and N by $N-1$.

The unbiased estimators of $D_P(\hat{w}_m)$ and $P(w_m)$ in (6.1) and (6.4) follow from the definition of $D_P(\hat{w}_m)$ and $P(w_m)$ in (3.3) and (4.1) and the expectation of $\hat{\mu}^\top\hat{\Sigma}^{-1}\hat{\mu}$ available in Kan and Zhou (2007), i.e.,

$$\mathbb{E}\left[\hat{\mu}^\top\hat{\Sigma}^{-1}\hat{\mu}\right] = \frac{T}{T-N-2}\left(\mu^\top\Sigma^{-1}\mu + \frac{N}{T}\right), \tag{A.35}$$

which implies that the unbiased estimator of $\mu^\top\Sigma^{-1}\mu$ is

$$\mathbb{E}\left[\frac{(T-N-2)\hat{\mu}^\top\hat{\Sigma}^{-1}\hat{\mu} - N}{T}\right] = \mu^\top\Sigma^{-1}\mu. \tag{A.36}$$

Then, to find the unbiased estimator of $M_I(\hat{w}_m)$, we write it from (A.13) as

$$M_I(\hat{w}_m) = \frac{1}{\gamma^2}\left[\frac{c_1}{T}\mathrm{tr}(\Sigma^{-1}) + \frac{c_1}{T-2}\mathrm{tr}(\Sigma^{-1})\mu^\top\Sigma^{-1}\mu + \left(\frac{T-N-2}{T-2}c_1 - 1\right)\mu^\top\Sigma^{-2}\mu\right]. \tag{A.37}$$

We now derive unbiased estimators of $\mathrm{tr}(\Sigma^{-1})$, $\mathrm{tr}(\Sigma^{-1})\mu^\top\Sigma^{-1}\mu$, and $\mu^\top\Sigma^{-2}\mu$. First, using the fact that $\mathbb{E}[\hat{\Sigma}^{-1}] = \frac{T}{T-N-2}\Sigma^{-1}$, the unbiased estimator of $\mathrm{tr}(\Sigma^{-1})$ is

$$\mathbb{E}\left[\frac{T-N-2}{T}\mathrm{tr}(\hat{\Sigma}^{-1})\right] = \mathrm{tr}(\Sigma^{-1}). \tag{A.38}$$

Second, the expectation of the plug-in estimator of $\mathrm{tr}(\Sigma^{-1})\mu^\top\Sigma^{-1}\mu$ is

$$\mathbb{E}\Big[\mathrm{tr}(\hat{\Sigma}^{-1})\hat{\mu}^\top\Sigma^{-1}\hat{\mu}\Big] = \mu^\top\mathbb{E}\Big[\mathrm{tr}(\hat{\Sigma}^{-1})\hat{\Sigma}^{-1}\Big]\mu + \frac{1}{T}\mathrm{tr}\Big(\Sigma\mathbb{E}\Big[\mathrm{tr}(\hat{\Sigma}^{-1})\hat{\Sigma}^{-1}\Big]\Big). \tag{A.39}$$

Using Equations (84)–(85) in Hillier and Kan (2024), we have

$$\begin{aligned}&\mathbb{E}\Big[\mathrm{tr}(\hat{\Sigma}^{-1})\hat{\Sigma}^{-1}\Big]\\ &= \frac{T^2}{(T-N-2)^2}\left[\left(1 + \frac{2}{(T-N-2)(T-N-4)}\right)\mathrm{tr}(\Sigma^{-1})\Sigma^{-1} + \frac{2\Sigma^{-2}}{T-N-4}\right].\end{aligned} \tag{A.40}$$

Therefore, (A.39) becomes

$$\begin{aligned}\mathbb{E}\Big[\mathrm{tr}(\hat{\Sigma}^{-1})\hat{\mu}^\top\Sigma^{-1}\hat{\mu}\Big] &= \frac{T^2}{(T-N-2)^2}\left[\left(1 + \frac{2}{(T-N-2)(T-N-4)}\right)\mathrm{tr}(\Sigma^{-1})\mu^\top\Sigma^{-1}\mu\right.\\ &\left. + \frac{2\mu^\top\Sigma^{-2}\mu}{T-N-4} + \left(\frac{N}{T} + \frac{2(T-2)}{T(T-N-2)(T-N-4)}\right)\mathrm{tr}(\Sigma^{-1})\right].\end{aligned} \tag{A.41}$$

That is, we have

$$\begin{aligned}&\mathbb{E}\left[\frac{(T-N-2)^2}{T^2}\mathrm{tr}(\hat{\Sigma}^{-1})\hat{\mu}^\top\hat{\Sigma}^{-1}\hat{\mu} - \frac{2\mu^\top\Sigma^{-2}\mu}{T-N-4} - \left(\frac{N}{T} + \frac{2(T-2)}{T(T-N-2)(T-N-4)}\right)\mathrm{tr}(\Sigma^{-1})\right]\\ &= \left(1 + \frac{2}{(T-N-2)(T-N-4)}\right)\mathrm{tr}(\Sigma^{-1})\mu^\top\Sigma^{-1}\mu.\end{aligned} \tag{A.42}$$

Replacing $\mu^\top\Sigma^{-2}\mu$ by (A.48) below, we find that

$$\mathbb{E}\left[\frac{(T-N-2)^2}{T^2}\mathrm{tr}(\hat{\Sigma}^{-1})\hat{\mu}^\top\hat{\Sigma}^{-1}\hat{\mu} - \frac{2(T-N-2)}{T^2}\hat{\mu}^\top\hat{\Sigma}^{-2}\hat{\mu} - \frac{N}{T}\mathrm{tr}(\Sigma^{-1})\right] = \mathrm{tr}(\Sigma^{-1})\mu^\top\Sigma^{-1}\mu, \tag{A.43}$$

which, after replacing $\mathrm{tr}(\Sigma^{-1})$ by its unbiased estimator in (A.38), means that the unbiased estimator of $\mathrm{tr}(\Sigma^{-1})\mu^\top\Sigma^{-1}\mu$ is

$$\mathbb{E}\left[\frac{(T-N-2)^2}{T^2}\mathrm{tr}(\hat{\Sigma}^{-1})\hat{\mu}^\top\Sigma^{-1}\hat{\mu} - \frac{2(T-N-2)}{T^2}\hat{\mu}^\top\hat{\Sigma}^{-2}\hat{\mu} - \frac{N(T-N-2)}{T^2}\mathrm{tr}(\hat{\Sigma}^{-1})\right]$$

$$= \mathrm{tr}(\Sigma^{-1})\mu^\top\Sigma^{-1}\mu. \tag{A.44}$$

Third, the expectation of the plug-in estimator of $\mu^\top\Sigma^{-2}\mu$ is

$$\mathbb{E}\Big[\hat{\mu}^\top\hat{\Sigma}^{-2}\hat{\mu}\Big] = \mu^\top\mathbb{E}\Big[\hat{\Sigma}^{-2}\Big]\mu + \frac{1}{T}\mathrm{tr}\Big(\Sigma\mathbb{E}\Big[\hat{\Sigma}^{-2}\Big]\Big). \tag{A.45}$$

Using Equations (84)–(85) in Hillier and Kan (2024), we have

$$\mathbb{E}\Big[\hat{\Sigma}^{-2}\Big] = \frac{T^2}{(T-N-2)(T-N-4)}\left[\Sigma^{-2} + \frac{\mathrm{tr}(\Sigma^{-1})\Sigma^{-1}}{T-N-2}\right]. \tag{A.46}$$

Therefore, (A.45) becomes

$$\mathbb{E}\Big[\hat{\mu}^\top\hat{\Sigma}^{-2}\hat{\mu}\Big] = \frac{T^2}{(T-N-2)(T-N-4)}\left[\mu^\top\Sigma^{-2}\mu + \frac{1}{T-N-2}\left(\mu^\top\Sigma^{-1}\mu + \frac{T-2}{T}\right)\mathrm{tr}(\Sigma^{-1})\right]. \tag{A.47}$$

That is, we have

$$\mathbb{E}\left[\frac{(T-N-2)(T-N-4)}{T^2}\hat{\mu}^\top\hat{\Sigma}^{-2}\hat{\mu} - \frac{1}{T-N-2}\left(\mu^\top\Sigma^{-1}\mu + \frac{T-2}{T}\right)\mathrm{tr}(\Sigma^{-1})\right] = \mu^\top\Sigma^{-2}\mu, \tag{A.48}$$

which, after replacing $\mathrm{tr}(\Sigma^{-1})$ and $\mathrm{tr}(\Sigma^{-1})\mu^\top\Sigma^{-1}\mu$ by their unbiased estimators in (A.38) and (A.44), respectively, means that the unbiased estimator of $\mu^\top\Sigma^{-2}\mu$ is

$$\mathbb{E}\left[\frac{(T-N-2)(T-N-4)+2}{T^2}\hat{\mu}^\top\hat{\Sigma}^{-2}\hat{\mu} - \frac{T-N-2}{T^2}\mathrm{tr}(\hat{\Sigma}^{-1})(1+\hat{\mu}^\top\hat{\Sigma}^{-1}\hat{\mu})\right] = \mu^\top\Sigma^{-2}\mu. \tag{A.49}$$

Plugging the unbiased estimators of $\mathrm{tr}(\Sigma^{-1})$, $\mathrm{tr}(\Sigma^{-1})\mu^\top\Sigma^{-1}\mu$, and $\mu^\top\Sigma^{-2}\mu$ in (A.38), (A.44), and (A.49) into (A.37) yields the desired unbiased estimator of $M_I(\hat{w}_m)$ in (6.5).

Finally, the unbiased estimator of $L_I(w_m) = \mu^\top\Sigma^{-2}\mu/\gamma^2$ in (6.7) directly follows from (A.49), which completes the proof.

A.5 Proof of Corollary 1

Just like for the proof of Proposition 4, we focus on deriving the unbiased estimator of κ_1. The proof for the unbiased estimator of κ_2 follows a similar reasoning, and this estimator admits a similar formula to that for the unbiased estimator of κ_1 with $\hat{\Sigma}^{-1}$ replaced by $\hat{Q}$ and N by $N-1$. From its definition in (4.7), κ_1 depends on μ and Σ as

$$\kappa_1 = \mu^\top\Sigma^{-2}\mu - \frac{1}{N}\mathrm{tr}(\Sigma^{-1})\mu^\top\Sigma^{-1}\mu. \tag{A.50}$$

Therefore, we can obtain an unbiased estimator of κ_1 by replacing $\mu^\top\Sigma^{-2}\mu$ and $\mathrm{tr}(\Sigma^{-1})\mu^\top\Sigma^{-1}\mu$ by their unbiased estimators in (A.44) and (A.49). This gives

$$\begin{aligned}\hat{\kappa}_1 = &\frac{(T-N-2)(T-N-4)+2}{T^2}\hat{\mu}^\top\hat{\Sigma}^{-2}\hat{\mu} - \frac{T-N-2}{T^2}\mathrm{tr}(\hat{\Sigma}^{-1})(1+\hat{\mu}^\top\hat{\Sigma}^{-1}\hat{\mu}) \\ &- \frac{1}{N}\left[\frac{(T-N-2)^2}{T^2}\mathrm{tr}(\hat{\Sigma}^{-1})\hat{\mu}^\top\Sigma^{-1}\hat{\mu} - \frac{2(T-N-2)}{T^2}\hat{\mu}^\top\hat{\Sigma}^{-2}\hat{\mu} - \frac{N(T-N-2)}{T^2}\mathrm{tr}(\hat{\Sigma}^{-1})\right],\end{aligned} \tag{A.51}$$

which simplifies to the desired result in (6.13), and this completes the proof.

References

Ao, M., Li, Y., Zheng, X., 2019. Approaching mean-variance efficiency for large portfolios. The Review of Financial Studies 32, 2890–2919.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. The Journal of Finance 52, 2003–2049.

Bams, D., Lehnert, T., Wolff, C., 2009. Loss functions in option valuation: A framework for selection. Management Science 55, 853–862.

Barroso, P., Saxena, K., 2021. Lest we forget: Using out-of-sample forecast errors in portfolio optimization. The Review of Financial Studies 35, 1222–1278.

Berger, J., 1980. Statistical Decision Theory: Foundations, Concepts, and Methods. Springer, New York.

Bodnar, T., Dette, H., Parolya, N., Thorsen, E., 2022. Sampling distributions of optimal portfolio weights and characteristics in small and large dimensions. Random Matrices: Theory and Applications 11.

Bodnar, T., Parolya, N., Schmid, W., 2018. Estimation of the global minimum variance portfolio in high dimensions. European Journal of Operational Research 266, 371–390.

Butler, A., Kwon, R. H., 2023. Integrating prediction in mean-variance portfolio optimization. Quantitative Finance 23, 429–452.

Cenesizoglu, T., Timmermann, A., 2012. Do return prediction models add economic value? Journal of Banking & Finance 36, 2974–2987.

Christoffersen, P., Jacobs, K., 2004. The importance of the loss function in option valuation. Journal of Financial Economics 72, 291–318.

Clark, B., Edirisinghe, C., Simaan, M., 2022. Estimation risk and the implicit value of index-tracking. Quantitative Finance 22, 303–319.

DeMiguel, V., Garlappi, L., Nogales, F. J., Uppal, R., 2009a. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. Management Science 55, 798–812.

DeMiguel, V., Garlappi, L., Uppal, R., 2009b. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? The Review of Financial Studies 22, 1915–1953.

DeMiguel, V., Martín-Utrera, A., Nogales, F., 2015. Parameter uncertainty in multiperiod portfolio optimization with transaction costs. Journal of Financial and Quantitative Analysis 50, 1443–1471.

DeMiguel, V., Nogales, F. J., 2009. Portfolio selection with robust estimation. Operations Research 57, 560–577.

El Ghaoui, L., Oks, M., Oustry, F., 2003. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. Operations Research 51, 543–556.

Elmachtoub, A., Grigas, P., 2022. Smart "predict, then optimize". Management Science 68, 9–26.

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

Frahm, G., Memmel, C., 2010. Dominating estimators for minimum-variance portfolios. Journal of Econometrics 159, 289–302.

Hillier, G., Kan, R., 2024. On the expectations of equivariant matrix-valued functions of Wishart and inverse Wishart matrices. Scandinavian Journal of Statistics 51, 697–723.

Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. The Journal of Finance 58, 1651–1684.

Javed, F., Mazur, S., Ngailo, E., 2021. Higher order moments of the estimated tangency portfolio weights. Journal of Applied Statistics 48, 517–535.

Kan, R., Wang, X., Zhou, G., 2022. Optimal portfolio choice with estimation risk: No risk-free asset case. Management Science 68, 2047–2068.

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

Kelly, B., Malamud, S., Zhou, K., 2024. The virtue of complexity in return prediction. The Journal of Finance 79, 459–503.

Khashanah, K., Simaan, M., Simaan, Y., 2022. Do we need higher-order comoments to enhance mean-variance portfolios? Evidence from a simplified jump process. International Review of Financial Analysis 81, 102068.

Kozak, S., Nagel, S., Santosh, S., 2018. Interpreting factor models. The Journal of Finance 73, 1183–1223.

Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. Journal of Financial Economics 135, 271–292.

Lassance, N., 2022. Reconciling mean-variance portfolio theory with non-Gaussian returns. European Journal of Operational Research 297, 729–740.

Lassance, N., DeMiguel, V., Vrins, F., 2022. Optimal portfolio diversification via independent component analysis. Operations Research 70, 55–72.

Lassance, N., Martìn-Utrera, A., Simaan, M., 2024. The risk of expected utility under parameter uncertainty. Management Science (forthcoming).

Lassance, N., Vanderveken, R., Vrins, F., 2023. On the combination of naive and mean-variance portfolio strategies. Journal of Business & Economic Statistics (forthcoming).

Ledoit, O., Wolf, M., 2003a. Honey, I shrunk the sample covariance matrix. The Journal of Portfolio Management 30, 110–119.

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

Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis 88, 365–411.

Ledoit, O., Wolf, M., 2017. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. The Review of Financial Studies 30, 4349–4388.

Liu, Y., Zhou, G., Zhu, Y., 2024. Maximizing the Sharpe ratio: A genetic programming approach. Available at SSRN 3726609.

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

Simaan, Y., 2014. The opportunity cost of mean-variance choice under estimation risk. European Journal of Operational Research 234, 382–391.

Tu, J., Zhou, G., 2011. Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. Journal of Financial Economics 99, 204–215.

Vanderschueren, T., Verdonck, T., Baesens, B., Verbeke, W., 2022. Predict-then-optimize or predict-and-optimize? An empirical evaluation of cost-sensitive learning strategies. Information Sciences 594, 400–415.

Yen, Y.-M., 2016. Sparse weighted-norm minimum variance portfolios. Review of Finance 20, 1259–1287.

Zhao, L., Chakrabarti, D., Muthuraman, K., 2019. Portfolio construction by mitigating error amplification: The bounded-noise portfolio. Operations Research 67, 965–983.