



Optimal portfolio choice: a minimum expected loss approach

Andrés Ramírez-Hassan¹ · Rosember Guerra-Urzola^{2,3}

Received: 28 February 2019 / Accepted: 13 August 2019 / Published online: 7 September 2019
© The Author(s) 2019

Abstract

The mainstream in finance tackles portfolio selection based on a plug-in approach without consideration of the main objective of the inferential situation. We propose minimum expected loss (MELO) estimators for portfolio selection that explicitly consider the trading rule of interest. The asymptotic properties of our MELO proposal are similar to the plug-in approach. Nevertheless, simulation exercises show that our proposal exhibits better finite sample properties when compared to the competing alternatives, especially when the tangency portfolio is taken as the asset allocation strategy. We have also developed a graphical user interface to help practitioners to use our MELO proposal.

Keywords Bayesian estimation · Minimum expected loss · Portfolio selection

JEL Classification C01 · C11 · G11

1 Introduction

The mainstream methods to estimate the optimal weights in the portfolio allocation problem is based on the *plug-in* approach; that is, individual location and scale estimates are simply plugged into the objective expression without explicit consideration of the main goal of the inferential situation. However, this approach has some shortcomings: it ignores parameter uncertainty [1–6], has infinite mean in some cases (tangency and Treynor–Black portfolios) [7], and has unbounded risks relative to quadratic loss functions [8].

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11579-019-00246-w>) contains supplementary material, which is available to authorized users.

✉ Rosember Guerra-Urzola
r.i.guerraurzola@uvt.nl; rguerra@eafit.edu.co

Andrés Ramírez-Hassan
aramir21@eafit.edu.co

¹ Department of Economics, School of Economics and Finance, Universidad EAFIT, Medellín, Colombia

² Department of Methodology and Statistics, School of Social and Behavioral Science, Tilburg University, Tilburg, The Netherlands

³ Department of Finance, School of Economics and Finance, Universidad EAFIT, Medellín, Colombia

To mitigate these issues, we follow a decision theory framework based on a Bayesian approach, the minimum expected loss approach (MELO), where the posterior expected value of a generalized quadratic loss function, which depends explicitly on the optimal assets weights (the main estimation goal), is minimized. We analyze the global minimum variance portfolio, tangency portfolio, and Treynor–Black portfolio.

Zellner [8] introduced the MELO approach in simultaneous equations models. He showed that the MELO estimator has, at least, finite first and second moments, and finite risk with respect to a generalized quadratic loss function. Further, Zellner [9] approximated the small sample moments and risk functions of the MELO estimators, and Zellner [10] showed that MELO estimates of structural parameters are weighted averages of direct least squares and two-stage least squares. It has been shown in Monte Carlo experiments that the MELO estimates have less mean squared errors than the two-stage least squares estimates [11]. However, the former has more bias than the latter [11]. In addition, Swamy and Mehta [12] showed that conditions for existence of the full information maximum likelihood estimator are more demanding than the conditions to get the MELO estimator in undersized sample conditions, that is, when the number of exogenous variables exceeds the sample size. Zellner [13] proposed the Bayesian method of moments, and related it to the MELO, extending his proposal to cases where there are just moment conditions for estimation. Recently, Ramirez-Hassan and Correa-Giraldo [14] proposed the MELO estimator using a generalized quadratic loss function focused on rational functions of parameters. Ramirez-Hassan and Correa-Giraldo [14] showed that the asymptotic properties of the MELO estimator are similar to the plug-in approach. Nevertheless, simulation exercises show that the MELO has better finite statistical properties than the *plug-in* approach. Moreover, the MELO estimator is rooted in a Bayesian framework, therefore, it takes into account estimation error by construction.

Estimation error is a huge concern in optimal portfolio strategies. In particular, it seems that expected returns are the primary source of estimation error [5] accounting 10 times more error than the covariance matrix [15] (Kan and Zhou [6] argue against this point). Therefore, many approaches have been proposed to mitigate this issue. The Bayesian approach, which explicitly takes into account parameter uncertainty, has played a prominent role; Barry [2], Brown [16] and Frost and Savarino [17] addressed this issue using diffuse priors for parameters, finding that their strategies drive to the same admissible set of optimal portfolios than using traditional analysis. Nevertheless, their portfolio weight estimators differ from those obtained by frequentist approaches. Meanwhile, Klein and Bawa [3] proposed an informative prior setting, which changes the admissible set of portfolios and optimal portfolio weights. The Bayes–Stein strategy is another important approach, which is an application of shrinkage estimators that produce biased estimators but with lower mean squared error [18]. This approach is used by Jorion [19] to estimate expected returns, Ledoit and Wolf [20,21] to estimate the covariance matrix, the former based on a linear shrinkage approach, whereas the latter in a non-linear fashion, and [22,23], where the shrinkage target depends on the investor's prior belief in an asset pricing model. Frahm and Memmel [24] promoted the global minimum variance portfolio using a shrinkage estimator for the variance matrix due to expected value accounting for most of the estimation error.

Recently, DeMiguel et al. [25] proposed to use naive (equal) weights to avoid the issue of estimation risk. They used 14 portfolios strategies and seven datasets to test the power of the naive strategy, showing that the naive portfolio outperforms most of the methodologies with different datasets. However, this idea was strongly criticized by Kritzman et al. [26] because it is based on particular designs of datasets. Kritzman et al. [26] showed examples where optimization strategies perform better than naive strategies. Another stream is based on combining portfolios from different optimal rules to diversify estimation errors. Kan and

Zhou [6] proposed optimal portfolio allocation estimators minimizing a risk function that depends on the out-of-sample performance of the expected investor's utility function. Kan and Zhou [6] focused on admissible trading strategies, admissibility is a minimum requirement on decision rules, proposing a “three-fund” portfolio rule composed by the risk-free asset, tangency portfolio, and the global minimum-variance. Tu and Zhou [27] proposed a combined portfolio between the naive strategy and one that comes from an optimization problem. They proposed the naive portfolio as a shrinkage target. Other strategies to mitigate estimation error are based on robust portfolios [28,29], and transforming the optimal weight estimation problem into linear regressions [30,31]. In the former approach, parameter uncertainty is taken into account in the optimization procedure. In the latter approach, Li [30] proposed a sparse and stable methodology based on lasso and ridge regressions with similar statistical characteristics than shrinkage estimators. Klimentka and Wolter [31], also proposed a regression framework that uses the focused information criterion [32], which is based on the trading strategy, and model averaging to take model uncertainty into account.

Our proposal has some characteristics from previous proposals due to being based on a Bayesian setting [2,3,16,17] under a decision theory framework focused on the final inferential goal [6,31]. However, our proposal is based on minimizing the posterior expected loss function rather than the frequentist risk function, and our loss function is based on the trading strategy rather than the utility function. In particular, we exploit the specific structure (rational functions) of the main objective of estimation in three well known portfolio optimization problems, and propose the MELO approach obtaining same asymptotic results as the *plug-in* approach, but showing that our proposal obtains better statistical properties in finite samples when compared to the competing alternatives, especially when the optimal trading rule is the tangency portfolio. To the best of our knowledge, this is the first time that the MELO estimator is used for these three optimal portfolio strategies.

The rest of this paper is structured as follows. Section 2 shows theoretical framework of different competing alternatives. Section 3 develops the MELO estimates for global minimum variance, tangency portfolio and Treynor–Black model. Section 4 exhibits the outcomes of the simulation exercises. In Sect. 5, we develop an empirical study. Finally, we make some conclusions.

2 Theoretical framework

Suppose that the investment universe consists on N assets. Denoting by \mathbf{R}_t the excess of returns of the N assets at time t , $\mathbf{R}_t = (r_{1t}, r_{2t}, \dots, r_{Nt})'$.¹ It is assumed that the excess of returns has a multivariate normal distribution $\mathbf{R}_t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The portfolio weights are the proportion of wealth invested in each of the N assets, $\mathbf{w} = (w_1, w_2, \dots, w_N)'$. We suppose that the investor has a portfolio holding period of length κ and that the investor wants to maximize their wealth at the end of the investment horizon, $T + \kappa$, where T is the last period for which return data is available (sample size).

¹ The excess of returns is defined as the asset returns minus the corresponding risk-free rate.

2.1 Trading strategies

2.1.1 Global minimum variance portfolio

The global minimum variance (GMV) is a portfolio whose weights represent the combination that gives the minimum variance between all possible portfolios. It is defined as the solution of the minimization problems,

$$\operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^N} \mathbf{w}' \boldsymbol{\Sigma}_{T+\kappa} \mathbf{w}; \quad s.t. \quad \mathbf{w}' \mathbf{1} = 1,$$

where $\mathbf{1}$ denotes a vector of ones. Because $\boldsymbol{\Sigma}$ is positive defined, the GMV is unique and the solution of the minimization problem is

$$\mathbf{w}_{vp} = \frac{\boldsymbol{\Sigma}_{T+\kappa}^{-1} \mathbf{1}}{\mathbf{1}' \boldsymbol{\Sigma}_{T+\kappa}^{-1} \mathbf{1}}. \tag{1}$$

2.1.2 Tangency portfolio

The tangency portfolio is defined as the portfolio that has the highest Sharpe ratio. The tangency portfolio solves the constrained maximization problem

$$\operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^N} \frac{\mathbf{w}' \boldsymbol{\mu}_{T+\kappa}}{\sqrt{\mathbf{w}' \boldsymbol{\Sigma}_{T+\kappa} \mathbf{w}}}; \quad s.t. \quad \mathbf{w}' \mathbf{1} = 1,$$

thus, the solution has the expression

$$\mathbf{w}_{tp} = \frac{\boldsymbol{\Sigma}_{T+\kappa}^{-1} \boldsymbol{\mu}_{T+\kappa}}{\mathbf{1}' \boldsymbol{\Sigma}_{T+\kappa}^{-1} \boldsymbol{\mu}_{T+\kappa}}. \tag{2}$$

2.1.3 Treynor–Black Model

Active management searches some sources of abnormal returns (alpha) to outperform a passive benchmark portfolio. The Treynor–Black model, which was proposed by Treynor and Black [33], tackled this problem by assuming an investor who considers that most securities are mis-priced with respect to an asset pricing model but who believes that they have information that can be used to predict the abnormal returns of a few of the securities.

Consider the following regression model,

$$r_{it} = \alpha_i + \beta_i r_{Mt} + e_{it}$$

where r_{Mt} is the excess of return of the benchmark portfolio, and $e_i \sim N(\mathbf{0}, \mathbf{H})$.

This strategy consists of investing in an active portfolio (A) containing the assets for which the investor has made a prediction about abnormal return and a passive portfolio (B, benchmark) containing all assets in proportion to their market value. Let's \mathbf{w}^* denote the weights for the active portfolio that maximize the information ratio.

$$\operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^N} \frac{\mathbf{w}' \boldsymbol{\alpha}_{T+\kappa}}{\sqrt{\mathbf{w}' \mathbf{H}_{T+\kappa} \mathbf{w}}} \quad s.t. \quad \mathbf{w}' \mathbf{1} = 1,$$

where $\alpha_{T+\kappa} = (\alpha_{1,T+\kappa}, \alpha_{2,T+\kappa}, \dots, \alpha_{N,T+\kappa})'$. The solution is given by

$$w^* = \frac{H_{T+\kappa}^{-1} \alpha_{T+\kappa}}{\mathbf{1}' H_{T+\kappa}^{-1} \alpha_{T+\kappa}} \tag{3}$$

The second stage is to construct an optimal mix of A and B to form a risky portfolio P. This is a standard two risk assets portfolio problem. Here, w_A and $1 - w_A$ denote the weights of wealth invested in A and B, respectively, where

$$w_A = \frac{w_0}{1 + (1 - \beta_A)w_0},$$

where $w_0 = \frac{\alpha_A/\sigma_A^2}{r_M/\sigma_M^2}$, $\sigma_A^2 = w^{*'} H_{T+\kappa} w^*$, $\alpha_A = w^{*'} \alpha_{T+\kappa}$, $\beta_A = \beta'_{T+\kappa} w^*$, and $\beta_{T+\kappa} = (\beta_{1,T+\kappa}, \beta_{2,T+\kappa}, \dots, \beta_{N,T+\kappa})'$.

Observe that all of the optimal weights depend on future expected returns at $T + \kappa$. As a consequence, they depend on parameter estimates.

2.2 Statistical strategies

In this subsection, we show different statistical strategies to estimate $\mu_{T+\kappa}$, $\Sigma_{T+\kappa}$, $\alpha_{T+\kappa}$, $\beta_{T+\kappa}$ and $H_{T+\kappa}$.

2.2.1 Plug-in approach

The classical approach estimates parameters using available sample information and then plugs these estimates in the optimal solutions omitting parameter uncertainty. In particular,

$$\begin{aligned} \hat{\mu} &= \frac{1}{T} \sum_{t=1}^T R_t, \\ \hat{\Sigma} &= \frac{1}{T-1} \sum_{t=1}^T (R_t - \hat{\mu})(R_t - \hat{\mu})', \\ \hat{B} &= (X'X)^{-1} X'R, \end{aligned}$$

and

$$\hat{H} = \frac{1}{T - N - 1} (R - X\hat{B})'(R - X\hat{B}),$$

where R is a $T \times N$ matrix of excess of returns, $X = [\mathbf{1} \ r_M]$ is a $T \times 2$ design matrix, and

$$\hat{B} = \begin{bmatrix} \hat{\alpha}' \\ \hat{\beta}' \end{bmatrix}.$$

2.2.2 Shrinkage approach

A *shrinkage estimator* is a weighted average of the sample estimator and the so-called *Bayes–Stein estimator* of the mean. Under this approach,

$$\hat{\mu}^{Sh} = (1 - \lambda)\hat{\mu} + \lambda \mathbf{1}\mu_0,$$

where μ_0 is the shrinkage target, and the shrinkage intensity λ is given by

$$\lambda = \min \left\{ 1, \frac{N - 2}{T(\hat{\boldsymbol{\mu}} - \mathbf{1}\mu_0)' \widehat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\mu}} - \mathbf{1}\mu_0)} \right\}.$$

Jorion [19] proposed as the shrinkage target the return on the global minimum variance portfolio,²

$$\mu_0 = \frac{\mathbf{1}' \widehat{\boldsymbol{\Sigma}}^{-1}}{\mathbf{1}' \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}} \hat{\boldsymbol{\mu}}.$$

2.2.3 Bayesian approach

The Bayesian approach accounts for parameter uncertainty. In particular, it expresses the investor’s problem in terms of the predictive distribution of the future excess returns. Denoting the unobserved κ next-periods excesses return data by $\mathbf{R}_{T+\kappa}$, the predictive return density is

$$p(\mathbf{R}_{T+\kappa} \mid \mathbf{R}) \propto \int \int p(\mathbf{R}_{T+\kappa} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{R}) d\boldsymbol{\mu} d\boldsymbol{\Sigma},$$

where $p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{R})$ is the joint posterior density, and $p(\mathbf{R}_{T+\kappa} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a multivariate normal density

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{R}) \propto \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{R}) \times p(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where $\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{R}) \propto |\boldsymbol{\Sigma}|^{-T/2} \exp\left(-\frac{1}{2} \sum_{t=1}^T (\mathbf{R}_t - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{R}_t - \boldsymbol{\mu})\right)$ is the likelihood function, and $p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the prior density.

In the following, we show the Bayesian solution under two situations: non-informative and informative priors (see supplementary material section 1).

Non-informative priors In this case, the investor is uncertain about the distribution of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and has no particular prior knowledge. This situation can be represented by a flat prior, which is typically taken to be the Jeffreys’ prior (see supplementary material subsection 1.2).

The estimates for $\boldsymbol{\mu}_{T+\kappa}$, $\boldsymbol{\Sigma}_{T+\kappa}$, $\boldsymbol{\alpha}_{T+\kappa}$, $\boldsymbol{\beta}_{T+\kappa}$ and $\mathbf{H}_{T+\kappa}$ are

$$\hat{\boldsymbol{\mu}}^{NB} = \hat{\boldsymbol{\mu}}, \tag{4}$$

$$\hat{\boldsymbol{\Sigma}}^{NB} = \frac{c_1^{-1}(T - 1)}{(T + \kappa - N - 3)} \widehat{\boldsymbol{\Sigma}}, \tag{5}$$

$$\hat{\mathbf{B}}^{NB} = \hat{\mathbf{B}}, \tag{6}$$

and

$$\hat{\mathbf{H}}^{NB} = \frac{c_2^{-1}(T - 1)}{(T + \kappa - N - 4)} \hat{\mathbf{H}}, \tag{7}$$

² Shrinkage estimators for the covariance matrix have also been developed. For example, Ledoit and Wolf [20,21] propose shrinkage estimators when the number of assets can be greater than the sample size. Frahm and Memmel [24] propose an estimator that dominates the traditional estimator with respect to the out-of-sample variance of the portfolio return.

where $c_1 = \frac{T(T+\kappa)(T+\kappa-1) - (\kappa-1)(T-\kappa+1)}{T(T+\kappa)^2}$, $C = I - Z(X'X + Z'Z)^{-1}Z' = \begin{bmatrix} C_{1:\kappa-1, 1:\kappa-1} & C_{1:\kappa-1, \kappa} \\ C_{\kappa, 1:\kappa-1} & C_{\kappa, \kappa} \end{bmatrix}$, $Z = [1 \ r_{M\kappa}]$ is a $\kappa \times 2$ matrix, $r_{M\kappa}$ is a forecast about future benchmark portfolio returns, $c_2 = C_{\kappa, \kappa} - C_{\kappa, 1:\kappa-1} C_{1:\kappa-1, 1:\kappa-1}^{-1} C_{1:\kappa-1, \kappa}$.³

Informative priors Now we suppose that the investor has information about parameters in the investment period. We get the following results using conjugate family priors (see supplementary material subsection 1.1),

$$\hat{\mu}^{IB} = \frac{\tau}{T + \tau} \eta + \frac{T}{T + \tau} \hat{\mu}, \tag{8}$$

$$\hat{\Sigma}^{IB} = \frac{c_3^{-1}}{(T + \kappa + \nu - N - 2)} \left(\Omega + (T - 1) \hat{\Sigma} + \frac{T\tau}{T + \tau} (\eta - \hat{\mu})(\eta - \hat{\mu})' \right), \tag{9}$$

$$\hat{B}^{IB} = (V_0^{-1} + X'X)^{-1} (V_0^{-1} B_0 + X'R), \tag{10}$$

and,

$$\hat{H}^{IB} = \frac{c_2^{-1}}{(T + \nu_0 + \kappa - N - 2)} \left(H_0 + S + B_0' V_0^{-1} B_0 + \hat{B}' X' X \hat{B} - \hat{B}^{IB'} \hat{V}^{-1} \hat{B}^{IB} \right), \tag{11}$$

where η is an N dimensional vector of prior mean returns, τ is a hyperparameter that defines prior precision, Ω and H_0 are prior scale matrices associated with the covariance matrix,

ν and ν_0 are degrees of freedom, $V_0 = \begin{bmatrix} \frac{1}{\tau_\alpha} & 0 \\ 0 & \frac{1}{\tau_\beta} \end{bmatrix}$ is a matrix that defines prior precision

in the Treynor–Black model (τ_α and τ_β define precision regarding α and β , respectively), $c_3 = \frac{(T+\tau)(T+\tau+\kappa)(T+\tau+\kappa-1) - (\kappa-1)(T+\tau-\kappa+1)}{(T+\tau)(T+\tau+\kappa)^2}$, $S = (R - X\hat{B})'(R - X\hat{B})$, and $\tilde{V} = (V_0^{-1} + X'X)^{-1}$.⁴

Note that $\hat{\mu}^{IB}$ and \hat{B}^{IB} are weighted averages between sample and prior information. More precision about prior information implies more weight associated with this source.

3 Minimum expected loss for trading strategies

Taking into account that the financial trading strategies (Eqs. 1, 2 and 3) are rational functions of parameters and that these are the final objective of estimation, we propose the following framework.

Suppose that the main concern of estimation is $\omega = g(\theta) : \Theta \subset \mathbb{R}^L \rightarrow \mathbb{R}^N$, $N \leq L$; that is, $\omega = (w_1, w_2, \dots, w_N)' = (g_1(\theta), g_2(\theta), \dots, g_N(\theta))'$, $g_i(\theta) = \frac{l_i(\theta)}{m(\theta)} : \mathbb{R}^L \rightarrow \mathbb{R}$, $i = 1, 2, \dots, N$, $l_i(\theta)$ and $m(\theta) \neq 0$ are polynomial functions in θ , such that $g_i(\theta)$ is a continuously differentiable constant order transformation.⁵

Setting $\omega = (w_1, w_2, \dots, w_N)$, the optimal portfolio weights, we propose to focus our inferential problem directly on our final objective; that is, the trading strategies. Therefore, we select as an estimator the Bayesian action that minimizes the posterior expected value of

³ In our simulation exercises, we get $r_{M\kappa}$ from a normal distribution. In our application, we use target price indices from Bloomberg to calculate the target returns.

⁴ $C = I - Z(V_0^{-1} + X'X + Z'Z)^{-1}Z'$, c_2 is defined as in the uninformative case.

⁵ Constant order means that g_i is bounded ($g_i(\theta) = O(1)$).

a generalized quadratic loss function focused on the optimal portfolio rules, let us say $\mathbf{g}(\boldsymbol{\theta})$,

$$\operatorname{argmin}_{\hat{\boldsymbol{\omega}} \in \mathbb{R}^N} \mathbb{E}_{\pi(\boldsymbol{\theta}|\text{Data})} \{ \mathcal{L}(\mathbf{g}(\boldsymbol{\theta}), \hat{\boldsymbol{\omega}}) \} = \operatorname{argmin}_{\hat{\boldsymbol{\omega}} \in \mathbb{R}^K} \int_{\Theta} \{ \mathcal{L}(\mathbf{g}(\boldsymbol{\theta}), \hat{\boldsymbol{\omega}}) \} \pi(\boldsymbol{\theta}|\text{Data}) d\boldsymbol{\theta},$$

where $\pi(\boldsymbol{\theta}|\text{Data})$ is the posterior distribution, $\mathcal{L}(\mathbf{g}(\boldsymbol{\theta}), \hat{\boldsymbol{\omega}}) = h(\boldsymbol{\theta})(\mathbf{g}(\boldsymbol{\theta}) - \hat{\boldsymbol{\omega}})'(\mathbf{g}(\boldsymbol{\theta}) - \hat{\boldsymbol{\omega}})$, where $h(\boldsymbol{\theta}) > 0$ is a case specific weighting function.

Proposition 1 *Let $\mathcal{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\omega}}) = h(\boldsymbol{\theta})(\hat{\boldsymbol{\omega}} - \mathbf{g}(\boldsymbol{\theta}))'(\hat{\boldsymbol{\omega}} - \mathbf{g}(\boldsymbol{\theta}))$ be a loss function, where $h(\boldsymbol{\theta}) : \mathbb{R}^L \rightarrow \mathbb{R}^{++}$ is a continuous constant order weighting function. Then, the MELO estimate is*

$$\begin{aligned} \hat{\boldsymbol{\omega}}^*(\mathbf{R}) &= \frac{\mathbb{E}_{\pi(\boldsymbol{\theta}|\mathbf{R})}[\mathbf{g}(\boldsymbol{\theta})h(\boldsymbol{\theta})]}{\mathbb{E}_{\pi(\boldsymbol{\theta}|\mathbf{R})}[h(\boldsymbol{\theta})]} \\ &= \int_{\Theta} \mathbf{g}(\boldsymbol{\theta}) \frac{h(\boldsymbol{\theta})}{\int_{\Theta} h(\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{R})d\boldsymbol{\theta}} \pi(\boldsymbol{\theta}|\mathbf{R})d\boldsymbol{\theta}, \end{aligned}$$

provided previous assumptions on $\mathbf{g}(\boldsymbol{\theta})$ and $h(\boldsymbol{\theta})$, and integration and differentiation can be interchanged (see assumptions **E** and **F** in supplementary material subsection 2.1 for details).

See the supplementary material for a proof (subsection 2.2).

Observe that Proposition 1 implies that the MELO is a kernel weighted average of $\mathbf{g}(\boldsymbol{\theta})$. These weights implicitly depend on the probability associated with each $\boldsymbol{\theta}$ in their parameter space, as well as their magnitude. When h does not depend on $\boldsymbol{\theta}$, which implies equal weight to each $\boldsymbol{\theta}$, the minimum expected loss estimate is the posterior mean; that is, $\hat{\boldsymbol{\omega}}^*(\mathbf{R}) = \mathbb{E}_{\pi(\boldsymbol{\theta}|\mathbf{R})}\mathbf{g}(\boldsymbol{\theta})$.

If our problem is to estimate the weights for the global minimum variance portfolio, $\mathbf{g}(\boldsymbol{\Sigma}) = \boldsymbol{\omega} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}$, where $\boldsymbol{\Sigma}$ is the covariance matrix of the excess of returns, then we have that $(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})\boldsymbol{\omega} - \boldsymbol{\Sigma}^{-1}\mathbf{1} = \mathbf{0}$. So, we set $\mathcal{L}(\boldsymbol{\Sigma}, \hat{\boldsymbol{\omega}}) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} = (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})\hat{\boldsymbol{\omega}} - \boldsymbol{\Sigma}^{-1}\mathbf{1} = (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})$ is the estimation error introduced by the estimate $\hat{\boldsymbol{\omega}}$. Then, the posterior expected loss function is

$$\begin{aligned} \mathbb{E}_{\pi(\boldsymbol{\Sigma}|\mathbf{R})}\mathcal{L}(\boldsymbol{\Sigma}, \hat{\boldsymbol{\omega}}) &= \mathbb{E}_{\pi(\boldsymbol{\Sigma}|\mathbf{R})}\{((\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})\hat{\boldsymbol{\omega}} - \boldsymbol{\Sigma}^{-1}\mathbf{1})'((\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})\hat{\boldsymbol{\omega}} - \boldsymbol{\Sigma}^{-1}\mathbf{1})\} \\ &= \mathbb{E}_{\pi(\boldsymbol{\Sigma}|\mathbf{R})}\{(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})'(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})\}. \end{aligned}$$

Corollary 1 *The MELO estimate for the weights associated with the minimum variance portfolio is given by*

$$\begin{aligned} \hat{\boldsymbol{\omega}}^* &= [\mathbb{E}_{\pi(\boldsymbol{\Sigma}|\mathbf{R})}\{(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2\}]^{-1}\mathbb{E}_{\pi(\boldsymbol{\Sigma}|\mathbf{R})}\{(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})\boldsymbol{\Sigma}^{-1}\mathbf{1}\} \\ &= \int \boldsymbol{\omega} \frac{(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}{\int (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2\pi(\boldsymbol{\Sigma}|\mathbf{R})d\boldsymbol{\Sigma}} \pi(\boldsymbol{\Sigma}|\mathbf{R})d\boldsymbol{\Sigma} \\ &= \int \boldsymbol{\omega} \frac{\frac{1}{(\sigma_{MVP}^2)^2}}{\int \frac{1}{(\sigma_{MVP}^2)^2}\pi(\boldsymbol{\Sigma}|\mathbf{R})d\boldsymbol{\Sigma}} \pi(\boldsymbol{\Sigma}|\mathbf{R})d\boldsymbol{\Sigma}, \end{aligned}$$

where $\sigma_{MVP}^2 = \frac{1}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}$ is the variance of the minimum-variance portfolio.

Proof This is an immediate consequence of Proposition 1 taking $\mathbf{g}(\boldsymbol{\theta}) = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}$ and $h(\boldsymbol{\theta}) = (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2$. □

We can see from Corollary 1 that the MELO estimate for the weights of the minimum variance portfolio is a weighted average, where the weights depend on the updated belief regarding the variance of the minimum variance portfolio. In particular, covariance matrices that imply larger portfolio’s variance have smaller weights to calculate the MELO estimates. This is consistent with the logic of the optimization problem from a financial theory perspective, whose concern is to minimize the variance of the portfolio.

If the main concern is an estimate of the weights associated with the tangency portfolio, $\omega = \frac{\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mu}$, where μ and Σ are the mean and covariance matrix of the excess of returns, then we set $\epsilon = (\mathbf{1}'\Sigma^{-1}\mu)\hat{\omega} - \Sigma^{-1}\mu$. Then, the loss function is $\mathcal{L}(\Sigma, \mu, \hat{\omega}) = \epsilon'\epsilon$, and the posterior expected loss,

$$\begin{aligned} \mathbb{E}_{\pi(\Sigma, \mu | R)} \mathcal{L}(\Sigma, \mu, \hat{\omega}) &= \mathbb{E}_{\pi(\Sigma, \mu | R)} \{((\mathbf{1}'\Sigma^{-1}\mu)\hat{\omega} - \Sigma^{-1}\mu)'((\mathbf{1}'\Sigma^{-1}\mu)\hat{\omega} - \Sigma^{-1}\mu)\} \\ &= \mathbb{E}_{\pi(\Sigma, \mu | R)} \{(\mathbf{1}'\Sigma^{-1}\mu)^2(\hat{\omega} - \omega)'\hat{\omega} - \omega)\}. \end{aligned}$$

Corollary 2 *The MELO estimate for the weights associated with the tangency portfolio is given by*

$$\begin{aligned} \hat{\omega}^* &= [\mathbb{E}_{\pi(\Sigma, \mu | R)} \{(\mathbf{1}'\Sigma^{-1}\mu)^2\}]^{-1} \mathbb{E}_{\pi(\Sigma, \mu | R)} \{(\mathbf{1}'\Sigma^{-1}\mu)\Sigma^{-1}\mu\} \\ &= \int \omega \int \frac{(\mathbf{1}'\Sigma^{-1}\mu)^2}{(\mathbf{1}'\Sigma^{-1}\mu)^2 \pi(\Sigma, \mu | R)} \pi(\Sigma, \mu | R) d\Sigma d\mu \\ &= \int \omega \frac{\left(\frac{\mu_{TP}}{\sigma_{TP}^2}\right)^2}{\int \left(\frac{\mu_{TP}}{\sigma_{TP}^2}\right)^2 \pi(\Sigma, \mu | R) d\Sigma d\mu} \pi(\Sigma, \mu | R) d\Sigma d\mu, \end{aligned}$$

where $\mu_{TP} = \frac{\mu'\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mu}$ and $\sigma_{TP}^2 = \frac{\mu'\Sigma^{-1}\mu}{(\mathbf{1}'\Sigma^{-1}\mu)^2}$ are the mean and variance of the tangency portfolio.

Proof This is a consequence from Proposition 1 taking $\mathbf{g}(\theta) = \frac{\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mu}$ and $h(\theta) = (\mathbf{1}'\Sigma^{-1}\mu)^2$. □

We can see from Corollary 2 that the MELO estimate for the weights of the tangency portfolio is a weighted average, where the weights depend on the updated belief regarding the ratio between the mean and the variance of the tangency portfolio. In particular, combinations of the mean and covariance matrices that imply larger portfolio’s ratios have larger weights to calculate the MELO estimate. This is consistent with the logic of the optimization problem from a financial theory perspective, whose concern is to maximize the Sharpe ratio.

In addition, we propose MELO estimates for the weights of the Treynor–Black model, whose optimal solution is $\omega = \frac{H^{-1}\alpha}{\mathbf{1}'H^{-1}\alpha}$, where α and H are the intercept and covariance matrix of the stochastic errors in the model $r_{it} = \alpha + \beta_i r_{Mt} + e_{it}$. In this framework, we set $\epsilon = (\mathbf{1}'H^{-1}\alpha)\hat{\omega} - H^{-1}\alpha$, such that the loss function is $\mathcal{L}(H, \alpha, \hat{\omega}) = (\mathbf{1}'H^{-1}\alpha)^2(\hat{\omega} - \omega)'\hat{\omega} - \omega)$, and the posterior expected loss,

$$\begin{aligned} \mathbb{E}_{\pi(H, \alpha | R, r_M)} \mathcal{L}(H, \alpha, \hat{H}) &= \mathbb{E}_{\pi(H, \alpha | R, r_M)} \{((\mathbf{1}'H^{-1}\alpha)\hat{\omega} - H^{-1}\alpha)'((\mathbf{1}'H^{-1}\alpha)\hat{\omega} - H^{-1}\alpha)\} \\ &= \mathbb{E}_{\pi(H, \alpha | R, r_M)} \{(\mathbf{1}'H^{-1}\alpha)^2(\hat{\omega} - \omega)'\hat{\omega} - \omega)\}. \end{aligned}$$

Corollary 3 *The MELO estimate for the weights associated with the Treynor–Black portfolio is given by*

$$\begin{aligned} \hat{\omega}^* &= [\mathbb{E}_{\pi(\mathbf{H}, \alpha | \mathbf{R}, r_M)} \{(\mathbf{1}' \mathbf{H}^{-1} \alpha)^2\}]^{-1} \mathbb{E}_{\pi(\mathbf{H}, \alpha | \mathbf{R}, r_M)} \{(\mathbf{1}' \mathbf{H}^{-1} \alpha) \mathbf{H}^{-1} \alpha\} \\ &= \int \omega \frac{(\mathbf{1}' \mathbf{H}^{-1} \alpha)^2}{\int (\mathbf{1}' \mathbf{H}^{-1} \alpha)^2 \pi(\mathbf{H}, \alpha | \mathbf{R}) d\mathbf{H} d\alpha} \pi(\mathbf{H}, \alpha | \mathbf{R}, r_M) d\mathbf{H} d\alpha \\ &= \int \omega \frac{\left(\frac{\alpha_{TB}}{\sigma_{TB}^2}\right)^2}{\int \left(\frac{\alpha_{TB}}{\sigma_{TB}^2}\right)^2 \pi(\mathbf{H}, \alpha | \mathbf{R}) d\mathbf{H} d\alpha} \pi(\mathbf{H}, \alpha | \mathbf{R}, r_M) d\mathbf{H} d\alpha, \end{aligned}$$

where $\alpha_{TB} = \frac{\alpha' \mathbf{H}^{-1} \alpha}{\mathbf{1}' \mathbf{H}^{-1} \alpha}$ and $\sigma_{TB}^2 = \frac{\alpha' \mathbf{H}^{-1} \alpha}{(\mathbf{1}' \mathbf{H}^{-1} \alpha)^2}$ are the weighted alpha and weighted stochastic error variance of the Treynor–Black portfolio.

Proof This is a consequence from Proposition 1 taking $\mathbf{g}(\theta) = \frac{\mathbf{H}^{-1} \alpha}{\mathbf{1}' \mathbf{H}^{-1} \alpha}$ and $h(\theta) = (\mathbf{1}' \mathbf{H}^{-1} \alpha)^2$. □

We observe in Corollary 3 that the MELO estimate for the weights of the Treynor–Black portfolio is a weighted average, where the weights depend on the updated belief regarding variables directly associated with the information ratio. This is consistent with the logic of maximizing the information ratio.

For the asymptotic results, we find that our MELO proposal has the same properties as the *plug-in* (ML) estimator.

Proposition 2 *Assuming that $\mathbf{g}(\theta)$ and $h(\theta)$ are continuous constant order functions having nonzero first order, then the density function, $f(\mathbf{R} | \theta)$, satisfies common assumptions of the maximum likelihood estimator [34], and $\pi(\theta)$ satisfies the Bernstein–von Mises theorem’s conditions [35] (see Assumptions in supplementary material, subsection 2.1 for details) then,*

1. $\hat{\omega}^* = \mathbf{g}(\hat{\theta}) + o(1)$.
2. $\sqrt{T}(\hat{\omega}^* - \mathbf{g}(\theta_0)) = \sqrt{T}(\mathbf{g}(\hat{\theta}) - \mathbf{g}(\theta_0)) + o_p(1)$.

where $\hat{\theta}$ and θ_0 are the maximum likelihood estimator and “true” parameter, respectively.

See supplementary material for a proof (subsection 2.3).

Consequently, $\frac{\mathbb{E}_{\pi(\Sigma | \mathbf{R})} \{(\mathbf{1}' \Sigma^{-1} \mathbf{1}) \Sigma^{-1} \mathbf{1}\}}{\mathbb{E}_{\pi(\Sigma | \mathbf{R})} \{(\mathbf{1}' \Sigma^{-1} \mathbf{1})^2\}} \xrightarrow{P} \frac{\hat{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}' \hat{\Sigma}^{-1} \mathbf{1}}$, $\frac{\mathbb{E}_{\pi(\Sigma, \mu | \mathbf{R})} \{(\mathbf{1}' \Sigma^{-1} \mu) \Sigma^{-1} \mu\}}{\mathbb{E}_{\pi(\Sigma, \mu | \mathbf{R})} \{(\mathbf{1}' \Sigma^{-1} \mu)^2\}} \xrightarrow{P} \frac{\hat{\Sigma}^{-1} \hat{\mu}}{\mathbf{1}' \hat{\Sigma}^{-1} \hat{\mu}}$ and $\frac{\mathbb{E}_{\pi(\mathbf{H}, \alpha | \mathbf{R})} \{(\mathbf{1}' \mathbf{H}^{-1} \alpha) \mathbf{H}^{-1} \alpha\}}{\mathbb{E}_{\pi(\mathbf{H}, \alpha | \mathbf{R})} \{(\mathbf{1}' \mathbf{H}^{-1} \alpha)^2\}} \xrightarrow{P} \frac{\hat{\mathbf{H}}^{-1} \hat{\alpha}}{\mathbf{1}' \hat{\mathbf{H}}^{-1} \hat{\alpha}}$ in the cases of the minimum variance portfolio, tangency portfolio estimator, and the Treynor–Black portfolio estimators. We find the same results for asymptotic distributions; that is, $\sqrt{T}(\hat{\omega}^* - \mathbf{g}(\theta_0)) \xrightarrow{d} \sqrt{T}(\mathbf{g}(\hat{\theta}) - \mathbf{g}(\theta_0))$. However, we should take into account that in the case of tangency portfolio, and probably Treynor–Black case, the moments of the exact distribution do not exist [7].

4 Simulation exercises

We conduct a synthetic data experiment to demonstrate the performance of our MELO proposal in finite samples. We set the population parameters for the excess of returns and

Table 1 Weights: global minimum variance

Method	Sample size	Min	1st Q	Median	Mean	3rd Q	Max
Stocks = 10							
Plug-in	120	0.0014	0.0060	0.0086	0.0093	0.0116	0.0303
Bayesian NI		0.0014	0.0060	0.0086	0.0093	0.0116	0.0303
Bayesian I		0.0015	0.0058	0.0084	0.0091	0.0113	0.0297
MELO NI		0.0012	0.0060	0.0085	0.0093	0.0115	0.0309
MELO I		0.0009	0.0057	0.0083	0.0092	0.0116	0.0319
Naive (1/N)		0.0066	0.0066	0.0066	0.0066	0.0066	0.0066
Plug-in	240	0.0005	0.0027	0.0039	0.0042	0.0052	0.0149
Bayesian NI		0.0005	0.0027	0.0039	0.0042	0.0052	0.0149
Bayesian I		0.0006	0.0027	0.0038	0.0041	0.0052	0.0147
MELO NI		0.0005	0.0028	0.0039	0.0042	0.0053	0.0150
MELO I		0.0004	0.0031	0.0044	0.0047	0.0059	0.0151
Naive (1/N)		0.0023	0.0023	0.0023	0.0023	0.0023	0.0023
Stocks = 50							
Plug-in	120	0.0068	0.0136	0.0168	0.0175	0.0206	0.0498
Bayesian NI		0.0068	0.0136	0.0168	0.0175	0.0206	0.0498
Bayesian I		0.0067	0.0130	0.0159	0.0167	0.0194	0.0448
MELO NI		0.0071	0.0136	0.0166	0.0174	0.0204	0.0465
MELO I		0.0070	0.0136	0.0165	0.0173	0.0201	0.0445
Naive (1/N)		0.0123	0.0123	0.0123	0.0123	0.0123	0.0123
Plug-in	240	0.0022	0.0043	0.0051	0.0059	0.0107	0.0052
Bayesian NI		0.0022	0.0043	0.0051	0.0059	0.0107	0.0052
Bayesian I		0.0022	0.0042	0.0050	0.0058	0.0105	0.0051
MELO NI		0.0022	0.0043	0.0051	0.0060	0.0105	0.0052
MELO I		0.0023	0.0045	0.0054	0.0062	0.0106	0.0055
Naive (1/N)		0.0056	0.0056	0.0056	0.0056	0.0056	0.0056

This table shows summary statistics for the squared error NI, non-informative priors; I, informative priors

covariance matrix. Then, we perform 1000 simulation exercises ($S = 1000$) using two sample sizes (120 and 240), and two portfolio sizes (10 and 50). We evaluate estimation errors of trading strategies (Sect. 2.1) using statistical strategies in Sect. 2.2 and MELO proposal (Sect. 3).⁶ In particular, we calculate squared errors,

$$SE_s = \sum_{i=1}^N (\hat{\omega}_i^s - \omega_i)^2, \quad i = 1, 2, \dots, N, \quad s = 1, 2, \dots, S.$$

where ω_i are optimal trading weights using population parameters and $\hat{\omega}_i^s$ are optimal trading weight estimates for each simulation using different statistical approaches.

Table 1 shows descriptive statistics of the squared error for the global minimum variance portfolios. As we can see, the results are almost the same for each methodology, except

⁶ We consider population parameters as hyperparameters in the case of informative Bayesian strategies, see subsection 2.4.1 in supplementary material for further details.

Table 2 Variance: global minimum variance

Method	Sample size	Min	1st Q	Median	Mean	3rd Q	Max
Stocks = 10; $\sigma_{120} = 0.3257$; $\sigma_{240} = 0.3212$							
Plug-in	120	0.3277	0.3343	0.3375	0.3386	0.3419	0.3671
Bayesian NI		0.3277	0.3343	0.3375	0.3386	0.3419	0.3671
Bayesian I		0.3276	0.3340	0.3372	0.3383	0.3414	0.3655
MELO NI		0.3275	0.3343	0.3375	0.3386	0.3419	0.3672
MELO I		0.3268	0.3341	0.3376	0.3385	0.3415	0.3704
Naive (1/N)		0.3349	0.3349	0.3349	0.3349	0.3349	0.3349
Plug-in	240	0.3220	0.3254	0.3270	0.3275	0.3290	0.3424
Bayesian NI		0.3220	0.3254	0.3270	0.3275	0.3290	0.3424
Bayesian I		0.3221	0.3253	0.3270	0.3274	0.3289	0.3421
MELO NI		0.3220	0.3254	0.3271	0.3275	0.3290	0.3425
MELO I		0.3219	0.3259	0.3278	0.3281	0.3300	0.3431
Naive (1/N)		0.3247	0.3247	0.3247	0.3247	0.3247	0.3247
Stocks = 50; $\sigma_{120} = 0.1225$; $\sigma_{240} = 0.1253$							
Plug-in	120	0.1388	0.1551	0.1610	0.1617	0.1672	0.2193
Bayesian NI		0.1388	0.1551	0.1610	0.1617	0.1672	0.2193
Bayesian I		0.1389	0.1537	0.1591	0.1599	0.1652	0.2131
MELO NI		0.1404	0.1547	0.1602	0.1615	0.1670	0.2163
MELO I		0.1405	0.1544	0.1597	0.1607	0.1656	0.2128
Naive (1/N)		0.1509	0.1509	0.1509	0.1509	0.1509	0.1509
Plug-in	240	0.1324	0.1387	0.1409	0.1433	0.1576	0.1412
Bayesian NI		0.1324	0.1387	0.1409	0.1433	0.1576	0.1412
Bayesian I		0.1324	0.1385	0.1406	0.1430	0.1561	0.1409
MELO NI		0.1325	0.1387	0.1409	0.1434	0.1572	0.1412
MELO I		0.1333	0.1395	0.1417	0.1439	0.1562	0.1419
Naive (1/N)		0.1422	0.1422	0.1422	0.1422	0.1422	0.1422

This table shows summary statistics for the squared error NI, non-informative priors; I, informative priors

naive weights, which gives the lowest mean squared errors. The fact that we got almost same results with most of the trading strategies is in agreement with the literature, given that the global minimum variance portfolio depends on only the covariance matrix and this does not introduce excessive estimation error [15]. Additionally, we cannot forget that the main concern of the global minimum variance portfolio is to minimize portfolio variance. Table 2 shows the descriptive statistics of portfolio variance of each methodology. We can observe that there are no meaningful differences between them.

The results change drastically when using the tangency portfolio due to including estimation of the expected return. We can see the outcomes of our simulation exercises in Table 3. In particular, mean squared error (MSE) and range of variability associated with MELO are lower than the other methodologies. As expected, the plug-in and non-informative Bayesian both obtain same results. They are also the worst estimators in these settings. Table 4 shows the Sharpe ratio for each methodology, where we observe that the two MELO methodologies

Table 3 Weights: tangency portfolio

Method	Sample size	Min	1st Q	Median	Mean	3rd Q	Max
Stocks = 10							
Plug-in	120	0.1637	0.8114	1.3279	2460.6289	9.1321	1,856,324.2834
Shrinkage		0.1458	0.8285	1.2193	1379.9003	4.7931	1,110,729.1815
Bayesian NI		0.1637	0.8114	1.3279	2460.6289	9.1321	1,856,324.2834
Bayesian I		0.0535	0.3037	0.4858	1086.8496	1.2657	927,980.8120
MELO NI		0.1549	0.4986	0.6033	0.6354	0.7250	2.7799
MELO I		0.0431	0.2516	0.3738	0.5009	0.5226	9.1157
Naive (1/N)		1.7360	1.7360	1.7360	1.7360	1.7360	1.7360
Plug-in	240	0.0255	0.3066	0.5035	2,268,339.9636	1.7101	2,266,316,679.0746
Shrinkage		0.0644	0.3095	0.4747	1,691,898.3817	0.9892	1,690,657,301.7466
Bayesian NI		0.0255	0.3066	0.5035	2,268,339.9635	1.7101	2,266,316,678.9866
Bayesian I		0.0141	0.0918	0.1443	5.8521	0.2631	5228.1172
MELO NI		0.0173	0.0896	0.1265	0.1331	0.1705	0.3964
MELO I		0.0110	0.0896	0.1388	0.2256	0.2239	10.6222
Naive (1/N)		1.1570	1.1570	1.1570	1.1570	1.1570	1.1570
Stocks = 50							
Plug-in	120	0.3312	0.7400	1.1142	417.4814	8.2190	12,3267.4803
Shrinkage		0.3857	0.7452	0.9432	216.2511	4.0306	62,733.5613
Bayesian NI		0.3312	0.7400	1.1142	417.4814	8.2190	123,267.4803
Bayesian I		0.2306	0.5214	0.7254	187.2023	2.5847	58,935.0646
MELO NI		0.2365	0.4686	0.5608	0.7322	0.6732	6.6383
MELO I		0.4169	0.6542	0.7339	0.8681	0.8622	3.9096
Naive (1/N)		1.5310	1.5310	1.5310	1.5310	1.5310	1.5310

Table 3 continued

Method	Sample size	Min	1st Q	Median	Mean	3rd Q	Max
Plug-in	240	0.1332	0.2592	0.3641	329.1857	1.1850	153,998.9689
Shrinkage		0.1307	0.2550	0.3199	219.7735	0.7016	101,804.3638
Bayesian NI		0.1332	0.2592	0.3641	329.1857	1.1850	153,998.9689
Bayesian I		0.0868	0.1567	0.2027	74.0546	0.4421	50,300.8170
MELO NI		0.0742	0.1256	0.1513	0.1654	0.1833	2.4186
MELO I		0.0690	0.1417	0.1679	0.1871	0.1998	2.6509
Naive (1/N)		0.7058	0.7058	0.7058	0.7058	0.7058	0.7058

This table shows summary statistics for the squared error NI, non-informative priors; I, informative priors

Table 4 Sharpe ratio: tangency portfolio

Method	Sample size	Min	1st Q	Median	Mean	3rd Q	Max
Stocks = 10; $SR_{120} = 0.3793$; $SR_{240} = 0.3867$							
Plug-in	120	-0.3530	0.2293	0.2860	0.1948	0.3177	0.3675
Shrinkage		-0.3526	0.2186	0.2791	0.1905	0.3134	0.3675
Bayesian NI		-0.3530	0.2293	0.2860	0.1948	0.3177	0.3675
Bayesian I		-0.3575	0.3405	0.3518	0.3363	0.3599	0.3760
MELO NI		-0.1854	0.3425	0.3507	0.3445	0.3576	0.3743
MELO I		-0.3321	0.3408	0.3524	0.3365	0.3601	0.3757
Naive (1/N)		0.1131	0.1131	0.1131	0.1131	0.1131	0.1131
Plug-in	240	-0.3578	0.3156	0.3375	0.2955	0.3539	0.3830
Shrinkage		-0.3576	0.3121	0.3355	0.2938	0.3530	0.3800
Bayesian NI		-0.3578	0.3156	0.3375	0.2955	0.3539	0.3830
Bayesian I		-0.3436	0.3676	0.3724	0.3707	0.3766	0.3853
MELO NI		0.3546	0.3741	0.3770	0.3764	0.3796	0.3856
MELO I		-0.3310	0.3675	0.3725	0.3707	0.3767	0.3851
Naive (1/N)		0.1167	0.1167	0.1167	0.1167	0.1167	0.1167
Stocks = 50; $SR_{120} = 1.2082$; $SR_{240} = 0.9470$							
Plug-in	120	-0.8998	0.6823	0.7548	0.5525	0.8133	0.9540
Shrinkage		-0.8995	0.6753	0.7488	0.5488	0.8069	0.9585
Bayesian NI		-0.8998	0.6823	0.7548	0.5525	0.8133	0.9540
Bayesian I		-0.9851	0.8528	0.8917	0.7860	0.9309	1.0331
MELO NI		-0.8826	0.8472	0.8897	0.8006	0.9271	1.0301
MELO I		-0.8143	0.7987	0.8527	0.7422	0.8933	1.0053
Naive (1/N)		0.1188	0.1188	0.1188	0.1188	0.1188	0.1188
Plug-in	240	-0.7954	0.7185	0.7464	0.7761	0.8526	0.6842
Shrinkage		-0.7946	0.7165	0.7450	0.7752	0.8505	0.6833
Bayesian NI		-0.7954	0.7185	0.7464	0.7761	0.8526	0.6842
Bayesian I		-0.8255	0.8144	0.8300	0.8454	0.8819	0.8150
MELO NI		-0.7597	0.8248	0.8405	0.8545	0.8851	0.8318
MELO I		-0.7527	0.8087	0.8257	0.8414	0.8847	0.8097
Naive (1/N)		0.1260	0.1260	0.1260	0.1260	0.1260	0.1260

This table shows summary statistics for the squared error NI, non-informative priors; I, informative priors

and the Bayesian with informative priors have the highest Sharpe ratios, whereas the naive weights get the worst performance on average.

Table 5 shows descriptive statistics of the squared error associated with Treynor–Black trading strategy. We observe that the plug-in approach has the highest MSE followed by the non-informative Bayesian. MELO informative presents the smallest MSE. Table 6 shows descriptive statistics of information ratios. Informative Bayesian and MELO, on average, have the best outcomes.

We also performed out of sample simulation exercises for tangency portfolio and Treynor–Black model taking 12 periods as the investment horizon and holding portfolios until the end of these periods. For this experiment, we consider the average of the portfolio returns in the out

Table 5 Weights: Treynor–Black

Method	Sample size	Min	1st Q	Median	Mean	3rd Q	Max
Stocks = 10							
Plug-in	120	0.3396	0.7902	1.6252	2365.4157	6.0915	1,061,714,0961
Bayesian NI		0.0020	0.0294	0.0751	58.7507	0.2215	38,031.8940
Bayesian I		0.0005	0.0051	0.0129	0.0320	0.0305	0.9191
MELO NI		0.0011	0.0195	0.0451	0.0758	0.0911	1.2610
MELO I		0.0003	0.0049	0.0108	0.0186	0.0232	0.2519
Naïve (1/N)		0.2802	0.2802	0.2802	0.2802	0.2802	0.2802
240							
Plug-in		0.2777	0.6662	1.2291	223,477,510.9376	4.1673	223,477,240,219.1140
Bayesian NI		0.0006	0.0156	0.0363	18.5595	0.1014	17,363,2625
Bayesian I		0.0001	0.0026	0.0061	0.0137	0.0148	0.4262
MELO NI		0.0010	0.0123	0.0251	0.0449	0.0511	1.6924
MELO I		0.0001	0.0024	0.0059	0.0103	0.0128	0.2220
Naïve (1/N)		0.2802	0.2802	0.2802	0.2802	0.2802	0.2802
Stocks = 50							
Plug-in	120	0.9634	1.2804	1.7577	179,2506	4.0986	81,425,0863
Bayesian NI		0.0048	0.0676	0.1573	0.2535	0.3377	2.3051
Bayesian I		0.0009	0.0049	0.0096	0.0171	0.0209	0.3626
MELO NI		0.0009	0.0068	0.0133	0.0236	0.0277	0.5404
MELO I		0.0010	0.0047	0.0096	0.0168	0.0201	0.3277
Naïve (1/N)		1.0380	1.0380	1.0380	1.0380	1.0380	1.0380
240							
Plug-in		0.9187	1.1249	1.3579	34,0615	2.3395	7694,2140
Bayesian NI		0.0045	0.1047	0.1741	0.2154	0.2822	1.5560
Bayesian I		0.0004	0.0017	0.0034	0.0061	0.0073	0.0561
MELO NI		0.0005	0.0023	0.0046	0.0082	0.0104	0.0667
MELO I		0.0003	0.0017	0.0034	0.0060	0.0075	0.0526
Naïve (1/N)		1.0380	1.0380	1.0380	1.0380	1.0380	1.0380

This table shows summary statistics for the squared error NI, non-informative priors; I, informative priors

Table 6 Information ratio: Treynor–Black

Method	Sample size	Min	1st Q	Median	Mean	3rd Q	Max
Stocks = 10; $IR = 0.3214$							
Plug-in	120	-0.0593	0.0095	0.0221	0.0226	0.0350	0.0964
Bayesian NI		-0.2187	0.1909	0.2275	0.2115	0.2550	0.3070
Bayesian I		0.2487	0.2886	0.2978	0.2957	0.3051	0.3171
MELO NI		-0.1155	0.2075	0.2401	0.2292	0.2648	0.3121
MELO I		0.2472	0.2901	0.2982	0.2967	0.3056	0.3182
Naive (1/N)		0.0490	0.0490	0.0490	0.0490	0.0490	0.0490
Plug-in	240	-0.0395	0.0157	0.0311	0.0296	0.0442	0.1041
Bayesian NI		-0.2279	0.2372	0.2620	0.2547	0.2812	0.3131
Bayesian I		0.2768	0.3036	0.3086	0.3077	0.3126	0.3211
MELO NI		-0.1117	0.2473	0.2691	0.2627	0.2854	0.3129
MELO I		0.2763	0.3042	0.3088	0.3080	0.3128	0.3212
Naive (1/N)		0.0490	0.0490	0.0490	0.0490	0.0490	0.0490
Stocks = 50; $IR = 1.9090$							
Plug-in	120	-0.0666	0.0197	0.0359	0.0315	0.0483	0.0893
Bayesian NI		0.7516	1.0297	1.0966	1.0919	1.1570	1.3577
Bayesian I		1.1797	1.4175	1.4733	1.4692	1.5271	1.6595
MELO NI		0.9856	1.2953	1.3668	1.3584	1.4282	1.6244
MELO I		1.1749	1.4151	1.4719	1.4677	1.5255	1.6808
Naive (1/N)		0.0882	0.0882	0.0882	0.0882	0.0882	0.0882
Plug-in	240	-0.0730	0.0322	0.0491	0.0442	0.0629	0.1034
Bayesian NI		1.1275	1.3191	1.3633	1.3619	1.4085	1.5466
Bayesian I		1.5599	1.6744	1.7015	1.6993	1.7290	1.7967
MELO NI		1.4489	1.6071	1.6418	1.6393	1.6767	1.7669
MELO I		1.5599	1.6720	1.7000	1.6977	1.7265	1.7950
Naive (1/N)		0.0882	0.0882	0.0882	0.0882	0.0882	0.0882

This table shows summary statistics for the squared error NI, non-informative priors; I, informative priors

of sample period as hyperparameter for the informative prior on the expected return, and the average of the difference between the return and benchmark returns as hyperparameter for the informative prior of the abnormal portfolio return in the Treynor–Black model. Information about the covariance matrix is not considered.⁷

We calculate the mean Sharpe ratio using 1,000 simulations for each of the sample periods. Figure 1 shows the results. The MELO and the Bayesian using informative priors have always the greater Sharpe ratios. However, the goodness of using informative priors depends on how well these priors are defined. For instance, in our experiments, we have the best possible priors that could be used due to using population parameters (in sample) and mean future returns (out of sample) as hyperparameters. Observe that the non-informative MELO is in third place when the sample size is 120, whereas this position is for the shrinkage estimator using 240 as sample size. Meanwhile, the non-informative Bayesian, which gives the same

⁷ The prior scale matrix is the sample covariance matrix.

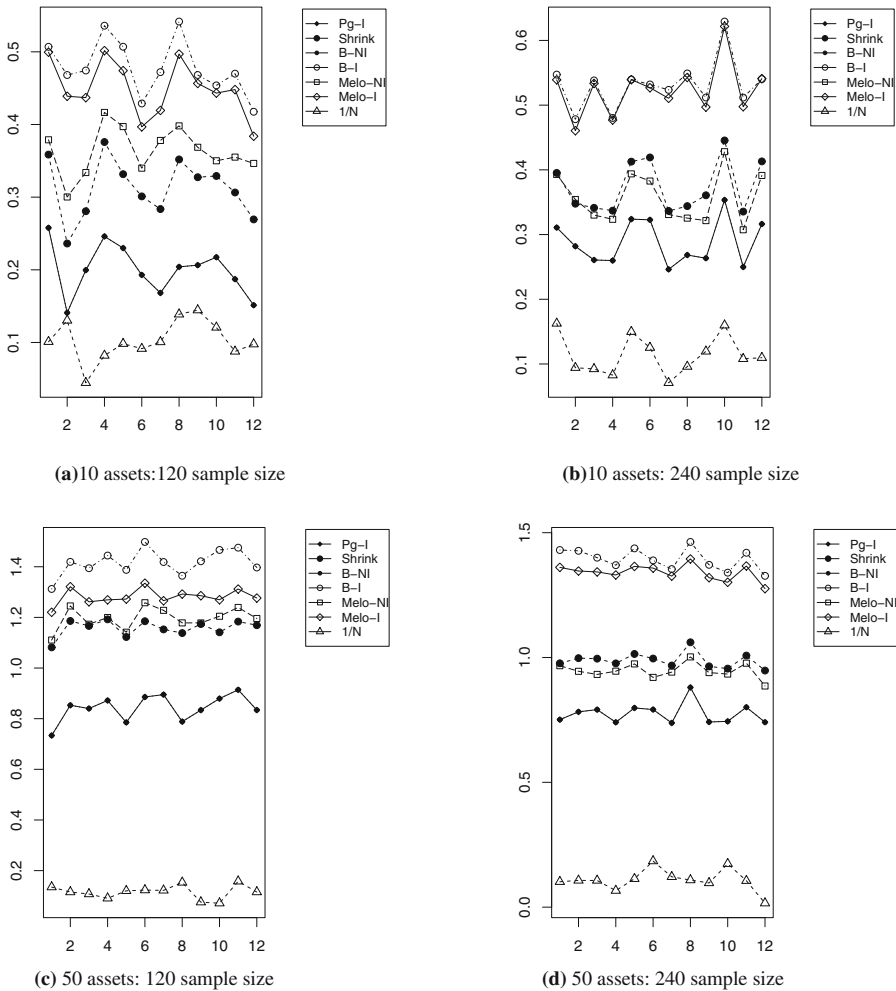


Fig. 1 Mean Sharpe ratio

results than the plug-in approach, obtains the second worst results, followed by the naive approach, which obtains the worst performance.

Figure 2 shows the mean of the information ratios using 1,000 simulations for each of the sample periods. We observe that MELO and Bayesian using informative priors have almost greater information ratios, except when using 50 assets with a sample size of 120—where only MELO using informative prior has the greatest information ratio. MELO and Bayesian using non-informative priors have almost same mean information ratios using 10 assets. MELO using non-informative priors has the second and third best information ratio when using 50 assets and a sample size of 120 and 240, respectively. Meanwhile, the plug-in and naives approaches obtain the worst results.

In subsection 2.4.2 in supplementary material, we show robustness checks regarding distributional assumptions of our previous results.

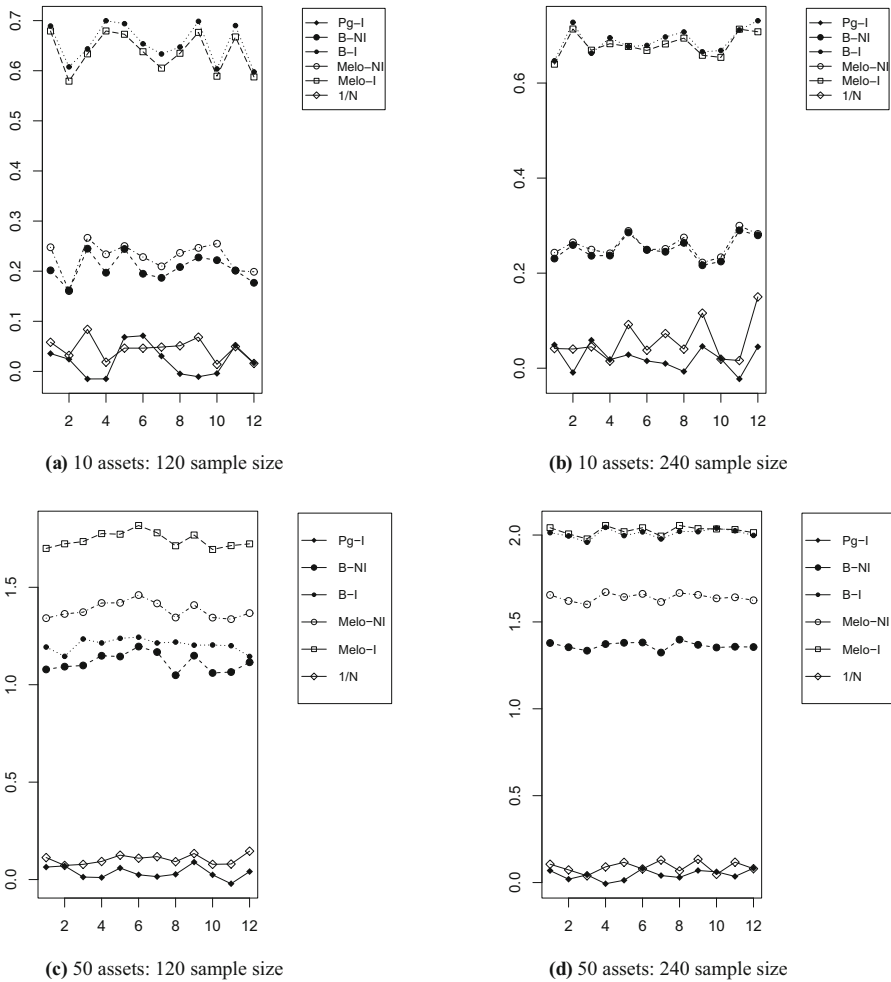


Fig. 2 Mean information ratio

5 Empirical study

We use weekly historical return of 21 MSCI international equity indices: Canada, United States, Austria, Belgium, Denmark, Finland, France, Germany, Israel, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, Australia, Japan, and Singapore. The index is adjusted by dividends and splits. We use weekly closing prices from June 2009 to June 2017.⁸ We calculate the excess of returns with respect to the interest rate of the 3-month US treasury bill. We get 417 weekly excess of returns.

⁸ The sample period is selected after the 2008 financial crisis to avoid introducing this abnormal period.

Table 7 Summary statistics: global minimum variance application

	Min (%)	1st Q (%)	Median (%)	Mean (%)	3rd Q (%)	Max (%)
5 stocks						
Plug-in	2.47	9.04	11.78	12.22	15.41	26.03
Bayesian NI	1.37	6.78	10.68	10.82	14.59	21.92
Bayesian I	0.55	5.41	8.49	9.11	11.85	23.01
MELO NI	27.40	41.85	46.03	45.50	49.04	56.71
MELO I	5.48	13.15	17.26	18.18	22.47	38.36
Naive (1/N)	0.822	2.740	4.110	4.170	5.479	9.315
10 stocks						
Plug-in	1.92	7.67	9.86	10.43	12.88	22.74
Bayesian NI	2.74	7.12	9.59	9.59	11.85	19.73
Bayesian I	0.00	7.05	9.04	9.10	11.03	20.82
MELO NI	36.71	44.38	46.99	47.09	49.59	56.16
MELO I	8.22	15.34	18.22	18.52	21.71	29.86
Naive (1/N)	1.10	4.11	5.21	5.28	6.30	10.68
15 stocks						
Plug-in	3.01	6.58	8.90	9.24	11.23	20.55
Bayesian NI	3.29	7.95	9.59	9.87	11.78	19.73
Bayesian I	3.56	9.59	11.37	11.42	13.97	18.63
MELO NI	41.37	44.86	46.85	47.05	49.11	55.34
MELO I	6.85	14.18	16.30	16.77	19.25	26.85
Naive (1/N)	2.47	4.38	5.48	5.65	6.64	11.23

This table shows the summary statistic of the percentage of the out of sample period that each methodology get the highest return

We use a one year band width rolling window to estimate all trading and statistical strategies. We re-balance trading strategies every three months. The first portfolio is set in June 2010, and is held constant up to September 2010. Then, out-of-sample returns are calculated during this period. The second estimation is done in September 2010 and held constant up to December 2010, and the out-of-sample returns are calculated during this period, and so on.⁹ Therefore, we obtain 365 (417-52) out of sample returns for each strategy. Then, we calculated the number of times that each strategy gets the highest out of sample return. Consequently, their relative frequencies are the most profitable. We repeat this process 100 times. Therefore we have 100 sets of relative frequencies counting the most profitable strategy. In each iteration, we randomly draw equity indices to have three portfolio sizes (5, 10, and 15 stocks).

⁹ In the cases of methodologies that use informative priors, we used the target prices given by Bloomberg to calculate the “target returns”, and these target returns were used as hyperparameters. We set hyperparameters such that sample and prior information have the same weight in posterior inference. This means that $\tau = \tau_\alpha = 52$, given 52 weeks in a year.

Table 8 Summary statistics: tangency portfolio application

	Min (%)	1st Q (%)	Median (%)	Mean (%)	3rd Q (%)	Max (%)
5 stocks						
Plug-in	6.58	12.26	15.21	16.24	20.27	34.52
Shrinkage	3.01	10.62	14.52	15.00	18.63	28.77
Bayesian NI	4.93	10.96	14.11	14.29	17.05	27.40
Bayesian I	5.21	18.36	22.47	22.77	26.92	48.49
MELO NI	12.60	21.58	24.93	24.85	28.49	37.53
MELO I	0.00	2.19	3.84	4.26	5.75	17.53
Naive (1/N)	0.55	1.64	2.19	2.59	3.56	6.85
10 stocks						
Plug-in	4.93	12.33	14.79	14.82	17.60	23.56
Shrinkage	6.58	12.33	14.66	14.76	16.99	24.11
Bayesian NI	2.47	9.79	12.74	13.59	16.78	28.49
Bayesian I	12.60	20.82	23.70	24.22	27.74	35.34
MELO NI	13.15	23.29	25.89	26.04	29.59	35.34
MELO I	0.00	1.64	2.88	3.43	4.66	14.52
Naive (1/N)	0.55	2.19	3.01	3.15	3.90	7.12
15 stocks						
Plug-in	4.66	11.64	13.97	13.95	16.51	24.11
Shrinkage	7.67	15.62	18.08	17.70	19.73	26.58
Bayesian NI	3.29	9.52	12.47	12.27	14.86	23.01
Bayesian I	10.41	17.81	21.51	21.77	24.66	34.25
MELO NI	18.90	25.14	28.22	27.73	29.93	37.53
MELO I	0.00	1.85	2.74	3.04	4.11	7.40
Naive (1/N)	0.82	2.40	3.29	3.54	4.38	9.59

This table shows the summary statistic of the percentage of the out of sample period that each methodology get the highest return

We can see in Table 7 that, on average, the non-informative MELO got the highest out of sample return in 45.50%, 47.09%, and 47.05% of times using portfolio sizes equal to 5, 10, and 15 assets, respectively. On the other hand, the naive weights got on average the worst out of sample performance.

Table 8 shows the results for the tangency portfolio. We can observe that, on average, non-informative MELO got the highest out of sample return in 24.85%, 26.04%, and 27.73% of times using 5, 10, and 15 assets, respectively. On the other hand, the naive weights got the worst performance on average with 5 and 10 stocks, and the informative MELO was the worst using 15 stocks.

In the Treynor–Black empirical study (see Table 9), the naive approach got the best out of sample performance, 22.59%, 24.39% and 27.48% of times using 5, 10, and 15 assets, respectively. The non-informative Bayesian got the second best performance using 5 stocks, and the non-informative MELO got this position using 10 and 15 stocks. The informative Bayesian got worst out of sample performance on average.

Table 9 Summary statistics: Treynor–Black application

	Min (%)	1st Q (%)	Median (%)	Mean (%)	3rd Q (%)	Max (%)
5 stocks						
Plug-in	3.29	10.34	13.15	13.01	15.62	28.22
Bayesian NI	7.95	15.07	20.96	20.80	25.27	34.52
Bayesian I	2.19	8.42	11.78	11.54	13.97	24.93
MELO NI	6.03	13.15	17.67	18.36	22.47	35.89
MELO I	2.74	8.77	12.60	13.70	17.33	35.62
Naive (1/N)	7.95	17.26	23.29	22.59	27.40	41.64
10 stocks						
Plug-in	6.03	13.42	16.16	16.45	19.18	27.95
Bayesian NI	5.48	11.78	17.26	17.10	21.23	32.60
Bayesian I	3.01	6.85	10.14	10.03	12.05	21.37
MELO NI	6.30	14.18	19.18	19.01	23.63	33.42
MELO I	3.56	10.00	12.88	13.01	16.23	25.48
Naive (1/N)	10.14	19.73	24.38	24.39	28.22	41.92
15 stocks						
Plug-in	9.32	15.82	18.36	18.47	21.16	27.12
Bayesian NI	2.74	10.62	14.11	14.37	18.36	25.21
Bayesian I	2.19	5.48	8.49	8.60	10.96	19.18
MELO NI	2.74	13.63	19.73	18.98	23.56	31.78
MELO I	1.64	8.97	11.78	12.09	14.38	27.12
Naive (1/N)	7.40	23.84	27.12	27.48	30.68	49.86

This table shows the summary statistic of the percentage of the out of sample period that each methodology get the highest return

6 Conclusions

In this paper, we proposed a decision theory framework to mitigate estimation risk. Our proposal has the same statistical properties as the delta method (maximum likelihood) estimator. However, it seems from our simulation exercises that our non-informative MELO proposal has better finite properties than competing alternatives. The degree of estimation improvement depends on the trading strategy. In particular, it seems that tangency can be better estimated using our approach. The non-informative MELO is the most realistic scenario in our simulation exercises, showing less degree of estimation variability and the lowest error. Our results are robust to heavy and serially correlated error distributions.

It seems from our empirical study that the non-informative MELO is the best statistical strategy when global minimum variance or tangency portfolios are used as the trading strategy. Meanwhile, the naive approach is the best in the case of Treynor–Black trading strategy. However, the naive weights have the worst performance in the other trading strategies. It seems that the non-informative MELO is robust to these three trading strategies, and the implicit data generating process of the returns [Student's t and autoregressive process, AR(1)].

We should note at this point that real world applications are surrounded by a lot of noise, which invalidates many of the implicit assumptions in portfolio selection methodologies at financial and statistical level. Consequently, our recommendation is to implement all of these

methodologies, then identify which generates the best outcomes in a cross-validation dataset, and finally make decisions based on these results. Therefore, we have developed a graphical user interface that helps to apply traditional approaches, as well as our proposal, which can be download at <https://besmarter-team.shinyapps.io/meloportfolio/>. See supplementary material, section 3.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Michaud, R.: The markowitz optimization enigma: is ‘optimized’ optimal? *Financ. Anal. J.* **45**(1), 31 (1989). <https://doi.org/10.2469/faj.v45.n1.31>
2. Barry, C.: Portfolio analysis under uncertain means, variances, and covariances. *J. Finance* **29**(2), 515 (1974). <https://doi.org/10.2307/2329297>
3. Klein, R.W., Bawa, V.S.: The effect of estimation risk on optimal portfolio choice. *J. Financ. Econ.* **3**(3), 215 (1976). [https://doi.org/10.1016/0304-405X\(76\)90004-0](https://doi.org/10.1016/0304-405X(76)90004-0)
4. Merton, R.C.: On estimating the expected return on the market: an exploratory investigation. *J. Financ. Econ.* **8**(4), 323 (1980). [https://doi.org/10.1016/0304-405X\(80\)90007-0](https://doi.org/10.1016/0304-405X(80)90007-0)
5. Chopra, V.K.: Sampling–resampling techniques for the computation of posterior. Densities in normal means problems. *J. Invest.* **2**(3), 51 (1993). <https://doi.org/10.3905/joi.2009.18.3.012>
6. Kan, R., Zhou, G.: Optimal portfolio choice with parameter uncertainty. *J. Financ. Quant. Anal.* **42**(3), 621 (2007). <https://doi.org/10.1017/S0022109000004129>
7. Okhrin, Y., Schmid, W.: Distributional properties of portfolio weights. *J. Econom.* **134**(1), 235 (2006). <https://doi.org/10.1016/j.jeconom.2005.06.022>
8. Zellner, A.: Estimation of functions of population means and regression coefficients including structural coefficients: a minimum expected loss (MELO) approach. *J. Econom.* **44**, 127 (1978). [https://doi.org/10.1016/0304-4076\(78\)90024-6](https://doi.org/10.1016/0304-4076(78)90024-6)
9. Zellner, A., Park, S.B.: Minimum expected loss (MELO) estimators for functions of parameters and structural coefficients of econometric models. *J. Am. Stat. Assoc.* **74**(365), 185 (1979). <https://doi.org/10.1080/01621459.1979.10481636>
10. Zellner, A., Park, S.B.: A note on the relationship of minimum expected loss (MELO) and other structural coefficient estimates. *Rev. Econ. Stat.* **62**(3), 482 (1980)
11. Park, Sb: Some sampling properties of minimum expected loss (MELO) estimators of structural coefficients. *J. Econom.* **18**, 295 (1982)
12. Swamy, P.A.V.B., Mehta, J.S.: Further results on Zellner’s minimum expected loss and full information maximum likelihood estimators for undersized samples. *J. Bus. Econ. Stat.* **1**(2), 154 (1983)
13. Zellner, A.: The finite sample properties of simultaneous equations’ estimates and estimators Bayesian and non-Bayesian approaches. *J. Econom.* **6**, 185 (1998)
14. Ramirez-Hassan, A., Correa-Giraldo, M.: Focused econometric estimation for noisy and small datasets: A Bayesian minimum expected loss estimator approach. *Aust. N. Z. J. Stat.* (**in press**)
15. Chopra, V.K., Ziemba, W.T.: The effect of errors in means, variances, and covariances on optimal portfolio choice. *J. Portf. Manag.* **19**(2), 6 (1993). <https://doi.org/10.3905/jpm.1993.409440>
16. Brown, S.: The effect of estimation risk on capital market equilibrium. *J. Financ. Quant. Anal.* **14**(2), 215 (1979). <https://doi.org/10.2307/2330499>
17. Frost, P.A., Savarino, J.E.: An empirical Bayes approach to efficient portfolio selection. *J. Financ. Quant. Anal.* **21**(3), 293 (1986). <https://doi.org/10.2307/2331043>
18. Stein, C.: Inadmissibility of usual estimator for the mean of multivariate normal distributions. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 4, p. 197 (1955)
19. Jorion, P.: Bayes Stein estimation for portfolio analysis. *J. Financ. Quant. Anal.* **21**(3), 279 (1986). <https://doi.org/10.2307/2331042>

20. Ledoit, O., Wolf, M.: Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empir. Finance* **10**(5), 603 (2003). [https://doi.org/10.1016/S0927-5398\(03\)00007-0](https://doi.org/10.1016/S0927-5398(03)00007-0)
21. Ledoit, O., Wolf, M.: Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Rev. Financ. Stud.* **30**(12), 4349 (2017)
22. Pástor, U.: Portfolio selection and asset pricing models. *J. Finance* **55**(1), 179 (2000)
23. Pástor, U., Stambaugh, R.F.: Comparing asset pricing models: an investment perspective. *J. Financ. Econ.* **56**(3), 335 (2000). [https://doi.org/10.1016/S0304-405X\(00\)00044-1](https://doi.org/10.1016/S0304-405X(00)00044-1)
24. Frahm, G., Memmel, C.: Dominating estimators for minimum-variance portfolios. *J. Econom.* **159**(2), 289 (2010). <https://doi.org/10.1016/j.jeconom.2010.07.007>
25. DeMiguel, V., Garlappi, L., Uppal, R.: Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? *Rev. Financ. Stud.* **22**(5), 1915 (2009). <https://doi.org/10.1093/rfs/hhm075>
26. Kritzman, M., Page, S., Turkington, D.: In defense of optimization: the fallacy of 1/N. *Financ. Anal. J.* **66**(2), 31 (2010). <https://doi.org/10.2469/faj.v66.n2.6>
27. Tu, J., Zhou, G.: Markowitz meets Talmud: a combination of sophisticated and naive diversification strategies. *J. Financ. Econ.* **99**(1), 204 (2011). <https://doi.org/10.1016/j.jfineco.2010.08.013>
28. Mustafa Ç, P.: On robust mean–variance portfolios. *Optimization* **65**(5), 1039 (2016). <https://doi.org/10.1080/02331934.2015.1132216>
29. Olivares-Nadal, A.V., DeMiguel, V.: Technical note a robust perspective on transaction costs in portfolio optimization. *Invest. Manag. Financ. Innov.* **66**(3), 733 (2018). <https://doi.org/10.1287/opre.2017.1699>
30. Li, J.: Sparse and stable portfolio selection with parameter uncertainty. *J. Bus. Econ. Stat.* **33**(3), 381 (2015). <https://doi.org/10.1080/07350015.2014.954708>
31. Klimenka, F., Wolter, J.: Multiple regression model averaging and the focused information criterion with an application to portfolio choice. *J. Bus. Econ. Stat.* (2018). <https://doi.org/10.1080/07350015.2017.1383262>
32. Claeskens, G., Hjort, N.L.: The focused information criterion. *J. Am. Stat. Assoc.* **98**(464), 900 (2003). <https://doi.org/10.1198/016214503000000819>
33. Treynor, J., Black, F.: How to use security analysis to improve portfolio selection. *J. Bus.* **46**(1), 66 (1973)
34. Lehmann, E., Casella, G.: *Theory of Point Estimation*, 2nd edn. Springer, Berlin (2003)
35. Bickel, P.J., Yahav, J.A.: Some contributions to the asymptotic theory of Bayes solutions. *Z. Wahrsch. Verw. Geb.* **11**, 257 (1969)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.