IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 June 2022, accepted 1 August 2022, date of publication 10 August 2022, date of current version 23 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3197896

RESEARCH ARTICLE

Portfolio Optimization Using a Consistent Vector-Based MSE Estimation Approach

MAAZ MAHADI[1,2], (Student Member, IEEE), TARIG BALLAL[3], (Member, IEEE), MUHAMMAD MOINUDDIN[1,2], (Member, IEEE), AND UBAID M. AL-SAGGAF[1,2], (Member, IEEE)

[1]Center of Excellence in Intelligent Engineering Systems (CEIES), King Abdulaziz University, Jeddah 21589, Saudi Arabia
[2]Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[3]Division of Computer, Electrical and Mathematical Sciences and Engineering (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia

Corresponding author: Maaz Mahadi (mmahadi@stu.kau.edu.sa)

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University (KAU), Jeddah, Saudi Arabia, under Grant KEP-PhD: 82-135-1443.

ABSTRACT This paper is concerned with optimizing the weights of the global minimum-variance portfolio (GMVP) in high-dimensional settings where both observation and population dimensions grow at a bounded ratio. Optimizing the GMVP weights is highly influenced by the data covariance matrix estimation. In a high-dimensional setting, it is well known that the sample covariance matrix is not a proper estimator of the true covariance matrix since it is not invertible when we have fewer observations than the data dimension. Even with more observations, the sample covariance matrix may not be well-conditioned. This paper determines the GMVP weights based on a regularized covariance matrix estimator to overcome the abovementioned difficulties. Unlike other methods, the proper selection of the regularization parameter is achieved by minimizing the mean squared error of an estimate of the noise vector that accounts for the uncertainty in the data mean estimation. Using random-matrix-theory tools, we derive a consistent estimator of the achievable mean squared error that allows us to find the optimal regularization parameter using a simple line search. Simulation results demonstrate the effectiveness of the proposed method when the data dimension is larger than, or of the same order of, the number of data samples.

INDEX TERMS Portfolio optimization, global minimum-variance portfolio, GMVP, random matrix theory, RMT, consistent estimator.

I. INTRODUCTION

Decision-making regarding investment in the stock market has become increasingly more complex because of the dynamic nature of the stocks available to investors and the advent of new unconventional and risky options [1]. Throughout the years, the portfolio optimization problem has attracted the attention of many signal-processing researchers due to its close relationship to the field. The portfolio optimization problem aims at achieving the maximum possible returns with the least volatility percentage [2]. The Economist, Harry Markowitz, introduced the modern portfolio theory, or mean-variance analysis (MVP), in [3]. Other portfolios, such as the global MVP and the maximum sharp ratio portfolio (MSRP) have been proposed as improvements to the MVP. Portfolio optimization utilizes the available financial data to reach conclusions regarding the allocation of wealth to each of the available stocks. The most important measurement in portfolio optimization is the data covariance matrix (CM).

CM estimation in the classical signal processing framework relies on asymptotic statistics of some observations, n, which is assumed to grow largely compared to the population dimension, p, i.e., $n/p \to \infty$ as $n \to \infty$ [4]. However, many practical applications, such as finance, bioinformatics and data classification, require an estimate of the CM when the data dimension is large compared to the sample size [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Tingwen Huang.

In such cases, it is well known that the default estimator, i.e., the empirical *sample covariance matrix* (SCM), is usually ill-conditioned, leading to poor performance.

In the case of $p > n$, the SCM is not invertible; whereas for $p < n$, the SCM is invertible but might be ill-conditioned, which substantially increases estimation errors. In other words, for a large p, it is not practically guaranteed that the number of observations is sufficient to develop a well-conditioned CM estimator [6]. Such scenarios have motivated researchers to look into estimation problems in the high-dimensional regime [4].

In scenarios with limited data, a regularized SCM (RSCM) estimator of the following general form is widely used [5]:

$$\widehat{\mathbf{\Sigma}}_{\beta,\gamma} = \beta\widehat{\mathbf{\Sigma}} + \gamma\mathbf{I}_p, \tag{1}$$

where $\widehat{\mathbf{\Sigma}}$ is the SCM defined in (7) (further ahead), β, $\gamma \in \mathbb{R}^+$ are the regularization, or *shrinkage*, parameters. These parameters can be determined based on minimizing the *mean squared error* (MSE), which results in oracle shrinkage parameters, β_o and γ_o, as follows [5], [6]:

$$(\beta_o, \gamma_o) = \underset{\beta,\gamma>0}{\arg\min}\, \mathbb{E}\left[\left\|\widehat{\mathbf{\Sigma}}_{\beta,\gamma} - \mathbf{\Sigma}\right\|_F^2\right], \tag{2}$$

where $\|.\|_F$ denotes the Frobenius matrix norm. The estimation of (β_o, γ_o) based on (2) depends on the true CM, $\mathbf{\Sigma}$. To circumvent this issue, Ledoit and Wolf [6] proposed a distribution-free consistent estimator of (β_o, γ_o) in high-dimensional settings. The work in [5] assumes that the observations are from unspecified elliptically symmetric distribution. The consistent estimator proposed in [7] uses a hybrid CM estimator based on the Taylor's M-estimator and Ledoit-Wolf shrinkage estimator, which suits a *global minimum variance portfolio* (GMVP) influenced by outliers. A similar approach based on the M-estimator is proposed in [8], considering $n > p$ with fully automated selection of the shrinkage parameters. The minimum variance portfolio estimator in [9] is based on certain sparsity assumptions imposed on the inverse of the CM. The work presented in [10] proposes a different RSCM estimator by manipulating the expression of the GMVP weights.

In this paper, we propose a single-parameter CM estimator. Instead of minimizing the MSE, as in (2), we minimize the MSE of the estimation of the sample noise vector. We utilize RMT tools to obtain a consistent estimator of this MSE. The value of the regularization parameter γ is selected as the one that minimizes the estimated MSE. By choosing to minimize the MSE of the noise vector's estimation, we implicitly consider the inaccuracy of estimating the true mean.

A. CONTRIBUTIONS OF THE PAPER

The contributions of this paper can be summarized as follows:

- We propose a regularized sample covariance matrix estimator of the covariance matrix for the portfolio optimization problem based on estimating a noise vector that accounts for the uncertainty in estimating the true mean.
- Under the assumption of the double asymptotic regime, we derive the asymptotic performance of the MSE of the estimated vector.
- We derive a general consistent estimator of the MSE.
- We utilize the derived consistent MSE estimator to optimally tune the regularization parameter associated with the regularized sample covariance matrix estimator. The application of the proposed estimator results in our VB-MSE estimator.

B. NOTATIONS

Uppercase boldface letters denote matrices, while lowercase boldface letters denote vectors. Scalars are denoted by normal lowercase letters. The superscript notation, $(.)^T$ denotes the transpose of a matrix or a vector, while $\mathbb{E}(.)$ denotes the expectation operator, and tr[.] is the trace of a matrix. $\mathbb{R}$, $\mathbb{R}^+$, and $\mathbb{C}$, respectively, denote real, positive-real, and complex fields of dimension specified by a superscript. The variable z denotes a complex variable. The notation $a \asymp b$ denotes that a and b are asymptotically equivalent, i.e., $|a-b| \xrightarrow{\text{a.s.}} 0$, where a.s. denotes almost-sure convergence. The l_2 norm (of a vector), or the 2-induced norm (of a matrix) is denoted by $\|.\|_2$, and the identity matrix of dimension n is denoted by $\mathbf{I}_n$.

II. GLOBAL MINIMUM VARIANCE PORTFOLIO

We consider a time series comprising $\mathbf{y}_1, \mathbf{y}_2 \cdots, \mathbf{y}_L$ logarithmic returns of p financial assets over a certain investment period. We assume that the elements of $\mathbf{y}_t$, $(t = 1, 2, \cdots, L)$ are independent and identically distributed (i.i.d.) and are generated according to the following stochastic model [11]:

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{\Sigma}_t^{\frac{1}{2}}\mathbf{x}_t, \tag{3}$$

where $\boldsymbol{\mu}_t \in \mathbb{R}^{n\times 1}$ and $\mathbf{\Sigma}_t \in \mathbb{R}^{p\times p}$ are the mean and the CM of the asset returns over the investment period, and the only information available about $\mathbf{x}_t$ is that it is an i.i.d. random noise vector of zero mean and identity CM. For simplicity, we drop the subscript t from $\boldsymbol{\mu}_t$ and $\mathbf{\Sigma}_t$. For the investment period of interest, we define $\mathbf{w} \in \mathbb{R}^p$ as the asset holdings vector, also known as the weight vector. The GMVP optimally minimizes the portfolio variance under single-period investment horizon, such that the weight vector is normalized by the outstanding wealth [11], i.e.,

$$\min_{\mathbf{w}\in\mathbb{R}^p} \mathbf{w}^T\mathbf{\Sigma}\mathbf{w} \quad \text{subject to } \mathbb{1}_p^T\mathbf{w} = 1, \tag{4}$$

where $\mathbb{1}_p$ is a column vector of p 1's. The solution of (4) can be obtained by using the Lagrange-multipliers method, which results in the optimum weights [7]:

$$\mathbf{w}_{\text{GMVP}} = \frac{\mathbf{\Sigma}^{-1}\mathbb{1}_p}{\mathbb{1}_p^T\mathbf{\Sigma}^{-1}\mathbb{1}_p}. \tag{5}$$

The CM in (5) is unknown and should be estimated. As stated earlier, the SCM estimate does not perform well because it is usually ill-conditioned; hence, we apply the RSCM estimator

and (5) becomes

$$\widehat{\mathbf{w}}_{\text{GMVP}} = \frac{\widehat{\mathbf{\Sigma}}_{\text{RSCM}}^{-1}\mathbb{1}_p}{\mathbb{1}_p^T\widehat{\mathbf{\Sigma}}_{\text{RSCM}}^{-1}\mathbb{1}_p}, \tag{6}$$

where $\widehat{\mathbf{\Sigma}}_{\text{RSCM}}$ is the RSCM which can take the form of (1), for example. In the following section, we develop a RSCM estimator method and properly set the value of its regularization parameter.

III. THE PROPOSED CONSISTENT VECTOR-BASED MSE ESTIMATOR

The SCM, $\widehat{\mathbf{\Sigma}}$, and the sample mean, $\widehat{\boldsymbol{\mu}}$, can be estimated from the n past return observations as follows:

$$\widehat{\mathbf{\Sigma}} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{y}_{t-j} - \widehat{\boldsymbol{\mu}})(\mathbf{y}_{t-j} - \widehat{\boldsymbol{\mu}})^T, \tag{7}$$

$$\widehat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{y}_{t-j}. \tag{8}$$

We notice that computing $\widehat{\mathbf{\Sigma}}$ using (7) involves evaluating the sample mean, not the true mean. This can worsen performance, especially for a small number of observations. Subtracting $\widehat{\boldsymbol{\mu}}$ from both sides of (3), we obtain

$$\mathbf{y}_t - \widehat{\boldsymbol{\mu}} = (\widehat{\mathbf{\Sigma}}^{\frac{1}{2}} + \mathbf{\Delta})\mathbf{x}_t + \boldsymbol{\delta}, \tag{9}$$

where $\boldsymbol{\delta} \triangleq \boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}$ and $\mathbf{\Delta} = \mathbf{\Sigma}^{\frac{1}{2}} - \widehat{\mathbf{\Sigma}}^{\frac{1}{2}}$. Eq. (9) can be viewed as a linear model with bounded uncertainties in both $\mathbf{\Sigma}^{\frac{1}{2}}$ and $\boldsymbol{\mu}$ [12]. We seek an estimate, $\widehat{\mathbf{x}}_t$ that performs well for any allowed perturbation $(\mathbf{\Delta}, \boldsymbol{\delta})$ by formulating the following min-max problem [12]:

$$\min_{\widehat{\mathbf{x}}_t}\max_{\mathbf{\Delta},\boldsymbol{\delta}}\left[\|(\widehat{\mathbf{\Sigma}}^{\frac{1}{2}} + \mathbf{\Delta})\widehat{\mathbf{x}}_t - (\mathbf{y}_t - \widehat{\boldsymbol{\mu}} - \boldsymbol{\delta})\|_2\right]. \tag{10}$$

A unique solution can exist which takes the form [12]

$$\widehat{\mathbf{x}}_t = (\widehat{\mathbf{\Sigma}} + \gamma\mathbf{I}_p)^{-1}\widehat{\mathbf{\Sigma}}^{\frac{1}{2}}(\mathbf{y}_t - \widehat{\boldsymbol{\mu}}). \tag{11}$$

$\widehat{\mathbf{x}}_t$ is a function of γ, which when properly set leads to the best estimate of $\mathbf{x}_t$. It is easy to recognize that $(\widehat{\mathbf{\Sigma}} + \gamma\mathbf{I}_p)^{-1}$ can be used as an estimator of the CM inverse, i.e., $\widehat{\mathbf{\Sigma}}_{\text{RSCM}}^{-1} = \widehat{\mathbf{\Sigma}}_{\gamma}^{-1} = (\widehat{\mathbf{\Sigma}} + \gamma\mathbf{I}_p)^{-1}$. Such estimator is widely used in the literature, e.g., [13], [14], [15], [16], [17], [18], [19], [20], [21]; to name a few. The optimal value of γ that estimates $\widehat{\mathbf{\Sigma}}_{\gamma}^{-1}$ is the one that minimizes the MSE for estimating $\mathbf{x}_t$. That is

$$\text{MSE}(\gamma) = \mathbb{E}\left[\|\mathbf{x}_t - \widehat{\mathbf{x}}_t\|_2^2\right] \tag{12}$$

$$= \mathbb{E}\left[\|\mathbf{x}_t - \widehat{\mathbf{\Sigma}}_{\gamma}^{-1}\widehat{\mathbf{\Sigma}}^{\frac{1}{2}}(\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{x}_t + \boldsymbol{\delta})\|_2^2\right]. \tag{13}$$

We choose the optimal γ_{o} as follows:

$$\gamma_{\text{o}} = \arg\min \text{MSE}(\gamma). \tag{14}$$

The choice of minimizing the MSE is reasonable because, under certain conditions, the minimization problem in (4) and the minimum MSE are equivalent [22], [23]. Unlike the other methods, it is remarkable that the uncertainty in estimating the mean is taken into account in (13). We expect the effect of the uncertainty in the mean estimation to be high when we have a limited number of observations. Also, unlike the methods that are based on (2), when we search for the optimal γ that minimizes (13), we actually estimate the inverse of the CM rather than estimating the CM itself. This is important because we use it in (6). We obtain the following normalized (by n) expression of the MSE (see Appendix B):

$$\text{MSE}(\gamma) = \frac{p}{n} + \frac{n+1}{n}\mathbb{E}\Big[\underbrace{\frac{1}{n}\text{tr}\big[\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}\big(\widehat{\mathbf{\Sigma}} + \gamma\mathbf{I}_p\big)^{-2}\big]}_{A(\gamma)}\Big]$$
$$-2\mathbb{E}\Big[\underbrace{\frac{1}{n}\text{tr}\big[\mathbf{\Sigma}^{\frac{1}{2}}\widehat{\mathbf{\Sigma}}^{\frac{1}{2}}\big(\widehat{\mathbf{\Sigma}} + \gamma\mathbf{I}_p\big)^{-1}\big]}_{B(\gamma)}\Big]. \tag{15}$$

We observe that (15) is expressed in terms of the unknown quantity, $\mathbf{\Sigma}$. In this case, using a direct plugin formula, i.e., substituting $\mathbf{\Sigma}$ with $\widehat{\mathbf{\Sigma}}$ results in

$$\widehat{\text{MSE}}_{\text{plugin}}(\gamma) = \frac{p}{n} + \frac{n+1}{n}\mathbb{E}\Big[\frac{1}{n}\text{tr}\big[\widehat{\mathbf{\Sigma}}^2\big(\widehat{\mathbf{\Sigma}} + \gamma\mathbf{I}_p\big)^{-2}\big]\Big]$$
$$-2\mathbb{E}\Big[\frac{1}{n}\text{tr}\big[\widehat{\mathbf{\Sigma}}\big(\widehat{\mathbf{\Sigma}} + \gamma\mathbf{I}_p\big)^{-1}\big]\Big]. \tag{16}$$

However, the estimator in (16) is an inconsistent estimator in the regime where n and p grow at constant rate [24]. To clarify, Fig. 1 plots an example of the derived MSE(γ) (15) and the plugin estimation method (16) versus a wide range values of γ. It is clear that using the plugin strategy does not help obtain the minimum MSE suitably. Instead, as the figure depicts, the plugin estimation method selects an improper γ that corresponds to a high MSE.

As an alternative remedy, we seek a consistent estimator of (15) by leveraging tools from RMT. To this end, we need to first obtain an asymptotic expression of (15). To do so, the following assumption should hold true.

Assumption 1: As $p,n \to \infty, p/n \to c \in (0, \infty)$.

Assumption 1 leads to the following theorem:

Theorem 1: Under Assumption 1, MSE(γ) in (15) asymptotically converges to

$$\text{MSE}(\gamma) \asymp \frac{p}{n} + \frac{n+1}{n}.\frac{1}{n}(\tilde{\delta}_1 + \gamma\tilde{\delta}_1')\text{tr}\Big[\mathbf{\Sigma}^2\big(\tilde{\delta}_1\mathbf{\Sigma} + \gamma\mathbf{I}_p\big)^{-2}\Big]$$
$$-\frac{2}{n}\Re\Big[\text{tr}\Big[\mathbf{\Sigma}^{\frac{1}{2}}\big(\tilde{\delta}_2\mathbf{\Sigma}^{\frac{1}{2}} - i\sqrt{\gamma}\,\mathbf{I}_p\big)^{-1}\Big]\Big], \tag{17}$$

where $\tilde{\delta}_1$ is the unique positive solution to the following system of equations:

$$\begin{cases}\delta_1 = \frac{1}{n}\text{tr}\big[\mathbf{\Sigma}\big(\tilde{\delta}_1\mathbf{\Sigma} + \gamma\mathbf{I}_p\big)^{-1}\big],\\ \tilde{\delta}_1 = \frac{1}{n}\text{tr}\big[\mathbf{T}\big(\delta_1\mathbf{T} + \mathbf{I}_n\big)^{-1}\big],\end{cases} \tag{18}$$

where $\mathbf{T} = \text{diag}([1, 1, \cdots, 1, 0]^T) \in \mathbb{R}^{n\times n}$; hence, $\tilde{\delta}_1$ can be written as follows:

$$\tilde{\delta}_1 = \frac{1}{1+\delta_1}. \tag{19}$$

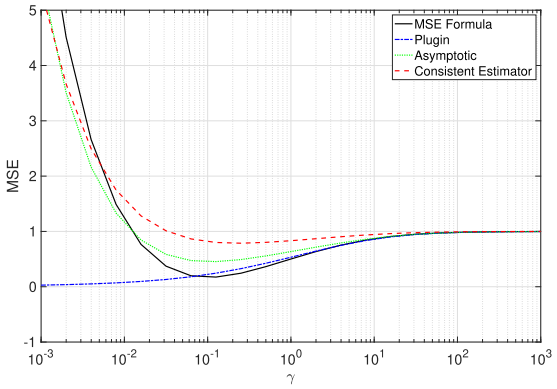


FIGURE 1. Different MSE curves versus the regularization parameter: the derived formula of the MSE, Eq. (15), the plugin estimator, Eq. (16), the asymptotic curve, Eq (17), and the consistent MSE Eq. (22). The results are generated from Gaussian data that follows the model (3), with $p = n = 300$, $[\Sigma]_{i,j} = 0.6^{|i-j|}$ and $\mu = \mathbb{1}_p$.

Similarly, $\tilde{\delta}_2$ is obtained by solving

$$\begin{cases} \delta_2 = \dfrac{1}{n}\mathrm{tr}\Big[\boldsymbol{\Sigma}^{\frac{1}{2}}\big(\tilde{\delta}_2\boldsymbol{\Sigma}^{\frac{1}{2}} - i\sqrt{\gamma}\,\mathbf{I}_p\big)^{-1}\Big], \\ \tilde{\delta}_2 = \dfrac{1}{n}\mathrm{tr}\Big[\mathbf{T}\big(\delta_2\mathbf{T} + \mathbf{I}_n\big)^{-1}\Big]. \end{cases} \tag{20}$$

and

$$\tilde{\delta}_2 = \frac{1}{1+\delta_2}. \tag{21}$$

Proof: see Appendix C.

Now, we are in a position to reveal the consistent estimator of (15).

Theorem 2: Under Assumption 1, the consistent estimator of (15) is given by (22)

$$\widehat{\mathrm{MSE}}(\gamma) \asymp \frac{p}{n} - 2\Re(\hat{\delta}_2) + \frac{n+1}{n}\frac{(1+\hat{\delta}_1)^2}{\hat{\delta}_1'} \cdot\Big[(\hat{\tilde{\delta}}_1 + \gamma\hat{\tilde{\delta}}_1')\hat{\delta}_1' - \frac{1}{n}\mathrm{tr}\Big[\widehat{\boldsymbol{\Sigma}}\big(\widehat{\boldsymbol{\Sigma}} + \gamma\mathbf{I}_p\big)^{-2}\Big]\Big], \tag{22}$$

where $\hat{\delta}_1$ and $\hat{\delta}_2$ are the consistent estimators of δ_1 and δ_2, respectively, and are given by

$$\hat{\delta}_1 = \frac{\frac{1}{n}\mathrm{tr}\Big[\widehat{\boldsymbol{\Sigma}}\big(\widehat{\boldsymbol{\Sigma}} + \gamma\mathbf{I}_p\big)^{-1}\Big]}{1 - \frac{1}{n}\mathrm{tr}\Big[\widehat{\boldsymbol{\Sigma}}\big(\widehat{\boldsymbol{\Sigma}} + \gamma\mathbf{I}_p\big)^{-1}\Big]}, \tag{23}$$

$$\hat{\delta}_2 = \frac{\frac{1}{n}\mathrm{tr}\Big[\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\big(\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}} - i\sqrt{\gamma}\,\mathbf{I}_p\big)^{-1}\Big]}{1 - \frac{1}{n}\mathrm{tr}\Big[\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\big(\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}} - i\sqrt{\gamma}\,\mathbf{I}_p\big)^{-1}\Big]}. \tag{24}$$

Proof: see Appendix D.

Back to Fig. 1, which compares the derived MSE with the asymptotic formula (17) and the consistently estimated MSE (22), it can be seen clearly that the consistent MSE is more suitable for obtaining the value of γ that minimizes (15).

A closed-form solution for γ in (22) is infeasible, so we rely on using a line search, where we search for γ that minimizes (22) within a predefined range.

A. SUMMARY OF THE PROPOSED VB-MSE (VECTOR BASED-MSE) METHOD FOR PORTFOLIO OPTIMIZATION

1) From the historical data estimate $\widehat{\boldsymbol{\Sigma}}$ using Eq. (7).
2) Find the regularization parameter value, γ_o, that minimizes $\widehat{\mathrm{MSE}}(\gamma)$ in (22) using a line search.
3) Use γ_o to compute $\widehat{\boldsymbol{\Sigma}}_{\gamma_o}^{-1} = (\widehat{\boldsymbol{\Sigma}} + \gamma_o\mathbf{I}_p)^{-1}$.
4) Calculate $\widehat{\mathbf{w}}_{\mathrm{GMVP}}$ from (6) by using $\widehat{\boldsymbol{\Sigma}}_{\mathrm{RSCM}}^{-1} = \widehat{\boldsymbol{\Sigma}}_{\gamma_o}^{-1}$.

IV. PERFORMANCE EVALUATION

In this section, we present a simulation study to shed some light on the performance of our proposed method. First, we provide a simulation result that relates the proposed approach to other loss functions that quantify the estimator. Then, the convergence and time complexity of the proposed VB-MSE method, along with other competitive methods, is presented. Lastly, a simulation of the portfolio optimization problem is considered using synthetic and real data.

A. THE PROPOSED VB-MSE CONSISTENT ESTIMATOR AND QUANTIFYING THE ESTIMATOR

We relate our proposed estimator to other loss functions used in quantifying the closeness of an estimate to the true covariance matrix. Specifically, we use the minimum-variance and the inverse-Frobenius loss functions. The minimum-variance loss function is defined as [25]

$$\mathcal{L}^{\mathrm{MV}}\left(\widehat{\boldsymbol{\Sigma}}_\gamma, \boldsymbol{\Sigma}\right) \triangleq \frac{\mathrm{tr}\left(\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\right)/p}{\left[\mathrm{tr}\left(\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\right)/p\right]^2} - \frac{1}{\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\right)/p}. \tag{25}$$

An important performance measure that involves this loss function is the *percentage relative improvement in average loss* (PRIAL) defined as

$$\mathrm{PRIAL}^{\mathrm{MV}}\left(\widehat{\boldsymbol{\Sigma}}_\gamma\right) := \frac{\mathbb{E}\left[\mathcal{L}^{\mathrm{MV}}\left(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}\right)\right] - \mathbb{E}\left[\mathcal{L}^{\mathrm{MV}}\left(\widehat{\boldsymbol{\Sigma}}_\gamma, \boldsymbol{\Sigma}\right)\right]}{\mathbb{E}\left[\mathcal{L}^{\mathrm{MV}}\left(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}\right)\right] - \mathbb{E}\left[\mathcal{L}^{\mathrm{MV}}\left(\widehat{\boldsymbol{\Sigma}}^*, \boldsymbol{\Sigma}\right)\right]} \times 100\%, \tag{26}$$

where $\widehat{\boldsymbol{\Sigma}}^*$ is the finite sample-optimal rotation-equivariant estimator, which is the closest estimator to $\boldsymbol{\Sigma}$ according to the minimum variance loss [25]. Based on the rotation-equivariant assumption, the eigenvectors of $\widehat{\boldsymbol{\Sigma}}^*$ and the SCM, $\widehat{\boldsymbol{\Sigma}}$, are the same but their eigenvalues differ. All the methods used in this paper, including ours, belong to this class of estimators, i.e., rotation-equivariant estimators. We obtain the eigenvalue decomposition of the SCM from

$$\widehat{\boldsymbol{\Sigma}} = \mathbf{U}\mathbf{D}\mathbf{U}^T. \tag{27}$$

Then, the finite sample-optimal estimator is [25]

$$\widehat{\boldsymbol{\Sigma}}^* = \mathbf{U}\mathbf{D}^*\mathbf{U}^T, \tag{28}$$

where $\mathbf{D}^*$ is the matrix of eigenvalues that minimizes the minimum variance loss function; it is calculated as [25]

$$\mathbf{D}^* = \mathbf{U}^T\boldsymbol{\Sigma}\mathbf{U}. \tag{29}$$

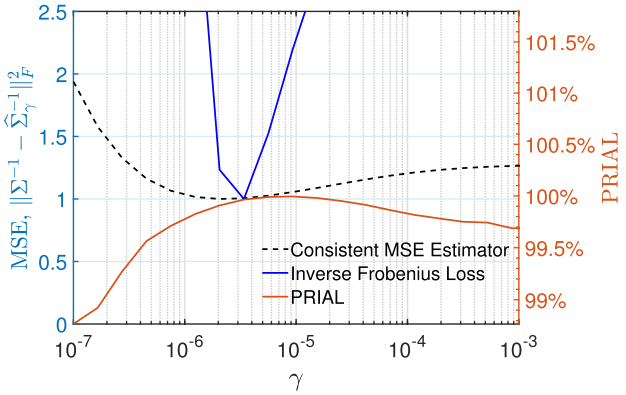


FIGURE 2. The normalized proposed MSE (consistent) estimation and the normalized inverse-Frobenius loss function (left y-axis) along with the PRIAL (right y-axis) versus the regularization parameter. The value of γ that minimizes the proposed consistent MSE estimator is shown to (almost) coincide with the optima of the two other criteria (the inverse-Frobenius loss and the PRIAL).

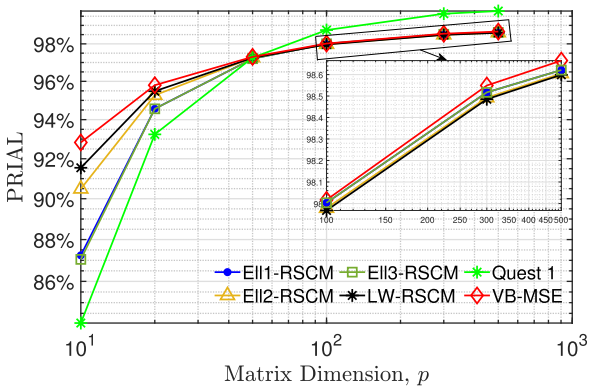


FIGURE 3. Convergence of various methods as the matrix dimension and the sample size grow large at a fixed ratio. The results are averaged across 100 Monte Carlo simulations.

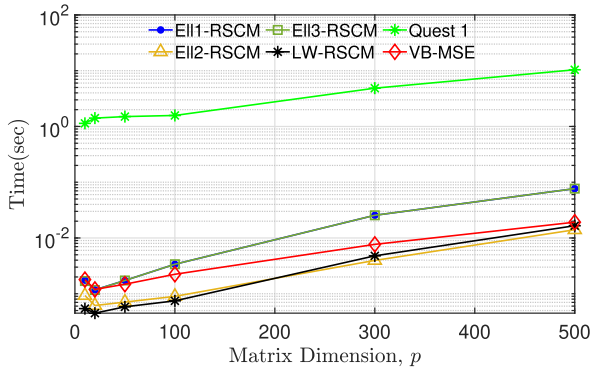


FIGURE 4. Computational speed of different estimators.

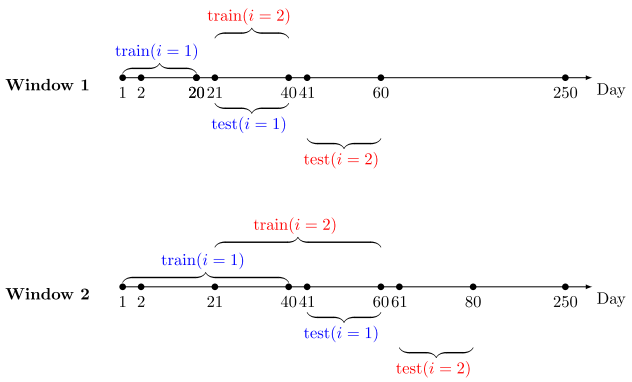


FIGURE 5. Out-of-sample method for portfolio optimization.

Note this finite-sample-optimal estimator is unattainable because it requires the knowledge of the true covariance matrix.

From the definition (26), $\text{PRIAL}^{\text{MV}}\left(\widehat{\mathbf{\Sigma}}\right) = 0\%$. This means that the SCM represents a reference against which any loss reduction is measured. Similarly, $\text{PRIAL}^{\text{MV}}\left(\widehat{\mathbf{\Sigma}}^*\right) = 100\%$ is the maximum amount of loss reduction that can be achieved under the rotation-equivariant assumption. The PRIAL estimates how much of the possibility for variance reduction is captured by any other estimator [25].

The second loss function is the inverse-Frobenius loss function defined as [25]

$$\mathcal{L}^{IF}\left(\widehat{\mathbf{\Sigma}}_\gamma, \mathbf{\Sigma}\right) = \|\widehat{\mathbf{\Sigma}}_\gamma^{-1} - \mathbf{\Sigma}^{-1}\|_F^2 \tag{30}$$

$$\triangleq \frac{1}{p}\text{tr}\left[\left(\widehat{\mathbf{\Sigma}}_\gamma^{-1} - \mathbf{\Sigma}^{-1}\right)^2\right] \tag{31}$$

In Fig. 2, we plot the proposed VB-MSE consistent estimator (normalized) along with the inverse Frobenius loss (normalized) and PRIAL versus the regularization parameter value. As can be seen from the figure, the value of γ that minimizes the VB-MSE consistent estimator almost coincides with the one that minimizes the inverse-Frobenius loss and maximizes the PRIAL.

It is remarkable that although our proposed consistent MSE estimator involves only estimated quantities, its performance almost matches both the PRIAL and Frobenius that rely on the true covariance matrix.

In the following subsections, we provide a Monte Carlo simulation study of the VB-MSE estimator against other estimators. The competitive methods are the elliptical estimators ELL1-RSCM, ELL2-RSCM and ELL3-RSCM [5], [26], the Ledoit-Wolf estimator, LW-RSCM [6], [26], and the nonlinear estimator Quest 1 [27].

The results are generated from Gaussian data that follows the model (3) with $[\mathbf{\Sigma}]_{i,j} = 0.6^{|i-j|}\sigma, (\sigma = 1 \times 10^{-4})$. We study convergence and time complexity.

B. CONVERGENCE

We consider the convergence under the large-dimensional asymptotic regime, where the data dimension, p, and the number of observations, n, grow to infinity with a ratio p/n that converges to some limit. In Fig. 3, we consider a ratio p/n that converges to $\frac{1}{2}$. We plot the PRIAL measure of each method versus p. It can be seen that the VB-MSE method has the highest PRIAL when $p = 10, 20$, and 50. For $p \geq 100$, the nonlinear estimator Quest 1 wins the comparison.

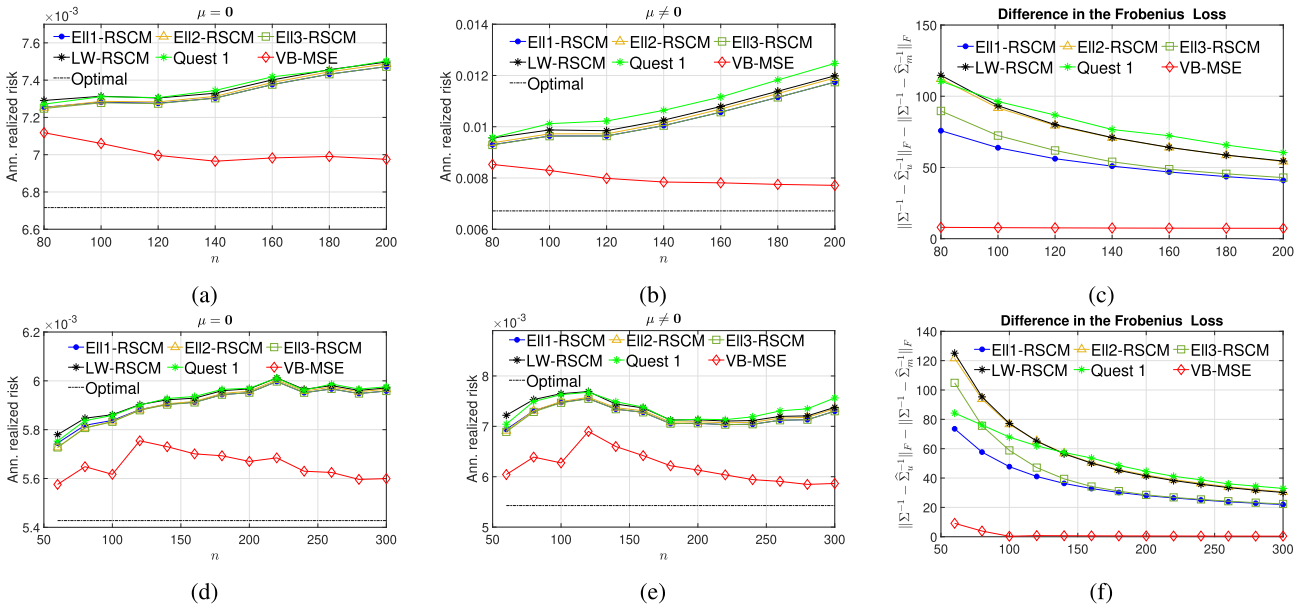


FIGURE 6. Annualized realized risk versus training window length using synthetic data generated from Gaussian distribution. (a) $\mu = 0$, $p = 200$. (b) $\mu \neq 0$, $p = 200$. (c) Difference in the Frobenius loss, $p = 200$. (d) $\mu = 0$, $p = 300$. (e) $\mu \neq 0$, $p = 300$. (f) Difference in the Frobenius loss, $p = 300$.

This result agrees with a previous study [25], which shows that the nonlinear estimators generally converge better than the linear estimators.

C. TIME COMPLEXITY

All the presented methods require computing the SCM, which is of complexity $\mathcal{O}(p^2n)$. Both LW-RSCM and Ell2-RSCM are computationally more efficient than Ell1-RSCM as pointed in [5], especially for the high-dimension setup. For the proposed method, we observe that implementing the VB-MSE estimator according to the steps presented in III-A requires computing (23), (24) and the last term in (22), each of $\mathcal{O}(p^3)$ complexity. Also, the line search method used to obtain the regularization parameter in the VB-MSE approach increases the time complexity. The nonlinear shrinkage estimator, Quest 1, requires a complexity of $\mathcal{O}(p^3)$ twice for extracting and recombining the eigenvalues and eigenvectors [25].

We can enhance the speed of the VB-MSE by observing that we can write

$$\text{tr}[(\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\Sigma}} + \gamma \mathbf{I}_p)^{-1}] = \sum_{j=1}^{d} \frac{d_j}{d_j + \gamma}, \tag{32}$$

$$\text{tr}[(\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}(\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}} - i\sqrt{\gamma}\, \mathbf{I}_p)^{-1}] = \sum_{j=1}^{d} \frac{\sqrt{d_j}}{\sqrt{d_j} - i\sqrt{\gamma}}, \tag{33}$$

$$\text{tr}[(\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\Sigma}} + \gamma \mathbf{I}_p)^{-2}] = \sum_{j=1}^{d} \frac{d_j}{(d_j + \gamma)^2}. \tag{34}$$

The formulas in (32)–(34) can be used in (23), (24) and (13). Note that we need the eigenvalue decomposition ($O(p^3)$ complexity) one time.

We consider a runtime example using Matlab R2019a running on a 64-bit, Core(TM) i7-2600K 3.40 GHz Windows PC. The result is plotted in Fig. 4 which shows that the VB-MSE method is faster than Quest 1, Ell1-RSCM and Ell3-RSCM when $50 < p < 500$.

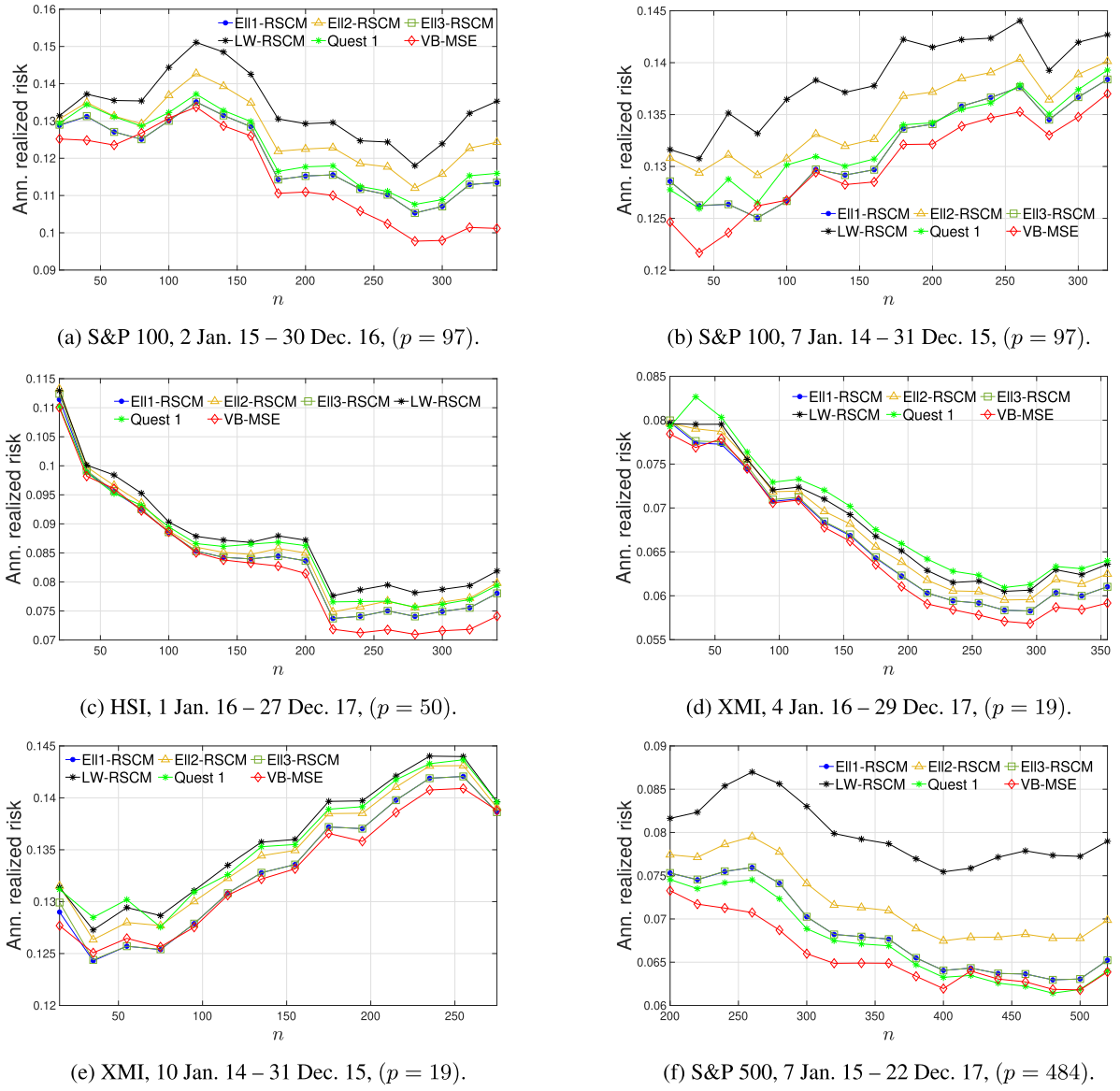
We also consider a simulation example for a very high-dimension scenario, $p = 5000$ and $n = 10000$, as shown in Table 1. While our proposed method (VB-MSE) performs slowly at this dimension, the PRIAL measure reveals a slight improvement compared to the remaining methods. Quest 1 method is not feasible for implementation at very high dimension [25].

TABLE 1. Monte Carlo simulations for $p = 5000$ and $n = 10000$.

Method	PRIAL (%)	Time (sec)
Ell1-RSCM	98.75	31.09
Ell2-RSCM	98.75	9.87
Ell3-RSCM	98.75	31.09
LW-RSCM	98.75	9.80
VB-MSE	98.82	38.51

D. PORTFOLIO OPTIMIZATION SIMULATION USING SYNTHETIC DATA

This section simulates the portfolio optimization problem using synthetic data generated from Gaussian distribution with $[\boldsymbol{\Sigma}]_{i,j} = 0.6^{|i-j|}\sigma$, ($\sigma = 1 \times 10^{-5}$). As conventionally described in the financial literature, we implement the *out-of-sample* strategy defined in terms of a rolling window method (see [7]). At a particular day t, the training window for CM estimation is formed from the previous n days, i.e., from



(a) S&P 100, 2 Jan. 15 – 30 Dec. 16, ($p = 97$).

(b) S&P 100, 7 Jan. 14 – 31 Dec. 15, ($p = 97$).

(c) HSI, 1 Jan. 16 – 27 Dec. 17, ($p = 50$).

(d) XMI, 4 Jan. 16 – 29 Dec. 17, ($p = 19$).

(e) XMI, 10 Jan. 14 – 31 Dec. 15, ($p = 19$).

(f) S&P 500, 7 Jan. 15 – 22 Dec. 17, ($p = 484$).

FIGURE 7. **Annualized realized risk versus training window length for different stock indices.**

$t - n$ to $t - 1$, to design the portfolio weights, $\widehat{\mathbf{w}}_{\text{GMVP}}$. The portfolio returns in the following 20 days are computed based on these weights. Next, the window is shifted 20 days forward and the returns for another 20 days are computed. The same procedure is repeated until the end of the data. Finally, the realized risk is computed as the standard deviation of the returns.

Fig. 5 illustrates an example of how to perform the out-of-sample procedure. Assume that we aim to study the performance of each method over a year (250 working days). In Window 1, we perform the first iteration ($i = 1$) that we train the first available 20 days (Day 1 – Day 20) and use the following 20 days as test data (Day 21 – Day 40). The next iteration in Window 1 ($i = 2$) considers Day 21 – Day 40 for the training phase and uses data of Day 41 – Day 60 for testing, and so on. For Window 2, train data size is increased to 40 in each iteration. For example, the first iteration ($i = 1$), train data is taken from Day 1 – Day 40 and test data from Day 41 – Day 60. The second iteration ($i = 2$) collects data of Day 21 – Day 60 for training, and data of Day 61 – Day 80 for testing, etc. The subsequent windows follow similarly.

Fig. 6 simulates the aforementioned out-of-sample procedure considering two scenarios. The first scenario assumes data in (3) has zero mean ($\boldsymbol{\mu} = \mathbf{0}$), and the second scenario assumes $\boldsymbol{\mu} \neq \mathbf{0}$. These two scenarios are shown in Fig. 6 (a), and Fig. 6 (b), respectively, with $p = 200$. As can be seen, the proposed VB-MSE method noticeably outperforms all the methods. The same observation can be reported when $p = 300$ in Fig. 6 (d), and Fig. 6 (e).To study the effect of the

presence of mean, we plot the difference in the Frobenius loss between the two scenarios as shown in Fig. 6 (c) ($p = 200$) and in Fig. 6 (f) ($p = 300$). Both figures reveal that the difference is the smallest when using the VB-MSE method. This result supports our previous claim that the proposed method should perform well because it incorporates the error in estimating the mean.

E. PORTFOLIO OPTIMIZATION SIMULATION USING REAL DATA

The following list describes the data from different stock market indices used in our evaluation:

- **Standard and Poor's 500 (S&P 500) index**: This index includes 500 companies. The net returns of 484 stocks ($p = 484$) are obtained for 784 working days between 7 Jan. 2015 and 22 Dec. 2017.
- **Standard and Poor's 100 (S&P 100) index**: The index is a subset of the S&P 500 that comprises 100 stocks. We consider two different periods to obtain the net returns from different stocks [28]. The first period is from 7 Jan. 2014 to 31 Dec. 2015 (501 trading days), where we fetch data of 97 stocks ($p = 97$). The second period is from 2 Jan., 2015 to 30 Dec. 2016 (504 trading days) that contains net returns of 97 stocks ($p = 97$).
- **NYSE Arca Major Market Index (XMI)**: This market index is made up of 20 Blue Chip industrial stocks of major U.S. corporations [29]. A full-length time series containing 503 working days from 4 Jan., 2016 to 29 Dec. 2017 is obtained for 19 stocks ($p = 19$). The second period is from 10 Jan. 2014 to 31 Dec. 2015 (498 working days).
- **Hang Seng Index (HSI)**: This market index comprises 50 stocks [30]. The returns of all the stocks ($p = 50$) is obtained from 1 Jan. 2016 to 27 Dec. 2017 (491 trading days).

Fig. 7 shows the annualized realized risk of the aforementioned market indices versus the number of training samples. We compare the proposed vector-based method, VB-MSE, against the elliptical estimators ELL1-RSCM, ELL2-RSCM and ELL3-RSCM [5], [26], the Ledoit-Wolf estimator, LW-RSCM [6], [26], the nonlinear estimator Quest 1 [27].

Fig. 7 (a) plots the result of the S&P 100 index from 2 Jan. 2015 to 30 Dec. 2016. As can be seen from the figure, the performance of the proposed VB-MSE method outperforms all other the methods except at $n = 80$ and 100, where it is slightly worse than Ell1-RSCM aand Ell3-RSCM. Similarly, VB-MSE has a superior performance in Fig. 7 (b), which plots the result from 7 Jan. 2014 to 31 Dec. 2015. However, at $n = 20$ and 80 Ell1-RSCM and Ell3-RSCM perform better. The realized risk for the HSI index is depicted in Fig. 7 (c) from 1 Jan. 2016 – 27 Dec. 2017. The proposed method has comparable performance to Quest 1, Ell1-RSCM and Ell3-RSCM at $n = 20, 40$ and 60 but it outperforms all the methods for $100 < n \leq 340$. The results of the XMI index from 4 Jan. 2016 – 29 Dec. 2017 and from 10 Jan. 2014 – 31 Dec. 2015 are shown in Fig. 7 (d) and Fig. 7 (e), respectively. Overall, in both figures, VB-MSE is the best performing method. Finally, Fig. 7 (f) plots the realized risk of the S&P 500 index from 10 Jan. 2015 – 31 Dec. 2017. The figure shows clearly that the proposed method outperforms the other methods when $200 \leq n \leq 400$.

From Fig. 7 (a) – (f), we can conclude that, on average, the proposed VB-MSE method compares favorably to all the benchmark methods tested in this paper. The method is also more consistent over the various datasets.

V. CONCLUSION

In this paper, we have proposed a regularized covariance matrix estimator under high-dimensionality settings. Unlike the competitive methods, the proposed method exploited a linear model with bounded uncertainties in estimating the true covariance matrix and the mean. Based on this model, the estimation problem is reduced to minimizing the MSE of the noise vector. The proposed method searches for the optimal regularization parameter based on a consistent estimator of the MSE of the estimated vector. Portfolio optimization results from real financial data show that the proposed method performs reasonably well and outperforms a host of benchmark methods.

APPENDIX A
MATHEMATICAL TOOLS

For convenience, we write Equation (3) in matrix form

$$\mathbf{Y} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{X} + \boldsymbol{\mu}\,\mathbb{1}_n, \tag{35}$$

where $\mathbf{X} = [\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n]$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$. We need to express the SCM in (7) in an appropriate matrix form as well, as follows:

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n-1}\mathbf{B}\mathbf{B}^T, \tag{36}$$

where $\mathbf{B} \in \mathbb{R}^{p\times n}$. It can be immediately recognized from (7) that $\mathbf{B}$ is

$$\mathbf{B} = \mathbf{Y} - \widehat{\boldsymbol{\mu}}\mathbb{1}_n^T. \tag{37}$$

Also, we can easily verify that

$$\widehat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \frac{1}{n}\mathbf{Z}\mathbb{1}_n, \tag{38}$$

where $\mathbf{Z} \triangleq \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{X}$. Finally, we perform the following operations to reach the model of $\widehat{\boldsymbol{\Sigma}}$ at the end:

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}} &= \frac{1}{n-1}\left(\mathbf{Z}\mathbf{Z}^T - \mathbf{Z}\frac{\mathbb{1}_n\mathbb{1}_n^T}{n}\mathbf{Z}^T\right)\\ &= \frac{1}{n-1}\mathbf{Z}\left(\mathbf{I}_n - \frac{\mathbb{1}_n\mathbb{1}_n^T}{n}\right)\mathbf{Z}^T\\ &= \frac{1}{n-1}\mathbf{Z}\mathbf{U}\mathbf{T}\mathbf{U}^T\mathbf{Z}^T\end{aligned}$$

$$= \frac{1}{n-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{X}\mathbf{U}\mathbf{T}\mathbf{U}^T\mathbf{X}^T\boldsymbol{\Sigma}^{\frac{1}{2}} \tag{39}$$

$$= \frac{1}{n-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\widetilde{\mathbf{X}}\mathbf{T}\widetilde{\mathbf{X}}^T\boldsymbol{\Sigma}^{\frac{1}{2}}, \tag{40}$$

where $\mathbf{U}$ and $\mathbf{T}$ are the matrices of eigenvalue vectors and eigenvalues, respectively, of $(\mathbf{I}_n - \frac{\mathbb{1}_n\mathbb{1}_n^T}{n})$ obtained using the eigenvalue decomposition. The Gaussian distribution is invariant when multiplying by a unitary matrix; hence, $\widetilde{\mathbf{X}}$ has the same distribution as $\mathbf{X}$ [14], [16].

The model (40) is a well-established model in is the RMT literature. Based on this model, for $z \in \mathbb{C} - \mathbb{R}^+$ and bounded $\boldsymbol{\Theta} \in \mathbb{R}^{p\times p}$ the following relations, which will be used throughout the derivations, hold true (under Assumption 1) [11]:

$$\mathrm{tr}\left[\boldsymbol{\Theta}\left(\widehat{\boldsymbol{\Sigma}} - z\mathbf{I}_p\right)^{-1}\right] \asymp \mathrm{tr}\left[\boldsymbol{\Theta}\left(\tilde{\delta}\boldsymbol{\Sigma} - z\mathbf{I}_p\right)^{-1}\right] \tag{41}$$

$$\mathrm{tr}\left[\boldsymbol{\Theta}\widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{\Sigma}} - z\mathbf{I}_p\right)^{-1}\right] \asymp \tilde{\delta}\mathrm{tr}\left[\boldsymbol{\Theta}\left(\tilde{\delta}\boldsymbol{\Sigma} - z\mathbf{I}_p\right)^{-1}\right] \tag{42}$$

$$\mathrm{tr}\left[\boldsymbol{\Theta}\left(\widehat{\boldsymbol{\Sigma}} - z\mathbf{I}_p\right)^{-2}\right] \asymp \mathrm{tr}\left[\boldsymbol{\Theta}\boldsymbol{\Sigma}\left(\tilde{\delta}\boldsymbol{\Sigma} - z\mathbf{I}_p\right)^{-2}\right] \tag{43}$$

$$\mathrm{tr}\left[\boldsymbol{\Theta}\widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{\Sigma}} - z\mathbf{I}_p\right)^{-2}\right] \asymp (\tilde{\delta} - z\tilde{\delta}')\mathrm{tr}\left[\boldsymbol{\Theta}\boldsymbol{\Sigma}\left(\tilde{\delta}\boldsymbol{\Sigma} - z\mathbf{I}_p\right)^{-2}\right]. \tag{44}$$

APPENDIX B
DERIVING THE MSE FORMULA

The MSE in (15) can be easily obtained from expanding (13) and computing the resulted terms. The first term results from $\mathbb{E}[\mathbf{x}_t\mathbf{x}_t^T] = \mathbf{I}_p$. The second term is computed as follows:

$$\mathrm{tr}\left[\mathbb{E}\left[\widehat{\mathbf{x}}_t\widehat{\mathbf{x}}_t^T\right]\right] = \mathrm{tr}\left[\mathbb{E}\left[\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t^T\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\right]\right]. \tag{45}$$

where $\tilde{\mathbf{y}}_t \triangleq (\mathbf{y}_t - \widehat{\boldsymbol{\mu}}) = \left(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{x}_t + \boldsymbol{\delta}\right)$. Using the fact that the expectation and the trace are interchangeable and the cyclic property of traces, we can write

$$\mathrm{tr}\left[\mathbb{E}\left[\widehat{\mathbf{x}}_t\widehat{\mathbf{x}}_t^T\right]\right] \tag{46}$$

$$= \mathrm{tr}\left[\mathbb{E}\left[\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{x}_t + \boldsymbol{\delta}\right)\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{x}_t + \boldsymbol{\delta}\right)^T\right]\right]. \tag{47}$$

Also, using the eigenvalue decomposition of $\widehat{\boldsymbol{\Sigma}}$, it is easy to prove that

$$\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\widehat{\boldsymbol{\Sigma}} = \left(\widehat{\boldsymbol{\Sigma}} + \gamma\mathbf{I}_p\right)^{-1}\widehat{\boldsymbol{\Sigma}} \tag{48}$$

$$= \widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{\Sigma}} + \gamma\mathbf{I}_p\right)^{-1} \tag{49}$$

$$= \widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\left(\widehat{\boldsymbol{\Sigma}} + \gamma\mathbf{I}_p\right)^{-1}\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}. \tag{50}$$

Hence,

$$\mathrm{tr}\left[\mathbb{E}\left[\widehat{\mathbf{x}}_t\widehat{\mathbf{x}}_t^T\right]\right] \tag{51}$$

$$= \mathrm{tr}\left[\mathbb{E}\left[\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_\gamma^{-2}\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{x}_t + \boldsymbol{\delta}\right)\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{x}_t + \boldsymbol{\delta}\right)^T\right]\right]. \tag{52}$$

Observing that $\mathbf{x}_t$, $\boldsymbol{\delta}$ and $\widehat{\boldsymbol{\Sigma}}$ are independent, and $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \frac{\boldsymbol{\Sigma}}{n})$, we obtain

$$\mathrm{tr}\left[\mathbb{E}\left[\widehat{\mathbf{x}}_t\widehat{\mathbf{x}}_t^T\right]\right] \tag{53}$$

$$= \mathrm{tr}\left[\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{\Sigma}} + \gamma\mathbf{I}_p\right)^{-2}\right] + \frac{1}{n}\mathrm{tr}\left[\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{\Sigma}} + \gamma\mathbf{I}_p\right)^{-2}\right] \tag{54}$$

$$= \frac{n+1}{n}\mathrm{tr}\left[\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{\Sigma}} + \gamma\mathbf{I}_p\right)^{-2}\right]. \tag{55}$$

Finally, the third term is obtained from

$$\mathrm{tr}\left[\mathbb{E}\left[\widehat{\mathbf{x}}_t\mathbf{x}_t^T\right]\right] \tag{56}$$

$$= \mathrm{tr}\left[\mathbb{E}\left[\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}(\mathbf{y} - \widehat{\boldsymbol{\mu}})\mathbf{x}_t^T\right]\right] \tag{57}$$

$$= \mathrm{tr}\left[\mathbb{E}\left[\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{x}_t + \boldsymbol{\delta}\right)\mathbf{x}_t^T\right]\right] \tag{58}$$

$$= \mathrm{tr}\left[\boldsymbol{\Sigma}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\right]. \tag{59}$$

APPENDIX C
PROOF OF THEOREM 1

$\mathbb{E}[A(\gamma)]$ in (15) can be directly obtained from (43) with setting $\boldsymbol{\Theta} = \boldsymbol{\Sigma}$ and $z = -\gamma$. The second term, $B(\gamma)$, resulted from adding and subtracting $i\sqrt{\gamma}\,\mathbf{I}_p$ with factoring $\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}$ as follows:

$$\mathrm{tr}\left[\boldsymbol{\Sigma}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\right]$$

$$= \mathrm{tr}\Bigg\{\boldsymbol{\Sigma}^{\frac{1}{2}}\left[\left(\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}} + i\sqrt{\gamma}\,\mathbf{I}_p\right) - i\sqrt{\gamma}\,\mathbf{I}_p\right] \tag{60}$$

$$\cdot\left[\left(\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}} + i\sqrt{\gamma}\,\mathbf{I}_p\right)\left(\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}} - i\sqrt{\gamma}\,\mathbf{I}_p\right)\right]^{-1}\Bigg\} \tag{61}$$

$$= \mathrm{tr}\left[\boldsymbol{\Sigma}^{\frac{1}{2}}\left(\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}} - i\sqrt{\gamma}\,\mathbf{I}_p\right)^{-1}\right] - i\sqrt{\gamma}\mathrm{tr}\left[\boldsymbol{\Sigma}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\right]. \tag{62}$$

We can further simplify (62) by noticing that the quantity on the left-hand side is a real quantity, so this implies

$$\Im\left[\mathrm{tr}\left[\boldsymbol{\Sigma}^{\frac{1}{2}}\left(\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}} - i\sqrt{\gamma}\,\mathbf{I}_p\right)^{-1}\right]\right] = \sqrt{\gamma}\mathrm{tr}\left[\boldsymbol{\Sigma}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\right]. \tag{63}$$

Thus, we can express $B(\gamma)$ equivalently as

$$\mathrm{tr}\left[\boldsymbol{\Sigma}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}_\gamma^{-1}\right] = \Re\left[\mathrm{tr}\left[\boldsymbol{\Sigma}^{\frac{1}{2}}\left(\widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}} - i\sqrt{\gamma}\,\mathbf{I}_p\right)^{-1}\right]\right], \tag{64}$$

so we can find $\mathbb{E}[B(\gamma)]$ easily from (41) with setting $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{\frac{1}{2}}$, $z = i\sqrt{\gamma}$, and the third term in (17) resulted.

APPENDIX D
PROOF OF THEOREM 2

The MSE(γ) expressed in (17) converges to a sum of deterministic terms. To find a consistent estimator of (17), it is sufficient to find a consistent estimator of each of these terms (Theorem 3.2.6 in [31]).

The consistent estimator, $\hat{\delta}_1$, can be derived using (18) and (19) in (42) (with $\boldsymbol{\Theta} = \mathbf{I}_p$),

$$\frac{1}{n}\mathrm{tr}\left[\widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{\Sigma}} + \gamma \mathbf{I}_p\right)^{-1}\right] \asymp \frac{\delta_1}{1+\delta_1}. \tag{65}$$

The consistent estimator, $\hat{\delta}_1$, in (23) (that satisfies $\hat{\delta} \asymp \delta$) results immediately after rearranging (65). The derivation of $\hat{\delta}_2$ follows similarly.

To derive the consistent estimator of $\phi \triangleq \frac{1}{n}\left[\boldsymbol{\Sigma}^2\left(\tilde{\delta}_1 \boldsymbol{\Sigma} + \gamma \mathbf{I}_p\right)^{-2}\right]$ we differentiate δ_1 to obtain

$$\delta_1' = \frac{\Psi}{1 - \phi(1+\delta_1)^{-2}}, \tag{66}$$

where $\Psi \triangleq \frac{1}{n}\mathrm{tr}[\boldsymbol{\Sigma}(\tilde{\delta}_1 \boldsymbol{\Sigma} + \gamma \mathbf{I}_p)^{-2}]$. Hence, we can estimate ϕ consistently, (i.e., $\hat{\phi} \asymp \phi$) as

$$\hat{\phi} = \frac{\hat{\delta}_1' - \widehat{\Psi}}{\hat{\delta}_1'(1+\hat{\delta}_1)^{-2}}, \tag{67}$$

where $\widehat{\Psi}$ is the consistent estimator of Ψ can be estimated from (44) when $\boldsymbol{\Theta} = \mathbf{I}_p$, $z = -\gamma$, as follows:

$$\widehat{\Psi} = \frac{\frac{1}{n}\mathrm{tr}\left[\widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{\Sigma}} + \gamma \mathbf{I}_p\right)^{-2}\right]}{(\hat{\tilde{\delta}}_1 + \gamma \hat{\tilde{\delta}}_1')}. \tag{68}$$

Substituting in (67), we can obtain the second term as

$$\frac{1}{n}\left[\boldsymbol{\Sigma}^2\left(\tilde{\delta}_1 \boldsymbol{\Sigma} + \gamma \mathbf{I}_p\right)^{-2}\right] \asymp \frac{(1+\hat{\delta}_1)^2}{\hat{\delta}_1'}\left[\hat{\delta}_1' - \frac{\frac{1}{n}\mathrm{tr}\left[\widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{\Sigma}} + \gamma \mathbf{I}_p\right)^{-2}\right]}{(\hat{\tilde{\delta}}_1 + \gamma \hat{\tilde{\delta}}_1')}\right]. \tag{69}$$

ACKNOWLEDGMENT

The authors thank the Deanship of Scientific Research for their support.

REFERENCES

[1] T. Bodnar, S. Dmytriv, N. Parolya, and W. Schmid, "Tests for the weights of the global minimum variance portfolio in a high-dimensional setting," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4479–4493, Sep. 2019.

[2] S. Huni and A. B. Sibindi, "An application of the Markowitz's mean-variance framework in constructing optimal portfolios using the Johannesburg securities exchange tradeable indices," *J. Accounting Manage.*, vol. 10, no. 2, pp. 41–57, 2020.

[3] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, 1952. [Online]. Available: http://www.jstor.org/stable/2975974

[4] R. Couillet and M. Debbah, "Signal processing in large systems: A new paradigm," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 24–39, Jan. 2012.

[5] E. Ollila and E. Raninen, "Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2707–2719, May 2019.

[6] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, 2004.

[7] L. Yang, R. Couillet, and M. R. McKay, "A robust statistics approach to minimum variance portfolio optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6684–6697, Dec. 2015.

[8] E. Ollila, D. P. Palomar, and F. Pascal, "Shrinking the eigenvalues of M-estimators of covariance matrix," *IEEE Trans. Signal Process.*, vol. 69, pp. 256–269, 2020.

[9] T. T. Cai, J. Hu, Y. Li, and X. Zheng, "High-dimensional minimum variance portfolio estimation based on high-frequency data," *J. Econometrics*, vol. 214, no. 2, pp. 482–494, Feb. 2020.

[10] T. Ballal, A. S. Abdelrahman, A. H. Muqaibel, and T. Y. Al-Naffouri, "An adaptive regularization approach to portfolio optimization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 5175–5179.

[11] F. Rubio, X. Mestre, and D. P. Palomar, "Performance analysis and optimal selection of large minimum variance portfolios under estimation risk," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 4, pp. 337–350, Aug. 2012.

[12] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, "Parameter estimation in the presence of bounded data uncertainties," *SIAM J. Matrix Anal. Appl.*, vol. 19, no. 1, pp. 235–252, Jan. 1998.

[13] Y. Guo, T. Hastie, and R. Tibshirani, "Regularized linear discriminant analysis and its application in microarrays," *Biostatistics*, vol. 8, no. 1, pp. 86–100, 2007.

[14] A. Zollanvari and E. R. Dougherty, "Generalized consistent error estimator of linear discriminant analysis," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2804–2814, Jun. 2015.

[15] T. Ballal, M. A. Suliman, and T. Y. Al-Naffouri, "Bounded perturbation regularization for linear least squares estimation," *IEEE Access*, vol. 5, pp. 27551–27562, 2017.

[16] K. Elkhalil, A. Kammoun, R. Couillet, T. Y. Al-Naffouri, and M.-S. Alouini, "A large dimensional study of regularized discriminant analysis," *IEEE Trans. Signal Process.*, vol. 68, pp. 2464–2479, 2020.

[17] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-24, no. 4, pp. 397–401, Jul. 1988.

[18] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.

[19] M. Mahadi, T. Ballal, M. Moinuddin, T. Y. Al-Naffouri, and U. Al-Saggaf, "Low-complexity robust beamforming for a moving source," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Jan. 2021, pp. 1846–1850.

[20] M. Mahadi, T. Ballal, M. Moinuddin, T. Y. Al-Naffouri, and U. Al-Saggaf, "A robust LCMP beamformer with limited snapshots," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Jan. 2021, pp. 1831–1835.

[21] M. A. Suliman, H. Sifaou, T. Ballal, M.-S. Alouini, and T. Y. Al-Naffouri, "Robust estimation in linear ILL-posed problems with adaptive regularization scheme," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 4504–4508.

[22] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[23] F. Rubio, "Generalized consistent estimation in arbitrarily high dimensional signal processing," Ph.D. dissertation, Universitat Politecnica de Catalunya, Barcelona, Spain, 2008.

[24] L. B. Niyazi, A. Kammoun, H. Dahrouj, M.-S. Alouini, and T. Y. Al-Naffouri, "Asymptotic analysis of an ensemble of randomly projected linear discriminants," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 3, pp. 914–930, Nov. 2020.

[25] O. Ledoit and M. Wolf, "Analytical nonlinear shrinkage of large-dimensional covariance matrices," *Ann. Statist.*, vol. 48, no. 5, pp. 3043–3065, Oct. 2020.

[26] E. Ollila and E. Raninen, *MATLAB Regularizedscm ToolBox Version 1.0*. Accessed: May 12, 2021. [Online]. Available: http://users.spa.aalto.fi/esollila/regscm/

[27] O. Ledoit and M. Wolf, "Numerical implementation of the QuEST function," *Comput. Statist. Data Anal.*, vol. 115, pp. 199–223, Nov. 2017.

[28] *S&P Dow Jones Indices*. Accessed: Jun. 15, 2021. [Online]. Available: https://www.spglobal.com/spdji/en/indices/equity/sp-100

[29] *NYSE: The New York Stock Exchange*. Accessed: Jun. 15, 2021. [Online]. Available: https://www.nyse.com/quote/index/XMI

[30] *Hang Seng Indexes*. Accessed: Jun. 15, 2021. [Online]. Available: https://www.hsi.com.hk/eng

[31] A. Takeshi, *Advanced Econometrics*. Cambridge, MA, USA: Harvard Univ. Press, 1985.

MAAZ MAHADI (Student Member, IEEE) received the B.Sc. degree (Hons.) in electrical engineering from the University of Khartoum, Sudan, and the M.Sc. degree in electrical engineering from International Islamic University Malaysia, Kuala Lumpur, Malaysia, in 2014. He is currently pursuing the Ph.D. degree with King Abdulaziz University, Jeddah, Saudi Arabia. His research interests include statistical signal processing applications, beamforming, discriminant analysis, and random matrix theory.

TARIG BALLAL (Member, IEEE) received the B.Sc. degree (Hons.) in electrical engineering from the University of Khartoum, Sudan, in 2001, the M.Sc. degree in telecommunications from the Blekinge Institute of Technology, Sweden, in 2005, and the Ph.D. degree from the School of Computer Science and Informatics, University College Dublin, Dublin, Ireland, in 2011. From 2011 to 2012, he was a Research Engineer with Resmed Sensor Technologies Ltd., and University College Dublin. During this period, he developed methods for non-contact breathing monitoring and classification from wireless biosensors data. Since 2012, he has been with the Electrical Engineering Department, King Abdullah University of Science and Technology, Saudi Arabia, where he is currently a Research Scientist. His current research focuses on regularization and robust estimation methods, image and signal processing, acoustic and electromagnetic sensing, and tracking and localization. His research interests include GNSS precise positioning and attitude determination, in addition to signal processing applications in wireless communications.

MUHAMMAD MOINUDDIN (Member, IEEE) received the B.S. degree in electrical engineering from the Nadirshaw Edulji Dinshaw University of Engineering and Technology, Karachi, Pakistan, in 1998, and the M.S. and Ph.D. degrees in electrical engineering from the King Fahd University of Petroleum and Minerals (KFUPM), Dharan, Saudi Arabia, in 2001 and 2006, respectively. He was an Assistant Professor with the Department of Electrical Engineering, KFUPM, and the Hafr Al-Batin Community College, Hafr Al-Batin, Saudi Arabia, from 2007 to 2010. He was an Associate Professor with the Telecommunication Engineering Department, Iqra University, Karachi, from 2010 to 2012. He was an Assistant Professor with the Department of Electrical Engineering, KAU, Jeddah, Saudi Arabia, from 2013 to 2015, where he is currently a Professor and is also associated with the Center of Excellence in Intelligent Engineering Systems. His research interests include adaptive filtering, wireless communications, and signal processing for communications.

UBAID M. AL-SAGGAF (Member, IEEE) received the dual B.Sc. degree (Hons.) in electrical engineering and mathematics from the King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia, in 1980, the M.Sc. degree, in 1983, and the Ph.D. degree in electrical engineering from Stanford University, USA, in 1986. He joined the KFUPM, as an Assistant Professor, in 1986. From 1992 to 2010, he was on leave working for the Ministry of Defense and Aviation as an Executive Technical Advisor with the Research and Development Department. He joined King Abdulaziz University (KAU), in September 2010, where he is currently a Professor. He is also the Founder and the Director of the Center of Excellence in Intelligent Engineering Systems (CEIES), and the Founder of the Innovation and Prototyping Center (IPC), KAU. His research interest includes broad spectrum from theoretical to practical aspects of engineering, including systems, control, communications, and signal processing.