Estimation of Linear Process Spectra with an Application to Determining the Mean-Variance Frontier for Time Series*

Haoyang Liu† Alexander Aue‡ Debashis Paul‡

November 10, 2016

Abstract

In finite samples, risk is known to be underestimated in the mean-variance frontier. This paper proposes a novel algorithm correcting this finite sample bias when asset returns follow a high-dimensional time series. Assuming a linear process formulation studied in Liu, Aue, and Paul (2015), an algorithm is proposed to estimate the spectral distributions of the coefficient matrices of the linear process by making use of the asymptotic behavior of the empirical spectral distributions of symmetrized autocovariance matrices. This leads to the formulation of a strategy for the estimation of the mean-variance frontier, utilizing the estimates of the coefficient matrix spectra. The proposed method is extended to a setting in which the returns are assumed to have a factor model structure with observed factors, while the unknown idiosyncratic terms are assumed to belong to the aforementioned class of linear processes. The performance of the proposed methods is examined through extensive simulation studies.

Keywords: Factor Model; High-dimensional Time Series; Linear Processes; Markowitz Portfolio Optimization; Random matrix theory; Risk Management; Stieltjes Transform.

1 Introduction

One persistent challenge in macroeconmics and finance consists of devising inference procedures for short time series data of length T with a large number N of cross sections. Cochrane (2005) argues: "Our econometric techniques all are designed for large time series and small cross-sections. Our data has a large cross section and short time series. A large unsolved problem in finance is the development of appropriate large-N small-T tools for evaluating asset pricing models." One particular large-N small-T impact mentioned in Cochrane (2005) is the underestimation of risk in the mean-variance frontier. This Mean-Variance Frontier (henceforth, MVF) measures the minimum risk to be taken in order to achieve a set of expected returns.

*This research was partially supported by NSF grants DMS 1305858 and DMS 1407530

†Haas School of Business, University of California, Berkeley, CA 94720, USA, email: liuhy@berkeley.edu

‡Department of Statistics, University of California, Davis, CA 95616, USA, email: [aaue,debpaul]@ucdavis.edu

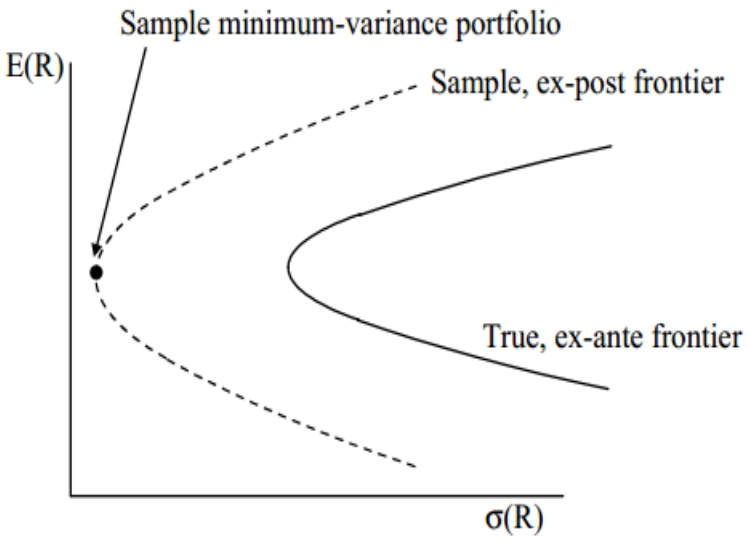

Figure 1: Ex-ante and Ex-Post frontiers. Source: Cochrane (2005).

To paraphrase Cochrane (2005), apart from statistical uncertainty due to sampling, the large dimensionality of the data leads to peculiar discrepancies between the sample MVF compared to its population counterpart. From Figure 1, we can see that the ex-post frontier, calculated from finite samples, significantly underestimates risk compared to the true, ex-ante frontier. Existing works solve this problem for time independent data (El Karoui, 2013; Kan and Zhou, 2007). In this paper, we propose a novel algorithm correcting the bias in MVF for time series returns.

Figure 2 illustrates the main result via Monte Carlo simulation. Dashed lines around each solid line constitute the one standard error bands around the respective mean estimates. The mean estimate of the proposed LinShrink algorithm closely matches the truth, outperforming the other two algorithms when the data is sampled from a high-dimensional, second-order moving average process. The magenta curve displays the result of the IndShrink method proposed by El Karoui (2013), stressing that his algorithm yields an unbiased estimator when the underlying data generating process is time independent. The difference between the magenta curve and the green curve is due to the model misspecification between a time independent model and a time series model. Given that most time series in finance and econometrics have important inter-temporal correlations, it is important to design an algorithm that accounts for the time dependence.

A simple justification of risk underestimation by the naive estimate of MVF, that is, the sample version of the minimum portfolio risk formula, follows from Jensen's inequality. The difference between the population and sample versions only gets larger with increasing N. To understand this intuitively, consider the extreme case of mutually uncorrelated assets. Here, the covariance matrix $\mathbf{\Sigma}$ is a diagonal matrix and the population

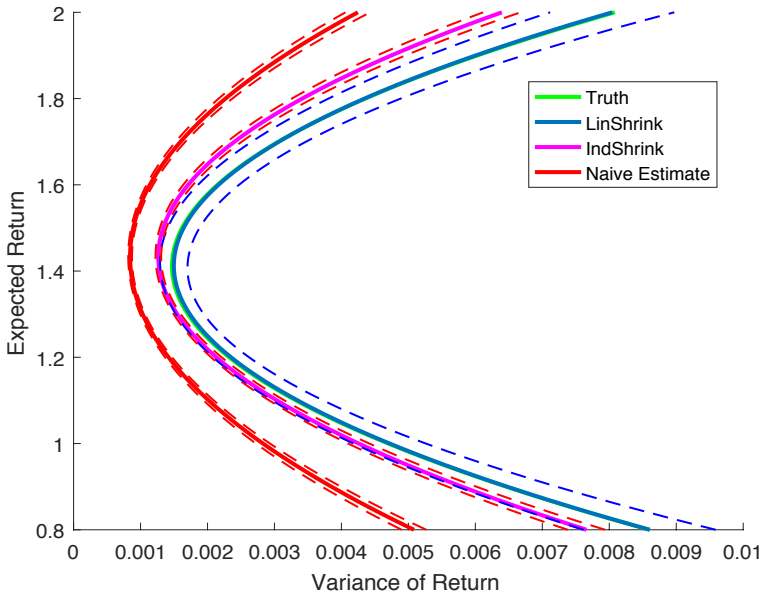

Figure 2: MVF estimation for an MA(2) process with $N = 1000$, $T = 3000$. LinShrink is the proposed algorithm, IndShrink stands for shrinkage assuming independence, and the naive estimate uses the standard sample covariance matrix estimate.

correlation of returns from any pair of assets is 0. However, the sample correlations could still be significantly positive or negative. These positive or negative sample correlations make assets look like good hedges for each other, despite their independence in the population. These pseudo hedges are the fundamental reason why risk is underestimated in finite samples. This is particularly an issue when the number of assets N is large. There are $N(N-1)/2$ number of distinct correlations, not to mention correlations between linear combinations, or portfolios, of assets. Thus, when N is large, some pairs of assets should appear to be correlated, or good hedges, with a very high probability despite the cross sectional independence assumption.

The issue of bias in the empirical estimate of MVF gets more intricate when, in addition, there is temporal dependence across observations. The aim of this paper, then, is to explore this aspect in detail by making use of a widely used class of time series – linear processes – as a benchmark for explaining the phenomena. In particular, Random Matrix Theory (RMT) is utilized using the results developed in Liu, Aue, and Paul (2015) as a starting point. Throughout the exposition, significant differences and intricacies associated with the temporal dependence in the returns are emphasized, in particular when related to issues regarding MVF estimation.

To articulate the main ideas, note that the basic problem of MVF can be formulated as that of estimating the quadratic form $\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B}$, where $\mathbf{\Sigma} = \mathrm{Cov}(X_t)$ denotes the $N \times N$ covariance matrix for asset return vector X_t, $t = 1, \ldots, T$ and $\mathbf{B}$ is an $N \times K$ matrix. In MVF, $\mathbf{B}$ is an $N \times 2$ matrix with expected returns of

the N assets as the first column, and every entry in the second column being equal to 1. As mentioned above, the "natural estimate" $\mathbf{B}^T\mathbf{S}^{-1}\mathbf{B}$ is biased, even if the population mean return is assumed to be known. In this paper, a correction for the finite sample bias is proposed for $(X_t\colon t \in \mathbb{Z})$ in the class of high-dimensional linear time series of the form $X_t = \sum_{\ell=0}^{\infty} \mathbf{A}_\ell Z_{t-\ell}$, where the coefficient matrices $(\mathbf{A}_\ell\colon \ell \in \mathbb{N}_0)$ are symmetric and simultaneously diagonalizable, and $(Z_t\colon t \in \mathbb{Z})$ an independent, identically distributed (i.i.d.) process of innovations with independent coordinates.[1] Under the stated assumptions, Liu, Aue, and Paul (2015) established the existence of a limiting spectral distribution for any finite order symmetrized sample autocovariance matrix. The latter result forms the cornerstone of the estimation procedures proposed in this paper. Note that the assumed process on one hand generalizes the i.i.d. observations that have mostly been assumed in the literature on (high-dimensional) portfolio optimization and encompasses the stationary Autoregressive Moving Average (ARMA) processes widely used in time series modeling, with the restriction of simultaneously diagonalizable coefficients. On the other hand, it also constitutes a nontrivial extension to existing results in RMT literature on the asymptotic behavior of spectra of sample covariance matrices in large dimensions, such as the works of El Karoui (2008) and Ledoit and Wolf (2012, 2015). As is demonstrated later, estimation of the eigenvalue distribution of these coefficient matrices is a challenge due to the highly nonlinear nature of the relationship between the sample spectra and the population parameters. The first contribution of this paper is to develop an algorithm for the estimation of these eigenvalue distributions. It is also described how solving this problem allows for designing an effective algorithm, termed LinShrink, that includes a model selection step for estimating the quantity of primary interest in this paper, namely the MVF, in a time series context. The second contribution of this paper is the extension of the algorithm to the scenario where asset returns follow a factor structure with the idiosyncratic term possessing the linear process structure specified above. Related work for high-dimensional factor models may be found in Bai (2003), Bai and Ng (2002), and Onatski (2009, 2012).

Although the results are illustrated in the classical MVF context, the proposed methodology has the potential for further applications. Quadratic forms of the type $\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B}$ arise in many fields of time series econometrics; for example, in computing standard errors in the generalized methods of moments and in linear discriminant analysis. Moreover, estimating the eigenvalue distribution of coefficients of the linear time series provides avenues for designing appropriate shrinkage rules for objects such as the spectral density matrix associated with it. The latter object, and functionals derived from it, are especially important in prediction problems. Extensions of the proposed method to these and other specific contexts constitute promising directions for future research.

Besides the papers mentioned above, our work contributes to the vast literature on asset pricing and optimal

[1] In this paper, $\mathbb{Z}$, $\mathbb{N}$ and $\mathbb{N}_0$ denote the integers, the positive integers and the nonnegative integers, respectively.

portfolio choice. The pathbreaking work of Markowitz Markowitz (1952) utilized optimization theory to determine the optimal MVF under the assumption of known mean and covariances of the returns. Since then, decades of work in econometrics and finance have addressed various facets of MVF estimation under statistical and parameter uncertainties. The importance of the latter was revealed by Barberis (2000), who emphasized that even after incorporating parameter uncertainty, there is sufficient predictability in returns, thus making it important to take estimation risk into account when allocating assets to stocks.

Chopra and Ziemba (1993) were among the earliest to study the phenomena associated with high dimensionality on the portfolio risk estimation. Kan and Zhou (2007) considered the effects of model uncertainty and studied the implications of combinations of portfolio choices. The bias in the empirical MVF can be significant even when T and N are of comparable sizes. The bias in this setting was quantified by Bai, Liu, and Wong (2009) and El Karoui (2010) through the use of the RMT framework, assuming that the returns across time are independent. The latter works, as well as El Karoui (2013), also suggested methods for correcting for this bias under the same assumptions. It should be mentioned here that the use of RMT in explaining behavior of high-dimensional financial data was pioneered by Laloux, Cizeau, Bouchaud, and Potters (1999); Laloux, Cizeau, Potters, and Bouchaud (2000). A broad overview of this growing literature can be found in Bouchaud and Potters (2009), and Paul and Aue (2014).

The approaches put forth in Bai, Liu, and Wong (2009) and El Karoui (2010, 2013) rely on various ways of reducing the bias in the estimation of quadratic forms of the type $\mathbf{a}^T\mathbf{\Sigma}^{-1}\mathbf{b}$, where $\mathbf{\Sigma}$ is the population return covariance matrix. The vectors $\mathbf{a}$ and $\mathbf{b}$ are typically either the population mean returns μ or some fixed vector such as the vector of ones. One redeeming feature of their work are the minimal structural assumptions on the population parameters. Recently, another line of study within the MVF estimation framework has come into the focus that seeks to modify prior work through imposing various constraints on the portfolio weights. Empirical research, for example in Brodie, Daubechies, De Mol, Giannone, and Loris (2009); DeMiguel, Garlappi, Nogales, and Uppal (2009); Jagannathan and Ma (2003), has shown that such constraints can enhance portfolio performance. Some of these empirical results were validated through theoretical justifications in Fan, Zhang, and Yu (2012). Very recently, Ao, Li, and Zheng (2016) proposed a new sparsity constraint-based estimation of MVF and showed its consistency.

The remainder of the paper is organized as the follows. Section 2 reviews existing results, sets the model and states the main assumptions. Section 3 describes in detail the proposed estimation algorithm, including a thresholding and model selection algorithm. Section 4 discusses the application of the proposed methodology to the Markowitz portfolio problem and an extension to factor models with known factors. Section 5 reports empirical results from extensive Monte Carlo simulations. Section 6 concludes.

2 Outline of Paper and Overview of Related Prior Work

The methodology developed in this paper is built upon the theoretical analysis of the behavior of empirical spectral distributions of symmetrized sample autocovariance matrices carried out in Liu, Aue, and Paul (2015), whose results are linked to the present work in the following way. Here, interest is in estimating a generalized quadratic form of the type $\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B}$ with a strategy based on the spectral decomposition of the covariance matrix $\mathbf{\Sigma}$. Denoting the distinct (ordered) eigenvalues of $\mathbf{\Sigma}$ by $(\sigma_j : j = 1, \ldots, J)$ and corresponding eigen-projection matrices $(\mathbf{P}_j : j = 1, \ldots, J)$, where $J \in \{1, \ldots, p\}$, the quadratic form can be expressed as

$$\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B} = \sum_{j=1}^{J} \frac{1}{\sigma_j}\mathbf{B}^T\mathbf{P}_j\mathbf{B}. \tag{2.1}$$

In a nutshell, the strategy is based upon splitting the estimation problem into two steps. The first step involves the estimation of the eigenvalues $(\sigma_j : j = 1, \ldots, J)$ together with their multiplicities. This can be equivalently expressed in the form of the *empirical spectral distribution (ESD)* of $\mathbf{\Sigma}$. The second step involves estimation of the parameters $\mathbf{\Theta}_j = \mathbf{B}^T\mathbf{P}_j\mathbf{B}$, $j = 1, \ldots, J$. The main results of Liu, Aue, and Paul (2015) provide the foundation for estimating the ESD of $\mathbf{\Sigma}$. Given the estimate of the ESD of $\mathbf{\Sigma}$, one can make use of a carefully constructed regression formulation, again based on the derivations in Liu, Aue, and Paul (2015), to estimate the parameters $(\mathbf{\Theta}_j : j = 1, \ldots, J)$. The final estimate of $\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B}$ is obtained by combining the two component estimates.

To elaborate further, Liu, Aue, and Paul (2015) established the existence of nonrandom limits of the ESD of the sample covariance matrix $\mathbf{S}$ and symmetrized autocovariance matrices for linear process of the kind described in Section 1. These results are expressed in terms of the *Stieltjes transform* of the ESDs of the sample autocovariance matrices and the coefficients of the linear process. The proposed estimation strategy for the ESD of $\mathbf{\Sigma}$ makes explicit use of the relationships between the Stieltjes transforms of the ESD of $\mathbf{S}$ and that of the coefficient matrices to formulate an optimization problem where the ESDs of the coefficient matrices are used as unknown parameters. Section 2.1 contains details on the model and the main notions, including the definition of the high-dimensional setting, symmetrized autocovariance matrices, Stieltjes transform and the description of the linear time series model. For convenience, Section 2.2 summarizes the main results of Liu, Aue, and Paul (2015).

2.1 Setting

The limiting scenario considered considered here is the classical high dimensional setting when the number of dimensions N grows with the number of observations T, so that $N = N(T)$ is assumed to be a function of the sample size satisfying

$$\lim_{T\to\infty} \frac{N}{T} = c \in (0, \infty). \tag{2.2}$$

A sequence of random vectors $(X_t\colon t\in\mathbb{Z})$ with values in $\mathbb{C}^N$ is called a linear process or moving average process of order infinity, abbreviated by the acronym MA(∞), if it has the representation

$$X_t=\sum_{\ell=0}^{\infty}\mathbf{A}_\ell Z_{t-\ell},\qquad t\in\mathbb{Z},\tag{2.3}$$

where $(Z_t\colon t\in\mathbb{Z})$ denotes a sequence of independent, identically distributed N-dimensional random vectors whose entries are independent and satisfy $\mathbb{E}[Z_{nt}]=0$, $\mathbb{E}[|Z_{nt}|^2]=1$, and $\mathbb{E}[|Z_{nt}|^4]<\infty$, where Z_{nt} denotes the nth coordinate of Z_t. In the complex-valued case this is meant as $\mathbb{E}[\mathrm{Re}(Z_{nt})^2]=\mathbb{E}[\mathrm{Im}(Z_{nt})^2]=1/2$. It is also assumed that real and imaginary parts are independent.

The symmetrized lag-τ sample autocovariance associated with the process $(X_t\colon t\in\mathbb{Z})$ is defined as

$$\mathbf{C}_\tau=\frac{1}{2T}\sum_{t=1}^{T-\tau}\big(X_tX_{t+\tau}^*+X_{t+\tau}X_t^*\big),\qquad \tau\in\mathbb{N}_0,$$

assuming $X_1,\dots,X_T$ have been observed. For $\tau=0$, this definition gives the covariance matrix $\mathbf{S}=\mathbf{C}_0$. Let

$$\hat{F}_\tau(\sigma)=\frac{1}{p}\sum_{n=1}^{N}\mathbb{I}_{\{\sigma_{n,\tau}\le\sigma\}},$$

denote the empirical spectral distribution (ESD) of $\mathbf{C}_\tau$, where $\sigma_{1,\tau},\dots,\sigma_{N,\tau}$ are the eigenvalues of $\mathbf{C}_\tau$.

Assumption 2.1 below lists several additional assumptions on the coefficient matrices $\mathbf{A}_\ell$ of the linear process in (2.3). The essence of this assumption is that, up to an unknown rotation matrix $\mathbf{U}$, the coordinates of the observation vector X_t, form uncorrelated stationary time series with the coefficients in the linear process representation being functionally related in a suitably smooth manner, as indicated by the behavior of a set of continuous functions f_ℓ.

Assumption 2.1. *(a) Set $\mathbf{A}_0=\mathbf{I}$, the $N\times N$ identity matrix.*

(b) The matrices $(\mathbf{A}_\ell\colon \ell\in\mathbb{N}_0)$ are simultaneously diagonalizable Hermitian matrices satisfying $\|\mathbf{A}_\ell\|\le\bar{\lambda}_{\mathbf{A}_\ell}$ for all $\ell\in\mathbb{N}_0$ and large N with

$$\sum_{\ell=0}^{\infty}\bar{\lambda}_{\mathbf{A}_\ell}\le\bar{\lambda}_{\mathbf{A}}<\infty\qquad\text{and}\qquad\sum_{\ell=0}^{\infty}\ell\bar{\lambda}_{\mathbf{A}_\ell}\le\bar{\lambda}'_{\mathbf{A}}<\infty.$$

Note that one can set $\bar{\lambda}_{\mathbf{A}_0}=1$.

(c) There are continuous functions $f_\ell\colon\mathbb{R}^m\to\mathbb{R}$, $\ell\in\mathbb{N}_0$, such that, for every N, there is a set of points $\lambda_1,\dots,\lambda_N\in\mathbb{R}^m$, not necessarily distinct, and a unitary $N\times N$ matrix $\mathbf{U}$ such that

$$\mathbf{U}^*\mathbf{A}_\ell\mathbf{U}=\mathrm{diag}(f_\ell(\lambda_1),\dots,f_\ell(\lambda_N)),\qquad\ell\in\mathbb{N},$$

and $f_0(\lambda)=1$. Note that the functions f_ℓ could also be allowed to depend on $N=N(T)$ as long as they converge uniformly to continuous functions as $T\to\infty$.

(d) With probability one, $F_p^{\mathbf{A}}$, the ESD of $\{\lambda_1, \ldots, \lambda_N\}$, converges weakly to a nonrandom probability distribution function $F^{\mathbf{A}}$ on $\mathbb{R}^m$ as $N \to \infty$.

One classical route to formulate results in the high-dimensional setting prescribed in (2.2) is through the use of the Stieltjes transform, which transforms a distribution to a function defined on $\mathbb{C}^+$, where $\mathbb{C}^+ = \{x + iy \colon x \in \mathbb{R}, y > 0\}$ denotes the upper complex half plane. The Stieltjes transform of a distribution function F on the real line is the function

$$s_F \colon \mathbb{C}^+ \to \mathbb{C}^+, \; z \mapsto s_F(z) = \int \frac{1}{\sigma - z} dF(\sigma).$$

It can be shown that s_F is analytic on $\mathbb{C}^+$ and that the distribution function F can be reconstructed from s_F using an inversion formula. See Bai and Silverstein (2010) or Paul and Aue (2014) for further descriptions on Stieltjes transforms and their usage in random matrix theory.

2.2 Large-Sample Spectral Behavior of $\mathbf{C}_\tau$

Denote by $\mathbf{A} = [\mathbf{A}_0 : \mathbf{A}_1 : \cdots]$ the matrix collecting the coefficient matrices of the linear process $(X_t \colon t \in \mathbb{Z})$. Define the transfer functions

$$\psi(\lambda, \nu) = \sum_{\ell=0}^{\infty} e^{i\ell\nu} f_\ell(\lambda) \qquad \text{and} \qquad \psi(\mathbf{A}, \nu) = \sum_{\ell=0}^{\infty} e^{i\ell\nu} \mathbf{A}_\ell, \tag{2.4}$$

and the power transfer functions

$$h(\lambda, \nu) = |\psi(\lambda, \nu)|^2 \qquad \text{and} \qquad \mathcal{H}(\mathbf{A}, \nu) = \psi(\mathbf{A}, \nu)\psi(\mathbf{A}, \nu)^*.$$

Note that the contribution of the temporal dependence of the underlying time series on the asymptotic behavior of $\hat{F}_\tau$ is quantified through $h(\lambda, \nu)$. Specifically, $h(\lambda_n, \nu)$ with λ_n as in part (c) of Assumption 2.1 is (up to normalization) the spectral density of the nth coordinate of the process rotated with the help of the unitary matrix $\mathbf{U}$. With these definitions, the main results of Liu, Aue, and Paul (2015) can be stated as follows.

Theorem 2.1. *If a complex-valued linear process $(X_t \colon t \in \mathbb{Z})$ with independent, identically distributed Z_{nt}, $\mathbb{E}[Z_{nt}] = 0$, $\mathbb{E}[\mathrm{Re}(Z_{nt})^2] = \mathbb{E}[\mathrm{Im}(Z_{nt})^2] = 1/2$, $\mathrm{Re}(Z_{nt})$ and $\mathrm{Im}(Z_{nt})$ independent, and $\mathbb{E}[|Z_{nt}|^4] < \infty$, satisfies Assumption 2.1, then, with probability one and in the high-dimensional setting (2.2), $\hat{F}_\tau$ converges to a nonrandom probability distribution F_τ with Stieltjes transform s_τ determined by the equation*

$$s_\tau(z) = \int \left[\frac{1}{2\pi} \int_0^{2\pi} \frac{\cos(\tau\nu) h(\lambda, \nu)}{1 + c\cos(\tau\nu) K_\tau(z, \nu)} d\nu - z \right]^{-1} dF^{\mathbf{A}}(\lambda), \tag{2.5}$$

where $K_\tau \colon \mathbb{C}^+ \times [0, 2\pi] \to \mathbb{C}^+$ is a Stieltjes kernel, that is, $K_\tau(\cdot, \nu)$ is the Stieltjes transform of a measure with total mass $m_\nu = \int h(\lambda, \nu) dF^{\mathbf{A}}(\lambda)$ for every fixed $\nu \in [0, 2\pi]$, whenever $m_\nu > 0$. Moreover, K_τ is the unique solution of

$$K_\tau(z, \nu) = \int \left[\frac{1}{2\pi} \int_0^{2\pi} \frac{\cos(\tau\nu') h(\lambda, \nu')}{1 + c\cos(\tau\nu') K_\tau(z, \nu')} d\nu' - z \right]^{-1} h(\lambda, \nu) \, dF^{\mathbf{A}}(\lambda) \tag{2.6}$$

subject to the restriction that K_τ is a Stieltjes kernel. Otherwise, if $m_\nu = 0$, then $K_\tau(z, \nu)$ is identically zero on $\mathbb{C}^+$ and so still satisfies (2.6)*.*

Theorem 2.2. *If a real-valued linear process $(X_t \colon t \in \mathbb{Z})$ with independent, identically distributed real-valued Z_{nt}, $\mathbb{E}[Z_{nt}] = 0$, $\mathbb{E}[Z_{nt}^2] = 1$ and $\mathbb{E}[Z_{nt}^4] < \infty$, satisfies Assumption 2.1 with real symmetric coefficient matrices $(\mathbf{A}_\ell \colon \ell \in \mathbb{N}_0)$, then the result of Theorem 2.1 is retained.*

3 Estimation Strategy

The estimation strategy for the quadratic form $\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B}$ is based on two steps that are summarized in the following. All details are given in subsequent sections.

(1) *Estimate the ESD of $\mathbf{\Sigma}$:* Since under the assumed model, the ESD of $\mathbf{\Sigma}$ is determined by $F^{\mathbf{A}}$ and the (known) functions $(f_\ell \colon \ell \in \mathbb{N})$, a strategy is formulated for the estimation of $F^{\mathbf{A}}$ within the linear process framework. For computational tractability, the assumed model is chosen to be a finite-order MA process that can serve as an approximation to the true, and typically unknown, infinite-order linear process. Once the process is specified, the system of equations (2.5) and (2.6) describing the Stieltjes transform of the limiting spectral distribution can be utilized for a collection of symmetrized sample autocovariance matrices, including the sample covariance matrix $\mathbf{S}$, to formulate an optimization problem involving a discrepancy measure between the empirical and limiting Stieltjes transforms, with $F^{\mathbf{A}}$ serving as the unknown parameter of interest. To enable this optimization, $F^{\mathbf{A}}$ is parametrized by treating it as a mixture of point masses, though more general formulations are feasible. The proposed approach to estimation of $F^{\mathbf{A}}$, and consequently the ESD of $\mathbf{\Sigma}$, is connected to, but significantly more involved than, the approach for estimating the distribution of eigenvalues of the population covariance matrix adopted by El Karoui (2008). However, this approach is quite distinct from other recent estimation procedures for the spectrum of the population covariance matrix, such as those by Bai, Chen, and Yao (2010) and Ledoit and Wolf (2015). Details of the first step are given in Section 3.1.

(2) *Estimate the parameters $\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_J$:* Recall that $\mathbf{\Theta}_j = \mathbf{B}^T\mathbf{P}_j\mathbf{B}$, where $\mathbf{P}_j$ is the eigen-projection matrix of $\mathbf{\Sigma}$ corresponding to its jth distinct eigenvalue σ_j, $j = 1, \ldots, J$. For the estimation of these parameters, make use of the *deterministic equivalent* of the resolvent $\mathbf{R}(z) = (\mathbf{S} - z\mathbf{I}_p)^{-1}$, $z \in \mathbb{C}^+$. This result is a key ingredient in Liu, Aue, and Paul (2015) for establishing the limiting spectral distribution (LSD) of $\mathbf{S}$ and allows for the use of a regression problem formulation, minimizing the sum of squared distances between the matrix-valued quantities $\mathbf{B}^T\mathbf{R}(z)\mathbf{B}$ and their limiting values, which can be expressed in terms of functionals of $F^{\mathbf{A}}$ (dependent on z) and the parameters $(\mathbf{\Theta}_j \colon j = 1, \ldots, J)$,

with the discrepancy measure between the empirical and limiting quantities then summed over a suitably dense set of values of $z \in \mathbb{C}^+$. Substituting the estimates of the ESDs of the coefficient matrices, this objective function is minimized with respect to the parameters $(\mathbf{\Theta}_j \colon j = 1, \ldots, J)$ to find their estimators. Details of the second estimation step are given in Section 3.2.

A few comments regarding the estimation procedure are in order. First, the MA representation of the linear time series formulation allows us to make meaningful approximations of the process by finite order MA processes. This kind of approximation is important for a stable implementation of the proposed algorithms and turns out to be quite effective, as is demonstrated through numerical studies. Second, the estimate of the ESD of $\mathbf{\Sigma}$ is derived from the estimate $F^{\mathbf{A}}$, even though the latter object could be a higher dimensional distribution if the assumed MA process is of order larger than one. Moreover, the quadratic form of interest, $\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B}$ is a lower dimensional estimation object. This suggests, and is supported by our numerical studies, that the estimation of the latter is simpler in the sense that even when the estimation of $F^{\mathbf{A}}$ is not very accurate, the estimate of the quadratic form could still remain quite precise. Moreover, when it comes to estimation of the quadratic form, there is some redundancy in that not all of the eigen-subspaces of Σ may contribute significantly to the object $\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B}$. This means that a model selection procedure choosing a set of significant $\mathbf{\Theta}_j$ can be more efficient. Indeed, such a model selection strategy is developed using the principle of cross validation. Finally, even though the description here focuses on the linear process formulation for the observed return, to deal with more realistic scenarios, the estimation strategy is extended to factor models whose idiosyncratic term follows the linear process structure described above. The corresponding description is given in Section 4.2 below.

3.1 Estimating the ESD of $\mathbf{\Sigma}$

The essence of step (1) is to invert equations (2.5) and (2.6). By this we mean that we observe $s_\tau(z)$ on the left-hand side of (2.5) and want to estimate $F^{\mathbf{A}}$ on the right-hand side. The following steps are involved in our estimation procedure.

(a) Discretization of $F^{\mathbf{A}}$. To estimate $F^{\mathbf{A}}$, use the discrete distribution

$$\hat{F}^{\mathbf{A}} = \sum_{j=1}^{J} \hat{\eta}_j \delta_{\alpha_j},$$

as an approximation, where $\hat{\eta}_j$ are weights, δ_{α_j} Dirac measures and J the number of grid points $\alpha_j \in \mathbb{R}^q$. For example, for an MA$(2)$ process with $q = 2$, the number of grid points is the product of the grid size for each coordinate corresponding to the MA coefficients. In this formulation, $\hat{\eta}_j$ are the objects to estimate.

(b) Picking a finite collection of $z \in \mathbb{C}^+$. Observe first that, for every given z, (2.5) and (2.6) constitute a separate system of equations. For fixed z and τ, compute as discrepancy measure the squared error loss

between the Stieltjes transform of the empirical and limiting ESDs, that is, the two sides of (2.5). Pick then a finite collection $\mathcal{D}$ of $z \in \mathbb{C}^+$ and a finite collection $\mathcal{T}$ of lag orders $\tau \geq 0$. The foregoing gives as optimization criterion the sum over $\mathcal{D}$ and $\mathcal{T}$ and the optimal $\hat{F}^{\mathbf{A}}$ is chosen as the one that minimizes this criterion.

The collection $\mathcal{D}$ is chosen to satisfy two criteria. The first criterion is that the real parts need to cover the range of empirical eigenvalues. The second criterion is that the imaginary parts need to be reasonably close to zero. The closer to zero the imaginary parts of z are, the more sensitive is the Stieltjes transform to the changes to the model parameters. However, when the imaginary parts are too close to zero, the Stieltjes transforms also become noisy. Thus, in the numerical studies, the real parts are chosen to be equally spaced over the range of the empirical distributions of the eigenvalues, while the imaginary parts are chosen to follow the geometric sequence of $1/4$, $1/2$, 1, 2 and 4.

(c) Reformulating (2.5) and (2.6) as a system of equations. To solve equations (2.5) and (2.6) for $F^{\mathbf{A}}$, solving for the intermediate object $K_\tau(z, \nu)$ defined in (2.6) is needed. For any given z, $K_\tau(z, \nu)$ is a function defined on $[0, 2\pi]$. Rewrite $K_\tau(z, \nu)$ so that it is a combination of a finite number of trigonometric functions. Thus, estimating the Stieltjes kernel $K_\tau(z, \nu)$ becomes equivalent to estimating a finite number of parameters. For example, for an MA(q) process, (2.6) can be written as

$$\begin{aligned} K_\tau(z, \nu) &= \int \frac{h(\lambda, \nu)}{\mathcal{M}_\tau(z, \lambda) - z} \, dF^{\mathbf{A}}(\lambda) \\ &= \sum_{\ell=0}^{q} \sum_{k=\ell+1}^{q} 2\cos((k-\ell)\nu) \int \frac{f_\ell(\lambda) f_k(\lambda)}{\mathcal{M}_\tau(z, \lambda) - z} \, dF^{\mathbf{A}}(\lambda) + \sum_{\ell=0}^{q} \int \frac{|f_\ell(\lambda)|^2}{\mathcal{M}_\tau(z, \lambda) - z} \, dF^{\mathbf{A}}(\lambda), \end{aligned}$$

where

$$\mathcal{M}_\tau(z, \lambda) = \frac{1}{2\pi} \int_0^{2\pi} \frac{\cos(\tau\theta) h(\lambda, \theta)}{1 + c\cos(\tau\theta) K_\tau(z, \theta)} d\theta, \tag{3.1}$$

and $z \in \mathbb{C}^+$. For $0 \leq \ell \leq q$ define

$$s_{\ell,\tau}(z) = \int \frac{|f_\ell(\lambda)|^2}{\mathcal{M}_\tau(z, \lambda) - z} \, dF^{\mathbf{A}}(\lambda),$$

and, for $0 \leq \ell < \ell' \leq q$, let

$$s_{\ell,\ell',\tau}(z) = s_{\ell',\ell,\tau}(z) = \int \frac{f_\ell(\lambda) f_{\ell'}(\lambda)}{\mathcal{M}_\tau(z, \lambda) - z} \, dF^{\mathbf{A}}(\lambda).$$

Then, (2.5) and (2.6) can be formulated as a system of equations involving $s_{\ell,\tau}(z)$ and $s_{\ell,\ell',\tau}(z)$, which are solved using Newton's method for any given value of $z \in \mathbb{C}^+$. The existence and uniqueness of the latter solution is guaranteed by the results in Liu, Aue, and Paul (2015).

3.2 Estimation of Quadratic Form

To accommodate step (2), use the fact that, for any $z \in \mathbb{C}^+$,

$$\mathbf{B}^T \mathbf{R}(z) \mathbf{B} \approx \sum_{j=1}^{J} \frac{1}{\mathcal{M}(z, \alpha_j) - z} \mathbf{\Theta}_j, \tag{3.2}$$

where J is the number of grid points defined in Section 3.1, $\mathbf{R}(z) = (\mathbf{S} - zI)^{-1}$, $\mathbf{\Theta}_j = \mathbf{B}^T\mathbf{P}_j\mathbf{B}$, and $\mathcal{M}(z, \alpha)$ is the kernel defined in (3.1), determined by the Stieltjes kernel describing the LSD of $\mathbf{S}$. It should be noted here that the approximation in (3.2) holds in a limiting sense, as articulated through the notion of *deterministic equivalent of the resolvent* $\mathbf{R}(z)$, the details of which can be found in Liu, Aue, and Paul (2015). Notice that $\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_J$ are symmetric (indeed, non-negative definite) and $\sum_{j=1}^{J} \mathbf{\Theta}_j = \mathbf{B}^T\mathbf{B}$. Equation (3.2) leads to the following estimation strategy for $\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B}$.

- *Step 1:* Let $\mathcal{D} \subset \mathbb{C}^+$ be the finite grid from Section 3.1. Then, estimate $\mathbf{\Theta} = (\mathbf{\Theta}_1 : \cdots : \mathbf{\Theta}_J)$ by

$$\hat{\mathbf{\Theta}} = \arg \min_{\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_J \in \mathcal{S}^{K \times K}} \sum_{z \in \mathcal{D}} \left\| \mathbf{B}^T\mathbf{R}(z)\mathbf{B} - \sum_{j=1}^{J} \frac{1}{\mathcal{M}(z, \alpha_j) - z} \mathbf{\Theta}_j \right\|_F^2, \tag{3.3}$$

  where $\mathcal{S}^{K \times K}$ denotes the class of $K \times K$ symmetric matrices and $\| \cdot \|_F$ Frobenius norm.

- *Step 2:* Estimate $\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B}$ by

$$\sum_{j=1}^{J} \frac{1}{\psi(\alpha_j, 0)} \hat{\mathbf{\Theta}}_j. \tag{3.4}$$

3.3 Thresholding and Model Selection

One challenge in the proposed algorithm is choosing how many grid points $\alpha_j \in \mathbb{R}^q$ to keep from step (1) of Section 3.1. On one hand, there should be sufficiently many grid points to ensure an accurate approximation of $F^{\mathbf{A}}$. On the other hand, there should be sufficiently few grid points so that their corresponding weights can be estimated with high precision. One natural solution is to threshold eigenvalues with weights below some pre-specified tuning parameter ξ. Below, an algorithm for a given value of ξ is described first and a discussion is then added on how to choose the tuning parameter based on a model selection approach.

Algorithm 3.1 (**Thresholding).** Perform the following three steps.

- *Thresholding.* Let $\xi \geq 0$ be a threshold. For each $j \in \{1, \ldots, J\}$, threshold the estimated weights $\hat{\eta}_j$ at ξ. Define $\mathcal{J}(\xi) = \{j \in \{1, \ldots, J\} \colon \hat{\eta}_j > \xi\}$ and drop the grid point α_j if $j \notin \mathcal{J}(\xi)$.

- *Reweighting.* For all $j \in \mathcal{J}(\xi)$, assign the updated weight

$$\hat{\eta}_j(\xi) = \frac{\hat{\eta}_j}{\sum_{j' \in \mathcal{J}(\xi)} \hat{\eta}_{j'}}.$$

  Then, the *thresholded estimate* of $F^{\mathbf{A}}$ is

$$\hat{F}_{\xi}^{\mathbf{A}} = \sum_{j \in \mathcal{J}(\xi)} \hat{\eta}_j(\xi) \delta_{\alpha_j}. \tag{3.5}$$

- *Estimation of* $\boldsymbol{\Theta}$*:* Estimate the redefined $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_j \colon j \in \mathcal{J}(\xi))$ by minimizing

$$\sum_{z \in \mathcal{D}} \left\| \mathbf{B}^T \mathbf{R}(z) \mathbf{B} - \sum_{j \in \mathcal{J}(\xi)} \frac{1}{\hat{\mathcal{M}}_\xi(z, \alpha_j) - z} \boldsymbol{\Theta}_j \right\|_F^2,$$

where $\hat{\mathcal{M}}_\xi(z, \alpha)$ is the kernel $\mathcal{M}_0(z, \alpha)$ in (3.1) determined by the distribution $\hat{F}_\xi^{\mathbf{A}}$. Here, the restriction that $\boldsymbol{\Theta}_j$, $j \in \mathcal{J}(\xi)$, are symmetric is imposed again. Finally, estimate $\mathbf{B}^T \boldsymbol{\Sigma}^{-1} \mathbf{B}$ by

$$\hat{\boldsymbol{\Phi}}(\xi) = \sum_{j \in \mathcal{J}(\xi)} \frac{1}{\psi(\alpha_j, 0)} \hat{\boldsymbol{\Theta}}_j(\xi). \tag{3.6}$$

The thresholding strategy outlined in Algorithm 3.1 could be performed for a sequence of nonnegative ξ (the maximal ξ being the point beyond which $\mathcal{J}(\xi)$ is the empty set). In practice, however, a particular value of the thresholding parameter ξ needs to be chosen. This is a model selection problem, and a simple cross-validation strategy is proposed to solve it. Accordingly, split the data $\{X_1, \ldots, X_T\}$ into two parts, a *training set* consisting of the first half of the observations $\{X_1, \ldots, X_{T/2}\}$ and a *test set* of the second half of the observations $\{X_{T/2+1}, \ldots, X_T\}$.

Algorithm 3.2 (**Model selection).** Perform the following five steps.

- Given a $\xi \geq 0$, obtain the estimate $\hat{F}_\xi^{\mathbf{A},\mathrm{Train}}$ based on the training data. Note that the dimension-to-sample size ratio needs to be adjusted to $2N/T$ rather than N/T while carrying out this procedure.
- Estimate $\boldsymbol{\Theta}_j$, $j \in \mathcal{J}^{\mathrm{Train}}(\xi)$, from the training data by the procedure described in Algorithm 3.1. Let the corresponding estimates be denoted by $(\hat{\boldsymbol{\Theta}}_j^{\mathrm{Train}}(\xi) \colon j \in \mathcal{J}^{\mathrm{Train}}(\xi))$.
- Let $\mathbf{S}_{0,\mathrm{Test}}$ denote the sample covariance matrix for the test data, and let $\mathbf{R}_{\mathrm{Test}}(z)$ be the corresponding resolvent. Let $\hat{\mathcal{M}}_\xi^{\mathrm{Test}}(z, \alpha)$ be the analog of $\mathcal{M}_0(z, \alpha)$ when using the ratio $2N/T$ in the computation for the LSD, while using $\hat{F}_\xi^{\mathbf{A},\mathrm{Train}}$ for the distribution. Compute then the *forward cross validation score*

$$\mathrm{CV}_f(\xi) = \sum_{z \in \mathcal{D}} \left\| \mathbf{B}^T \mathbf{R}_{\mathrm{Test}}(z) \mathbf{B} - \sum_{j \in \mathcal{J}^{\mathrm{Train}}(\xi)} \frac{1}{\hat{\mathcal{M}}_\xi^{\mathrm{Test}}(z, \alpha_j) - z} \hat{\boldsymbol{\Theta}}_j^{\mathrm{Train}}(\xi) \right\|_F^2.$$

- Flip the training data set and the test data set to calculate the *backward cross validation score* $\mathrm{CV}_b(\xi)$. Take the sum $\mathrm{CV}(\xi) = \mathrm{CV}_f(\xi) + \mathrm{CV}_b(\xi)$ as the score for model selection.
- Define ξ_{opt} to be the value of ξ for which $\mathrm{CV}_f(\xi)$ is minimized subject to the restriction that $\hat{\Phi}(\xi)$ defined in (3.6) is nonnegative definite. The latter restriction helps to further narrow down the scope of plausible models and thereby improves statistical efficiency.

Note that, although in the simulation study below $F^{\mathbf{A}}$ is set up as a discrete distribution, the proposed algorithm can be extended to continuous $F^{\mathbf{A}}$; for example through an approximation with a set of spline

functions. Hence, estimating a continuous $F^{\mathbf{A}}$ becomes equivalent to estimating the coefficients in the spline representation. This, however, can be implemented using numerical methods similar to the ones utilized for the discrete $F^{\mathbf{A}}$ under consideration here.

4 Extensions

4.1 Application to the Markowitz Portfolio Problem

As documented in El Karoui El Karoui (2010), the Markowitz portfolio problem can be formulated as a special case of the quadratic program

$$\min_{w\in\mathbb{R}^p} \frac{1}{2} w^T \mathbf{\Sigma} w \qquad \text{subject to } \; w^T v_k = u_k, \quad k = 1, \ldots, K. \tag{4.1}$$

Let $\mathbf{V} = [v_1 : \cdots : v_K]$ be the $p \times K$ matrix whose kth column is v_k and $U = (u_1, \ldots, u_K)^T$ the K-dimensional vector whose kth entry is u_k. Define the $K \times K$ matrix

$$\mathbf{Q} = \mathbf{V}^T \mathbf{\Sigma}^{-1} \mathbf{V}, \tag{4.2}$$

assuming that the columns of $\mathbf{V}$ are such that $\mathbf{Q}$ is invertible. The solution of the quadratic program with linear equality constraints (4.1) is

$$w_{\text{optimal}} = \mathbf{\Sigma}^{-1} \mathbf{V}^T \mathbf{Q}^{-1} \mathbf{V} \tag{4.3}$$

and

$$w_{\text{optimal}}^T \Sigma w_{\text{optimal}} = U^T \mathbf{Q}^{-1} U. \tag{4.4}$$

The Markowitz portfolio problem fits into the above framework of quadratic programs. Its aim is to find the minimum risk that one has to absorb in order to achieve an expected portfolio return. For each expected portfolio return, there is one minimum risk determined by the smallest variance of portfolio returns, so it is standard to plot the variance of portfolio returns against the expected portfolio returns, as in Figures 1 and 2. The Markowitz portfolio problem has two linear constraints, so that $K = 2$. The first constraint is that the total weight is 1, while the second constraint is that the expected portfolio return is equal to μ_P, that is,

$$w^T e = 1,$$
$$w^T \mu = \mu_P,$$

where e is a p-dimensional vector whose entries are all equal to 1, μ is a vector of expected asset returns in the portfolio and μ_P is a particular expected portfolio return. Consequently, the respective quantities in (4.3) and (4.4) become

$$U_P = \begin{bmatrix} 1 \\ \mu_P \end{bmatrix} \qquad \text{and} \qquad \mathbf{V} = [\, e : \mu \,],$$

noting that the dependence of U on μ_P will be expressed by the subscript P. Expected returns μ are typically unknown and commonly estimated by the sample mean $\hat{\mu}$. For each fixed μ_P, the goal is then to estimate the minimum risk in (4.4). The algorithm for this is listed in the following.

(1) Since the matrix $\mathbf{V}$ involves unknown expected asset returns, the first step is to calculate $\hat{\mathbf{V}} = [\, e : \hat{\mu}\,]$, where $\hat{\mu}$ is the sample mean.

(2) The proposed algorithm provides the estimate $\hat{\mathbf{Q}} = \hat{\mathbf{V}}^T \hat{\mathbf{\Sigma}}^{-1} \hat{\mathbf{V}}$ for $\mathbf{Q}$ in (4.2).

(3) The estimate for the minimum risk in (4.4) is $U_P^T \hat{\mathbf{Q}}^{-1} U_P$.

Note that, for each expected portfolio return, a minimum risk needs to be estimated. However, all estimated minimum risks share the same $\hat{\mathbf{Q}}$, which hence has to be estimated only once.

4.2 Extension to Factor Models

In this section, an extension of the proposed algorithm to incorporate factor structures is discussed. In the framework to be introduced, it is assumed that there are M known factors, the context being situations for which the leading factors in asset returns (for example the market return factor, "small minus big" and "high minus low" factors in the Fama–French three-factor model, see Fama and French (1993)) can reasonably be considered as known. When the leading factors are unknown, they can be estimated from the leading eigenvectors of the sample covariance matrix with well established procedures described in the literature; for example, in Onatski (2009, 2010). Focusing here on the known factors framework, asset returns may be written as

$$Y_t = \sum_{m=1}^{M} L_m f_{m,t} + X_t, \qquad X_t = \sum_{\ell=0}^{\infty} \mathbf{A}_\ell Z_{t-\ell}, \qquad t \in \mathbb{Z},$$

where $f_{m,t}$ is the observable return of factor m at time t, and L_m the corresponding unknown $N \times 1$ factor loading, $m = 1, \ldots, M$. Assume that $\mathbf{L} = [L_1 : \cdots : L_M]$ is orthogonal to the eigenvectors of the linear process coefficient matrices $(\mathbf{A}_\ell \colon \ell \in \mathbb{N})$. The estimation procedure for the quadratic form $\mathbf{B}^T \mathbf{\Sigma}^{-1} \mathbf{B}$ may be adjusted to this factor model setting in the following way.

(1) Estimate the matrix of factor loadings $\mathbf{L}$ by regressing Y_t on the factor vector $\mathbf{f}_t = (f_{1,t}, \ldots, f_{M,t})^T$. This yields the least squares estimate $\hat{\mathbf{L}} = [\hat{L}_1 : \cdots : \hat{L}_M]$.

(2) Estimate the covariance matrix of factor returns by $\sum_{m=1}^{M} \hat{\sigma}^2(f_m) \hat{L}_m \hat{L}_m^T$, where $\hat{L}_m$ is the estimated loading vector on factor m and $\hat{\sigma}(f_m)$ its estimated standard deviation. In matrix notation, the estimated covariance matrix for the factor structure is therefore

$$\hat{\mathbf{L}} \hat{\mathbf{\Delta}} \hat{\mathbf{L}}^T,$$

where $\hat{\mathbf{\Delta}} = \operatorname{diag}(\hat{\sigma}(f_1), \ldots, \hat{\sigma}(f_M))$ is an $M \times M$ diagonal matrix.

(3) Note that

$$\mathbf{B}^T\boldsymbol{\Sigma}^{-1}\mathbf{B} = \mathbf{B}^T\left(\mathbf{L}(\mathbf{I}_M+\boldsymbol{\Delta})^{-1}\mathbf{L}^T + \left(\mathbf{P}_L^{\perp} + \sum_{\ell=0}^{\infty}\mathbf{A}_\ell\mathbf{A}_\ell'\right)^{-}\right)^{-1}\mathbf{B}$$
$$= \mathbf{B}^T\mathbf{L}(\mathbf{I}_M+\boldsymbol{\Delta})^{-1}\mathbf{L}^T\mathbf{B} + \mathbf{B}_L^T\left(\mathbf{I}_N + \sum_{\ell=0}^{\infty}\mathbf{A}_\ell\mathbf{A}_\ell'\right)^{-1}\mathbf{B}_L, \tag{4.5}$$

where $\mathbf{P}_L^{\perp} = \mathbf{I}_N - \mathbf{L}\mathbf{L}^T$, with $\mathbf{C}^-$ denoting the Moore–Penrose generalized inverse of the symmetric matrix $\mathbf{C}$, and $\mathbf{B}_L = \mathbf{P}_L^{\perp}\mathbf{B}$. In the above calculation it was used that, since $\mathbf{P}_L^{\perp}\mathbf{P}_L^{\perp} = \mathbf{P}_L^{\perp}$ and $\mathbf{P}_L^{\perp}(\sum_{\ell=0}^{\infty}\mathbf{A}_\ell\mathbf{A}_\ell')\mathbf{P}_L^{\perp} = \sum_{\ell=0}^{\infty}\mathbf{A}_\ell\mathbf{A}_\ell'$, we have $(\mathbf{P}_L^{\perp}+\sum_{\ell=0}^{\infty}\mathbf{A}_\ell\mathbf{A}_\ell')^- = \mathbf{P}_L^{\perp}(\mathbf{I}_N+\sum_{\ell=0}^{\infty}\mathbf{A}_\ell\mathbf{A}_\ell')^{-1}\mathbf{P}_L^{\perp}$.

(4) Using the calculations in (3), the first term in (4.5) concerning the factor structure can be estimated by $\mathbf{B}^T\hat{\mathbf{L}}(\mathbf{I}_K+\hat{\boldsymbol{\Delta}})^{-1}\hat{\mathbf{L}}^T\mathbf{B}$. The second term concerning with the idiosyncratic time series component can be estimated using the estimation strategy for quadratic forms introduced in the previous sections.

5 Empirical Results

In this section, results are reported of a set of simulation studies designed to show the effectiveness of the proposed methodology for estimating the MVF in the time series framework presented. This is done in two settings: (i) when the observations follow a stationary linear process model, following the basic structural assumptions outlined in Section 2.1; and (ii) when the observations follow a factor model structure with known factors, and the idiosyncratic term constitutes a linear process satisfying analogous assumptions. In addition, we demonstrate the effectiveness of the proposed strategy for estimating the eigenvalue distributions of the coefficient matrices of the linear process, even though the latter is not the primary concern of this paper.

Section 5.1 reports results for MVF estimation as well as estimation of eigenvalue distribution of coefficient matrices when the underlying data generating process is an MA(2) time series. Section 5.2 reports results for MVF estimation when the underlying data generating process is an AR(1) time series. In both cases, the simulated processes are estimated as an MA(2) process, leading to a model misspecification for the latter case. Finally, Section 5.3 demonstrates the performance of MVF estimation within the factor model framework of Section 4.2.

The performance of the proposed method, referred to as *LinShrink* (standing for shrinkage under linear process structure), is compared to two alternative approaches. The first one, referred to as *Naive Estimate*, is based on simply replacing the population covariance with the sample covariance. The second method is the shrinkage approach based on independent observations, as proposed in El Karoui (2010), and is referred to as *IndShrink*.

5.1 Estimation for MA(2) Processes

First a description of the simulation setting is given. We considered the setting, where the two MA(2) coefficient matrices, A_1 and A_2, were symmetric and simultaneously diagonalizable, with the joint bivariate spectral distribution given in Table 1 below. The common eigenbasis for A_1 and A_2 was chosen to be uniformly distributed on the space of $N \times N$ orthogonal matrices. The innovation process consisted of i.i.d. vectors with independent standard normal entries.

Eigenvalue pair	Probability
$(0.1, 0.2)$	0.3
$(0.4, 0.5)$	0.3
$(0.7, 0.9)$	0.4

Table 1: Joint eigenvalue distribution of (A_1, A_2) for the MA(2) processes of Section 5.1.

We consider two scenarios with $N/T = 1/3$, namely *Case 1:* $N = 1000, T = 3000$; *Case 2:* $N = 2000$, $T = 6000$. In each case, a square grid $\{(i/10, j/10) \colon i, j = 1, \ldots, 10\}$ was used as candidate eigenvalue pairs for the estimation of the joint empirical eigenvalue distributions of A_1 and A_2. For the MVF estimation, the expected return μ_P was chosen to be the $N \times 1$ vector with j-th coordinate $\mu_j = 1 + j/N$ for all $j = 1, \ldots, N$.

The result of the MVF estimation in *Case 1* are displayed in Figure 3. The left panel shows the plot of the return variance against the expected return, while the right panel shows the frequency distribution of the number of eigenvalue pairs receiving positive weights under the thresholding-based model selection approach described in Algorithms 3.1 and 3.2. The true MVF is indicated by the solid green line, while the pointwise mean of LinShrink, IndShrink and Naive Estimate are shown in solid blue, magenta and red, respectively. The broken curves of the respective colors represent $+/-$ one standard deviation bands. It can be seen that the LinShrink estimate is nearly unbiased, exhibiting slightly higher variability than the IndShrink estimate and the Naive Estimate, both of which are biased. The IndShrink estimate has smaller bias than the Naive Estimate, indicating that simply taking into account the dimensionality effect, while ignoring the temporal correlation, can still improve the estimate. The model selection performance shows that the proposed method largely avoids overfitting.

Results for the MVF estimation for *Case 2* are reported in Figure 4. Even though the sample size is twice as large as before, the only noticeable qualitative change in the performance of the LinShrink method is a slight reduction in variability and essentially zero bias, as reflected by the near complete overlap of the blue and green solid lines. Also, the estimation bias the two competing methods does not decrease. This is expected, since the dimension-to-sample size ratio remains the same. The model selection performance of the proposed method seems to improve slightly, with fewer models possessing a larger number of eigenvalue pairs getting selected compared with *Case 1*.

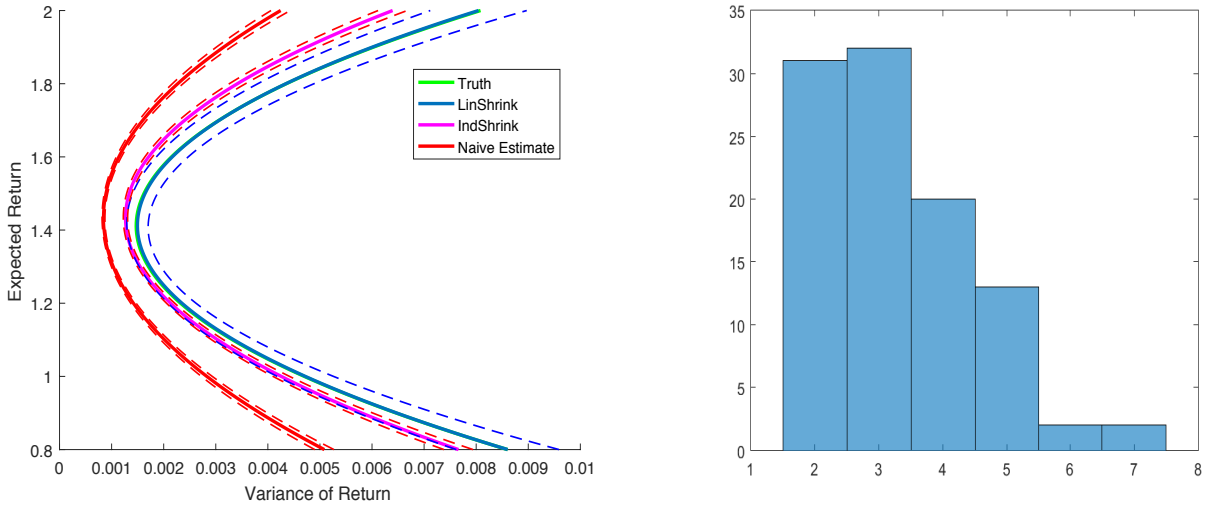

Figure 3: MVF estimation for an MA(2) process with $N = 1000$, $T = 3000$. **Left panel:** True and estimated MVF (solid line), with one standard deviation band (broken line). LinShrink: blue; IndShrink: magenta; Naive estimate: red. **Right panel:** Frequency distribution of the number of eigenpairs chosen by the model selection procedure.

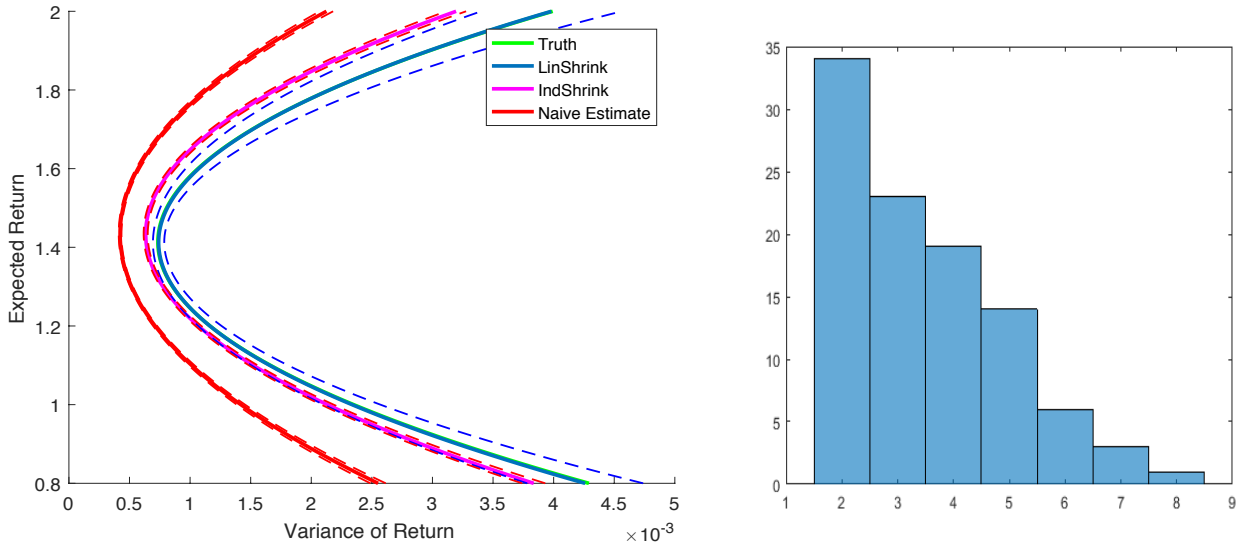

Figure 4: MVF estimation for an MA(2) process with $N = 2000$, $T = 6000$. **Left panel:** True and estimated MVF (solid line), with one standard deviation band (broken line). LinShrink: blue; IndShrink: magenta; Naive estimate: red. **Right panel:** Frequency distribution of the number of eigenpairs chosen by the model selection procedure.

Eigenvalue	Probability
0.1	0.3
0.2	0.3
0.4	0.4

Table 2: Eigenvalue distribution of A for the AR(1) processes of Section 5.2.

The true and estimated cumulative distribution functions (CDFs) of the marginal eigenvalue distributions of the coefficient matrices A_1 and A_2 in *Case 2* are displayed in Figure 5. It can be seen from these plots that the estimated marginal CDFs are quite reasonable approximations to the true marginal CDFs.

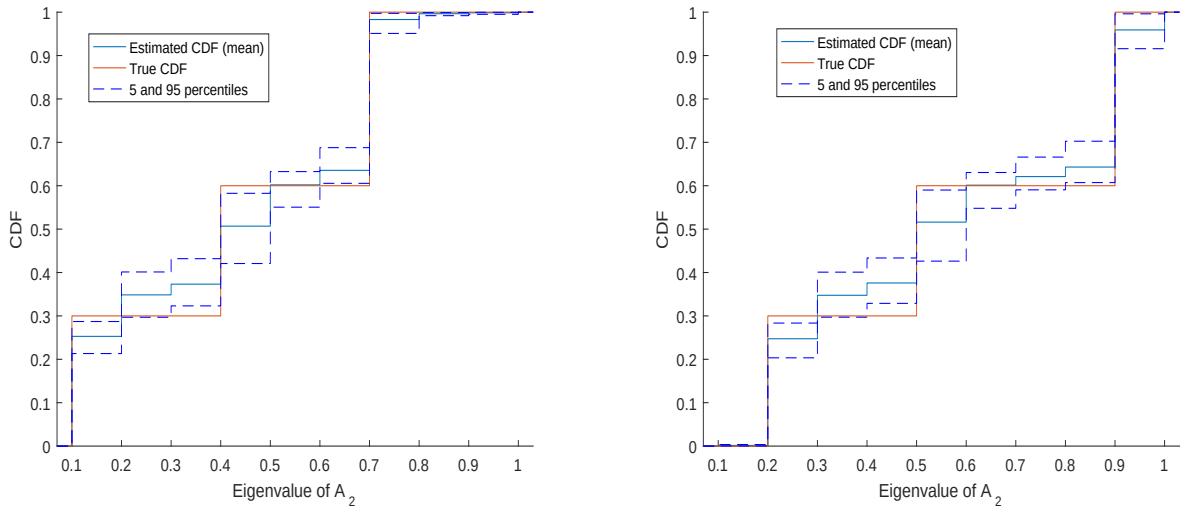

Figure 5: Estimation of the CDF of the eigenvalue distribution of coefficient matrices for the MA(2) process with $N = 2000, T = 6000$. True CDF: solid red; Mean of estimated CDF: solid blue; 5-th and 95-th pointwise percentiles of estimated CDF: broken blue. **Left panel:** CDF of eigenvalue distribution of A_1; **Right panel:** CDF of eigenvalue distribution of A_2.

5.2 Estimation for AR(1) Processes

This section considers the case of simulated AR(1) process, with the symmetric AR coefficient matrix A having the eigenvalue distribution given in Table 2. The innovation process was again chosen to be i.i.d. with independent standard normal entries.

Recall that the proposed method is designed to estimate the eigenvalue distribution of MA processes. Therefore, this example was used as test case for the effect of model misspecification. In the proposed MVF estimation procedure, the AR(1) process was approximated by an MA(2) process. This may be viewed as fitting to a truncated MA(∞) representation of the AR(1) process and can be interpreted as a model class restriction with the aim of securing an increased estimation efficiency. Accordingly, the MA coefficients in

this MA(2) representation were set as $A_1 = B$ and $A_2 = B^2$, for some unknown symmetric matrix B. Thus, while estimating the joint eigenvalue distribution of (A_1, A_2), the domain of the eigenvalue pairs was restricted to the grid $\{(j/10, (j/10)^2) : j = 1, \ldots, 10\}$. The parameter μ_P (the expected return) was chosen to be the same as in Section 5.1. The results for the MVF estimation are displayed in Figure 6. What is apparent from this plot is that, now all three estimators — LinShrink, IndShrink and Naive Estimate — are biased, though the bias in the proposed LinShrink method is significantly smaller compared to the other two methods. This bias in the proposed method is due to the bias in the approximation of the true AR(1) model by an MA(2) model. However, the relative performance of IndShrink and LinShrink, both of which are computed under model misspecification, also indicates that, with a finer approximation, for example by increasing the order of the MA process, the proposed method should be able to obtain a nearly unbiased estimator of the MVF.

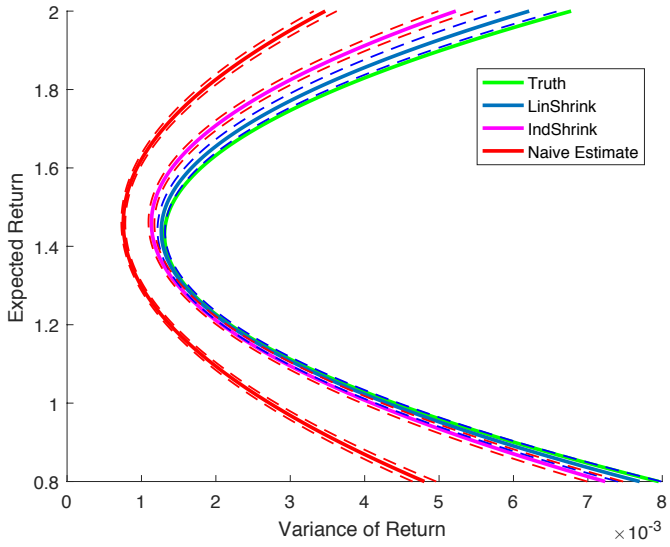

Figure 6: MVF estimation for an AR(1) process with $N = 1000$, $T = 3000$. True and estimated MVF (solid line), with one standard deviation band (broken line). LinShrink: blue; IndShrink: magenta; Naive estimate: red.

5.3 Estimation for Factor Models

In the final experiment, the returns were assumed to follow the factor model

$$Y_t = \mu_t + f_{1t}\mathbf{a}_1 + f_{2t}\mathbf{a}_2 + f_{3t}\mathbf{a}_1 + \varepsilon_t, \tag{5.1}$$

where $\mu_t = 1$; $\mathbf{a}_k$, $k = 1, \ldots, 3$, were three randomly generated, mutually orthonormal vectors; and, for each k, f_{kt} had i.i.d. $N(\nu_k, \sigma_k^2)$ entries with $\nu_k = 1.5$ for all k, and $\sigma_1 = 0.5$, $\sigma_2 = 0.625$ and $\sigma_3 = 0.7$. The ε_k were set up as the MA(2) process of Section 5.1, independent of the f_{kt}. Estimates of the eigenvalue

distributions of the MA(2) coefficients were obtained from an application of the proposed spectrum estimation procedure to the residuals from the regression fit for $Y_1, \ldots, Y_T$, using model (5.1), and treating the f_{kt} as known. In the next step, the procedure described in Section 4.2 was applied to estimate the MVF, using estimated mean vectors. The result is shown in Figure 7. The relative ordering of the three methods, in terms of degree of MVF estimation bias, are similar as in the settings considered previously. The difference between the estimates based on IndShrink and LinShrink, however, is smaller compared to the pure MA(2) time series case. The latter is due to the concentration of information in a few coordinates of the covariance matrix $\mathbf{\Sigma}$ as a result of the strong factor structure.

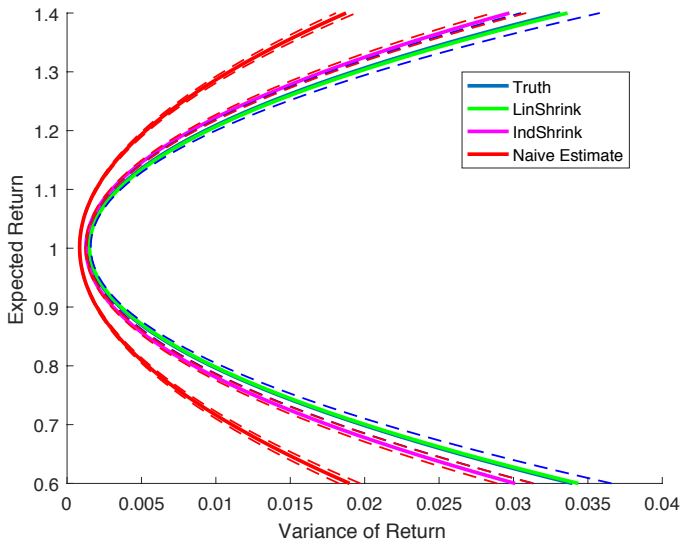

Figure 7: MVF estimation for a factor model with MA(2) idiosyncratic terms and $N = 1000$, $T = 3000$. True and estimated MVF (solid line), with one standard deviation band (broken line). LinShrink: blue; IndShrink: magenta; Naive Estimate: red.

5.4 Conclusions

One important conclusion that can be derived from this numerical study is that even when the estimation of coefficient spectra of the constituent linear process is not perfect — an admittedly difficult goal to achieve—, the quality of MVF estimation is still rather impressive, even when there is a model mismatch as in Section 5.2. More importantly, the results demonstrate the benefit of addressing the temporal dependence explicitly, as is done in the proposed methodology, by way of obtaining a nearly unbiased estimate of the MVF when the model is correct. This is a significant improvement over the empirical MVF (based on using the sample covariance matrix) and representative methods that only take into account spatial (coordinatewise) dependence, but do not address dependence across time.

6 Discussion

In this paper, the model for high-dimensional, stationary time series proposed in Liu, Aue, and Paul (2015) is utilized to address the question of MVF estimation within the framework of Markowitz portfolio optimization, with returns exhibiting time series structure. This investigation is motivated by the established fact that the empirical MVF estimate is significantly biased when the dimensionality of the returns is not negligible when measured against sample size. The proposed method relies on the characterization of the limiting behavior of the empirical eigenvalue distributions of symmetrized sample autocovariance matrices of the time series. We draw on the characterization established in Liu, Aue, and Paul (2015) to formulate an estimator of the eigenvalue distribution of the coefficient matrices for high-dimensional MA processes with simultaneously diagonalizable coefficients. These estimates are then utilized as input in an optimization procedure aimed at estimating quadratic forms involving the inverse population covariance matrix of the time series. The latter allows for the computation of MVF estimates in two scenarios: (i) when the time series of returns belong to the class of linear processes considered in Liu, Aue, and Paul (2015); and (ii) when the returns follow a factor structure with known, finite-dimensional factors, and the idiosyncratic terms are orthogonal to the factor space and follow the linear process model as in (i).

A set of numerical simulations was performed as a proof-of-concept, making performance comparisons between the proposed MVF estimator and the empirical estimator and the state-of-the-art MVF estimator proposed by El Karoui (2010). The empirical results demonstrate superior performance of the proposed method in terms of reducing the MVF estimation bias, thereby pointing to the key role played by the time series structure and its utilization in the proposed method.

There are several directions in which the current work can be extended. First, we aim to enhance the estimation method for quadratic forms by implementing modifications that will allow us to tackle ARMA-type linear processes of a given order, where the coefficients satisfy a simultaneous diagonalizability condition. Secondly, we would like to extend this estimation procedure to the setting of factor models with relatively strong, but unknown factors, with time-dependent idiosyncratic terms. Thirdly, we aim to carry out an extensive analysis based on stock price data to obtain a meaningful estimate of the MVF and to quantify its uncertainty. Finally, we would like to carry out a thorough mathematical investigation aiming to establish consistency of the proposed procedure for finite order ARMA-type time series with the required structural assumptions.

Acknowledgement

The research of Paul and Aue is partially supported by the National Science Foundation grant DMS-1407530.

References

AO, M., Y. LI, AND X. ZHENG (2016): "Solving the high-dimensional Markowitz optimization problem: When sparse regression meets random matrix theory," Discussion paper, Hong Kong University of Science and Technology, Technical Report.

BAI, J. (2003): "Inferential theory for factor models for large dimensions," *Econometrica*, 71, 135–171.

BAI, J., AND S. NG (2002): "Determining the number of factors in approximate factor models," *Econometrika*, 70, 191–221.

BAI, Z., H. LIU, AND W.-K. WONG (2009): "On the Markowitz meanvariance analysis of self-financing portfolios," *Risk and Decision Analysis*, 1(1), 35–42.

BAI, Z. D., D. CHEN, AND J.-F. YAO (2010): "On estimation of the population spectral distribution from a high-dimensional sample covariance matrix," *Australian and New Zealand Journal of Statistics*, 52, 423–437.

BAI, Z. D., AND J. W. SILVERSTEIN (2010): *Spectral Analysis of Large Dimensional Random Matrices*. Springer, New York.

BARBERIS, N. (2000): "Investing for the long run when returns are predictable," *The Journal of Finance*, 55(1), 225–264.

BOUCHAUD, J.-P., AND M. POTTERS (2009): "Financial applications of random matrix theory : a short review," Discussion paper, arXiv:0910.1205.

BRODIE, J., I. DAUBECHIES, C. DE MOL, D. GIANNONE, AND I. LORIS (2009): "Sparse and stable Markowitz portfolios," *Proceedings of the National Academy of Sciences*, 106, 12267–12272.

CHOPRA, V. K., AND W. T. ZIEMBA (1993): "The effect of errors in means, variances, and covariances on optimal portfolio choice," .

COCHRANE, J. (2005): *Asset Pricing*. Princeton University Press.

DEMIGUEL, V., L. GARLAPPI, F. J. NOGALES, AND R. UPPAL (2009): "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms," .

EL KAROUI, N. (2008): "Spectrum estimation for large dimensional covariance matrices using random matrix theory," *The Annals of Statistics*, 36(6), 2757–2790.

——— (2010): "High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation," .

——— (2013): "On the realized risk of high-dimensional Markowitz portfolios," *SIAM Journal on Financial Mathematics*, 4(1), 737–783.

FAMA, E. F., AND K. R. FRENCH (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33(1), 3 – 56.

FAN, J., J. ZHANG, AND K. YU (2012): "Vast portfolio selection with gross-exposure constraints," .

JAGANNATHAN, R., AND T. MA (2003): "Risk reduction in large portfolios: Why imposing the wrong constraints helps," .

KAN, R., AND G. ZHOU (2007): "Optimal Portfolio Choice with Parameter Uncertainty," *The Journal of Financial and Quantitative Analysis*, 42(3), 621–656.

LALOUX, L., P. CIZEAU, J.-P. BOUCHAUD, AND M. POTTERS (1999): "Noise dressing of financial correlation matrices," .

LALOUX, L., P. CIZEAU, M. POTTERS, AND J.-P. BOUCHAUD (2000): "Random Matrix Theory and Financial Correlations," .

LEDOIT, O., AND M. WOLF (2012): "Nonlinear shrinkage estimation of large-dimensional covariance matrices," *The Annals of Statistics*, 40, 1024–1060.

——— (2015): "Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions," *Journal of Multivariate Analysis*, 139, 360–384.

LIU, H., A. AUE, AND D. PAUL (2015): *The Annals of Statistics*(2), 675–712.

MARKOWITZ, H. (1952): "Portfolio Selection," *The Journal of Finance*, 7, 77–91.

ONATSKI, A. (2009): "Testing hypotheses about the number of factors in large factor models," *Econometrica*, 77, 1447–1479.

——— (2010): "Determining the number of factors from empirical distribution of eigenvalues," *Review of Economics and Statistics*, 92, 1004–1016.

——— (2012): "Asymptotics of the principal components estimator of large factor models with weakly influential factors," *Journal of Econometrics*, 168, 244–258.

PAUL, D., AND A. AUE (2014): "Random matrix theory in statistics: A review," *Journal of Statistical Planning and Inference*, 150, 1 – 29.