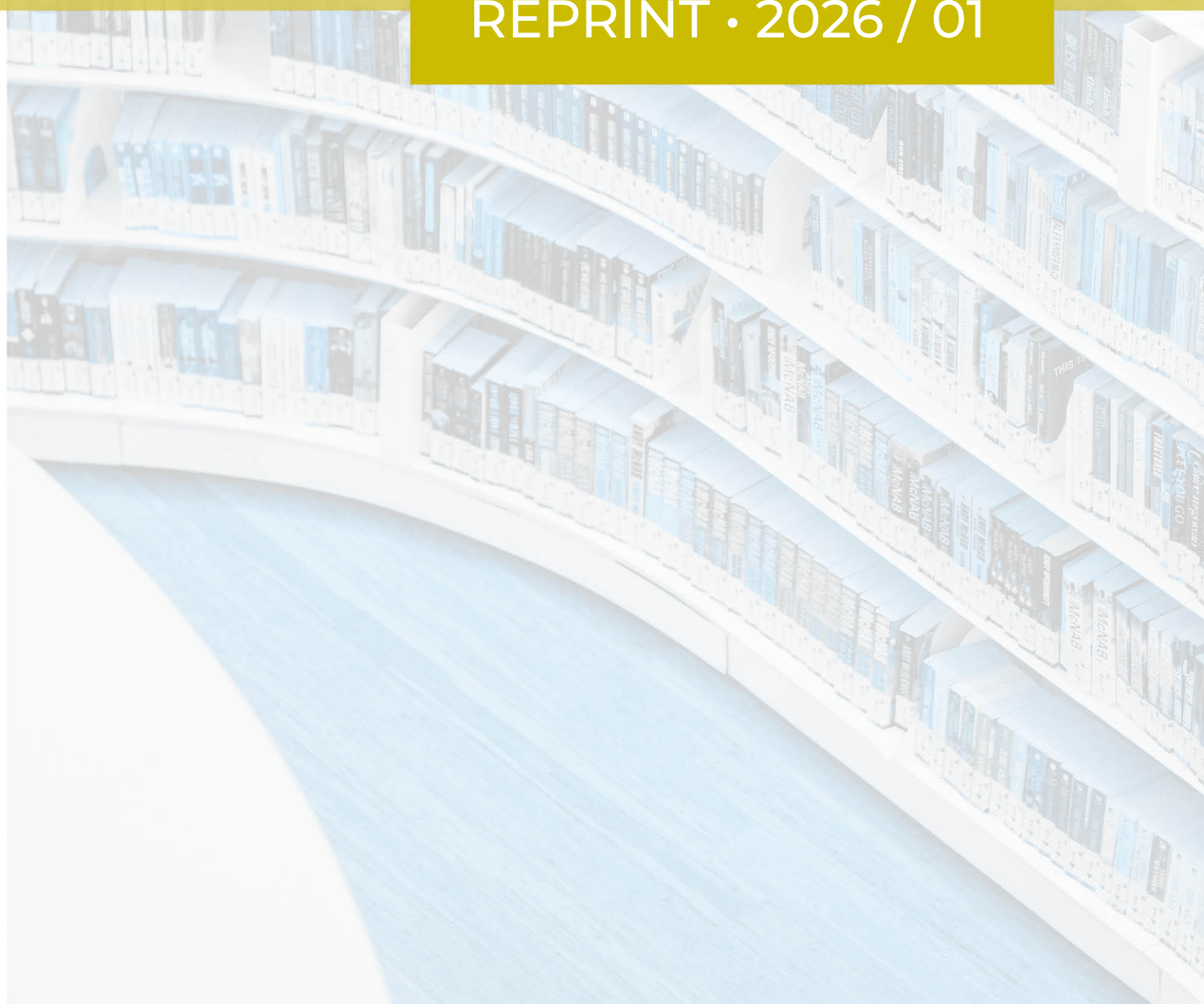
DISTRIBUTION-FREE SHRINKAGE OF HIGH-DIMENSIONAL MEAN VECTOR

Vali Asimit, Ziwei Chen, Nathan Lassance

REPRINT • 2026 / 01



UCLouvain

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

LFIN

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

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

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

Distribution-free shrinkage of high-dimensional mean vector*

Vali Asimit[1] Ziwei Chen[1] Nathan Lassance[2]

Forthcoming at *Journal of Business & Economic Statistics*

February 22, 2026

Abstract

We introduce shrinkage estimators of the sample mean vector in high dimension. Our estimators share desirable properties relative to existing methods: they are distribution-free, consider bona-fide target estimators, and use simple estimators of the shrinkage intensities independent of the precision matrix which are L_2 consistent in high dimension. Unlike existing estimators that impose that the whitened data be an i.i.d. matrix, we only impose i.i.d. across sample observations. We require uniform boundedness of the first four moments, and the high-dimensional asymptotics are in a general Kolmogorov setting where $N/T = O(1)$ as $T \to \infty$, with N the mean-vector dimension and T the sample size. We consider as a target estimator either zero or the grand mean. Simulations show that our shrinkage estimators are competitive with a range of benchmark estimators, both when the theoretical assumptions are satisfied and violated. Finally, we apply our estimators to constructing mean-variance portfolios of a large number of stocks, and find that they deliver robust out-of-sample Sharpe ratios.

Keywords: bona-fide estimator, high-dimensional asymptotics, mean-variance portfolio, mean-vector estimation, shrinkage estimation.

JEL classification: C13, C14, C55, G11.

*Nathan Lassance is corresponding author. (1) Bayes Business School, City St George's, University of London, United Kingdom. (2) LIDAM/LFIN, UCLouvain, Belgium. Email addresses: alexandru.asimit.1@citystgeorges.ac.uk, ziwei.chen.3@citystgeorges.ac.uk, nathan.lassance@uclouvain.be. Matlab, R, and Python codes implementing our estimators are available at github.com/Ziwei-ChenChen/shrinkage_mean_estimators.

1 Introduction

Modern high-dimensional statistics has reconsidered classic statistical problems when both the dimension N and the sample size T become large. For example, a classic result in robust linear regression is that we obtain a lower prediction error if least absolute deviations replaces ordinary least squares when the number of predictors N is either fixed or asymptotically negligible relative to T. However, in high-dimensional settings where N and T increase at the same rate, this result is reversed (Bean et al. 2013). Similarly, the standard unbiased sample covariance-matrix estimator is no longer effective for large N, and the celebrated shrinkage covariance estimator of Ledoit & Wolf (2004) rectifies its drawbacks. Similar problems occur when estimating the mean vector via the usual sample unbiased mean estimator (hereafter, sample mean) in high dimension (Bodnar et al. 2019).

Our objective in this paper is to design improved shrinkage estimators of the sample mean vector in high-dimensional settings. Since Stein (1956) and James & Stein (1961), it is well-known that the sample-mean estimator is inadmissible and that shrinkage toward a given target estimator helps reduce its mean squared error (MSE). This result sparked a first era of James-Stein and Bayesian shrinkage estimators of the mean vector that are parametric (often under the normal or elliptical distribution) and with finite-sample guarantees; see Thompson (1968), Efron & Morris (1972, 1973), Berger & Bock (1976), Gleser (1986), Jorion (1986), Green & Strawderman (1991), and Fourdrinier et al. (2003).

More recently, progress in random matrix theory started a second era of shrinkage estimators that are distribution-free and focus on high-dimensional guarantees, i.e., N is large; see Bodnar et al. (2022) for a review. Chételat & Wells (2012) design improved shrinkage estimators of the sample mean vector when $N > T$ and both N and T are fixed, but under multivariate normality. Wang et al. (2014) design a distribution-free shrinkage estimator of the sample mean toward the unit vector, where the shrinkage intensities are estimated with U-statistics and improvement over the sample mean is guaranteed as $N \to \infty$. Xie et al.

(2016) introduce parametric and semiparametric shrinkage estimators of the sample mean vector toward an optimized target estimator or the grand mean that are optimal as $N \to \infty$. However, they assume independent sample means and a parametric family of distributions. Bodnar et al. (2019) introduce a distribution-free shrinkage estimator of the sample mean toward an arbitrary constant target that they apply to predicting the equally weighted portfolio expected return, where the bona-fide shrinkage intensities are almost-surely consistent under the usual Kolmogorov asymptotics, i.e., $N/T \to c > 0$ ($c \neq 1$) as $T \to \infty$.

In this context, we introduce shrinkage estimators of the sample mean vector, toward zero or the grand mean, which are applied to finance and share desirable properties relative to these existing methods. First, they are distribution-free. We only impose the usual i.i.d. assumption on the observations, and uniform boundedness of the first four moments. In contrast, Chételat & Wells (2012) and Xie et al. (2016) directly impose a parametric assumption on the distribution, and Wang et al. (2014) and Bodnar et al. (2019) assume a stochastic representation for the data that is more restrictive than ours. Indeed, whereas we impose the standard i.i.d. assumption across sample observations, they impose that the whitened data be an i.i.d. matrix, thereby restricting the stochastic structure. Specifically, they use a stochastic representation for the distribution of the i.i.d. observations $\mathbf{X}_i = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\epsilon}_i + \boldsymbol{\mu}$, $i = 1, \ldots, T$, where $\boldsymbol{\epsilon}_i$ is a random vector of size N whose entries are i.i.d. with zero mean and unit variance (as Bodnar et al. (2019) explain, the independence between $\boldsymbol{\epsilon}_i$ and $\boldsymbol{\epsilon}_j$ can be partially relaxed in their high-dimensional asymptotics, but the more restrictive assumption is the entries of each $\boldsymbol{\epsilon}_i$ being i.i.d.). This assumption is more restrictive than only assuming an i.i.d. sample, which is sufficient for our focus on L_2 convergence instead of almost-sure convergence. Finally, unlike Xie et al. (2016), we do not assume that the sample means are independent because we do not restrict the covariance matrix to be diagonal.

Second, we shrink toward a feasible bona-fide target estimator. Specifically, we shrink the sample mean toward zero or the grand sample mean. The latter is the optimal choice under the MSE, and the resulting oracle shrinkage intensity accounts for estimation errors

in the grand mean. In comparison, Wang et al. (2014), Xie et al. (2016), and Bodnar et al. (2019) optimize the shrinkage intensity and the target estimator at the same time, so that the oracle shrinkage intensity is optimal for the oracle unfeasible target. This idea of using a sequential two-step approach, where the target estimator is optimized and replaced with a bona-fide estimator before deriving the oracle shrinkage intensity, instead of the usual one-step approach, is advocated in Lassance (2025). In particular, Lassance (2025) considers shrinking the sample mean vector under the invariant loss as one application, and finds that the two-step approach can significantly decrease the loss when N and T are close.

Third, our bona-fide shrinkage intensities are simple plug-in estimators that do not depend on the precision matrix (i.e., inverse covariance matrix) and satisfy high-dimensional L_2 consistency in a general Kolmogorov asymptotics where $N/T = O(1)$ as $T \to \infty$. We use plug-in counterparts to the oracle finite-sample shrinkage intensities, unlike the use of U-statistics in Wang et al. (2014) and counterparts to the oracle asymptotic shrinkage intensities in Bodnar et al. (2019). The shrinkage intensities do not depend on the precision matrix, which carries substantial estimation errors when N/T is large, because our loss function is the MSE as in Xie et al. (2016), unlike Chételat & Wells (2012), Wang et al. (2014), and Bodnar et al. (2019) who use the invariant loss (Wang et al. (2014) use the quadratic loss where the weighting matrix is the identity matrix or the diagonal precision matrix, i.e., the invariant loss under uncorrelated variables). Both choices are common, but the MSE is justified when the N variables have a comparable scale, which is the case for stock returns in our empirical analysis. Finally, our bona-fide shrinkage intensities satisfy high-dimensional Kolmogorov L_2 consistency. Specifically, the bona-fide shrinkage mean estimator, its MSE, and the sample estimate of the MSE are all L_2-consistent estimators of their oracle counterpart.

We evaluate the performance of our four novel estimators relative to those in James & Stein (1961), Jorion (1986), Chételat & Wells (2012), Wang et al. (2014), and Bodnar et al. (2019, 2022) using simulated data and empirical stock-return data. In the simulations, we draw data from a multivariate normal distribution satisfying the theoretical assumptions,

and a multivariate t-distribution with diverging fourth moment violating the assumptions, using a Toeplitz correlation matrix. We consider different sample sizes T, ratios N/T, and correlation strength. We measure the performance of an estimator as its L_2 error relative to that of the sample mean. We find that our estimators consistently improve upon the sample mean, are overall on par with the benchmark estimators, and in many cases deliver a greater accuracy than the benchmarks. We run several robustness tests considering alternative loss functions, heterogeneous variances, and autocorrelated data.

In the empirical analysis, we consider daily excess returns of 441 S&P500 individual stocks spanning the period from January 2000 to December 2023. We evaluate the out-of-sample Sharpe ratios that the mean-vector estimators deliver when applied for mean-variance portfolio selection, which is a popular application of shrinkage estimators (DeMiguel et al. 2009, Barroso & Saxena 2021, Kazak & Pohlmeier 2022, Bodnar et al. 2023, Lassance, Martín-Utrera & Simaan 2024). Using training windows of two years and testing windows of six months, we find that our estimators deliver Sharpe ratios that often exceed the benchmarks and are more robust across the full period, the financial crisis, and the covid-19 pandemic period. We also test the robustness of our results to using weekly returns.

In closing, while our estimators contribute to the literature on distribution-free high-dimensional shrinkage estimators of the mean vector under the i.i.d. assumption, there is also a large literature on dynamic estimators that avoid this restrictive assumption. Such dynamic estimators, which are beyond the scope of this paper, are particularly popular for asset returns in financial econometrics. Nonetheless, i.i.d. models can be useful when applied to stock returns—the application we consider in this paper—because simpler estimators involving fewer parameters may deliver better superior predictions and portfolio strategies (Welch & Goyal 2008, DeMiguel et al. 2014).

The rest of the paper is structured as follows. In Section 2, we present the general problem. In Sections 3 and 4, we derive the theoretical results for the case where we shrink toward a zero vector and the grand sample mean, respectively. In Section 5, we conduct the simulation

and empirical analyses. Section 6 concludes. The supplementary material contains robustness tests for our simulation and empirical results as well as the proofs of all theoretical results.

2 Presentation of the problem

In this section, we present some useful notation, our main assumptions, and the types of shrinkage estimators of the sample mean vector that we consider.

2.1 Notation

Given $\mathbf{A}, \mathbf{A}_1, \mathbf{A}_2$ square matrices of size T, we define the (standardized) Frobenius norm as $||\mathbf{A}||_T := \sqrt{\text{tr}(\mathbf{A}\mathbf{A}')/T}$, where $\mathbf{A}'$ is the transpose of $\mathbf{A}$. The trace operator is $\text{tr}(\boldsymbol{A})$.

Given $\mathbf{a}$ a column vector of size T, we define the (standardized) L_2 norm as $||\mathbf{a}||_T := \sqrt{\mathbf{a}'\mathbf{a}/T}$. The operator $\text{diag}(\mathbf{a})$ creates a diagonal matrix from the vector $\mathbf{a}$.

Given $\{a_T : T \in \mathbb{N}^\star\}$ a sequence of real numbers, we say that $a_T = O(1)$ as $T \to \infty$ if there exists $M > 0$ such that $\text{limsup}_{T\to\infty} |a_T| \leq M$, whereas $a_T = o(1)$ as $T \to \infty$ means that $\text{limsup}_{T\to\infty} a_T = 0$.

Given $\hat{\boldsymbol{\theta}}$ an estimator of $\boldsymbol{\theta} \in \mathbb{R}^N$, we define the mean squared error (MSE) of $\hat{\boldsymbol{\theta}}$ as

$$\text{MSE}(\hat{\boldsymbol{\theta}}) := \mathbb{E}\big[||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||_N^2\big] = \frac{1}{N}\Big[\big(\mathbb{E}[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta}\big)'\big(\mathbb{E}[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta}\big) + \text{tr}\big(\mathbb{V}[\hat{\boldsymbol{\theta}}]\big)\Big],$$

where $\mathbb{E}[\hat{\boldsymbol{\theta}}]$ and $\mathbb{V}[\hat{\boldsymbol{\theta}}]$ are the expectation and covariance matrix operators, respectively.

2.2 Assumptions

Consider a matrix of sample data $\mathbf{X} \in \mathbb{R}^{T\times N_T}$, where T is the sample size and N_T the number of variables. We use a subscript T in N_T because, in our high-dimensional asymptotics, we have $N_T/T = O(1)$ as $T \to \infty$, i.e., N_T varies with T. We use subscripts or superscripts T elsewhere. We present the data generating process for $\mathbf{X}$ in Assumption 1.

Assumption 1 *For any sample size $T \geq 1$, the sample of data $\mathbf{X} := \big(\mathbf{X}_1, \ldots, \mathbf{X}_T\big)$ consists of i.i.d. random vectors $\mathbf{X}_i \in \mathbb{R}^{N_T}$, $i = 1, \ldots, T$, drawn from a sampling distribution with mean vector $\boldsymbol{\mu}_{N_T} \in \mathbb{R}^{N_T}$ and positive-definite covariance matrix $\boldsymbol{\Sigma}_{N_T} \in \mathbb{R}^{N_T \times N_T}$.*

We do not assume any parametric form for the sampling distribution, and thus, our shrinkage estimators are distribution-free. We impose the i.i.d. assumption only across observations. As discussed in Section 1, Wang et al. (2014) and Bodnar et al. (2019) also impose that the whitening transform of the data be a matrix of i.i.d. entries, which is more restrictive. This stochastic representation is also used in, e.g., Bai & Saranadasa (1996), Chen et al. (2010), and Ledoit & Wolf (2022). We do not need this particular stochastic structure in our asymptotic results because we focus on L_2 convergence instead of almost-sure convergence. The latter is a powerful type of convergence that implies convergence in probability, which is equivalent to L_2 convergence if uniform integrability is guaranteed. Even if the latter holds, our set of assumptions is still more general than that imposed by the whitening transform of the data being an i.i.d. matrix. Thus, our setting is more general than that in Wang et al. (2014) and Bodnar et al. (2019), which helps to obtain elegant almost-sure convergence.

In Assumption 2, we present the assumptions used in our high-dimensional asymptotics.

Assumption 2 *The following holds:*

i) $N_T/T = O(1)$ *as* $T \to \infty$.

ii) The absolute value of the first moment and the fourth-order comoments of the sampling distribution are uniformly bounded, i.e., there exists a constant $M > 0$ *independent of* T, k, *and* l *such that* $|\mu_{k,N_T}| \leq M$ *and* $\mu^{(4)}_{kl,N_T} := \mathbb{E}[(X_{1k} - \mu_{k,N_T})^2 (X_{1l} - \mu_{l,N_T})^2] \leq M$ *for all* $k, l = 1, \ldots, N_T$ *and any* (T, N_T).

Although we assume bounded fourth-order comoments in Assumption 2, we present simulation evidence in Section 5.2 that our estimators also perform well when the marginal distributions have unbounded fourth-order moments.

2.3 Shrinkage estimators of the sample mean vector

Our objective is to estimate the mean vector $\boldsymbol{\mu}_{N_T}$. The usual estimator is the sample mean vector, defined as

$$\overline{\mathbf{X}}^{(T)} := \Big(\overline{X}_1^{(T)}, \ldots, \overline{X}_{N_T}^{(T)}\Big)', \quad \text{where} \quad \overline{X}_k^{(T)} := \frac{1}{T}\sum_{i=1}^{T} X_{ik}, \tag{1}$$

which is unbiased under Assumption 1 and the maximum-likelihood estimator under the multivariate normal distribution. Under Assumption 1, the MSE of $\overline{\mathbf{X}}^{(T)}$ is given by

$$\mathrm{MSE}\big(\overline{\mathbf{X}}^{(T)}\big) = \frac{1}{TN_T}\mathrm{tr}\big(\boldsymbol{\Sigma}_{N_T}\big). \tag{2}$$

We aim to reduce the MSE of $\overline{\mathbf{X}}^{(T)}$ using different shrinkage estimators. First, we consider two multiplicative shrinkage estimators that shrink $\overline{\mathbf{X}}^{(T)}$ toward zero, which are of the form

$$\hat{\boldsymbol{\mu}}^{StM}(c) := c\overline{\mathbf{X}}^{(T)}, \quad c \in \mathbb{R}, \tag{3}$$

and

$$\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}) := \mathrm{diag}(\mathbf{c})\overline{\mathbf{X}}^{(T)}, \quad \mathbf{c} = (c_1, \ldots, c_{N_T})' \in \mathbb{R}^{N_T}, \tag{4}$$

which we call the *Stein Multiplicative Shrinkage (St-MSh)* and *Diagonal Multiplicative Shrinkage (D-MSh)* estimators. The D-MSh estimator is appealing when the variables $\mathbf{X}_k$, $k = 1, \ldots, N_T$, have different scales because each variable is shrunk with its own intensity. Second, we consider two linear shrinkage estimators that shrink $\overline{\mathbf{X}}^{(T)}$ toward its grand mean,

$$\overline{X}_g^{(T)} := \frac{1}{N_T}\mathbf{1}_{N_T}'\overline{\mathbf{X}}^{(T)}. \tag{5}$$

This bona-fide target estimator is optimal under the MSE in the sense that the best constant estimator of the mean vector is $c^\star\mathbf{1}_{N_T}$ with $c^\star := \mathrm{argmin}_{c\in\mathbb{R}}\,\mathrm{MSE}(c\mathbf{1}_{N_T}) = \mathbf{1}_{N_T}'\boldsymbol{\mu}_{N_T}/N_T$. The two linear shrinkage estimators, called *One-Parameter Linear Shrinkage (O-LSh)* and *Two-*

Parameter Linear Shrinkage (T-LSh), are of the form

$$\hat{\boldsymbol{\mu}}^{OL}(\delta) := (1-\delta)\overline{\mathbf{X}}^{(T)} + \delta\overline{X}_g^{(T)}\mathbf{1}_{N_T}, \quad \delta \in \mathbb{R}, \tag{6}$$

and

$$\hat{\boldsymbol{\mu}}^{TL}(\xi,\delta) := (1-\delta)\overline{\mathbf{X}}^{(T)} + \delta\xi\overline{X}_g^{(T)}\mathbf{1}_{N_T}, \quad (\xi,\delta) \in \mathbb{R}^2. \tag{7}$$

O-LSh in (6) optimizes $\delta_1\overline{\mathbf{X}}^{(T)} + \delta_2\overline{X}_g^{(T)}\mathbf{1}_{N_T}$ with $\delta_1 + \delta_2 = 1$, whereas T-LSh in (7) optimizes $\delta_1\overline{\mathbf{X}}^{(T)} + \delta_2\overline{X}_g^{(T)}\mathbf{1}_{N_T}$ with no restriction. T-LSh extends O-LSh by adding a scaling adjustment to the target estimator, which accounts for the fact that $\overline{X}_g^{(T)}$ is only a bona-fide estimator of the oracle constant target estimator, $\mathbf{1}'_{N_T}\boldsymbol{\mu}_{N_T}/N_T$. By shrinking directly toward a bona-fide target estimator, we follow the two-step shrinkage approach advocated by Lassance (2025) where the shrinkage intensity accounts for estimation errors in the target-estimator scaling. This is in contrast with the standard one-step approach where the optimal shrinkage intensity is determined under the oracle target-estimator scaling; see, e.g., Thompson (1968), Wang et al. (2014), and Bodnar et al. (2019) for one-step shrinkage estimators of the sample mean.

For each of these four shrinkage estimators (St-MSh, D-MSh, O-LSh, T-LSh), we proceed in a similar manner. That is, we derive the optimal oracle shrinkage parameters (i.e., c, $\mathbf{c}$, δ, or ξ) for a fixed (T, N_T) under Assumption 1, we propose bona-fide estimators of the oracle shrinkage parameters, and we show, under Assumption 2, that the bona-fide shrinkage mean-vector estimators satisfy high-dimensional L_2 consistency as $N_T/T = O(1)$ with $T \to \infty$. In particular, for each bona-fide shrinkage mean-vector estimator, we establish that i) it is an L_2-consistent estimator of its oracle estimator, ii) it has the same asymptotic MSE as the oracle estimator, iii) the MSE of the bona-fide estimator is estimated consistently.

Two other commonly used forms of consistency than L_2 consistency rely on almost-sure and in-probability convergence. Compared to convergence in probability, L_2 convergence is stronger as it implies convergence in probability but the converse is not true, in general. Moreover, almost-sure convergence, a strong form of convergence, is not directly comparable

to L_2 convergence as one does not always imply the other, in general. Note also that L_2 convergence implies L_p convergence for $p \in (0, 2)$, which are more robust loss functions.

3 Multiplicative shrinkage toward zero

In Theorem 1, we derive the oracle shrinkage parameters, i.e., c for St-MSh in (3) and $\mathbf{c}$ for D-MSh in (4), that minimize the MSE. We also derive the MSE attained by the oracles.

Theorem 1 *Let Assumption 1 hold and the pair (T, N_T) be fixed and finite.*

i) The optimal oracle St-MSh estimator is unique and given by $\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)$ in (3) with

$$c_T^\star := \underset{c\in\mathbb{R}}{\operatorname{argmin}} \operatorname{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c)\big) = \frac{\big|\big|\boldsymbol{\mu}_{N_T}\big|\big|_{N_T}^2}{\big|\big|\boldsymbol{\mu}_{N_T}\big|\big|_{N_T}^2 + \operatorname{MSE}\big(\overline{\mathbf{X}}^{(T)}\big)} \in [0, 1), \tag{8}$$

where $\operatorname{MSE}\big(\overline{\mathbf{X}}^{(T)}\big)$ is given by (2). The MSE of $\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)$ is given by

$$\operatorname{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\big) = c_T^\star \operatorname{MSE}\big(\overline{\mathbf{X}}^{(T)}\big). \tag{9}$$

ii) The optimal oracle D-MSh estimator is unique and given by $\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)$ in (4) with the k^{th} entry of $\boldsymbol{c}_T^\star$ being equal to

$$c_{k,T}^\star := \left(\underset{c\in\mathbb{R}}{\operatorname{argmin}} \operatorname{MSE}\big(\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c})\big)\right)_k = \frac{\mu_{k,N_T}^2}{\mu_{k,N_T}^2 + \operatorname{MSE}\big(\overline{X}_k^{(T)}\big)} \in [0, 1), \tag{10}$$

where $\operatorname{MSE}\big(\overline{X}_k^{(T)}\big) = \frac{1}{T}\Sigma_{kk,N_T}$. The MSE of $\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)$ is given by

$$\operatorname{MSE}\big(\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)\big) = \frac{1}{N_T}\sum_{k=1}^{N_T} c_{k,T}^\star \operatorname{MSE}\big(\overline{X}_k^{(T)}\big). \tag{11}$$

iii) It holds that

$$\operatorname{MSE}\big(\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)\big) \leq \operatorname{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\big) < \operatorname{MSE}\big(\overline{\mathbf{X}}^{(T)}\big), \tag{12}$$

where the first inequality is an identity if and only if $\mu_{k,N_T}^2/\Sigma_{kk,N_T}$ is the same for all k.

Theorem 1 iii) shows that, both in St-MSh and D-MSh, it is optimal to shrink the sample mean $\overline{\mathbf{X}}^{(T)}$ closer to zero: $c_T^\star \in [0,1)$ and $\mathbf{c}_T^\star \in [0,1)^{N_T}$. The resulting MSE of the oracle St-MSh and D-MSh estimators are strictly lower than the MSE of the sample mean $\overline{\mathbf{X}}^{(T)}$. One should shrink less intensively when the true $\boldsymbol{\mu}_{N_T}$ is further away from zero, i.e., when $\big|\big|\boldsymbol{\mu}_{N_T}\big|\big|_{N_T}^2$ and μ_{k,N_T}^2 are larger. In contrast, one should shrink more intensively when the sample mean has a larger MSE, i.e., when $\text{MSE}\big(\overline{\mathbf{X}}^{(T)}\big)$ and $\text{MSE}\big(\overline{X}_k^{(T)}\big)$ are larger. Moreover, the optimal oracle D-MSh estimator, $\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)$, always attains a lower MSE than the optimal oracle St-MSh estimator, $\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)$, and these two estimators have the same MSE if and only if each variable has the same population coefficient of variation.

Theorem 1 establishes the optimal shrinkage estimators $\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)$ and $\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)$. However, both are oracle estimators because they depend on unknown population parameters. Specifically, we need to estimate $||\boldsymbol{\mu}_{N_T}||_{N_T}^2 = \frac{1}{N_T}\boldsymbol{\mu}'_{N_T}\boldsymbol{\mu}_{N_T}$ and $\text{MSE}\big(\overline{\mathbf{X}}^{(T)}\big) = \frac{1}{TN_T}\text{tr}\big(\boldsymbol{\Sigma}_{N_T}\big)$ for the St-MSh estimator, and μ_{k,N_T}^2 and $\text{MSE}\big(\overline{X}_k^{(T)}\big) = \frac{1}{T}\Sigma_{kk,N_T}$ for the D-MSh estimator. Replacing these quantities by their sample plug-in estimators, we obtain the bona-fide St-MSh estimator as $\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star)$ with

$$\hat{c}_T^\star := \frac{\big|\big|\overline{\mathbf{X}}^{(T)}\big|\big|_{N_T}^2}{\big|\big|\overline{\mathbf{X}}^{(T)}\big|\big|_{N_T}^2 + \frac{1}{TN_T}\text{tr}(\mathbf{S}_{N_T})}, \tag{13}$$

where, for any $T \geq 2$,

$$\mathbf{S}_{N_T} := \frac{1}{T-1}\Big(\mathbf{X}^{(T)} - \overline{\mathbf{X}}^{(T)}\mathbf{1}'_{N_T}\Big)\Big(\mathbf{X}^{(T)} - \overline{\mathbf{X}}^{(T)}\mathbf{1}'_{N_T}\Big)'$$

is the sample unbiased covariance matrix, and the bona-fide D-MSh estimator as $\hat{\boldsymbol{\mu}}^{DM}(\hat{\mathbf{c}}_T^\star)$ with the k^{th} entry of $\hat{\mathbf{c}}_T^\star$ being equal to

$$\hat{c}_{k,T}^\star := \frac{\big(\overline{X}_k^{(T)}\big)^2}{\big(\overline{X}_k^{(T)}\big)^2 + \frac{1}{T}S_{kk,N_T}}. \tag{14}$$

In Theorem 2, we study the high-dimensional L_2 consistency of the bona-fide St-MSh estimator, $\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star)$. We require Assumption 2 in addition to Assumption 1. The overall conclusion is that the bona-fide estimator shares the essential properties of the oracle estimator when

$N_T/T = O(1)$ as $T \to \infty$.

Theorem 2 *Let Assumptions 1 and 2 hold. The following holds concerning the bona-fide St-MSh estimator* $\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star)$ *in* (13) *and the oracle estimator* $\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)$ *in* (8)*.*

i) The bona-fide estimator is an L_2*-consistent estimator of the oracle estimator:*

$$\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star) - \hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\right|\right|_{N_T}^2\right] \to 0 \quad as \;\; T \to \infty. \tag{15}$$

ii) The bona-fide and oracle estimators have the same asymptotic MSE:

$$\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star)\big) - \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\big) \to 0 \quad as \;\; T \to \infty. \tag{16}$$

iii) The MSE of the bona-fide estimator is estimated consistently:

$$\mathbb{E}\left[\left|\widehat{\mathrm{MSE}}\big(\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star)\big) - \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\big)\right|\right] \to 0 \quad as \;\; T \to \infty, \tag{17}$$

where $\widehat{\mathrm{MSE}}$ *is the estimator obtained by plugging elements of* $(\overline{\mathbf{X}}^{(T)}, \mathbf{S}_{N_T})$ *in place of the corresponding elements of* $(\boldsymbol{\mu}_{N_T}, \boldsymbol{\Sigma}_{N_T})$*.*

Theorem 2 tells us that the bona-fide St-MSh estimator and its oracle counterpart are asymptotically interchangeable in terms of L_2, expected loss, and MSE. In Theorem 3, we establish the same properties for the bona-fide D-MSh estimator, $\hat{\boldsymbol{\mu}}^{DM}(\hat{\mathbf{c}}_T^\star)$.

Theorem 3 *Let Assumptions 1 and 2 hold. The following holds concerning the bona-fide D-MSh estimator* $\hat{\boldsymbol{\mu}}^{DM}(\hat{\mathbf{c}}_T^\star)$ *in* (14) *and the oracle estimator* $\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)$ *in* (10)*.*

i) The bona-fide estimator is an L_2*-consistent estimator of the oracle estimator:*

$$\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{DM}(\hat{\mathbf{c}}_T^\star) - \hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)\right|\right|_{N_T}^2\right] \to 0 \quad as \;\; T \to \infty. \tag{18}$$

ii) The bona-fide and oracle estimators have the same asymptotic MSE:

$$\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{DM}(\hat{\mathbf{c}}_T^\star)\big) - \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)\big) \to 0 \quad as \;\; T \to \infty. \tag{19}$$

iii) The MSE of the bona-fide estimator is estimated consistently:

$$\mathbb{E}\Big[\Big|\widehat{\mathrm{MSE}}\big(\hat{\boldsymbol{\mu}}^{DM}(\hat{\mathbf{c}}_T^\star)\big) - \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)\big)\Big|\Big] \to 0 \quad as \;\; T \to \infty. \tag{20}$$

4 Linear shrinkage toward the grand sample mean

In Theorem 4, we derive the oracle shrinkage parameters, i.e., δ for O-LSh in (6) and (ξ, δ) for T-LSh in (7), that minimize the MSE. We also derive the MSE attained by the oracles.

Theorem 4 *Let Assumption 1 hold and the pair (T, N_T) be fixed and finite. Define*

$$d_{1T} := \mathrm{MSE}\big(\overline{\mathbf{X}}^{(T)}\big) = \frac{1}{TN_T}\mathrm{tr}(\boldsymbol{\Sigma}_{N_T}), \tag{21}$$

$$d_{2T} := \mathrm{MSE}\big(\overline{X}_g^{(T)}\big) = \frac{1}{TN_T^2}\mathbf{1}_{N_T}'\boldsymbol{\Sigma}_{N_T}\mathbf{1}_{N_T}, \tag{22}$$

$$d_{3T} := \frac{1}{N_T}\boldsymbol{\mu}_{N_T}'\boldsymbol{\mu}_{N_T} - \frac{1}{N_T^2}(\mathbf{1}_{N_T}'\boldsymbol{\mu}_{N_T})^2, \tag{23}$$

$$d_{4T} := \frac{1}{N_T^2}(\mathbf{1}_{N_T}'\boldsymbol{\mu}_{N_T})^2, \tag{24}$$

which satisfy $d_{1T} > d_{2T} > 0$, $d_{3T} \geq 0$, and $d_{4T} \geq 0$.

i) The optimal oracle O-LSh estimator is unique and given by $\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)$ in (6) with

$$\delta_T^\star := \operatorname*{argmin}_{\delta\in\mathbb{R}} \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\delta)\big) = \frac{d_{1T} - d_{2T}}{d_{1T} - d_{2T} + d_{3T}} \in (0, 1]. \tag{25}$$

The MSE of $\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)$ is given by

$$\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)\big) = wd_{1T} + (1-w)d_{2T} \quad with \quad w := \frac{d_{3T}}{d_{1T} - d_{2T} + d_{3T}} \in [0, 1). \tag{26}$$

ii) The optimal oracle T-LSh estimator is unique and given by $\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star, \delta_T^\star)$ in (7) with

$$(\xi_T^\star, \delta_T^\star) := \operatorname*{argmin}_{(\xi,\delta)\in\mathbb{R}^2} \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\xi, \delta)\big) = \left(1 - \frac{d_{2T}/\delta_T^\star}{d_{2T} + d_{4T}}, \frac{d_{1T} - d_{2T}}{d_{1T} - d_{2T} + d_{3T}}\right). \tag{27}$$

The MSE of $\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star, \delta_T^\star)$ is given by

$$\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star, \delta_T^\star)\big) = \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)\big) - \frac{d_{2T}^2}{d_{2T} + d_{4T}}. \tag{28}$$

iii) It holds that

$$\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star, \delta_T^\star)\big) < \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)\big) < \mathrm{MSE}\big(\overline{\mathbf{X}}^{(T)}\big). \tag{29}$$

Theorem 4 shows that, in O-LSh, it is optimal to shrink the sample mean $\overline{\mathbf{X}}^{(T)}$ closer to its grand mean $\overline{X}_g^{(T)}$ because $\delta_T^\star \in (0, 1]$. A similar result holds in T-LSh, where one shrinks $\overline{\mathbf{X}}^{(T)}$ with the same intensity $\delta_T^\star$ toward $\xi_T^\star \overline{X}_g^{(T)}$. One should shrink less intensively when $\boldsymbol{\mu}_{N_T}$ is further away from its grand mean, i.e., when d_{3T} in (23) is larger. In contrast, one should shrink more intensively when $\overline{\mathbf{X}}^{(T)}$ has a larger MSE, i.e., when d_{1T} in (21) is larger. Moreover, the optimal oracle T-LSh estimator, $\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star, \delta_T^\star)$, always attains a strictly lower MSE than the optimal oracle O-LSh estimator, $\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)$.

Now, as in Section 3, $\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)$ and $\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star, \delta_T^\star)$ are unfeasible oracles because d_{1T}, d_{2T}, d_{3T}, and d_{4T} in (21)–(24) are scalar functions of the population $\boldsymbol{\mu}_{N_T}$ and $\boldsymbol{\Sigma}_{N_T}$. Replacing these four quantities by their sample plug-in estimators, we obtain the bona-fide O-LSh and T-LSh estimators as $\hat{\boldsymbol{\mu}}^{OL}(\hat{\delta}_T^\star)$ and $\hat{\boldsymbol{\mu}}^{TL}(\hat{\xi}_T^\star, \hat{\delta}_T^\star)$, where

$$\hat{\delta}_T^\star := \frac{\hat{d}_{1T} - \hat{d}_{2T}}{\hat{d}_{1T} - \hat{d}_{2T} + \hat{d}_{3T}}, \quad \hat{\xi}_T^\star := 1 - \frac{\hat{d}_{2T}/\hat{\delta}_T^\star}{\hat{d}_{2T} + \hat{d}_{4T}}, \tag{30}$$

and each $\hat{d}_{kT}$ is obtained by plugging $(\overline{\mathbf{X}}^{(T)}, \mathbf{S}_{N_T})$ in place of $(\boldsymbol{\mu}_{N_T}, \boldsymbol{\Sigma}_{N_T})$ into (21)–(24). In Theorem 5, we study the high-dimensional L_2 consistency of the bona-fide O-LSh estimator, $\hat{\boldsymbol{\mu}}^{OL}(\hat{\delta}_T^\star)$. Similar to St-MSh and D-MSh, we conclude that the bona-fide estimator and its oracle counterpart are asymptotically interchangeable in terms of L_2, expected loss, and MSE when $N_T/T = O(1)$ as $T \to \infty$.

Theorem 5 *Let Assumptions 1 and 2 hold. The following holds concerning the bona-fide O-LSh estimator* $\hat{\boldsymbol{\mu}}^{OL}(\hat{\delta}_T^\star)$ *in* (30) *and the oracle estimator* $\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)$ *in* (25).

i) The bona-fide estimator is an L_2*-consistent estimator of the oracle estimator:*

$$\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{OL}(\hat{\delta}_T^\star) - \hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)\right|\right|_{N_T}^2\right] \to 0 \quad \textit{as} \quad T \to \infty. \tag{31}$$

ii) The bona-fide and oracle estimators have the same asymptotic MSE:

$$\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\hat{\delta}_T^\star)\big) - \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)\big) \to 0 \quad as \;\; T \to \infty. \tag{32}$$

iii) The MSE of the bona-fide estimator is estimated consistently:

$$\mathbb{E}\Big[\Big|\widehat{\mathrm{MSE}}\big(\hat{\boldsymbol{\mu}}^{OL}(\hat{\delta}_T^\star)\big) - \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)\big)\Big|\Big] \to 0 \quad as \;\; T \to \infty. \tag{33}$$

In Theorem 6, we establish similar high-dimensional asymptotic properties for the bona-fide T-LSh estimator, $\hat{\boldsymbol{\mu}}^{TL}(\hat{\xi}_T^\star, \hat{\delta}_T^\star)$.

Theorem 6 *Let Assumptions 1 and 2 hold. The following holds concerning the bona-fide T-LSh estimator* $\hat{\boldsymbol{\mu}}^{TL}(\hat{\xi}_T^\star, \hat{\delta}_T^\star)$ *in* (30) *and the oracle estimator* $\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star, \delta_T^\star)$ *in* (27).

i) The bona-fide estimator is an L_2*-consistent estimator of the oracle estimator:*

$$\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{TL}(\hat{\xi}_T^\star, \hat{\delta}_T^\star) - \hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star, \delta_T^\star)\right|\right|_{N_T}^2\right] \to 0 \quad as \;\; T \to \infty. \tag{34}$$

ii) The bona-fide and oracle estimators have the same asymptotic MSE:

$$\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\hat{\xi}_T^\star, \hat{\delta}_T^\star)\big) - \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star, \delta_T^\star)\big) \to 0 \quad as \;\; T \to \infty. \tag{35}$$

iii) The MSE of the bona-fide estimator is estimated consistently:

$$\mathbb{E}\Big[\Big|\widehat{\mathrm{MSE}}\big(\hat{\boldsymbol{\mu}}^{TL}(\hat{\xi}_T^\star, \hat{\delta}_T^\star)\big) - \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star, \delta_T^\star)\big)\Big|\Big] \to 0 \quad as \;\; T \to \infty, \tag{36}$$

provided that one of these two conditions is satisfied about d_{3T} *and* d_{4T} *in* (23)–(24)*:*

a) there exist constants $M_1 > 0$ *and* $M_2 > 0$ *that are independent of* (T, N_T) *such that* $M_1 d_{4T} \leq d_{3T} \leq M_2 d_{4T}$ *holds for all* T*, or*

b) $\mathrm{limsup}_{T\to\infty}\, d_{3T} + d_{4T} = \mathrm{limsup}_{T\to\infty}\, \frac{1}{N_T} \boldsymbol{\mu}'_{N_T} \boldsymbol{\mu}_{N_T} = 0$.

We note that it seems difficult to find a plausible model of the population mean vector $\boldsymbol{\mu}_{N_T}$ for which none of the last two conditions is satisfied.

5 Numerical analysis

We now evaluate the performance of the bona-fide St-MSh, D-MSh, O-LSh, and T-LSh estimators introduced in Sections 3 and 4. In Section 5.1, we detail the benchmark estimators we consider. In Section 5.2, we conduct a simulation analysis using data generated from multivariate normal and t distributions, where the latter has diverging fourth moments, and thus, violates the theoretical assumptions. In Section 5.3, we conduct an empirical analysis where we construct mean-variance portfolios of a large number of S&P500 individual stocks. The base-case simulation setting satisfies the theoretical assumption of i.i.d. sampling, whereas daily empirical stock returns typically violate the i.i.d. assumption.

5.1 Benchmark estimators

We compare our four proposed estimators to six benchmark estimators of the mean vector.

1. The *sample mean vector*, $\overline{\mathbf{X}}^{(T)}$ in (1), which is well-defined for all N_T and T.

2. The positive-part James & Stein (1961) estimator under a common variance, which we call *PJS estimator*. It is designed to minimize the MSE, is well-defined for all $N_T \geq 2$ and T, and is given by

$$\hat{\boldsymbol{\mu}}_{PJS} := \left(1 - \frac{(N_T - 2)\hat{\sigma}^2}{T(\overline{\mathbf{X}}^{(T)})'\overline{\mathbf{X}}^{(T)}}\right)_+ \overline{\mathbf{X}}^{(T)},$$

   where $x_+ := \max(x, 0)$ and $\hat{\sigma}^2 := \frac{1}{N_T}\text{tr}(\mathbf{S}_{N_T})$.

3. The shrinkage estimator in Wang et al. (2014), which we call *Wang estimator*. It is designed for all N_T and $T \geq 2$. They consider a quadratic loss with a general weighting matrix $\mathbf{Q}$ and, setting $\mathbf{Q} = \mathbf{I}_{N_T}$, i.e., the MSE, the Wang estimator is defined as

$$\hat{\boldsymbol{\mu}}_{Wang} := (1 - \hat{\delta}^{\star}_{Wang})\overline{\mathbf{X}}^{(T)} + \hat{\delta}^{\star}_{Wang}\overline{X}_g^{(T)}\mathbf{1}_{N_T},$$

where $\overline{X}_g^{(T)}$ is the grand sample mean in (5),

$$\hat{\delta}^{\star}_{Wang} := \frac{Y_{2T}}{Y_{1T} + Y_{2T} - Y_{3T}},$$

and

$$\begin{aligned}
Y_{1T} &:= \frac{1}{N_T(T-1)} \sum_{i=1}^{T} \sum_{j \neq i}^{T} \mathbf{X}_i' \mathbf{X}_j, \\
Y_{2T} &:= \frac{1}{N_T T} \left(\sum_{i=1}^{T} \mathbf{X}_i' \mathbf{X}_i - \frac{1}{T-1} \sum_{i=1}^{T} \sum_{j \neq i}^{T} \mathbf{X}_i' \mathbf{X}_j \right), \\
Y_{3T} &:= \frac{1}{N_T^2(T-1)} \sum_{i=1}^{T} \sum_{j \neq i}^{T} (\mathbf{1}_{N_T}' \mathbf{X}_i)(\mathbf{X}_j' \mathbf{1}_{N_T}).
\end{aligned}$$

4. The shrinkage estimator in Bodnar et al. (2019) minimizing the invariant quadratic loss, which we call *BOP estimator*. It requires either $T > N_T$ or $N_T > T \geq 2$. The case $N_T = T$ is not covered because $N_T/T \to 1$ as $N_T, T \to \infty$ is excluded in Section 3 of Bodnar et al. (2019). They consider a general nonrandom target estimator and, considering the unit vector as a target, the BOP estimator is defined as

$$\hat{\boldsymbol{\mu}}_{BOP} := (1 - \hat{\delta}^{\star}_{BOP}) \overline{\mathbf{X}}^{(T)} + \hat{\delta}^{\star}_{BOP} \frac{\mathbf{1}_{N_T}' \mathbf{S}_{N_T}^{+} \overline{\mathbf{X}}^{(T)}}{\mathbf{1}_{N_T}' \mathbf{S}_{N_T}^{+} \mathbf{1}_{N_T}} \mathbf{1}_{N_T},$$

where $\mathbf{S}_{N_T}^{+}$ denotes the Moore–Penrose inverse of $\mathbf{S}_{N_T}$, which corresponds to the usual inverse when $T > N_T$, and

$$\hat{\delta}^{\star}_{BOP} := \frac{\min\{T,\, N_T\}}{|T - N_T|} \times \frac{\mathbf{1}_{N_T}' \mathbf{S}_{N_T}^{+} \mathbf{1}_{N_T}}{\left((\overline{\mathbf{X}}^{(T)})' \mathbf{S}_{N_T}^{+} \overline{\mathbf{X}}^{(T)} \mathbf{1}_{N_T}' \mathbf{S}_{N_T}^{+} \mathbf{1}_{N_T} - \left(\mathbf{1}_{N_T}' \mathbf{S}_{N_T}^{+} \overline{\mathbf{X}}^{(T)} \right)^2 \right)^{-1}}.$$

5. The Bayes-Stein shrinkage estimator in Jorion (1986), which we call *Jorion estimator*. It is designed for settings where $T > N_T + 2$ and is defined as

$$\hat{\boldsymbol{\mu}}_{Jorion} := (1 - \hat{\delta}_{Jorion}) \overline{\mathbf{X}}^{(T)} + \hat{\delta}_{Jorion} \frac{\mathbf{1}_{N_T}' \tilde{\mathbf{S}}_{N_T}^{-1} \overline{\mathbf{X}}^{(T)}}{\mathbf{1}_{N_T}' \tilde{\mathbf{S}}_{N_T}^{-1} \mathbf{1}_{N_T}} \mathbf{1}_{N_T},$$

where $\tilde{\mathbf{S}}_{N_T} = \frac{T-1}{T-N_T-2}\mathbf{S}_{N_T}$ and

$$\hat{\delta}_{Jorion} := \frac{N_T + 2}{N_T + 2 + T\left(\overline{\mathbf{X}}^{(T)} - \frac{\mathbf{1}'_{N_T}\tilde{\mathbf{S}}^{-1}_{N_T}\overline{\mathbf{X}}^{(T)}}{\mathbf{1}'_{N_T}\tilde{\mathbf{S}}^{-1}_{N_T}\mathbf{1}_{N_T}}\mathbf{1}_{N_T}\right)'\tilde{\mathbf{S}}^{-1}_{N_T}\left(\overline{\mathbf{X}}^{(T)} - \frac{\mathbf{1}'_{N_T}\tilde{\mathbf{S}}^{-1}_{N_T}\overline{\mathbf{X}}^{(T)}}{\mathbf{1}'_{N_T}\tilde{\mathbf{S}}^{-1}_{N_T}\mathbf{1}_{N_T}}\mathbf{1}_{N_T}\right)}.$$

6. The positive-part shrinkage estimator in Chételat & Wells (2012) minimizing the invariant quadratic loss, which we call *CW estimator*. It is designed for settings where $N_T > T \geq 3$ and is defined as

$$\hat{\boldsymbol{\mu}}_{CW} := \left(\mathbf{I}_{N_T} - \mathbf{S}_{N_T}\mathbf{S}^{+}_{N_T}\right)\overline{\mathbf{X}}^{(T)} + \left(1 - \frac{a}{(\overline{\mathbf{X}}^{(T)})'\mathbf{S}^{+}_{N_T}\overline{\mathbf{X}}^{(T)}}\right)_{+}\mathbf{S}_{N_T}\mathbf{S}^{+}_{N_T}\overline{\mathbf{X}}^{(T)},$$

where we set $a = (T-3)/(N_T - T + 4) \geq 0$ because it ensures that $\hat{\boldsymbol{\mu}}_{CW}$ dominates the James–Stein shrinkage estimator theoretically (Bodnar et al. 2022).

5.2 Simulated data

In Section 5.2.1, we detail the data generating process. In Section 5.2.2, we explain the performance measure we use to compare the estimators. In Section 5.2.3, we discuss the results.

5.2.1 Simulated data generating process

In each simulation, we draw a $T \times N_T$ data matrix $\mathbf{X}$. We consider sample sizes $T \in \{50, 100, 250, 500\}$ and, to distinguish between low and high-dimensional settings, we fix $N_T/T \in \{0.25, 0.5\}$ when $T > N_T$ and $N_T/T \in \{2, 4\}$ when $N_T > T$. We use 1,000 simulations.

The rows $\mathbf{X}_i$ of $\mathbf{X}$, $i = 1, \ldots, T$, are i.i.d. draws from a particular sampling distribution. We set the population mean vector $\boldsymbol{\mu}_{N_T}$ of this distribution as $\mu_{k,N_T} \sim U(-0.1, 0.3)$, $k = 1, \ldots, N_T$, where $U(a, b)$ denotes a uniform distribution over $[a, b]$. The range between -10% and 30% is realistic for annualized stock returns, in line with our empirical analysis. For simplicity, we set unit population variances, i.e., $\Sigma_{kk,N_T} = 1$ for all k. In Section B of the Supplementary Material, we test the robustness of the results to using heterogeneous variances. Moreover, in Section C of the Supplementary Material, we test all estimators by

simulating non-i.i.d. data from a normal vector-autoregressive model.

Then, we consider two different sampling distributions: a multivariate normal distribution and a multivariate t-distribution with $\nu = 3$ degrees of freedom. The first one satisfies the theoretical assumptions of all estimators under consideration, whereas the second one violates the common assumption of bounded fourth moments. In both cases, we use a Toeplitz correlation matrix, i.e., $\Sigma_{kl,N_T}(\rho) = \rho^{|k-l|}$, $k, l = 1, \ldots, N_T$, which corresponds to the covariance matrix of an AR(1) process. We set $\rho \in \{0, 0.5\}$.

5.2.2 Performance measure for simulated data

For a given mean-vector estimate $\hat{\boldsymbol{\mu}}_{N_T}$ in a given simulation, we measure the performance using the L_2-error ratio relative to the sample mean:

$$L_2\text{-error ratio}(\hat{\boldsymbol{\mu}}_{N_T}) := \frac{\left|\left|\hat{\boldsymbol{\mu}}_{N_T} - \boldsymbol{\mu}_{N_T}\right|\right|_2}{\left|\left|\overline{\mathbf{X}}^{(T)} - \boldsymbol{\mu}_{N_T}\right|\right|_2}. \tag{37}$$

We then report the average of the L_2-error ratios obtained across all simulations. In unreported results, we also consider the median and find similar conclusions. The lower the ratio the better, and a ratio lower than one indicates that the considered estimator is more accurate than the sample mean. In Section A of the Supplementary Material, we show that our results are robust to using the L_1 error and the invariant (i.e., Mahalanobis) loss function.

5.2.3 Discussion of simulation results

In Tables 1 and 2, we report the average L_2-error ratios of the different mean-vector estimators we consider for the cases $T > N_T$ and $N_T > T$, respectively. We make the following observations. First, all considered estimators improve upon the sample mean across all parameter values, i.e., the average L_2-error ratios are smaller than one.

Second, our O-LSh and T-LSh estimators that shrink toward the grand mean deliver a lower error than the St-MSh and D-MSh estimators that shrink toward zero, highlighting the importance of the shrinkage target.

Table 1: Average L_2-error ratio in simulated data for $T > N_T$

	Multivariate normal				Multivariate t with $\nu = 3$			
N_T/T	0.25	0.25	0.5	0.5	0.25	0.25	0.5	0.5
ρ	0	0.5	0	0.5	0	0.5	0	0.5
A) $T = 50$								
St-MSh	0.786	**0.797**	0.773	0.785	0.791	0.802	0.778	0.791
D-MSh	0.866	0.866	0.864	0.867	0.850	0.853	0.846	0.849
O-LSh	**0.744**	**0.796**	0.719	**0.756**	**0.750**	**0.797**	0.725	**0.757**
T-LSh	0.753	0.803	0.724	0.771	0.758	0.805	0.729	0.772
PJS	0.794	0.820	0.763	0.792	0.796	0.820	0.767	0.794
Wang	0.768	0.828	**0.698**	**0.757**	0.769	0.825	**0.703**	**0.751**
BOP	0.836	0.853	0.774	0.799	0.799	0.828	0.745	0.777
Jorion	**0.729**	0.798	**0.710**	0.771	**0.732**	**0.793**	**0.718**	0.770
B) $T = 100$								
St-MSh	0.857	0.864	0.853	0.858	0.862	0.866	0.855	0.858
D-MSh	0.928	0.927	0.924	0.926	0.913	0.912	0.908	0.908
O-LSh	**0.804**	**0.829**	0.794	**0.809**	**0.810**	**0.829**	**0.797**	**0.809**
T-LSh	0.807	0.843	0.795	0.815	0.813	0.842	0.798	0.814
PJS	0.858	0.870	0.849	0.859	0.863	0.871	0.852	0.858
Wang	0.805	0.836	**0.783**	**0.803**	0.812	0.834	**0.787**	**0.802**
BOP	0.821	**0.832**	0.805	0.821	0.816	**0.832**	0.798	0.821
Jorion	**0.799**	0.837	**0.793**	0.830	**0.807**	0.840	0.801	0.834
C) $T = 250$								
St-MSh	0.926	0.932	0.926	0.927	0.926	0.931	0.926	0.927
D-MSh	0.973	0.977	0.975	0.975	0.963	0.966	0.965	0.963
O-LSh	0.888	**0.895**	**0.884**	**0.889**	0.887	**0.895**	0.884	**0.889**
T-LSh	0.888	0.897	0.884	0.889	0.888	0.896	0.884	0.889
PJS	0.926	0.934	0.925	0.927	0.926	0.932	0.925	0.926
Wang	**0.886**	**0.896**	**0.881**	**0.887**	**0.886**	**0.894**	**0.881**	**0.886**
BOP	0.888	0.904	0.884	0.900	**0.887**	0.904	**0.884**	0.901
Jorion	**0.887**	0.908	0.884	0.905	0.889	0.908	0.887	0.906
D) $T = 500$								
St-MSh	0.962	0.963	0.960	0.961	0.962	0.963	0.960	0.961
D-MSh	1.001	1.000	0.999	1.000	0.993	0.994	0.994	0.993
O-LSh	0.937	**0.938**	**0.934**	**0.936**	**0.937**	**0.939**	**0.934**	**0.936**
T-LSh	0.937	0.938	0.934	0.936	0.937	0.939	0.935	0.936
PJS	0.962	0.963	0.960	0.961	0.962	0.963	0.960	0.961
Wang	**0.937**	**0.938**	**0.933**	**0.936**	**0.937**	**0.939**	**0.934**	**0.936**
BOP	0.937	0.945	0.934	0.944	0.937	0.945	0.935	0.944
Jorion	**0.937**	0.946	0.934	0.946	0.938	0.947	0.936	0.946

Notes. This table reports the average of the L_2-error ratios in (37) across 1,000 simulations for different estimators of the mean vector and different parameter values. Panels A to D consider a sample size $T \in \{50, 100, 250, 500\}$ while $N_T/T \in \{0.25, 0.5\}$. The data generating process is described in Section 5.2.1. The data is simulated either from a multivariate normal distribution or from a multivariate t-distribution with $\nu = 3$ degrees of freedom. In both cases, the dependence is governed by a Toeplitz correlation matrix with $\rho \in \{0, 0.5\}$. The proposed estimators are St-MSh and D-MSh in Section 3, and O-LSh and T-LSh in Section 4. The PJS, Wang, BOP, and Jorion benchmark estimators are described in Section 5.1. The smallest average L_2-error ratios are in underlined bold, and the second-smallest in bold.

Table 2: Average L_2-error ratio in simulated data for $N_T > T$

	Multivariate normal				Multivariate t with $\nu = 3$			
N_T/T	2	2	4	4	2	2	4	4
ρ	0	0.5	0	0.5	0	0.5	0	0.5
A) $T = 50$								
St-MSh	0.767	0.769	0.766	0.766	0.770	0.772	0.768	0.767
D-MSh	0.864	0.866	0.865	0.864	0.841	0.840	0.839	0.836
O-LSh	**0.702**	0.712	0.699	0.704	0.705	0.713	0.701	0.705
T-LSh	0.702	0.715	0.699	0.705	0.706	0.716	0.702	0.706
PJS	0.742	0.749	0.739	0.741	0.746	0.751	0.742	0.742
Wang	**0.649**	**0.666**	**0.641**	**0.651**	**0.654**	**0.667**	**0.645**	**0.651**
BOP	0.710	**0.696**	**0.687**	**0.678**	**0.693**	**0.700**	**0.675**	**0.690**
CW	0.891	0.864	0.945	0.927	0.867	0.854	0.919	0.914
B) $T = 100$								
St-MSh	0.847	0.848	0.847	0.848	0.848	0.849	0.848	0.849
D-MSh	0.923	0.924	0.924	0.924	0.902	0.902	0.902	0.902
O-LSh	0.783	0.788	0.782	0.785	0.785	0.788	0.784	0.786
T-LSh	0.783	0.788	0.782	0.785	0.785	0.788	0.784	0.786
PJS	0.838	0.841	0.838	0.840	0.840	0.841	0.839	0.841
Wang	**0.761**	**0.768**	**0.759**	**0.762**	**0.763**	**0.768**	**0.760**	**0.764**
BOP	**0.778**	**0.781**	**0.771**	**0.775**	**0.776**	**0.787**	**0.770**	**0.782**
CW	0.925	0.908	0.963	0.952	0.912	0.904	0.949	0.945
C) $T = 250$								
St-MSh	0.926	0.926	0.925	0.925	0.926	0.926	0.926	0.925
D-MSh	0.976	0.976	0.975	0.976	0.963	0.962	0.962	0.962
O-LSh	0.883	**0.883**	0.882	**0.882**	0.884	**0.883**	0.882	**0.883**
T-LSh	0.883	0.883	0.882	0.882	0.884	0.883	0.882	0.883
PJS	0.925	0.925	0.924	0.924	0.925	0.925	0.924	0.924
Wang	**0.879**	**0.879**	**0.877**	**0.878**	**0.880**	**0.879**	**0.878**	**0.878**
BOP	**0.881**	0.887	**0.879**	0.883	**0.882**	0.888	**0.880**	0.884
CW	0.964	0.955	0.982	0.976	0.959	0.954	0.976	0.974
D) $T = 500$								
St-MSh	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960
D-MSh	0.999	0.999	0.999	0.999	0.992	0.991	0.991	0.991
O-LSh	0.934	**0.934**	0.934	**0.933**	0.934	**0.934**	0.933	**0.934**
T-LSh	0.934	0.934	0.934	0.933	0.934	0.934	0.933	0.934
PJS	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960
Wang	**0.933**	**0.933**	**0.933**	**0.933**	**0.933**	**0.933**	**0.933**	**0.933**
BOP	**0.933**	0.937	**0.933**	0.935	**0.934**	0.938	**0.933**	0.935
CW	0.981	0.976	0.990	0.987	0.978	0.976	0.988	0.987

Notes. This table reports the average of the L_2-error ratios in (37) across 1,000 simulations for different estimators of the mean vector and different parameter values. Panels A to D consider a sample size $T \in \{50, 100, 250, 500\}$ while $N_T/T \in \{2, 4\}$. The data generating process is described in Section 5.2.1. The data is simulated either from a multivariate normal distribution or from a multivariate t-distribution with $\nu = 3$ degrees of freedom. In both cases, the dependence is governed by a Toeplitz correlation matrix with $\rho \in \{0, 0.5\}$. The proposed estimators are St-MSh and D-MSh in Section 3, and O-LSh and T-LSh in Section 4. The PJS, Wang, BOP, and CW benchmark estimators are described in Section 5.1. The smallest average L_2-error ratios are in underlined bold, and the second-smallest in bold.

Third, the O-LSh, T-LSh, Wang, BOP, and Jorion estimators have a similar performance overall, outperforming St-MSh, D-MSh, PJS, and CW. Although these estimators perform similarly, the Wang estimator deliver a smaller error when $N_T > T$, which may be attributed to the use of U-statistics in the shrinkage intensity that converge more rapidly than sample average estimators when N_T is large (Chen et al. 2010). We also note that, when $N_T > T$, the CW estimator performs poorly with an L_2-error ratio not much lower than one.

Overall, these simulations show that our shrinkage estimators are competitive with a range of benchmark estimators.

5.3 Empirical stock return data

We now apply our proposed shrinkage mean-vector estimators (St-MSh, D-MSh, O-LSh, T-LSh) to portfolio selection based on stock-return data, and compare them to the benchmarks (Sample, PJS, Wang, BOP, Jorion, CW). Stock returns differ in volatility and correlation, offering a more challenging setting for MSE-based estimators, like the ones we develop, relative to those based on the invariant loss. Shrinkage methods, and robust methods generally, are popular in portfolio selection for alleviating the impact of parameter uncertainty on the out-of-sample performance of optimal portfolios; see, e.g., Ledoit & Wolf (2003), DeMiguel et al. (2009), Tu & Zhou (2011), Barroso & Saxena (2021), Callot et al. (2021), Kan et al. (2021), Kazak & Pohlmeier (2022), Lassance et al. (2022), Bodnar et al. (2023), and Lassance, Vanderveken & Vrins (2024). In Section 5.3.1, we describe our dataset. In Section 5.3.2, we explain how we evaluate the out-of-sample Sharpe ratio of a given mean-variance portfolio estimator. In Section 5.3.3, we discuss the results.

5.3.1 Empirical dataset

Our dataset is composed of daily total returns in excess of the risk-free rate on 441 stocks continuously listed in the S&P500 index from January 3, 2000 to December 29, 2023. These stocks are selected from the 1,070 stocks that appeared in the S&P500 index at any point

during this observation period. We download the stock returns from Wharton Research Data Services (WRDS) and the risk-free rate from Kenneth French's website. In total, we have 6,037 daily excess-return observations for each stock. Focusing on firms with uninterrupted listing reduces structural breaks caused by firm turnover, but introduces selection bias. However, selection bias is not a major concern for us because we care about the relative performance of our estimators relative to the benchmarks, not the absolute performance.

Although we focus on daily returns in the main text, which allows a better-estimated covariance matrix and a setting where both N_T and T are large, we test the robustness of our results to using weekly returns in Section D of the Supplementary Material.

In Figure 1, we depict the time series of daily returns and cumulative daily returns for the S&P500 index over the time period considered. We also depict structural breakpoints identified by the Bai & Perron (2003) test. These breakpoints identify two periods of interest that we will use in our analysis: i) the period during and after the 2008 financial crisis (October 6, 2008 to August 3, 2011), ii) the period during and after the covid-19 pandemic (February 26, 2020 to December 29, 2023).

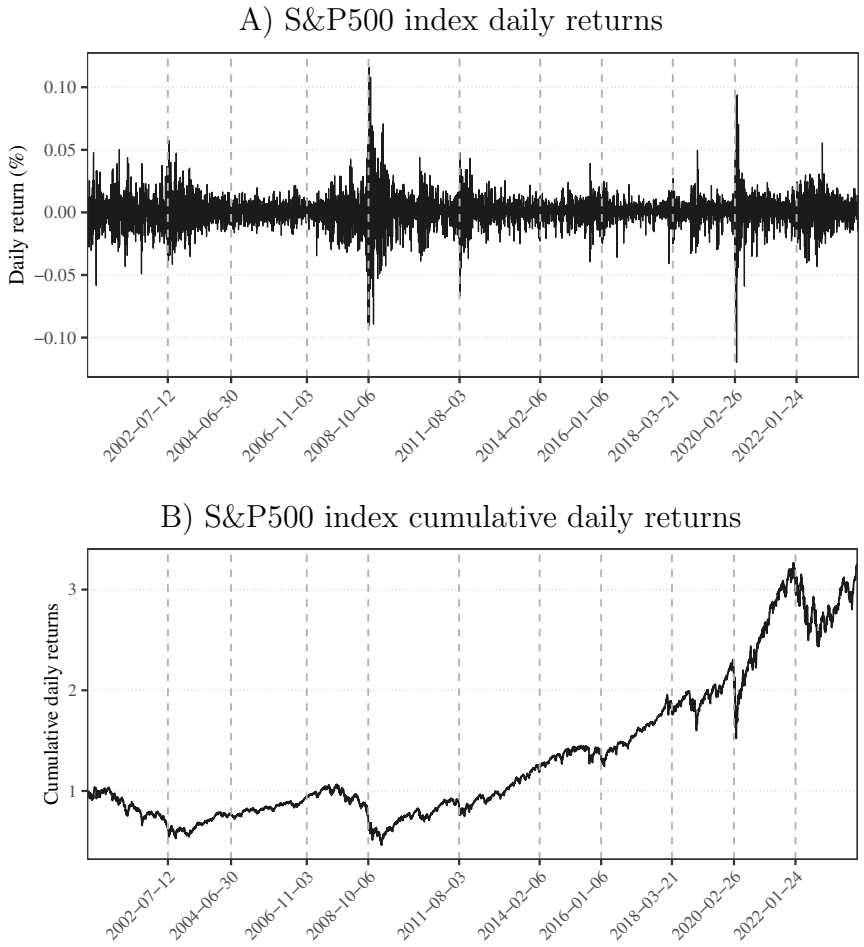
5.3.2 Performance measure for empirical data

We measure the performance of the expected-return estimators with the out-of-sample economic gains they deliver when applied by a mean-variance investor. To achieve this, we adopt a rolling-window framework with a training-window length of two years (504 days) and a testing-window length of six months (126 days). In each training window, we obtain an estimate of the expected-excess-return vector, $\hat{\boldsymbol{\mu}}_{N_T}$, depending on the selected estimator. Given $\hat{\boldsymbol{\mu}}_{N_T}$, we obtain an estimate of the Sharpe-ratio-optimal mean-variance portfolio:

$$\hat{\mathbf{w}}^{\star}_{N_T} := \frac{1}{\gamma}\mathbf{S}^{-1}_{N_T}\hat{\boldsymbol{\mu}}_{N_T}, \tag{38}$$

where γ is the risk-aversion coefficient, set in two different ways: i) a constant, whose value is without loss of generality because it does not affect the Sharpe ratio, ii) $\gamma = \mathbf{1}'_{N_T}\mathbf{S}^{-1}_{N_T}\hat{\boldsymbol{\mu}}_{N_T}$,

Figure 1: Time series of S&P500 index



Notes. This figure depicts the time series of S&P500 index daily total returns from January 3, 2000 to December 29, 2023. Panel A depicts returns in percentage. Panel B depicts cumulative returns. The dotted black vertical lines depict structural breakpoints identified by the Bai & Perron (2003) test.

in which case we recover the tangency portfolio that does not invest in the risk-free asset. This means that $T = 504$, $N_T = 441$, and $N_T/T = 0.875$, which is a challenging setting for portfolio selection. Note that $\mathbf{S}_{N_T}^{-1}$ exists because $T > N_T$, and the CW estimator does not work as it requires $N_T > T$. Given $\hat{\mathbf{w}}_{N_T}^{\star}$, we compute the out-of-sample portfolio excess returns over the following six-month testing window and evaluate the Sharpe ratio, defined as the ratio between the sample mean and sample standard deviation of the out-of-sample portfolio excess returns. In unreported results, we find that the comparison of our expected-return estimators to the benchmarks in terms of out-of-sample Sharpe ratio is similar when using a testing window shorter than six months, but the performance is more variable across

testing windows as they get shorter. We then roll the training and testing windows by one month. Finally, we report the average and standard deviation of the annualized out-of-sample Sharpe ratios across testing windows, for the full out-of-sample period and during and after the 2008 financial crisis and covid-19 pandemic periods, identified in Section 5.3.1. We report the standard deviation of out-of-sample Sharpe ratios following Lassance, Martín-Utrera & Simaan (2024) who argue that the risk of out-of-sample performance matters for uncertainty-averse investors. In total, we have 258, 34, and 40 testing windows for the full period, the financial crisis, and the pandemic period, respectively.

5.3.3 Discussion of empirical results

In Table 3, we report the average and standard deviation of the annualized out-of-sample Sharpe ratios delivered by the mean-variance portfolio $\hat{\mathbf{w}}^{\star}_{N_T}$ in (38) for the different expected-return estimators and time periods considered. Panel A considers the mean-variance portfolio with a risk-free asset investment (i.e., the risk-aversion coefficient γ is constant) and Panel B considers the tangency portfolio without a risk-free asset investment (i.e., $\gamma = \mathbf{1}'_{N_T}\mathbf{S}^{-1}_{N_T}\hat{\boldsymbol{\mu}}_{N_T}$). We do not include the PJS benchmark estimator because there are many training windows where it yields a zero mean-return vector, making the Sharpe ratio undefined. The main observation we draw from Table 3 is that, relative to the benchmarks, our novel expected-return estimators deliver an out-of-sample portfolio performance that is overall superior and more robust across periods.

In Panel A, D-MSh and O-LSh provide the most robust average out-of-sample Sharpe ratios for the full period and during the financial and pandemic crises, albeit with a slightly larger standard deviation than the benchmarks. Remarkably, during the pandemic period, D-MSh and O-LSh deliver an average out-of-sample Sharpe ratio of 0.616 and 0.448, respectively, versus 0.333 for the best benchmark, i.e., Jorion. The Wang estimator performs best for the full period and during the financial crisis, but delivers a negative average out-of-sample Sharpe ratio during the pandemic period.

Table 3: Out-of-sample Sharpe ratio statistics in empirical daily stock-return data

	Full period		Financial crisis		Pandemic	
	Mean	Std dev.	Mean	Std dev.	Mean	Std dev.
A) Mean-variance portfolio with a risk-free asset investment						
St-MSh	0.274	1.352	0.207	1.433	0.300	1.257
D-MSh	0.284	1.415	0.443	1.724	**<u>0.616</u>**	**1.241**
O-LSh	**0.505**	1.367	**0.502**	1.488	**0.448**	1.253
T-LSh	0.369	1.353	0.243	1.531	0.356	1.266
Sample	0.274	1.352	0.207	1.433	0.300	1.257
Wang	**<u>0.591</u>**	**<u>1.325</u>**	**<u>0.800</u>**	**<u>1.230</u>**	−0.081	1.246
BOP	0.007	**1.332**	−0.137	1.485	−0.211	**<u>1.160</u>**
Jorion	0.458	1.406	0.281	**1.400**	0.333	1.296
B) Tangency portfolio without a risk-free asset investment						
St-MSh	0.385	1.325	0.457	1.372	0.309	1.255
D-MSh	0.266	1.419	0.142	1.776	**<u>0.616</u>**	**1.241**
O-LSh	**<u>0.610</u>**	1.323	**<u>0.800</u>**	**<u>1.346</u>**	**0.448**	1.253
T-LSh	0.478	**<u>1.318</u>**	**0.711**	1.373	0.419	1.246
Sample	0.385	1.325	0.457	1.372	0.309	1.255
Wang	**0.601**	**1.321**	0.176	1.463	−0.136	1.242
BOP	−0.119	1.326	−0.458	1.417	−0.202	**<u>1.162</u>**
Jorion	0.551	1.373	0.460	**1.350**	0.369	1.286

Notes. This table reports the average and standard deviation of the annualized out-of-sample Sharpe ratios across all testing windows for different estimators, mean-variance portfolios, and time periods. We use a dataset of daily excess returns on $N_T = 441$ S&P500 individual stocks spanning the period from January 3, 2000 to December 29, 2023; see Section 5.3.1 for details. The proposed estimators are St-MSh and D-MSh in Section 3, and O-LSh and T-LSh in Section 4. The Sample, Wang, BOP, and Jorion benchmark estimators are described in Section 5.1. We do not include the PJS benchmark estimator because there are many training windows where it yields a zero mean-return vector, making the Sharpe ratio undefined. We adopt a rolling-window framework with a training-window length of two years (504 trading days) in which we estimate the vector of expected returns and the resulting mean-variance portfolio in (38), followed by a six-month (126 days) testing period in which we evaluate the Sharpe ratio. We then roll the training and testing windows by one month and proceed again. Finally, we report the average and standard deviation of the out-of-sample Sharpe ratios, for the full out-of-sample period but also during and after the 2008 financial crisis (October 6, 2008 to August 3, 2011) and covid-19 pandemic (February 26, 2020 to December 29, 2023), which we identify with the Bai & Perron (2003) test; see Figure 1. In total, we have 258, 34, and 40 testing windows for the full period, the financial crisis, and the pandemic period, respectively. In Panel A, we set a constant risk-aversion coefficient γ in the mean-variance portfolio (38), whose value does not affect the Sharpe ratio. In Panel B, we set $\gamma = \mathbf{1}_{N_T}' \mathbf{S}_{N_T}^{-1} \hat{\boldsymbol{\mu}}_{N_T}$, in which case we recover the tangency portfolio that does not invest in the risk-free asset. The highest average Sharpe ratios and lowest standard deviation of Sharpe ratios are in underlined bold, and the second-best in bold.

In Panel B, the results are more variable due to the time-varying γ, but our estimators still compare favorably. Specifically, O-LSh delivers the largest average out-of-sample Sharpe ratio for the full period and during the financial crisis, with a standard deviation similar to that of the other estimators. During the pandemic period, O-LSh provides the second-highest average out-of-sample Sharpe ratio, and D-MSh the largest one. The improvements compared to the benchmarks in Panel B can be substantial. For example, during the financial crisis, O-LSh and T-LSh deliver an average out-of-sample Sharpe ratio of 0.800 and 0.711, respectively, versus 0.457, 0.176, -0.458, and 0.460 for Sample, Wang, BOP, and Jorion, respectively. Overall, the results in Table 3 suggest that our shrinkage expected-return estimators deliver valuable economic gains when used in the construction of optimal mean-variance portfolios.

6 Conclusion & Discussion

Estimating the mean vector of a multivariate distribution is a central problem in statistics, with widespread applications in finance. This problem is particularly challenging in high dimension where the number of variables N is not negligible relative to the sample size T. In this context, we introduce four bona-fide shrinkage estimators of the mean vector that share several desirable properties not commonly satisfied by existing comparable methods. First, they are distribution-free, only requiring i.i.d. observations with uniformly bounded first four moments. Second, they use shrinkage intensities that are optimal for the bona-fide target estimator, not the oracle one. Third, they are optimal under the MSE loss function, and thus, do not depend on an estimated precision matrix. Fourth, they are L_2-consistent estimators of their oracle counterpart in a general Kolmogorov setting where $N/T = O(1)$ as $T \to \infty$. Simulation and empirical tests emphasize the value of our estimators relative to the sample mean and the estimators in James & Stein (1961), Jorion (1986), Chételat & Wells (2012), Wang et al. (2014), and Bodnar et al. (2019) for fitting the true mean vector and constructing mean-variance portfolios of a large number of S&P500 individual stocks.

While our proposed estimators are distribution-free and enjoy desirable high-dimensional consistency, they also have limitations and a specific scope of applicability. In particular, we rely on the L_2 loss function to derive the oracle shrinkage intensities and to establish consistency, and we assume i.i.d. sampling with bounded fourth moments. These assumptions may be restrictive in some applications. However, we have conducted several robustness checks of our simulation results under alternative settings, including the invariant (Mahalanobis) loss function, heterogeneous variances, and non-i.i.d. autoregressive data. Moreover, our empirical application uses S&P500 stock returns at different frequencies, which typically violate the i.i.d. assumption. Although these tests tend to confirm the benefits of our estimators beyond their theoretical scope, future research may extend our work to relax the assumptions, strengthen robustness, and broaden the empirical tests.

A relevant extension of our work is to adopt the invariant loss function, which is often preferred when variables differ in scale and correlation. The question is whether one can continue to avoid the specific stochastic structure used in Wang et al. (2014) and Bodnar et al. (2019), which, as discussed, is more restrictive than only assuming i.i.d. observations like we do. Another related direction is to employ more robust loss functions than L_2, such as Huber's or Tukey's loss. Moreover, extending our framework to accommodate more complex data panels with serial dependence, heteroskedasticity, or missing observations, which are common in financial time series, would broaden its practical scope. This represents a demanding but valuable research agenda, as most existing shrinkage estimators of a form similar to ours rely on i.i.d. panels (see Böhm & von Sachs (2009) for an exception). Finally, while our empirical analysis using S&P500 stocks shows promising results, additional studies across different asset classes and financial datasets would help assess the generalization of our findings.

Acknowledgments

This work was supported by the Fonds de la Recherche Scientifique (FNRS) under Grant Number J.0135.25. We thank conference participants at the 2025 INFORMS Annual Meeting (Atlanta) and the 3rd International Conference on Actuarial Science, Quantitative Finance and Risk Management (Beijing), and seminar participants at Georgia State University and Soochow University for their comments.

Data availability statement

The empirical data and replication codes are available as an online supplement to the paper.

Disclosure statement

The authors report there are no competing interests to declare.

References

Bai, J. & Perron, P. (2003), 'Computation and analysis of multiple structural change models', *Journal of Applied Econometrics* **18**(1), 1–22.

Bai, Z. & Saranadasa, H. (1996), 'Effect of high dimension: By an example of a two sample problem', *Statistica Sinica* **6**(2), 311–329.

Barroso, P. & Saxena, K. (2021), 'Lest we forget: Using out-of-sample forecast errors in portfolio optimization', *Review of Financial Studies* **35**(3), 1222–1278.

Bean, D., Bickel, P. J., El Karoui, N. & Yu, B. (2013), 'Optimal M-estimation in high-dimensional regression', *Proceedings of the National Academy of Sciences* **110**(36), 14563–14568.

Berger, J. O. & Bock, M. E. (1976), 'Combining independent normal mean estimation problems with unknown variances', *Annals of Statistics* **4**(3), 642–648.

Bodnar, O., Bodnar, T. & Parolya, N. (2022), 'Recent advances in shrinkage-based high-dimensional inference', *Journal of Multivariate Analysis* **188**.

Bodnar, T., Okhrin, O. & Parolya, N. (2019), 'Optimal shrinkage estimator for high-dimensional mean vector', *Journal of Multivariate Analysis* **170**, 63–79.

Bodnar, T., Okhrin, Y. & Parolya, N. (2023), 'Optimal shrinkage-based portfolio selection in high dimensions', *Journal of Business & Economic Statistics* **41**(1), 140–156.

Böhm, H. & von Sachs, R. (2009), 'Shrinkage estimation in the frequency domain of multivariate time series', *Journal of Multivariate Analysis* **100**(5), 913–935.

Callot, L., Caner, M., Önder, O. & Ulaşan, E. (2021), 'A nodewise regression approach to estimating large portfolios', *Journal of Business & Economic Statistics* **39**(2), 520–531.

Chen, S. X., Li-Xin, Z. & Zhong, P.-S. (2010), 'Tests for high-dimensional covariance matrices', *Journal of the American Statistical Association* **105**(490), 810–819.

Chételat, D. & Wells, M. T. (2012), 'Improved multivariate normal mean estimation with unknown covariance when p is greater than n', *Annals of Statistics* **40**(6), 3137–3160.

DeMiguel, V., Garlappi, L., Nogales, F. & Uppal, R. (2009), 'A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms', *Management Science* **55**(5), 798–812.

DeMiguel, V., Nogales, F. & Uppal, R. (2014), 'Stock return serial dependence and out-of-sample portfolio performance', *Review of Financial Studies* **27**(4), 1031–1073.

Efron, B. & Morris, C. (1972), 'Limiting the risk of Bayes and empirical Bayes estimators–Part II: The empirical Bayes case', *Journal of the American Statistical Association* **67**(337), 130–139.

Efron, B. & Morris, C. (1973), 'Stein's estimation rule and its competitors–An empirical Bayes approach', *Journal of the American Statistical Association* **68**(341), 117–130.

Fourdrinier, D., Strawderman, W. E. & Wells, M. T. (2003), 'Robust shrinkage estimation for elliptically symmetric distributions with unknown covariance matrix', *Journal of Multivariate Analysis* **85**, 24–39.

Gleser, L. J. (1986), 'Minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix', *Annals of Statistics* **14**(4), 1625–1633.

Green, E. & Strawderman, W. E. (1991), 'A James-Stein type estimator for combining unbiased and possibly biased estimators', *Journal of the American Statistical Association* **86**(416), 1001–1006.

James, W. & Stein, C. (1961), 'Estimation with quadratic loss', *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1**, 361–379.

Jorion, P. (1986), 'Bayes-Stein estimation for portfolio analysis', *Journal of Financial and Quantitative Analysis* **21**(3), 279–292.

Kan, R., Wang, X. & Zhou, G. (2021), 'Optimal portfolio choice with estimation risk: No risk-free asset case', *Management Science* **68**(3), 2047–2068.

Kazak, E. & Pohlmeier, W. (2022), 'Bagged pretested portfolio selection', *Journal of Business & Economic Statistics* **41**(4), 1116–1131.

Lassance, N. (2025), 'A sequential approach to shrinkage estimation', *SSRN working paper*.

Lassance, N., DeMiguel, V. & Vrins, F. (2022), 'Optimal portfolio diversification via independent component analysis', *Operations Research* **70**(1), 55–72.

Lassance, N., Martín-Utrera, A. & Simaan, M. (2024), 'The risk of expected utility under parameter uncertainty', *Management Science* **70**(11), 7644–7663.

Lassance, N., Vanderveken, R. & Vrins, F. (2024), 'On the combination of naive and mean-variance portfolio strategies', *Journal of Business & Economic Statistics* **42**(3), 875–889.

Ledoit, O. & Wolf, M. (2003), 'Honey, I shrunk the sample covariance matrix', *Journal of Portfolio Management* **30**(4), 110–119.

Ledoit, O. & Wolf, M. (2004), 'A well-conditioned estimator for large-dimensional covariance matrices', *Journal of Multivariate Analysis* **88**(2), 365–411.

Ledoit, O. & Wolf, M. (2022), 'Quadratic shrinkage for large covariance matrices', *Bernoulli* **28**(3), 1519–1547.

Stein, C. (1956), 'Inadmissibility of the usual estimator for the mean of a multivariate distribution', *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1**, 197–206.

Thompson, J. (1968), 'Some shrinkage techniques for estimating the mean', *Journal of the American Statistical Association* **63**(321), 113–122.

Tu, J. & Zhou, G. (2011), 'Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies', *Journal of Financial Economics* **99**(1), 204–215.

Wang, C., Tong, T., Cao, L. & Miao, B. (2014), 'Non-parametric shrinkage mean estimation for quadratic loss functions with unknown covariance matrices', *Journal of Multivariate Analysis* **125**, 222–232.

Welch, I. & Goyal, A. (2008), 'A comprehensive look at the empirical performance of equity premium prediction', *Review of Financial Studies* **21**(4), 1455–1508.

Xie, X., Kou, S. C. & Brown, L. (2016), 'Optimal shrinkage estimation of mean parameters in family of distributions with quadratic variance', *Annals of Statistics* **44**(2), 564–597.

Supplementary Material to

Distribution-free shrinkage of high-dimensional mean vector

Vali Asimit Ziwei Chen Nathan Lassance

This Supplementary Material contains six sections. In Section A, we report simulation results under the L_1 and invariant loss functions. In Section B, we report simulation results using heterogeneous variances. In Section C, we report simulation results using a non-i.i.d. VAR(1) model. In Section D, we report empirical results using weekly stock returns. In Section E, we provide useful lemmas for proving our theorems. In Section F, we detail the proofs of our theorems.

A Alternative loss functions

In this section, we test the robustness of our simulation results in Tables 1 and 2 to using loss functions based on the L_1 norm and the invariant (i.e., Mahalanobis) norm.

First, in Tables A.1 and A.2, we replicate the simulation results in Tables 1 and 2 using the L_1 norm. Specifically, in a given simulation and for a given mean-vector estimate $\hat{\boldsymbol{\mu}}_{N_T}$, we measure the performance using the L_1-error ratio relative to the sample mean:

$$L_1\text{-error ratio}(\hat{\boldsymbol{\mu}}_{N_T}) := \frac{\left|\left|\hat{\boldsymbol{\mu}}_{N_T} - \boldsymbol{\mu}_{N_T}\right|\right|_1}{\left|\left|\overline{\mathbf{X}}^{(T)} - \boldsymbol{\mu}_{N_T}\right|\right|_1}, \tag{A1}$$

which we then average across all simulations. We find that the performance of our mean-vector estimators relative to the benchmarks is similar to the one under the average L_2-error ratio in Tables 1 and 2, with O-LSh and T-LSh being relatively better ranked. This is a sensible result given that L_2 consistency implies L_1 consistency, as mentioned in Section 2.3.

Second, in Tables A.3 and A.4, we replicate the simulation results in Tables 1 and 2 using the invariant (i.e., Mahalanobis) loss function. Specifically, in a given simulation and for a given mean-vector estimate $\hat{\boldsymbol{\mu}}_{N_T}$, we measure the performance using the invariant-error ratio relative to the sample mean:

$$\text{Invariant-error ratio}(\hat{\boldsymbol{\mu}}_{N_T}) := \frac{(\hat{\boldsymbol{\mu}}_{N_T} - \boldsymbol{\mu}_{N_T})'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\mu}}_{N_T} - \boldsymbol{\mu}_{N_T})}{(\overline{\mathbf{X}}^{(T)} - \boldsymbol{\mu}_{N_T})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{X}}^{(T)} - \boldsymbol{\mu}_{N_T})}, \tag{A2}$$

which we then average across all simulations. We find that there are now larger differences

in ranking among estimators relative to the case with the average L_2-error ratio in Tables 1 and 2. However, the conclusions drawn from the L_2-error ratio are overall robust to using the invariant-error ratio. Still, one distinguishing difference is that the Jorion and BOP estimators, which are designed under the invariant loss, perform relatively better.

Overall, the results contained in this section indicate a robustness of the simulation results to the choice of the loss function.

B Heterogeneous variances

In the simulation analysis of Section 5.2, we use variables with unit variances. In this section, we test the robustness of the results to using variances randomly drawn from a uniform distribution. Specifically, we use $\Sigma_{kk,N_T} \sim U(0.5, 5)$ for all k. The rest of the simulation setting remains the same as that described in Section 5.2.1.

We report the results in Tables A.5 and A.6 for $T > N_T$ and $N_T > T$, respectively. Relative to the case with unit variances in Tables 1 and 2, the main observation is that the BOP and Jorion estimators, which are designed to minimize the invariant loss and thus account for heterogeneous variances unlike our MSE-based estimators, perform relatively better. However, our proposed estimators minimizing the MSE continue to have benefits even in this adversarial setting with highly heterogeneous variances. In particular, O-LSh is often among the best two estimators.

C Simulated data from a VAR(1) model

In this section, we test the robustness of the simulation results in Tables 1 and 2 to using non-i.i.d. data simulated from a normal vector-autoregressive model of order one, i.e., a VAR(1) model. The general form of the VAR(1) model is as follows:

$$\mathbf{X}_t = \mathbf{c}_{N_T} + \mathbf{A}_{N_T}\mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}_{N_T}, \boldsymbol{\Omega}_{N_T}), \tag{A3}$$

where $\mathbf{c}_{N_T} \in \mathbb{R}^{N_T}$ is the intercept, $\mathbf{A}_{N_T} \in \mathbb{R}^{N_T \times N_T}$ is the VAR(1) coefficient matrix, and $\boldsymbol{\epsilon}_t$ is the residual term with covariance matrix $\boldsymbol{\Omega}_{N_T} \in \mathbb{R}^{N_T \times N_T}$. We need the eigenvalues of $\mathbf{A}$ to be strictly smaller than one to ensure the VAR(1) process is stationary. We have that $\mathbf{X}_t$ is unconditionally multivariate normal with unconditional mean $\boldsymbol{\mu}_{N_T} = \mathbb{E}[\mathbf{X}_t]$ given by

$$\boldsymbol{\mu}_{N_T} = (\mathbf{I}_{N_T} - \mathbf{A}_{N_T})^{-1}\mathbf{c}_{N_T}, \tag{A4}$$

and an unconditional covariance matrix $\boldsymbol{\Sigma}_{N_T} = \mathbb{V}[\mathbf{X}_t]$ being the solution to

$$\boldsymbol{\Sigma}_{N_T} = \mathbf{A}_{N_T}\boldsymbol{\Sigma}_{N_T}\mathbf{A}'_{N_T} + \boldsymbol{\Omega}_{N_T}. \tag{A5}$$

For simplicity, we set $\mathbf{A}_{N_T} = \varphi\mathbf{I}_{N_T}$ with $\varphi \in (0, 1)$. In that case, given $\boldsymbol{\mu}_{N_T}$ and $\boldsymbol{\Sigma}_{N_T}$ calibrated as in the base-case simulation setting of Section 5.2.1, we have from (A4)–(A5) that we need to set $\mathbf{c}_{N_T}$ and $\boldsymbol{\Omega}_{N_T}$ equal to

$$\mathbf{c}_{N_T} = (1 - \varphi)\boldsymbol{\mu}_{N_T}, \tag{A6}$$

$$\boldsymbol{\Omega}_{N_T} = (1 - \varphi^2)\boldsymbol{\Sigma}_{N_T}. \tag{A7}$$

Finally, in each simulation, we randomly draw $\varphi \sim U(0, 1)$ and we generate a sample of size T from the VAR(1) model in (A3)–(A6)–(A7) using $\mathbf{X}_0 = \boldsymbol{\mu}_{N_T}$.

We report the results in Table A.7. The variables are unconditionally multivariate normal in the VAR(1) model, and we model the unconditional correlation matrix as a Toeplitz matrix with correlation $\rho \in \{0, 0.5\}$, in line with the base-case simulation setting. Given that the model is unconditionally multivariate normal, the relevant comparison is with the results in the panel "Multivariate normal" of Tables 1 and 2.

The results in Table A.7 are consistent with the panel "Multivariate normal" of Tables 1 and 2 given that O-LSh and Wang are the best two estimators. However, PJS also performs well when $T = 50$ is small. Therefore, the main distinguishing difference we observe relative to the base-case simulation setting is that when the i.i.d. assumption is violated, more parsimonious estimators, such as PJS, perform relatively better. This estimator is a simple

scaling of the sample mean estimator.

D Weekly stock returns

In this section, we test the robustness of the empirical results in Table 3 to using weekly excess returns, instead of daily. It is valuable to study the effect of reducing the return frequency because it has two main effects. On the one hand, it reduces the sample size, and thus, reduces estimation accuracy. On the other hand, it tends to render the stock returns closer to i.i.d. because of time aggregation, and thus, fits more closely the i.i.d. assumption underlying our mean-vector estimators and the benchmarks we consider.

In Table A.8, we replicate the out-of-sample Sharpe ratio statistics in Table 3 using weekly data. We do not consider monthly data because the sample size T would be too small relative to the number of assets N_T to construct optimal mean-variance portfolios in (38). We use a training-window length of ten years, i.e., $T = 520$, which is similar to $T = 504$ for the case of daily returns in Table 3. We also use testing windows of one year (52 weeks) to evaluate the out-of-sample Sharpe ratios. Overall, we find that the conclusions drawn from daily returns in Table 3 are robust to using weekly returns. In particular, D-MSh and O-LSh are the two estimators that perform most consistently well. Note that, unlike Table 3, the Sharpe ratios are negative during the financial crisis for all estimators except BOP. However, one ought to be cautious with these results because we only have 21 testing windows spanning the financial crisis due to the use of a lower return frequency and an out-of-sample period that starts later due to the use of a 10-year training-window length.

Table A.1: Average L_1-error ratio in simulated data for $T > N_T$

	Multivariate normal				Multivariate t with $\nu = 3$			
N_T/T	0.25	0.25	0.5	0.5	0.25	0.25	0.5	0.5
ρ	0	0.5	0	0.5	0	0.5	0	0.5
A) $T = 50$								
St-MSh	0.789	**0.797**	0.774	0.789	0.798	**0.808**	0.786	0.798
D-MSh	0.862	0.860	0.858	0.863	0.854	0.858	0.852	0.855
O-LSh	**0.749**	**0.803**	0.723	**0.761**	**0.757**	0.811	**0.735**	**0.766**
T-LSh	0.757	0.808	0.727	0.776	0.765	0.816	0.740	0.781
PJS	0.799	0.820	0.767	0.797	0.807	0.829	0.783	0.807
Wang	0.788	0.845	**0.719**	**0.775**	0.797	0.853	0.736	0.779
BOP	0.858	0.869	0.799	0.816	0.825	0.847	0.777	0.795
Jorion	**0.736**	0.807	**0.716**	0.778	**0.742**	**0.806**	**0.731**	**0.779**
B) $T = 100$								
St-MSh	0.862	0.866	0.853	0.860	0.869	0.871	0.860	0.867
D-MSh	0.926	0.921	0.919	0.922	0.918	0.913	0.912	0.912
O-LSh	**0.809**	**0.832**	**0.796**	**0.811**	**0.817**	**0.837**	**0.803**	**0.817**
T-LSh	0.813	0.848	0.796	0.817	0.821	0.849	**0.804**	**0.822**
PJS	0.865	0.873	0.852	0.863	0.874	0.879	0.862	0.871
Wang	0.820	0.847	**0.795**	**0.814**	0.835	0.855	0.808	0.824
BOP	0.840	**0.838**	0.820	0.825	0.837	**0.841**	0.817	0.830
Jorion	**0.806**	0.840	0.797	0.832	**0.816**	0.846	0.807	0.839
C) $T = 250$								
St-MSh	0.927	0.935	0.927	0.927	0.929	0.937	0.930	0.930
D-MSh	0.975	0.978	0.975	0.974	0.967	0.972	0.970	0.967
O-LSh	**0.890**	**0.897**	**0.885**	**0.890**	**0.892**	**0.902**	**0.890**	**0.893**
T-LSh	0.890	**0.900**	0.885	0.890	**0.892**	**0.903**	**0.890**	**0.893**
PJS	0.927	0.937	0.926	0.927	0.930	0.940	0.930	0.930
Wang	0.891	0.900	**0.884**	**0.890**	0.895	0.906	0.890	0.894
BOP	0.893	0.905	0.888	0.900	0.894	0.909	0.891	0.903
Jorion	**0.889**	0.909	0.886	0.906	0.893	0.913	0.892	0.908
D) $T = 500$								
St-MSh	0.963	0.963	0.961	0.962	0.963	0.965	0.961	0.963
D-MSh	1.003	1.001	1.001	1.000	0.998	0.999	0.998	0.996
O-LSh	**0.938**	**0.939**	**0.935**	**0.936**	**0.939**	**0.942**	**0.937**	**0.938**
T-LSh	**0.938**	**0.939**	0.935	0.936	**0.939**	**0.942**	**0.937**	**0.939**
PJS	0.963	0.964	0.961	0.962	0.963	0.966	0.962	0.963
Wang	0.938	0.939	**0.935**	**0.936**	0.940	0.943	0.937	0.939
BOP	0.939	0.945	0.936	0.944	0.939	0.947	0.937	0.945
Jorion	0.938	0.946	0.935	0.946	0.939	0.948	0.937	0.947

Notes. This table reports the average of the L_1-error ratios in (A1) across 1,000 simulations for different estimators of the mean vector and different parameter values. Panels A to D consider a sample size $T \in \{50, 100, 250, 500\}$ while $N_T/T \in \{0.25, 0.5\}$. The data generating process is described in Section 5.2.1. The data is simulated either from a multivariate normal distribution or from a multivariate t-distribution with $\nu = 3$ degrees of freedom. In both cases, the dependence is governed by a Toeplitz correlation matrix with $\rho \in \{0, 0.5\}$. The proposed estimators are St-MSh and D-MSh in Section 3, and O-LSh and T-LSh in Section 4. The PJS, Wang, BOP, and Jorion benchmark estimators are described in Section 5.1. The smallest average L_1-error ratios are in underlined bold, and the second-smallest in bold.

Table A.2: Average L_1-error ratio in simulated data for $N_T > T$

	Multivariate normal				Multivariate t with $\nu = 3$			
N_T/T	2	2	4	4	2	2	4	4
ρ	0	0.5	0	0.5	0	0.5	0	0.5
A) $T = 50$								
St-MSh	0.768	0.771	0.767	0.768	0.779	0.780	0.777	0.777
D-MSh	0.856	0.857	0.856	0.856	0.849	0.848	0.848	0.847
O-LSh	**0.704**	0.713	**0.700**	0.706	**0.714**	0.721	0.709	0.714
T-LSh	0.704	0.716	0.701	0.707	0.714	0.724	0.709	0.714
PJS	0.748	0.755	0.745	0.747	0.766	0.771	0.762	0.764
Wang	**0.668**	**0.683**	**0.659**	**0.668**	**0.689**	**0.700**	**0.679**	**0.686**
BOP	0.739	**0.708**	0.714	**0.687**	0.728	**0.721**	**0.705**	**0.706**
CW	0.892	0.863	0.945	0.927	0.881	0.866	0.935	0.925
B) $T = 100$								
St-MSh	0.847	0.849	0.848	0.849	0.854	0.856	0.854	0.855
D-MSh	0.919	0.919	0.919	0.919	0.910	0.910	0.910	0.910
O-LSh	**0.785**	0.789	**0.784**	0.786	**0.792**	**0.796**	0.791	0.792
T-LSh	0.785	0.790	0.784	0.786	0.792	0.797	0.791	0.793
PJS	0.841	0.844	0.841	0.842	0.851	0.854	0.851	0.853
Wang	**0.770**	**0.778**	**0.768**	**0.771**	**0.784**	**0.790**	**0.782**	**0.784**
BOP	0.793	**0.786**	0.784	**0.779**	0.795	0.799	**0.789**	**0.791**
CW	0.925	0.908	0.963	0.952	0.921	0.912	0.958	0.951
C) $T = 250$								
St-MSh	0.927	0.926	0.926	0.926	0.929	0.929	0.929	0.928
D-MSh	0.975	0.975	0.975	0.974	0.968	0.968	0.968	0.968
O-LSh	**0.884**	**0.885**	0.883	**0.883**	**0.888**	**0.888**	0.887	0.887
T-LSh	0.884	0.885	0.883	0.883	0.888	0.888	0.887	**0.887**
PJS	0.926	0.925	0.925	0.925	0.929	0.929	0.929	0.928
Wang	**0.882**	**0.882**	**0.880**	**0.881**	**0.887**	**0.887**	**0.886**	**0.886**
BOP	0.885	0.888	**0.883**	0.884	0.889	0.892	**0.886**	0.888
CW	0.965	0.955	0.982	0.976	0.963	0.957	0.980	0.977
D) $T = 500$								
St-MSh	0.960	0.960	0.960	0.960	0.962	0.962	0.962	0.961
D-MSh	1.000	1.000	1.000	1.000	0.996	0.996	0.996	0.996
O-LSh	**0.934**	**0.934**	**0.934**	**0.934**	**0.936**	**0.936**	**0.936**	**0.936**
T-LSh	0.934	0.934	0.934	0.934	0.936	**0.936**	0.936	0.936
PJS	0.960	0.960	0.960	0.960	0.962	0.962	0.962	0.962
Wang	**0.934**	**0.934**	**0.934**	**0.934**	**0.936**	0.937	**0.936**	**0.936**
BOP	0.934	0.938	0.934	0.935	0.936	0.939	0.936	0.937
CW	0.981	0.976	0.990	0.987	0.980	0.977	0.990	0.988

Notes. This table reports the average of the L_1-error ratios in (A1) across 1,000 simulations for different estimators of the mean vector and different parameter values. Panels A to D consider a sample size $T \in \{50, 100, 250, 500\}$ while $N_T/T \in \{2, 4\}$. The data generating process is described in Section 5.2.1. The data is simulated either from a multivariate normal distribution or from a multivariate t-distribution with $\nu = 3$ degrees of freedom. In both cases, the dependence is governed by a Toeplitz correlation matrix with $\rho \in \{0, 0.5\}$. The proposed estimators are St-MSh and D-MSh in Section 3, and O-LSh and T-LSh in Section 4. The PJS, Wang, BOP, and CW benchmark estimators are described in Section 5.1. The smallest average L_1-error ratios are in underlined bold, and the second-smallest in bold.

Table A.3: Average invariant-error ratio in simulated data for $T > N_T$

	Multivariate normal				Multivariate t with $\nu = 3$			
N_T/T	0.25	0.25	0.5	0.5	0.25	0.25	0.5	0.5
ρ	0	0.5	0	0.5	0	0.5	0	0.5
A) $T = 50$								
St-MSh	0.628	**0.652**	0.601	0.626	0.638	0.656	0.610	0.628
D-MSh	0.758	0.799	0.751	0.797	0.735	0.761	0.721	0.747
O-LSh	**0.562**	**0.653**	0.520	**0.604**	**0.572**	**0.653**	0.528	**0.602**
T-LSh	0.576	0.656	0.526	**0.613**	0.585	0.656	0.535	**0.610**
PJS	0.658	0.733	0.594	0.663	0.664	0.720	0.603	0.651
Wang	0.657	0.912	**0.507**	0.756	0.666	0.883	**0.516**	0.722
BOP	0.830	0.870	0.651	0.746	0.745	0.799	0.589	0.684
Jorion	**0.544**	0.655	**0.509**	0.618	**0.549**	**0.650**	**0.521**	0.616
B) $T = 100$								
St-MSh	0.740	0.759	0.729	0.748	0.748	0.754	0.734	0.740
D-MSh	0.866	0.909	0.857	0.908	0.841	0.865	0.829	0.854
O-LSh	**0.650**	**0.743**	**0.632**	**0.723**	**0.660**	**0.735**	**0.638**	**0.711**
T-LSh	0.655	0.753	0.634	**0.726**	0.665	0.743	0.639	**0.714**
PJS	0.746	0.786	0.726	0.761	0.754	0.772	0.731	0.743
Wang	0.664	0.877	**0.620**	0.820	0.676	0.845	**0.626**	0.785
BOP	0.696	0.774	0.658	0.754	0.683	0.753	0.645	0.730
Jorion	**0.644**	**0.743**	0.632	0.733	**0.657**	**0.739**	0.644	0.727
C) $T = 250$								
St-MSh	0.859	0.876	0.858	0.869	0.860	0.867	0.858	0.861
D-MSh	0.949	1.008	0.952	1.003	0.930	0.970	0.933	0.959
O-LSh	0.790	0.871	**0.782**	0.864	0.789	0.855	0.783	**0.849**
T-LSh	0.790	0.872	0.782	0.864	0.789	0.856	0.783	0.849
PJS	0.859	0.880	0.857	0.871	0.859	0.868	0.857	0.861
Wang	**0.788**	0.920	**0.777**	0.905	**0.787**	0.893	**0.777**	0.880
BOP	0.792	**0.865**	0.783	**0.859**	**0.789**	**0.854**	**0.783**	**0.848**
Jorion	**0.789**	**0.861**	0.783	**0.857**	0.792	**0.854**	0.788	0.850
D) $T = 500$								
St-MSh	0.926	0.932	0.922	0.930	0.926	0.927	0.923	0.924
D-MSh	1.002	1.053	0.999	1.051	0.988	1.025	0.988	1.022
O-LSh	**0.879**	0.937	**0.873**	0.934	**0.879**	0.926	**0.874**	0.923
T-LSh	0.879	0.937	0.873	0.934	**0.879**	0.927	0.874	0.923
PJS	0.926	0.933	0.922	0.931	0.926	0.927	0.922	0.924
Wang	0.879	0.956	**0.872**	0.951	0.879	0.942	**0.873**	0.936
BOP	0.880	**0.923**	0.874	**0.921**	0.879	**0.918**	0.874	**0.915**
Jorion	**0.879**	**0.922**	0.874	**0.921**	0.880	**0.918**	0.876	**0.915**

Notes. This table reports the average of the invariant-error ratios in (A2) across 1,000 simulations for different estimators of the mean vector and different parameter values. Panels A to D consider a sample size $T \in \{50, 100, 250, 500\}$ while $N_T/T \in \{0.25, 0.5\}$. The data generating process is described in Section 5.2.1. The data is simulated either from a multivariate normal distribution or from a multivariate t-distribution with $\nu = 3$ degrees of freedom. In both cases, the dependence is governed by a Toeplitz correlation matrix with $\rho \in \{0, 0.5\}$. The proposed estimators are St-MSh and D-MSh in Section 3, and O-LSh and T-LSh in Section 4. The PJS, Wang, BOP, and Jorion benchmark estimators are described in Section 5.1. The smallest average invariant-error ratios are in underlined bold, and the second-smallest in bold.

Table A.4: Average invariant-error ratio in simulated data for $N_T > T$

	Multivariate normal				Multivariate t with $\nu = 3$			
N_T/T	2	2	4	4	2	2	4	4
ρ	0	0.5	0	0.5	0	0.5	0	0.5
A) $T = 50$								
St-MSh	0.589	0.602	0.587	0.598	0.594	0.597	0.591	0.591
D-MSh	0.748	0.790	0.748	0.789	0.709	0.721	0.705	0.712
O-LSh	**0.493**	**0.560**	0.489	**0.554**	0.498	**0.552**	0.492	**0.544**
T-LSh	0.494	**0.562**	0.489	**0.554**	0.498	**0.553**	0.493	**0.544**
PJS	0.553	0.587	0.547	0.574	0.560	0.569	0.552	0.555
Wang	**0.424**	0.604	**0.413**	0.584	**0.431**	0.575	**0.418**	0.554
BOP	0.516	0.608	**0.479**	0.571	**0.488**	0.574	**0.459**	0.549
CW	0.796	0.832	0.893	0.925	0.753	0.780	0.846	0.874
B) $T = 100$								
St-MSh	0.717	0.731	0.718	0.730	0.720	0.721	0.719	0.721
D-MSh	0.853	0.901	0.854	0.901	0.816	0.835	0.814	0.833
O-LSh	0.614	**0.697**	0.612	**0.693**	0.617	**0.682**	0.614	**0.680**
T-LSh	0.614	**0.697**	0.612	**0.694**	0.617	**0.682**	0.614	**0.680**
PJS	0.704	0.728	0.703	0.724	0.707	0.710	0.705	0.708
Wang	**0.580**	0.752	**0.576**	0.744	**0.584**	0.718	**0.578**	0.714
BOP	**0.608**	0.719	**0.595**	0.710	**0.604**	0.695	**0.594**	0.688
CW	0.856	0.882	0.928	0.949	0.833	0.852	0.901	0.921
C) $T = 250$								
St-MSh	0.858	0.867	0.856	0.865	0.858	0.858	0.857	0.857
D-MSh	0.953	1.005	0.952	1.004	0.928	0.955	0.927	0.953
O-LSh	0.780	**0.858**	0.777	**0.856**	0.781	**0.842**	0.778	**0.840**
T-LSh	0.780	0.858	0.777	**0.856**	0.781	**0.842**	0.778	**0.840**
PJS	0.856	0.868	0.854	0.866	0.856	0.856	0.855	0.855
Wang	**0.773**	0.892	**0.770**	0.889	**0.774**	0.867	**0.771**	0.865
BOP	**0.777**	**0.857**	**0.773**	0.859	**0.778**	0.843	**0.774**	0.842
CW	0.929	0.940	0.964	0.975	0.920	0.927	0.954	0.965
D) $T = 500$								
St-MSh	0.922	**0.928**	0.922	**0.927**	0.922	0.922	0.922	0.922
D-MSh	0.998	1.050	0.998	1.051	0.983	1.017	0.982	1.015
O-LSh	0.872	0.931	0.872	0.930	0.872	**0.919**	0.871	**0.918**
T-LSh	0.872	0.931	0.872	0.930	0.872	0.919	0.871	0.918
PJS	0.922	0.928	0.922	0.928	0.922	0.922	0.921	0.922
Wang	**0.870**	0.946	**0.870**	0.945	**0.870**	0.931	**0.870**	0.930
BOP	**0.871**	**0.922**	**0.871**	**0.926**	**0.872**	**0.914**	**0.871**	**0.916**
CW	0.962	0.968	0.980	0.986	0.957	0.963	0.976	0.982

Notes. This table reports the average of the invariant-error ratios in (A2) across 1,000 simulations for different estimators of the mean vector and different parameter values. Panels A to D consider a sample size $T \in \{50, 100, 250, 500\}$ while $N_T/T \in \{2, 4\}$. The data generating process is described in Section 5.2.1. The data is simulated either from a multivariate normal distribution or from a multivariate t-distribution with $\nu = 3$ degrees of freedom. In both cases, the dependence is governed by a Toeplitz correlation matrix with $\rho \in \{0, 0.5\}$. The proposed estimators are St-MSh and D-MSh in Section 3, and O-LSh and T-LSh in Section 4. The PJS, Wang, BOP, and CW benchmark estimators are described in Section 5.1. The smallest average invariant-error ratios are in underlined bold, and the second-smallest in bold.

Table A.5: Average L_2-error ratio in simulated data for $T > N_T$ with heterogeneous variances

	Multivariate normal				Multivariate t with $\nu = 3$			
N_T/T	0.25	0.25	0.5	0.5	0.25	0.25	0.5	0.5
ρ	0	0.5	0	0.5	0	0.5	0	0.5
A) $T = 50$								
St-MSh	0.673	**0.682**	0.663	0.670	0.681	0.686	0.669	0.674
D-MSh	0.768	0.767	0.771	0.770	0.757	0.755	0.755	0.755
O-LSh	**0.649**	0.706	0.629	0.661	**0.655**	0.705	0.633	0.661
T-LSh	0.655	0.693	0.635	0.666	0.660	0.696	0.639	0.667
PJS	0.657	0.686	0.618	0.647	0.664	**0.683**	0.625	0.647
Wang	0.675	0.743	**0.580**	**0.642**	0.679	0.724	**0.587**	**0.635**
BOP	0.709	0.737	**0.616**	**0.645**	0.682	0.696	**0.606**	**0.637**
Jorion	**0.624**	**0.676**	0.620	0.667	**0.633**	**0.678**	0.633	0.677
B) $T = 100$								
St-MSh	0.749	0.756	0.743	0.746	0.753	0.757	0.747	0.750
D-MSh	0.827	0.828	0.826	0.828	0.815	0.814	0.814	0.816
O-LSh	0.697	0.728	0.684	**0.702**	0.701	0.726	0.688	**0.703**
T-LSh	0.703	0.741	0.686	0.710	0.706	0.739	0.690	0.710
PJS	0.740	0.759	0.721	0.732	0.745	0.756	0.725	0.734
Wang	**0.679**	**0.726**	**0.639**	**0.670**	**0.685**	**0.717**	**0.644**	**0.670**
BOP	**0.676**	**0.720**	**0.658**	0.702	**0.682**	**0.722**	**0.666**	0.715
Jorion	0.705	0.751	0.707	0.749	0.716	0.757	0.719	0.761
C) $T = 250$								
St-MSh	0.849	0.854	0.848	0.849	0.852	0.854	0.849	0.849
D-MSh	0.903	0.907	0.906	0.905	0.894	0.895	0.894	0.892
O-LSh	**0.789**	**0.803**	**0.785**	**0.791**	**0.792**	**0.802**	**0.787**	**0.790**
T-LSh	0.789	0.806	0.785	0.791	0.792	0.805	0.787	0.791
PJS	0.846	0.854	0.841	0.844	0.849	0.852	0.843	0.844
Wang	**0.776**	**0.797**	**0.766**	**0.774**	**0.781**	**0.793**	**0.769**	**0.774**
BOP	0.795	0.841	0.794	0.842	0.805	0.844	0.805	0.847
Jorion	0.818	0.858	0.821	0.860	0.827	0.860	0.830	0.864
D) $T = 500$								
St-MSh	0.912	0.912	0.908	0.910	0.913	0.912	0.909	0.910
D-MSh	0.954	0.954	0.951	0.952	0.946	0.945	0.942	0.943
O-LSh	**0.864**	**0.866**	**0.859**	**0.862**	**0.865**	**0.867**	**0.860**	**0.862**
T-LSh	0.864	0.867	0.859	0.862	0.866	0.867	0.860	0.863
PJS	0.911	0.911	0.906	0.909	0.913	0.911	0.907	0.909
Wang	**0.860**	**0.862**	**0.852**	**0.856**	**0.861**	**0.863**	**0.853**	**0.856**
BOP	0.881	0.912	0.879	0.913	0.887	0.912	0.884	0.912
Jorion	0.891	0.918	0.889	0.919	0.896	0.918	0.894	0.918

Notes. This table reports the average of the L_2-error ratios in (37) across 1,000 simulations for different estimators of the mean vector and different parameter values. Panels A to D consider a sample size $T \in \{50, 100, 250, 500\}$ while $N_T/T \in \{0.25, 0.5\}$. The data generating process is described in Section 5.2.1, with the difference that the variances of the variables are randomly drawn from a uniform distribution in the interval $[0.5, 5]$. The data is simulated either from a multivariate normal distribution or from a multivariate t-distribution with $\nu = 3$ degrees of freedom. In both cases, the dependence is governed by a Toeplitz correlation matrix with $\rho \in \{0, 0.5\}$. The proposed estimators are St-MSh and D-MSh in Section 3, and O-LSh and T-LSh in Section 4. The PJS, Wang, BOP, and Jorion benchmark estimators are described in Section 5.1. The smallest average invariant-error ratios are in underlined bold, and the second-smallest in bold.

Table A.6: Average L_2-error ratio in simulated data for $N_T > T$ with heterogeneous variances

	Multivariate normal				Multivariate t with $\nu = 3$			
N_T/T	2	2	4	4	2	2	4	4
ρ	0	0.5	0	0.5	0	0.5	0	0.5
A) $T = 50$								
St-MSh	0.657	0.657	0.655	0.656	0.660	0.661	0.658	0.657
D-MSh	0.776	0.775	0.777	0.775	0.756	0.756	0.755	0.752
O-LSh	0.609	0.619	0.605	0.611	0.611	0.621	0.607	0.611
T-LSh	0.610	0.624	0.605	0.613	0.613	0.625	0.607	0.613
PJS	0.580	0.586	0.572	0.575	0.586	0.592	0.576	**0.576**
Wang	**0.489**	**0.513**	**0.474**	**0.486**	**0.495**	**0.517**	**0.478**	**0.486**
BOP	**0.550**	**0.573**	**0.528**	**0.570**	**0.555**	**0.583**	**0.545**	0.595
CW	0.797	0.776	0.893	0.879	0.778	0.770	0.868	0.863
B) $T = 100$								
St-MSh	0.736	0.737	0.735	0.736	0.738	0.739	0.736	0.737
D-MSh	0.827	0.828	0.828	0.827	0.811	0.810	0.810	0.809
O-LSh	0.672	0.676	0.670	0.673	0.674	0.677	0.671	0.674
T-LSh	0.672	0.677	0.670	0.673	0.674	0.678	0.671	0.674
PJS	0.697	0.702	0.694	0.698	0.700	0.703	0.696	0.699
Wang	**0.598**	**0.607**	**0.591**	**0.599**	**0.601**	**0.608**	**0.593**	**0.600**
BOP	**0.624**	**0.660**	**0.615**	**0.658**	**0.635**	**0.669**	**0.630**	**0.670**
CW	0.837	0.825	0.914	0.904	0.828	0.824	0.898	0.895
C) $T = 250$								
St-MSh	0.845	0.845	0.845	0.845	0.846	0.846	0.845	0.845
D-MSh	0.905	0.904	0.905	0.905	0.891	0.891	0.890	0.890
O-LSh	0.780	**0.781**	0.779	**0.780**	0.781	**0.782**	0.780	**0.780**
T-LSh	0.780	0.781	0.779	0.780	0.781	0.782	0.780	0.780
PJS	0.836	0.837	0.835	0.835	0.837	0.837	0.836	0.835
Wang	**0.756**	**0.758**	**0.754**	**0.755**	**0.757**	**0.758**	**0.755**	**0.756**
BOP	**0.767**	0.794	**0.762**	0.784	**0.773**	0.796	**0.768**	0.787
CW	0.904	0.898	0.948	0.943	0.901	0.898	0.940	0.940
D) $T = 500$								
St-MSh	0.909	0.909	0.908	0.908	0.909	0.909	0.909	0.909
D-MSh	0.952	0.953	0.952	0.952	0.942	0.942	0.942	0.941
O-LSh	0.858	**0.859**	0.858	**0.858**	**0.859**	**0.859**	0.858	**0.858**
T-LSh	0.858	0.859	0.858	0.858	0.859	0.859	0.858	0.858
PJS	0.907	0.907	0.906	0.906	0.907	0.907	0.906	0.906
Wang	**0.851**	**0.852**	**0.851**	**0.851**	**0.851**	**0.852**	**0.851**	**0.851**
BOP	**0.857**	0.875	**0.854**	0.865	0.860	0.875	**0.856**	0.865
CW	0.944	0.940	0.969	0.965	0.942	0.941	0.965	0.964

Notes. This table reports the average of the L_2-error ratios in (37) across 1,000 simulations for different estimators of the mean vector and different parameter values. Panels A to D consider a sample size $T \in \{50, 100, 250, 500\}$ while $N_T/T \in \{2, 4\}$. The data generating process is described in Section 5.2.1, with the difference that the variances of the variables are randomly drawn from a uniform distribution in the interval $[0.5, 5]$. The data is simulated either from a multivariate normal distribution or from a multivariate t-distribution with $\nu = 3$ degrees of freedom. In both cases, the dependence is governed by a Toeplitz correlation matrix with $\rho \in \{0, 0.5\}$. The proposed estimators are St-MSh and D-MSh in Section 3, and O-LSh and T-LSh in Section 4. The PJS, Wang, BOP, and CW benchmark estimators are described in Section 5.1. The smallest average invariant-error ratios are in underlined bold, and the second-smallest in bold.

Table A.7: Average L_2-error ratio in simulated VAR(1) data

	$T > N_T$				$N_T > T$			
N_T/T	0.25	0.25	0.5	0.5	2	2	4	4
ρ	0	0.5	0	0.5	0	0.5	0	0.5
A) $T = 50$								
St-MSh	0.840	0.837	0.837	0.837	0.843	0.843	0.838	0.839
D-MSh	0.896	0.892	0.897	0.895	0.902	0.902	0.900	0.900
O-LSh	0.826	0.847	0.817	0.832	0.818	0.824	0.812	0.815
T-LSh	0.828	0.843	0.819	0.834	0.819	0.826	0.812	0.816
PJS	0.833	**0.832**	0.818	**0.820**	**0.814**	**0.815**	**0.807**	**0.808**
Wang	**0.790**	**0.805**	**0.772**	**0.788**	**0.770**	**0.777**	**0.762**	**0.765**
BOP	0.818	0.846	**0.809**	0.843	0.815	0.845	0.814	0.850
Jorion	**0.815**	0.849	0.820	0.852	/	/	/	/
CW	/	/	/	/	0.936	0.930	0.966	0.963
B) $T = 100$								
St-MSh	0.877	0.880	0.877	0.879	0.876	0.877	0.875	0.875
D-MSh	0.924	0.925	0.926	0.926	0.926	0.927	0.927	0.926
O-LSh	0.850	**0.866**	**0.849**	**0.858**	**0.845**	**0.848**	**0.843**	**0.845**
T-LSh	0.851	0.872	0.850	0.862	0.845	0.848	0.843	0.845
PJS	0.869	0.873	0.864	0.867	0.860	0.861	0.858	0.858
Wang	**0.822**	**0.838**	**0.820**	**0.829**	**0.815**	**0.818**	**0.813**	**0.814**
BOP	**0.842**	0.876	0.852	0.878	0.858	0.883	0.859	0.886
Jorion	0.853	0.884	0.867	0.889	/	/	/	/
CW	/	/	/	/	0.951	0.947	0.976	0.973
C) $T = 250$								
St-MSh	0.927	0.930	0.926	0.928	0.926	0.926	0.927	0.927
D-MSh	0.959	0.961	0.959	0.961	0.960	0.959	0.960	0.961
O-LSh	**0.901**	**0.907**	**0.899**	**0.902**	**0.898**	**0.898**	**0.898**	**0.898**
T-LSh	0.901	0.908	0.899	0.903	0.898	0.898	0.898	0.898
PJS	0.924	0.927	0.921	0.924	0.921	0.921	0.921	0.921
Wang	**0.890**	**0.895**	**0.888**	**0.890**	**0.886**	**0.887**	**0.887**	**0.887**
BOP	0.907	0.929	0.915	0.934	0.920	0.935	0.921	0.934
Jorion	0.913	0.932	0.920	0.938	/	/	/	/
CW	/	/	/	/	0.974	0.971	0.987	0.986
D) $T = 500$								
St-MSh	0.955	0.955	0.954	0.954	0.954	0.954	0.954	0.954
D-MSh	0.977	0.977	0.978	0.977	0.978	0.977	0.978	0.977
O-LSh	**0.933**	**0.935**	**0.932**	**0.934**	**0.932**	**0.932**	**0.932**	**0.932**
T-LSh	0.933	0.935	0.932	0.934	0.932	0.932	0.932	0.932
PJS	0.953	0.954	0.952	0.953	0.952	0.952	0.952	0.952
Wang	**0.929**	**0.929**	**0.927**	**0.929**	**0.928**	**0.928**	**0.927**	**0.927**
BOP	0.942	0.957	0.946	0.960	0.952	0.961	0.952	0.959
Jorion	0.945	0.958	0.949	0.962	/	/	/	/
CW	/	/	/	/	0.985	0.983	0.992	0.992

Notes. This table reports the average of the L_2-error ratios in (37) across 1,000 simulations for different estimators of the mean vector and different parameter values. Panels A to D consider a sample size $T \in \{50, 100, 250, 500\}$ while $N_T/T \in \{0.25, 0.5, 2, 4\}$. The data generating process is described in Section 5.2.1, with the difference that we use non-i.i.d. data generated from a VAR(1) model; see Section C for details. The variables are unconditionally multivariate normal in the VAR(1) model. The dependence is governed by a Toeplitz correlation matrix with $\rho \in \{0, 0.5\}$. The proposed estimators are St-MSh and D-MSh in Section 3, and O-LSh and T-LSh in Section 4. The PJS, Wang, BOP, Jorion, and CW benchmark estimators are described in Section 5.1. The smallest average L_2-error ratios are in underlined bold, and the second-smallest in bold.

Table A.8: Out-of-sample Sharpe ratio statistics in empirical weekly stock-return data

	Full period		Financial crisis		Pandemic	
	Mean	Std dev.	Mean	Std dev.	Mean	Std dev.
A) Mean-variance portfolio with a risk-free asset investment						
St-MSh	0.042	0.914	−0.693	**0.610**	0.319	**0.626**
D-MSh	**<u>0.234</u>**	0.952	−0.731	0.717	0.395	0.679
O-LSh	**0.099**	**<u>0.874</u>**	−0.592	0.616	**0.425**	0.717
T-LSh	0.061	0.920	−0.723	0.629	0.420	0.674
Sample	0.042	0.914	−0.693	0.610	0.319	0.626
Wang	0.079	**0.871**	**−0.435**	0.628	**<u>0.586</u>**	0.750
BOP	−0.045	0.966	**<u>0.934</u>**	0.722	−0.187	**<u>0.564</u>**
Jorion	0.047	0.878	−0.544	**<u>0.559</u>**	0.308	0.719
B) Tangency portfolio without a risk-free asset investment						
St-MSh	0.071	0.912	−0.693	**0.610**	0.292	**0.640**
D-MSh	**<u>0.324</u>**	0.925	−0.741	0.707	0.158	0.772
O-LSh	**0.099**	**0.874**	−0.592	0.616	**0.425**	0.717
T-LSh	0.068	0.919	−0.723	0.629	0.420	0.674
Sample	0.071	0.912	−0.693	0.610	0.292	0.640
Wang	0.079	**<u>0.871</u>**	**−0.435**	0.628	**<u>0.586</u>**	0.750
BOP	−0.068	0.964	**<u>0.934</u>**	0.722	−0.147	**<u>0.576</u>**
Jorion	0.082	0.876	−0.544	**<u>0.559</u>**	0.285	0.729

Notes. This table reports the average and standard deviation of the annualized out-of-sample Sharpe ratios across all testing windows for different estimators, mean-variance portfolios, and time periods. We use a dataset of weekly excess returns on $N_T = 441$ S&P500 individual stocks spanning the period from January 3, 2000 to December 29, 2023; see Section 5.3.1 for details. The proposed estimators are St-MSh and D-MSh in Section 3, and O-LSh and T-LSh in Section 4. The Sample, Wang, BOP, and Jorion benchmark estimators are described in Section 5.1. We do not include the PJS benchmark estimator because there are many training windows where it yields a zero mean-return vector, making the Sharpe ratio undefined. We adopt a rolling-window framework with a training-window length of ten years (520 trading weeks) in which we estimate the vector of expected returns and the resulting mean-variance portfolio in (38), followed by a one-year testing period in which we evaluate the Sharpe ratio. We then roll the training and testing windows by one month and proceed again. Finally, we report the average and standard deviation of the out-of-sample Sharpe ratios, for the full out-of-sample period but also during and after the 2008 financial crisis (October 6, 2008 to August 3, 2011) and covid-19 pandemic (February 26, 2020 to December 29, 2023), which we identify with the Bai & Perron (2003) test; see Figure 1. In total, we have 170, 21, and 37 testing windows for the full period, the financial crisis, and the pandemic period, respectively. In Panel A, we set a constant risk-aversion coefficient γ in the mean-variance portfolio (38), whose value does not affect the Sharpe ratio. In Panel B, we set $\gamma = \mathbf{1}_{N_T}' \mathbf{S}_{N_T}^{-1} \hat{\boldsymbol{\mu}}_{N_T}$, in which case we recover the tangency portfolio that does not invest in the risk-free asset. The highest average Sharpe ratios and lowest standard deviation of Sharpe ratios are in underlined bold, and the second-best in bold.

E Useful Lemmas

In this section, we provide lemmas that are useful for proving our theorems in Section F. We first present the lemma statements and then provide the proofs.

E.1 Lemma statements

Throughout, given θ_T and $\hat{\theta}_T$ a sequence of scalars and random variables indexed by $T \in \mathbb{N}^\star$, we denote $\hat{\theta}_T - \theta_T \xrightarrow{L_2} 0$ as $T \to \infty$ when $\mathbb{E}[\hat{\theta}_T] - \theta_T \to 0$ and $\mathbb{V}[\hat{\theta}_T] \to 0$ as $T \to \infty$.

Lemma 1 concerns the variance of the entries of the sample covariance matrix $\mathbf{S}_{N_T}$.

Lemma 1 *The variances of the entries of the sample covariance matrix $\mathbf{S}_{N_T}$ are given by*

$$\mathbb{V}\big[S_{kl,N_T}\big] = \frac{1}{T}\bigg(\mu^{(4)}_{kl,N_T} - \frac{T-2}{T-1}\Sigma^2_{kl,N_T} + \frac{1}{T-1}\Sigma_{kk,N_T}\Sigma_{ll,N_T}\bigg), \tag{A8}$$

for all $k, l = 1, \ldots, N_T$ and $T \geq 2$, where $\mu^{(4)}_{kl,N_T}$ is defined in Assumption 2.

Lemma 2 is inspired by Ledoit & Wolf (2004, Lemma A.1).

Lemma 2 *Let Δ_T be a sequence of non-negative random variables satisfying*

$$\mathbb{E}[\Delta_T] \to 0 \quad \textit{as} \quad T \to \infty. \tag{A9}$$

Further, let a_T and $\hat{a}_T$ be a sequence of scalars and random variables satisfying

$$\hat{a}_T - a_T \xrightarrow{L_2} 0 \ \textit{as} \ \ T \to \infty \quad \textit{and} \quad 0 < a_T \leq \delta \ \ \textit{for all} \ \ T \geq 1, \tag{A10}$$

for a given scalar $\delta > 0$ independent of T. Then, if there exist scalars $\tau_1, \tau_2, M_1, M_2 \geq 0$ such that

$$\frac{\Delta_T}{\hat{a}_T^{\tau_1} a_T^{\tau_2}} \leq M_1 \hat{a}_T + M_2 a_T \ \ \textit{almost surely for all} \ T \geq 1, \tag{A11}$$

it follows that

$$\mathbb{E}\left[\frac{\Delta_T}{\hat{a}_T^{\tau_1} a_T^{\tau_2}}\right] \to 0 \quad as \;\; T \to \infty. \tag{A12}$$

Lemma 3 is concerned with quantities appearing in the St-MSh oracle shrinkage intensity, $c_T^\star$ in (8), and its bona-fide estimator, $\hat{c}_T^\star$ in (13).

Lemma 3 *Let Assumptions 1 and 2 hold. Define*

$$a_{1T} := \left|\left|\boldsymbol{\mu}_{N_T}\right|\right|_{N_T}^2, \quad a_{2T} := a_{1T} + \frac{1}{TN_T}\mathrm{tr}\big(\boldsymbol{\Sigma}_{N_T}\big), \tag{A13}$$

$$\hat{a}_{1T} := \left|\left|\overline{\mathbf{X}}^{(T)}\right|\right|_{N_T}^2, \quad \hat{a}_{2T} := \hat{a}_{1T} + \frac{1}{TN_T}\mathrm{tr}(\mathbf{S}_{N_T}). \tag{A14}$$

i) There exists a constant M independent of T such that $\max\big(a_{1T}, a_{2T}\big) \leq M$ for any (T, N_T).

ii) $\hat{a}_{sT} - a_{sT} \xrightarrow{L_2} 0$ as $T \to \infty$ for any $s \in \{1, 2\}$.

iii) $0 \leq \hat{a}_{1T} \leq \hat{a}_{2T}$ almost surely and $0 \leq a_{1T} \leq a_{2T}$ for any (T, N_T).

Lemma 4 is concerned with quantities appearing in the D-MSh oracle shrinkage intensities, $\mathbf{c}_T^\star$ in (10), and their bona-fide estimators, $\hat{\mathbf{c}}_T^\star$ in (14).

Lemma 4 *Let Assumptions 1 and 2 hold. Define*

$$a_{k1,T} := \mu_{k,N_T}^2, \quad a_{k2,T} := a_{k1,T} + \frac{1}{T}\Sigma_{kk,N_T}, \tag{A15}$$

$$\hat{a}_{k1,T} := \big(\overline{X}_k^{(T)}\big)^2, \quad \hat{a}_{k2,T} := \hat{a}_{k1,T} + \frac{1}{T}S_{kk,N_T}. \tag{A16}$$

i) There exists a constant M independent of T and k such that $\max\big(a_{k1,T}, a_{k2,T}\big) \leq M$ for any (T, N_T), $k = 1, \ldots, N_T$, and $s \in \{1, 2\}$.

ii) $\hat{a}_{ks,T} - a_{ks,T} \xrightarrow{L_2} 0$ as $T \to \infty$ for any $k = 1, \ldots, N_T$ and $s \in \{1, 2\}$.

iii) $0 \leq \hat{a}_{k1,T} \leq \hat{a}_{k2,T}$ almost surely and $0 \leq a_{k1,T} \leq a_{k2,T}$ for any (T, N_T) and $k = 1, \ldots, N_T$.

Lemma 5 is concerned with quantities appearing in the O-LSh and T-LSh oracle shrinkage parameters, $\xi_T^\star$ and $\delta_T^\star$ in (27), and their bona-fide estimators, $\hat{\xi}_T^\star$ and $\hat{\delta}_T^\star$ in (30).

Lemma 5 *Let Assumptions 1 and 2 hold. Define*

$$\tilde{a}_{1T} := d_{1T} - d_{2T} = a_{2T} - a_{1T} - d_{2T}, \quad \tilde{a}_{2T} := \tilde{a}_{1T} + d_{3T}, \tag{A17}$$

$$\hat{\tilde{a}}_{1T} := \hat{a}_{2T} - \hat{a}_{1T} - \hat{d}_{2T}, \quad \hat{\tilde{a}}_{2T} := \hat{\tilde{a}}_{1T} + \hat{d}_{3T}, \tag{A18}$$

where a_{1T} and a_{2T} are defined in (A13)*, d_{1T}, d_{2T}, and d_{3T} in (21)–(22)–(23), and $\hat{d}_{1T}$, $\hat{d}_{2T}$, and $\hat{d}_{3T}$ are obtained by plugging $(\overline{\mathbf{X}}^{(T)}, \mathbf{S}_{N_T})$ in place of $(\boldsymbol{\mu}_{N_T}, \boldsymbol{\Sigma}_{N_T})$ into (21)–(22)–(23).*

i) There exists a constant M independent of T such that $\max\left(\tilde{a}_{1T}, \tilde{a}_{2T}\right) \leq M$ for any (T, N_T).

ii) $\hat{a}_{sT} - a_{sT} \xrightarrow{L_2} 0$ as $T \to \infty$ for any $s \in \{1, 2\}$.

iii) $0 \leq \hat{\tilde{a}}_{1T} \leq \hat{\tilde{a}}_{2T}$ almost surely and $0 \leq \tilde{a}_{1T} \leq \tilde{a}_{2T}$ for any (T, N_T).

E.2 Proof of Lemma 1

Equation (A8) appears in Severini (2005) when $k = l$, and thus, we now assume that $k \neq l$. Without loss of generality, we also assume that $\mu_{k,N_T} = 0$ for all $k = 1, \ldots, N_T$ because we deal with centered moments. First, we have that

$$\begin{aligned}
\mathbb{V}\big[S_{kl,N_T}\big] &= \mathbb{V}\left[\frac{1}{T}\left(\sum_{i=1}^{T} X_{ik}X_{il} - \frac{1}{T-1}\sum_{i\neq j}^{T} X_{ik}X_{jl}\right)\right] \\
&= \frac{1}{T^2}\left(\sum_{i=1}^{T}\mathbb{V}\big[X_{ik}X_{il}\big] + \frac{1}{(T-1)^2}\sum_{i\neq j}^{T}\mathbb{V}\big[X_{ik}X_{jl}\big] + \sum_{i\neq j}^{T}\mathrm{Cov}\big(X_{ik}X_{il}, X_{jk}X_{jl}\big)\right. \\
&\quad - \frac{1}{T-1}\sum_{i=1}^{T}\sum_{i_1\neq j_1}^{T}\mathrm{Cov}\big(X_{ik}X_{il}, X_{i_1k}X_{j_1l}\big) \\
&\quad \left. + \frac{1}{(T-1)^2}\sum_{\substack{i_1\neq j_1 \\ i_2\neq j_2}}^{T}\mathrm{Cov}\big(X_{i_1k}X_{j_1l}, X_{i_2k}X_{j_2l}\big)\right).
\end{aligned} \tag{A19}$$

It holds that

$$\mathbb{V}\big[X_{ik}X_{il}\big] = \mu^{(4)}_{kl,N_T} - \Sigma^2_{kl,N_T} \quad \text{for all} \;\; i = 1, \ldots, T \tag{A20}$$

and

$$\mathbb{V}\big[X_{ik}X_{jl}\big] = \Sigma_{kk,N_T}\Sigma_{ll,N_T} \quad \text{for all} \;\; i, j = 1, \ldots, T, \;\; i \neq j. \tag{A21}$$

Further, the i.i.d. assumption and the null-population-means assumption imply that

$$\mathrm{Cov}\big(X_{ik}X_{il}, X_{jk}X_{jl}\big) = 0 \quad \text{for all} \;\; i, j = 1, \ldots, T, \;\; i \neq j \tag{A22}$$

and

$$\mathrm{Cov}\big(X_{ik}X_{il}, X_{i_1k}X_{j_1l}\big) = 0 \quad \text{for all} \;\; i, i_1, j_1 = 1, \ldots, T, \;\; i_1 \neq j_1. \tag{A23}$$

Finally, the last set of covariances in (A19) are non-zero if and only if $(i_1, j_1) = (j_2, i_2)$, and we have that

$$\mathrm{Cov}\big(X_{i_1k}X_{j_1l}, X_{i_2k}X_{j_2l}\big) = \Sigma^2_{kl,N_T} \quad \text{for all} \;\; i_1, j_1 = 1, \ldots, T, \;\; i_1 \neq j_1, \;\; (i_1, j_1) = (j_2, i_2). \tag{A24}$$

Finally, we obtain (A8) by combining (A19)–(A24). This completes the proof.

E.3 Proof of Lemma 2

Let $\epsilon > 0$ be an arbitrary scalar and denote by $\mathcal{T}_1$ and $\mathcal{T}_2$ two disjoint index sets such that $\mathcal{T}_1 \cup \mathcal{T}_2 = \mathbb{N}^\star$ and $\mathcal{T}_1 := \{T \geq 1 : a_T \leq \epsilon\}$.

We first deal with indices $T \in \mathcal{T}_1$. (A10) implies that there exists $T_1 \geq 1$ such that

$$\begin{aligned}
\mathbb{E}\left[\frac{\Delta_T}{\hat{a}_T^{\tau_1} a_T^{\tau_2}}\right] &\leq M_1\mathbb{E}\big[\hat{a}_T\big] + M_2 a_T \\
&\leq M_1\mathbb{E}\big[\big|\hat{a}_T - a_T\big|\big] + \big(M_2 + 1\big)a_T \\
&\leq M_1\sqrt{\mathbb{E}\big[(\hat{a}_T - a_T)^2\big]} + \big(M_2 + 1\big)a_T,
\end{aligned}$$

$$\leq \quad \big(M_1 + M_2 + 1\big)\epsilon \quad \text{for all } T \geq T_1 \text{ such that } T \in \mathcal{T}_1, \tag{A25}$$

where the first and third inequality are due to (A11) and Lyapunov's inequality, respectively.

We focus next on indices $T \in \mathcal{T}_2$. Note that (A10) implies $\hat{a}_T - a_T \xrightarrow{p} 0$, which in turn means that there exists $T_2 \geq 1$ such that

$$\mathbb{P}\big(\big|\widehat{a}_T - a_T\big| \geq \epsilon/2\big) \leq \frac{\epsilon/2}{M_1\epsilon/2 + M_2\delta} \quad \text{for all } T \geq T_2 \text{ such that } T \in \mathcal{T}_2, \tag{A26}$$

with the caveat that the case $M_1 = M_2 = 0$ trivially implies (A12), which is excluded in (A26). Now, (A9) implies that there exists $T_3 \geq 1$ such that

$$\mathbb{E}[\Delta_T] \leq (\epsilon/2)^{\tau_1+\tau_2+1} \quad \text{for all } T \geq T_3 \text{ such that } T \in \mathcal{T}_2. \tag{A27}$$

Clearly, (A10), (A11) and (A26) yield

$$\begin{aligned}
\mathbb{E}\left[\frac{\Delta_T}{\hat{a}_T^{\tau_1} a_T^{\tau_2}} \mathbb{1}_{\{\hat{a}_T \leq \epsilon/2\}}\right] &\leq \mathbb{E}\big[\big(M_1\hat{a}_T + M_2 a_T\big)\mathbb{1}_{\{\hat{a}_T \leq \epsilon/2\}}\big] \\
&\leq (M_1\epsilon/2 + M_2\delta)\mathbb{P}\big(\hat{a}_T \leq \epsilon/2\big) \\
&\leq (M_1\epsilon/2 + M_2\delta)\mathbb{P}\big(\big|\widehat{a}_T - a_T\big| \geq \epsilon/2\big) \\
&\leq \epsilon/2 \quad \text{for all } T \geq T_2 \text{ such that } T \in \mathcal{T}_2,
\end{aligned} \tag{A28}$$

where $\mathbb{1}_A$ is the indicator function that is equal to 1 or 0 if condition A is true or false, respectively. Keeping in mind that $\tau_1, \tau_2 \geq 0$, the complement case of (A28) implies that

$$\begin{aligned}
\mathbb{E}\left[\frac{\Delta_T}{\hat{a}_T^{\tau_1} a_T^{\tau_2}} \mathbb{1}_{\{\hat{a}_T > \epsilon/2\}}\right] &\leq \mathbb{E}\big[\Delta_T \mathbb{1}_{\{\hat{a}_T > \epsilon/2\}}\big](\epsilon/2)^{-\tau_1}\epsilon^{-\tau_2} \\
&\leq \mathbb{E}[\Delta_T](\epsilon/2)^{-\tau_1}\epsilon^{-\tau_2} \\
&\leq \epsilon/2 \quad \text{for all } T \geq T_3 \text{ such that } T \in \mathcal{T}_2,
\end{aligned} \tag{A29}$$

where the last step is due to (A27). Combining (A28) and (A29) gives

$$\mathbb{E}\left[\frac{\Delta_T}{\hat{a}_T^{\tau_1} a_T^{\tau_2}}\right] \leq \epsilon \quad \text{for all } T \geq \max(T_2, T_3) \text{ such that } T \in \mathcal{T}_2. \tag{A30}$$

One can set ϵ small enough in (A25) and (A30) to conclude (A12). This completes the proof.

E.4 Proof of Lemma 3

Part i). The boundedness property of a_{1T} and a_{2T} are straightforward implications of Assumption 2.

Part ii). The estimator $\hat{a}_{1T}$ is asymptotically unbiased because

$$\begin{aligned}
\mathbb{E}\big[\hat{a}_{1T}\big] - a_{1T} &= \frac{1}{N_T}\sum_{k=1}^{N_T}\mathbb{E}\Big[\big(\overline{X}_k^{(T)}\big)^2\Big] - a_{1T} \\
&= \frac{1}{N_T}\sum_{k=1}^{N_T}\Big(\mathbb{V}\big[\overline{X}_k^{(T)}\big] + \Big(\mathbb{E}\big[\overline{X}_k^{(T)}\big]\Big)^2\Big) - a_{1T} \\
&= \frac{1}{N_T}\sum_{k=1}^{N_T}\Big(\frac{1}{T}\Sigma_{kk,N_T} + \mu_{k,N_T}^2\Big) - a_{1T} \\
&= \frac{1}{TN_T}\mathrm{tr}\big(\mathbf{\Sigma}_{N_T}\big), \\
&\to \ 0 \quad \text{as} \quad T \to \infty
\end{aligned} \tag{A31}$$

where the last step is due to Assumption 2. Moreover, the variance of $\hat{a}_{1T}$ asymptotically vanishes because

$$\begin{aligned}
\mathbb{V}\big[\hat{a}_{1T}\big] &= \frac{1}{N_T^2}\mathbb{V}\left[\sum_{k=1}^{N_T}\Big(\overline{X}_k^{(T)}\Big)^2\right] \\
&\leq \frac{1}{N_T}\sum_{k=1}^{N_T}\mathbb{V}\left[\Big(\overline{X}_k^{(T)}\Big)^2\right] \\
&= \frac{4}{TN_T}\sum_{k=1}^{N_T}\Sigma_{kk,N_T}\mu_{k,N_T}^2 + \frac{2}{T^2N_T}\sum_{k=1}^{N_T}\Big(\Sigma_{kk,N_T}^2 + 2\mu_{k,N_T}^{(3)}\mu_{k,N_T}\Big) \\
&\quad + \frac{1}{T^3N_T}\sum_{k=1}^{N_T}\Big(\mu_{kk,N_T}^{(4)} - 3\Sigma_{kk,N_T}^2\Big) \\
&\to \ 0 \quad \text{as} \quad T \to \infty,
\end{aligned} \tag{A32}$$

where $\mu_{k,N_T}^{(3)} := \mathbb{E}\big[(X_{1k} - \mu_{k,N_T})^3\big]$, the formula for the variance of the squared sample mean of i.i.d. random variables used in the third step can be found in Severini (2005), and the last step is due to Assumption 2. Combining (A31) and (A32) yields $\hat{a}_{1T} - a_{1T} \xrightarrow{L_2} 0$ as $T \to \infty$.

The proof for $\hat{a}_{2T}$ is similar. Indeed, we have

$$\mathbb{E}\big[\hat{a}_{2T}\big] - a_{2T} = \frac{1}{TN_T}\mathrm{tr}\big(\mathbf{\Sigma}_{N_T}\big) \to 0 \quad \text{as} \;\; T \to \infty, \tag{A33}$$

and

$$\begin{aligned}
\mathbb{V}\big[\hat{a}_{2T}\big] &= \mathbb{V}\bigg[\hat{a}_{1T} + \frac{1}{TN_T}\mathrm{tr}\big(\mathbf{S}_{N_T}\big)\bigg] \\
&\leq 2\mathbb{V}\big[\hat{a}_{1T}\big] + \frac{2}{T^2N_T^2}\mathbb{V}\big[\mathrm{tr}\big(\mathbf{S}_{N_T}\big)\big] \\
&\leq 2\mathbb{V}\big[\hat{a}_{1T}\big] + \frac{2}{T^2N_T}\sum_{k=1}^{N_T}\mathbb{V}\big[S_{kk,N_T}\big] \\
&= \mathbb{V}\big[\hat{a}_{1T}\big] + \frac{2}{T^3N_T}\sum_{k=1}^{N_T}\bigg(\mu^{(4)}_{kk,N_T} - \frac{T-3}{T-1}\Sigma^2_{kk,N_T}\bigg) \\
&\to 0 \quad \text{as} \;\; T \to \infty
\end{aligned} \tag{A34}$$

where Lemma 1 is used in the third step, and the last step is due to Assumption 2. Combining (A33) and (A34) yields $\hat{a}_{2T} - a_{2T} \xrightarrow{L_2} 0$ as $T \to \infty$.

Part iii). The proof follows from (A13)–(A14). This completes the proof.

E.5 Proof of Lemma 4

Part i). The proof is similar to that of Lemma 3i.

Part ii). The proof is similar to that of Lemma 3ii. We just detail a few steps related to the vanishing asymptotic variances. Using similar derivations to those in (A32), we find that

$$\begin{aligned}
\mathbb{V}\big[\hat{a}_{k1,T}\big] &= \mathbb{V}\Big[\big(\overline{X}_k^{(T)}\big)^2\Big] \\
&= \frac{4}{T}\mu^2_{k,N_T}\Sigma_{kk,N_T} + \frac{2}{T^2}\Big(\Sigma^2_{kk,N_T} + 2\mu^{(3)}_{k,N_T}\mu_{k,N_T}\Big) + \frac{1}{T^3}\Big(\mu^{(4)}_{kk,N_T} - 3\Sigma^2_{kk,N_T}\Big) \\
&\leq \frac{M_1}{T} = o(1) \quad \text{as} \;\; T \to \infty,
\end{aligned} \tag{A35}$$

where the constant $M_1 > 0$ is independent of (T, N_T, k), and the last step is due to Assumption 2. Further, as in (A34), we can obtain $\mathbb{V}\big[\hat{a}_{k2,T}\big] \to 0$ as $T \to \infty$; we skip the details for

conciseness. Therefore, $\hat{a}_{ks,T} - a_{ks,T} \xrightarrow{L_2} 0$ as $T \to \infty$ for any $s \in \{1, 2\}$.

Part iii). The proof follows from (A15)–(A16). This completes the proof.

E.6 Proof of Lemma 5

Part i). This result follows from Lemma 3i and the fact that there exists a constant M independent of T such that

$$\max\big(d_{2T}, d_{3T}\big) \leq M \quad \text{for all} \;\; (T, N_T). \tag{A36}$$

The boundedness of d_{2T} in (A36) follows from Assumption 2, and that of d_{3T} because the sum of the absolute value of all entries in the matrix $\mathbf{I}_{N_T} - \frac{1}{N_T}\mathbf{1}_{N_T}\mathbf{1}'_{N_T}$ is equal to $2(N_T - 1)$.

Part ii). We have $\mathbb{E}\big[\hat{\tilde{a}}_{1T}\big] - \tilde{a}_{1T} \to 0$ as $T \to \infty$ due to Lemma 3ii and the fact that S_{kl,N_T} is an unbiased estimator. Similarly, $\mathbb{E}\big[\hat{\tilde{a}}_{2T}\big] - \tilde{a}_{2T} \to 0$ as $T \to \infty$ holds too because

$$\begin{aligned}
\mathbb{E}\big[\hat{d}_{3T}\big] - d_{3T} &= \mathbb{E}\big[\hat{a}_{1T}\big] - a_{1T} - \frac{1}{N_T^2}\Big(\mathbb{E}\Big[\big(\mathbf{1}'_{N_T}\overline{\mathbf{X}}^{(T)}\big)^2\Big] - \big(\mathbf{1}'_{N_T}\boldsymbol{\mu}_{N_T}\big)^2\Big) \\
&= \mathbb{E}\big[\hat{a}_{1T}\big] - a_{1T} - \frac{1}{N_T^2}\mathbb{V}\Big[\mathbf{1}'_{N_T}\overline{\mathbf{X}}^{(T)}\Big] \\
&= \mathbb{E}\big[\hat{a}_{1T}\big] - a_{1T} - \frac{1}{TN_T^2}\mathbf{1}'_{N_T}\boldsymbol{\Sigma}_{N_T}\mathbf{1}_{N_T} \\
&\to 0 \quad \text{as} \;\; T \to \infty,
\end{aligned} \tag{A37}$$

by recalling Lemma 3ii once again and Assumption 2.

Turning to the asymptotically vanishing variances, we have

$$\mathbb{V}\big[\hat{\tilde{a}}_{1T}\big] \leq 3\mathbb{V}\big[\hat{a}_{1T}\big] + 3\mathbb{V}\big[\hat{a}_{2T}\big] + 3\mathbb{V}\big[\hat{d}_{2T}\big]. \tag{A38}$$

Similar steps to those in (A34) yield

$$\begin{aligned}
\mathbb{V}\big[\hat{d}_{2T}\big] &= \frac{1}{T^2N_T^4}\mathbb{V}\big[\mathbf{1}'_{N_T}\mathbf{S}_{N_T}\mathbf{1}_{N_T}\big] \\
&\leq \frac{1}{T^2N_T^2}\sum_{k,l=1}^{N_T}\mathbb{V}[S_{kl,N_T}]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T^3 N_T^2} \sum_{k,l=1}^{N_T} \left(\mu^{(4)}_{kl,N_T} - \frac{T-2}{T-1} \Sigma^2_{kl,N_T} + \frac{1}{T-1} \Sigma_{kk,N_T} \Sigma_{ll,N_T} \right), \\
&\leq \frac{M_1}{T^3}, \qquad \text{(A39)}
\end{aligned}
$$

where Lemma 1 is used in the third step, and the existence of $M_1 > 0$ independent of T is justified by Assumption 2. By combining (A38), (A39), and the results in Lemma 3ii, we have $\mathbb{V}\big[\hat{\tilde{a}}_{1T}\big] \to 0$ as $T \to \infty$. A similar result holds for $\hat{\tilde{a}}_{2T}$ by first noting that

$$
\begin{aligned}
\mathbb{V}\big[\hat{\tilde{a}}_{2T}\big] &\leq 2\mathbb{V}\big[\hat{\tilde{a}}_{1T}\big] + 2\mathbb{V}\big[\hat{d}_{3T}\big] \\
&= 2\mathbb{V}\big[\hat{\tilde{a}}_{1T}\big] + 2\mathbb{V}\left[\hat{a}_{1T} - \frac{1}{N_T^2} \big(\mathbf{1}'_{N_T} \overline{\mathbf{X}}^{(T)} \big)^2 \right] \\
&\leq 2\mathbb{V}\big[\hat{\tilde{a}}_{1T}\big] + 4\mathbb{V}\big[\hat{a}_{1T}\big] + \frac{4}{N_T^4} \mathbb{V}\Big[\big(\mathbf{1}'_{N_T} \overline{\mathbf{X}}^{(T)} \big)^2 \Big]. \qquad \text{(A40)}
\end{aligned}
$$

Now, the following holds for all $T \geq 1$:

$$
\frac{1}{N_T^4} \mathbb{V}\Big[\big(\mathbf{1}'_{N_T} \overline{\mathbf{X}}^{(T)} \big)^2 \Big] = \frac{1}{N_T^4} \mathbb{V}\left[\sum_{k,l=1}^{N_T} \overline{X}_k^{(T)} \, \overline{X}_l^{(T)} \right] \leq \frac{1}{N_T^2} \sum_{k,l=1}^{N_T} \mathbb{V}\Big[\overline{X}_k^{(T)} \, \overline{X}_l^{(T)} \Big] \leq \frac{M_2}{T^2}, \qquad \text{(A41)}
$$

where the existence of $M_2 > 0$ independent of T is guaranteed by Assumption 2 and some similar derivations to those in (A19) for which we only outline the main steps:

$$
\begin{aligned}
\mathbb{V}\Big[\overline{X}_k^{(T)} \, \overline{X}_l^{(T)} \Big] &= \frac{1}{T^2} \sum_{i,j=1}^{N_T} \left(\mathbb{E}\left[\left(\overline{X}_k^{(T)} \right)^2 X_{il} X_{jl} \right] - \mathbb{E}\Big[\overline{X}_k^{(T)} X_{il} \Big] \mathbb{E}\Big[\overline{X}_k^{(T)} X_{jl} \Big] \right) \\
&= \frac{1}{T^2} \sum_{i,j=1}^{N_T} \left(\frac{1}{T^2} \sum_{i_1,j_1=1}^{T} \mathbb{E}[X_{i_1 k} X_{j_1 k} X_{il} X_{jl}] - \frac{1}{T^2} \Sigma^2_{kl,N_T} \right) \\
&= \frac{1}{T^4} \sum_{i,j=1}^{N_T} \Big(\mu^{(4)}_{kl,N_T} \mathbb{1}_{\{i=j\}} + 2\Sigma^2_{kl,N_T} \mathbb{1}_{\{i \neq j\}} \Big) - \frac{1}{T^2} \Sigma^2_{kl,N_T} \\
&= \frac{1}{T^2} \left(\frac{1}{T} \mu^{(4)}_{kl,N_T} + \frac{T-2}{T} \Sigma^2_{kl,N_T} \right). \qquad \text{(A42)}
\end{aligned}
$$

By combining (A40), (A41), the results in Lemma 3ii, and the fact that $\mathbb{V}\big[\hat{\tilde{a}}_{1T}\big] \to 0$ as $T \to \infty$, we have $\mathbb{V}\big[\hat{\tilde{a}}_{2T}\big] \to 0$ as $T \to \infty$.

Part iii). The proof follows from (A17)–(A18). This completes the proof.

F Proofs of Theorems

F.1 Proof of Theorem 1

Part i). For a given $c \in \mathbb{R}$, the MSE of $\hat{\boldsymbol{\mu}}^{StM}(c)$ in (3) simplifies to

$$\begin{aligned} \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c)\big) &= \frac{1}{N_T}\Big(c^2\mathrm{tr}\big(\mathbb{V}\big(\overline{\mathbf{X}}^{(T)}\big)\big) + (1-c)^2\boldsymbol{\mu}_{N_T}'\boldsymbol{\mu}_{N_T}\Big) \\ &= c^2\mathrm{MSE}\big(\overline{\mathbf{X}}^{(T)}\big) + (1-c)^2\big|\big|\boldsymbol{\mu}_{N_T}\big|\big|_{N_T}^2, \end{aligned} \tag{A43}$$

which is strictly convex. Therefore, there is a unique $c \in \mathbb{R}$ minimizing $\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c)\big)$, which is given by $c_T^\star \in [0,1)$ in (8), where the value of one cannot be attained because $\mathrm{MSE}\big(\overline{\mathbf{X}}^{(T)}\big) > 0$ by assumption that $\boldsymbol{\Sigma}_{N_T}$ is positive definite. Finally, plugging $c_T^\star$ into (A43) yields the expression for $\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\big)$ in (9).

Part ii). For a given $\mathbf{c} \in \mathbb{R}^{N_T}$, the MSE of $\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c})$ in (4) simplifies to

$$\begin{aligned} \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c})\big) &= \frac{1}{N_T}\mathrm{tr}\big(\mathbb{V}\big(\,\mathrm{diag}(\mathbf{c})\overline{\mathbf{X}}^{(T)}\big)\big) + \big|\big|\,\mathrm{diag}(\mathbf{c})\boldsymbol{\mu}_{N_T} - \boldsymbol{\mu}_{N_T}\big|\big|_{N_T}^2 \\ &= \frac{1}{N_T}\mathbf{c}'\mathbb{V}\big(\overline{\mathbf{X}}^{(T)}\big)\mathbf{c} + \frac{1}{N_T}\sum_{k=1}^{N}(1-c_k)^2\mu_{k,N_T}^2 \\ &= \frac{1}{N_T}\sum_{k=1}^{N}\left(\frac{c_k^2}{T}\Sigma_{kk,N_T} + (1-c_k)^2\mu_{k,N_T}^2\right), \end{aligned} \tag{A44}$$

which is a sum of separable (with respect to each c_k) strictly convex quadratic functions. Therefore, for all $k = 1, \ldots, N_T$, there is a unique $c_k \in \mathbb{R}$ minimizing the k^{th} summand in (A44), which is given by $c_{k,T}^\star \in [0,1)$ in (10), where the value of one cannot be attained because $\Sigma_{kk,N_T} > 0$ by assumption that $\boldsymbol{\Sigma}_{N_T}$ is positive definite. Finally, plugging $\mathbf{c}_T^\star$ into (A44) yields the expression for $\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)\big)$ in (11).

Part iii). The inequalities $\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)\big) \leq \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\big) \leq \mathrm{MSE}\big(\overline{\mathbf{X}}^{(T)}\big)$ are clear because the underlying feasibility sets are increasingly restrictive. Moreover, $\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\big) < \mathrm{MSE}\big(\overline{\mathbf{X}}^{(T)}\big)$ because $\mathrm{MSE}\big(\overline{\mathbf{X}}^{(T)}\big) > 0$. Finally, we can show that $\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)\big) < \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\big)$ in (12) holds unless $\mu_{k,N_T}^2/\Sigma_{kk,N_T}$ is the same for all

$k = 1, \ldots, N_T$ by using Asimit et al. (2025, Lemma A.1). This completes the proof.

F.2 Proof of Theorem 2

We start by writing $c_T^\star$ in (8) and $\hat{c}_T^\star$ in (13) as

$$c_T^\star = \frac{a_{1T}}{a_{2T}} \quad \text{and} \quad \hat{c}_T^\star = \frac{\hat{a}_{1T}}{\hat{a}_{2T}}, \tag{A45}$$

where a_{1T}, a_{2T}, $\hat{a}_{1T}$, and $\hat{a}_{2T}$ are defined in (A13)–(A14).

Part i). We have

$$\left|\left|\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star) - \hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\right|\right|_{N_T}^2 = \left(\frac{\hat{a}_{1T}}{\hat{a}_{2T}} - \frac{a_{1T}}{a_{2T}}\right)^2 \hat{a}_{1T} \leq \frac{\left(\hat{a}_{1T}a_{2T} - \hat{a}_{2T}a_{1T}\right)^2}{\hat{a}_{2T}a_{2T}^2} \leq \hat{a}_{2T}, \tag{A46}$$

where the inequality holds almost surely for any (T, N_T) due to Lemma 3iii. We can now apply Lemma 2 with $\Delta_T = \left(\hat{a}_{1T}a_{2T} - \hat{a}_{2T}a_{1T}\right)^2$, $\hat{a}_T = \hat{a}_{2T}$, $a_T = a_{2T}$, $\tau_1 = 1$, $\tau_2 = 2$, $M_1 = 1$, and $M_2 = 0$, which implies (15). Note that the three conditions in Lemma 2 are satisfied due to Lemma 3ii, Lemma 3iii, and (A46). In particular, $\mathbb{E}[\Delta_T] \to 0$ as $T \to \infty$ holds because

$$\hat{a}_{1T}a_{2T} - \hat{a}_{2T}a_{1T} = (\hat{a}_{1T} - a_{1T})a_{2T} - (\hat{a}_{2T} - a_{2T})a_{1T}, \tag{A47}$$

and a linear combination of L_2-convergent-to-zero sequences weighted with bounded weights is L_2-convergent-to-zero too.

Part ii). First, we have that

$$\begin{aligned}
&\mathbb{E}\left[\left|\left|\left|\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star) - \boldsymbol{\mu}_{N_T}\right|\right|_{N_T}^2 - \left|\left|\hat{\boldsymbol{\mu}}^{StM}(c_T^\star) - \boldsymbol{\mu}_{N_T}\right|\right|_{N_T}^2\right|\right] \\
&\leq \sqrt{\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star) - \hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\right|\right|_{N_T}^2\right]} \sqrt{\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star) + \hat{\boldsymbol{\mu}}^{StM}(c_T^\star) - 2\boldsymbol{\mu}_{N_T}\right|\right|_{N_T}^2\right]},
\end{aligned} \tag{A48}$$

which is a consequence of the Cauchy-Schwarz inequality. Now, the first term in the right-hand side of (A48) converges to zero due to (15), and the second term in the right-hand side

of (A48) is uniformly bounded for all $T \geq 2$ because of Assumption 2 as well as

$$\begin{aligned}\mathbb{E}\Big[\Big|\Big|\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star) - \boldsymbol{\mu}_{N_T}\Big|\Big|_{N_T}^2\Big] &\leq \frac{2}{N_T}\sum_{k=1}^{N_T}\Big(\mathbb{E}\Big[\big(\hat{c}_T^\star \overline{X}_k^{(T)}\big)^2\Big] + \mu_{k,N_T}^2\Big) \\ &\leq \frac{2}{TN_T}\mathrm{tr}\big(\boldsymbol{\Sigma}_{N_T}\big) + \frac{4}{N_T}\boldsymbol{\mu}_{N_T}'\boldsymbol{\mu}_{N_T} \qquad \text{(A49)}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\Big[\Big|\Big|\hat{\boldsymbol{\mu}}^{StM}(c_T^\star) - \boldsymbol{\mu}_{N_T}\Big|\Big|_{N_T}^2\Big] &\leq \frac{2}{N_T}\sum_{k=1}^{N_T}\Big(\mathbb{E}\Big[\big(c_T^\star \overline{X}_k^{(T)}\big)^2\Big] + \mu_{k,N_T}^2\Big) \\ &\leq \frac{2}{TN_T}\mathrm{tr}\big(\boldsymbol{\Sigma}_{N_T}\big) + \frac{4}{N_T}\boldsymbol{\mu}_{N_T}'\boldsymbol{\mu}_{N_T}. \qquad \text{(A50)}\end{aligned}$$

Thus, the left-hand side of (A48) converges to zero, which proves (16).

Part iii). First, using (9), we denote

$$\begin{aligned}\Big|\widehat{\mathrm{MSE}}\big(\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star)\big) - \mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\big)\Big| &= \left|\frac{\hat{a}_{1T}\big(\hat{a}_{2T} - \hat{a}_{1T}\big)}{\hat{a}_{2T}} - \frac{a_{1T}\big(a_{2T} - a_{1T}\big)}{a_{2T}}\right| \\ &= \frac{\Big|\hat{a}_{1T}\big(\hat{a}_{2T} - \hat{a}_{1T}\big)a_{2T} - a_{1T}\hat{a}_{2T}\big(a_{2T} - a_{1T}\big)\Big|}{\hat{a}_{2T}a_{2T}} \\ &=: \frac{\widetilde{\Delta}_T}{\hat{a}_{2T}a_{2T}}. \qquad \text{(A51)}\end{aligned}$$

We apply Lemma 2 with $\Delta_T = \widetilde{\Delta}_T$, $\hat{a}_T = \hat{a}_{2T}$, $a_T = a_{2T}$, and $\tau_1 = \tau_2 = M_1 = M_2 = 1$, and thus, we only need to justify conditions (A9) and (A11). The latter follows from

$$\frac{\widetilde{\Delta}_T}{\hat{a}_{2T}a_{2T}} \leq \frac{\hat{a}_{1T}}{\hat{a}_{2T}}\big(\hat{a}_{2T} - \hat{a}_{1T}\big) + \frac{a_{1T}}{a_{2T}}\big(a_{2T} - a_{1T}\big) \leq \hat{a}_{2T} - \hat{a}_{1T} + a_{2T} - a_{1T} \leq \hat{a}_{2T} + a_{2T}, \quad \text{(A52)}$$

which holds almost surely for all $T \geq 2$. Regarding condition (A9), i.e., $\mathbb{E}[\widetilde{\Delta}_T] \to 0$ as $T \to \infty$, some cumbersome derivations yield $\widetilde{\Delta}_T = A_{1T} + A_{2T} + A_{3T} - A_{4T} - A_{5T}$, where

$$\begin{aligned}&A_{1T} := a_{2T}(\hat{a}_{1T} - a_{1T})(\hat{a}_{2T} - a_{2T}), \quad A_{2T} := a_{1T}a_{2T}\big((\hat{a}_{2T} - a_{2T}) - \big(\hat{a}_{1T} - a_{1T}\big)\big), \\ &A_{3T} := a_{2T}(a_{2T} - a_{1T})(\hat{a}_{1T} - a_{1T}), \quad A_{4T} := a_{1T}(a_{2T} - a_{1T})(\hat{a}_{2T} - a_{2T}), \\ &A_{5T} := a_{2T}(\hat{a}_{1T} - a_{1T})^2. \qquad \text{(A53)}\end{aligned}$$

Now,

$$\mathbb{E}\big[|A_{1T}|\big] \leq a_{2T}\sqrt{\mathbb{E}\Big[\big(\hat{a}_{1T}-a_{1T}\big)^2\Big]}\sqrt{\mathbb{E}\Big[\big(\hat{a}_{2T}-a_{2T}\big)^2\Big]} \to 0 \quad \text{as} \quad T\to\infty, \tag{A54}$$

due to Lemma 3i, Lemma 3ii, and Hölder's inequality,

$$\begin{aligned}\mathbb{E}\big[|A_{2T}|\big] &\leq a_{1T}a_{2T}\Big(\mathbb{E}\big|\hat{a}_{1T}-a_{1T}\big| + \mathbb{E}\big|\hat{a}_{2T}-a_{2T}\big|\Big)\\ &\leq a_{1T}a_{2T}\left(\sqrt{\mathbb{E}\Big[\big(\hat{a}_{1T}-a_{1T}\big)^2\Big]} + \sqrt{\mathbb{E}\Big[\big(\hat{a}_{2T}-a_{2T}\big)^2\Big]}\right) \to 0 \quad \text{as} \quad T\to\infty,\end{aligned} \tag{A55}$$

due to Lemma 3i, Lemma 3ii, and Lyapunov's inequality,

$$\begin{aligned}\mathbb{E}\big[|A_{3T}|\big] &\leq a_{2T}(a_{2T}-a_{1T})\sqrt{\mathbb{E}\Big[\big(\hat{a}_{1T}-a_{1T}\big)^2\Big]} \to 0 \quad \text{as} \quad T\to\infty,\\ \mathbb{E}\big[|A_{4T}|\big] &\leq a_{1T}(a_{2T}-a_{1T})\sqrt{\mathbb{E}\Big[\big(\hat{a}_{2T}-a_{2T}\big)^2\Big]} \to 0 \quad \text{as} \quad T\to\infty,\end{aligned} \tag{A56}$$

due to Lemma 3i, Lemma 3ii, Lemma 3iii, and Lyapunov's inequality, and finally,

$$\mathbb{E}\big[|A_{5T}|\big] = a_{2T}\mathbb{E}\Big[\big(\hat{a}_{2T}-a_{2T}\big)^2\Big] \to 0 \quad \text{as} \quad T\to\infty, \tag{A57}$$

due to Lemma 3i and Lemma 3ii. By combining (A54)–(A57), we obtain $\mathbb{E}[\widetilde{\Delta}_T] \to 0$ as $T\to\infty$. This yields (17) and completes the proof.

F.3 Proof of Theorem 3

We start by writing $c^\star_{k,T}$ in (10) and $\hat{c}^\star_{k,T}$ in (14) as

$$c^\star_{k,T} = \frac{a_{k1,T}}{a_{k2,T}} \quad \text{and} \quad \hat{c}^\star_{k,T} = \frac{\hat{a}_{k1,T}}{\hat{a}_{k2,T}}, \tag{A58}$$

where $a_{k1,T}$, $a_{k2,T}$, $\hat{a}_{k1,T}$, and $\hat{a}_{k2,T}$ are defined in (A15)–(A16).

Part i). We have

$$\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{DM}(\hat{\mathbf{c}}^\star_T) - \hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}^\star_T)\right|\right|^2_{N_T}\right] = \frac{1}{N_T}\sum_{k=1}^{N_T}\mathbb{E}\left[\left(\frac{\hat{a}_{k1,T}}{\hat{a}_{k2,T}} - \frac{a_{k1,T}}{a_{k2,T}}\right)^2 \hat{a}_{k2,T}\right], \tag{A59}$$

and there exists $M_1 > 0$ such that

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\hat{a}_{k1,T}}{\hat{a}_{k2,T}} - \frac{a_{k1,T}}{a_{k2,T}}\right)^2 \hat{a}_{k2,T}\right] &\leq \sqrt{\mathbb{E}\left[\left(\frac{\hat{a}_{k1,T}}{\hat{a}_{k2,T}} - \frac{a_{k1,T}}{a_{k2,T}}\right)^2\right]}\sqrt{\mathbb{E}\left[\left(\hat{a}_{k1,T} - \frac{a_{k1,T}}{a_{k2,T}}\hat{a}_{k2,T}\right)^2\right]} \\
&\leq \sqrt{\mathbb{E}\left[\left(\hat{a}_{k1,T} - \frac{a_{k1,T}}{a_{k2,T}}\hat{a}_{k2,T}\right)^2\right]} \\
&\leq \sqrt{\mathbb{V}\left[\left(\hat{a}_{k1,T} - \frac{a_{k1,T}}{a_{k2,T}}\hat{a}_{k2,T}\right)\right] + \left(\frac{1}{T}\Sigma_{kk,N_T}\right)^2} \\
&\leq \sqrt{2\mathbb{V}\big[\hat{a}_{k1,T}\big] + 2\mathbb{V}\big[\hat{a}_{k2,T}\big] + \left(\frac{1}{T}\Sigma_{kk,N_T}\right)^2} \\
&\leq \frac{M_1}{\sqrt{T}} \to 0 \quad \text{as} \quad T \to \infty. \qquad \text{(A60)}
\end{aligned}$$

The first inequality is due to Hölder's inequality, while the second, third, and fourth inequalities are due to Lemma 4iii. In the third step, we also use the fact that

$$\mathbb{E}\big[\hat{a}_{k1,T}\big] - a_{k1,T} = \mathbb{E}\big[\hat{a}_{k2,T}\big] - a_{k2,T} = \frac{1}{T}S_{kk,N_T}. \qquad \text{(A61)}$$

The final step follows from (A8), (A35), and the fact that $\mathbb{V}\big[\hat{a}_{k2,T}\big] \leq 2\mathbb{V}\big[\hat{a}_{k1,T}\big] + \frac{2}{T^2}\mathbb{V}\big[S_{kk,N_T}\big]$, while the existence of $M_1 > 0$ independent of (T, N_T, k) is guaranteed by Assumption 2.

Part ii). The Cauchy-Schwarz inequality implies that

$$\begin{aligned}
&\mathbb{E}\left[\left|\left|\left|\hat{\boldsymbol{\mu}}^{DM}(\hat{\mathbf{c}}_T^\star) - \boldsymbol{\mu}_{N_T}\right|\right|\right|_{N_T}^2 - \left|\left|\left|\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star) - \boldsymbol{\mu}_{N_T}\right|\right|\right|_{N_T}^2\right] \\
&\leq \sqrt{\mathbb{E}\left[\left|\left|\left|\hat{\boldsymbol{\mu}}^{DM}(\hat{\mathbf{c}}_T^\star) - \hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)\right|\right|\right|_{N_T}^2\right]} \times \sqrt{\mathbb{E}\left[\left|\left|\left|\hat{\boldsymbol{\mu}}^{DM}(\hat{\mathbf{c}}_T^\star) + \hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star) - 2\boldsymbol{\mu}_{N_T}\right|\right|\right|_{N_T}^2\right]}. \qquad \text{(A62)}
\end{aligned}$$

Now, the first term in the right-hand side of (A62) converges to zero due to (18), and the second term in the right-hand side of (A62) is uniformly bounded for all $T \geq 2$ because of Assumption 2 as well as

$$\begin{aligned}
\mathbb{E}\left|\left|\left(\hat{\boldsymbol{\mu}}^{DM}(\hat{\mathbf{c}}_T^\star) - \boldsymbol{\mu}_{N_T}\right|\right|\right|_{N_T}^2 &\leq \frac{2}{N_T}\sum_{k=1}^{N_T}\Big(\mathbb{E}\Big[\big(\hat{c}_{k,T}^\star \overline{X}_k^{(T)}\big)^2\Big] + \mu_{k,N_T}^2\Big) \\
&\leq \frac{2}{TN_T}\mathrm{tr}\big(\boldsymbol{\Sigma}_{N_T}\big) + \frac{4}{N_T}\boldsymbol{\mu}_{N_T}'\boldsymbol{\mu}_{N_T} \qquad \text{(A63)}
\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\Big|\Big|\hat{\boldsymbol{\mu}}^{DM}(\mathbf{c}_T^\star)-\boldsymbol{\mu}_{N_T}\Big|\Big|_{N_T}^2 &\leq \frac{2}{N_T}\sum_{k=1}^{N_T}\Big(\mathbb{E}\Big[\big(c_{k,T}^\star\overline{X}_k^{(T)}\big)^2\Big]+\mu_{k,N_T}^2\Big)\\ &\leq \frac{2}{TN_T}\mathrm{tr}\big(\boldsymbol{\Sigma}_{N_T}\big)+\frac{4}{N_T}\boldsymbol{\mu}_{N_T}'\boldsymbol{\mu}_{N_T}, \qquad \text{(A64)}\end{aligned}$$

which are due to Lemma 4iii. Thus, the left-hand side of (A62) converges to zero, which proves (19).

Part iii). First, using (11), we have that

$$\begin{aligned}&\mathbb{E}\Big[\Big|\widehat{\mathrm{MSE}}\big(\hat{\boldsymbol{\mu}}^{StM}(\hat{c}_T^\star)\big)-\mathrm{MSE}\big(\hat{\boldsymbol{\mu}}^{StM}(c_T^\star)\big)\Big|\Big]\\ =\;&\mathbb{E}\Bigg[\Bigg|\sum_{k=1}^{N_T}\frac{\hat{c}_{k,T}^\star}{TN_T}S_{kk,N_T}-\sum_{k=1}^{N_T}\frac{c_{k,T}^\star}{TN_T}\Sigma_{kk,N_T}\Bigg|\Bigg]\\ \leq\;&\frac{1}{TN_T}\sum_{k=1}^{N_T}\mathbb{E}\big[\big|\hat{c}_{k,T}^\star S_{kk,N_T}-c_{k,T}^\star\Sigma_{kk,N_T}\big|\big]\\ \leq\;&\frac{1}{N_T}\sum_{k=1}^{N_T}\mathbb{E}\Big[\Big|\big(\overline{X}_k^{(T)}\big)^2-\mu_{k,N_T}^2\Big|\Big]+\frac{1}{N_T}\sum_{k=1}^{N_T}\mathbb{E}[|S_{kk,N_T}-\Sigma_{kk,N_T}|]\\ \leq\;&\frac{1}{N_T}\sum_{k=1}^{N_T}\Bigg(\sqrt{\mathbb{E}\Big[\Big(\big(\overline{X}_k^{(T)}\big)^2-\mu_{k,N_T}^2\Big)^2\Big]}+\sqrt{\mathbb{V}[S_{kk,N_T}]}\Bigg). \qquad \text{(A65)}\end{aligned}$$

The second inequality is due to some simple derivations that require the direct definitions of $\hat{c}_{k,T}^\star$ and $\hat{c}_{k,T}^\star$, as well as Lemma 4iii. The last inequality is due to Lyapunov's inequality.

Second, using the formula for the variance of the squared sample mean of i.i.d. random variables in Severini (2005), we have

$$\begin{aligned}\mathbb{E}\Big[\Big(\big(\overline{X}_k^{(T)}\big)^2-\mu_{k,N_T}^2\Big)^2\Big] &= \frac{4}{T}\Sigma_{kk,N_T}\mu_{k,N_T}^2+\frac{1}{T^2}\Big(3\Sigma_{kk,N_T}^2+4\mu_{k,N_T}^{(3)}\mu_{k,N_T}\Big)\\ &\quad+\frac{1}{T^3}\Big(\mu_{kk,N_T}^{(4)}-3\Sigma_{kk,N_T}^2\Big)\\ &\leq \frac{M_2}{T} \quad \text{for all } T\geq 1, \qquad \text{(A66)}\end{aligned}$$

where the constant $M_2>0$ is independent of (T,N_T,k) and its existence is guaranteed by Assumption 2, e.g., take $M_2=4M^3+7M^2+M$ with the constant $M>0$ being as in

Assumption 2. Following Lemma 1, we also obtain

$$\mathbb{V}[S_{kk,N_T}] = \frac{1}{T}\left(\mu^{(4)}_{kk,N_T} - \frac{T-3}{T-1}\Sigma^2_{kk,N_T}\right) \leq \frac{M_3}{T} \quad \text{for all } T \geq 2, \tag{A67}$$

where the constant $M_3 > 0$ is independent of (T, N_T, k). Equations (A65)–(A67) and Assumption 2i yield (20). This completes the proof.

F.4 Proof of Theorem 4

Part i). For a given $\delta \in \mathbb{R}$, we obtain the MSE of $\hat{\boldsymbol{\mu}}^{OL}(\delta)$ in (6) by setting $\xi = 1$ in (A72), which gives

$$\text{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\delta)\big) = (1-\delta)^2 d_{1T} + \delta(2-\delta)d_{2T} + \delta^2 d_{3T}. \tag{A68}$$

There is a unique $\delta \in \mathbb{R}$ minimizing $\text{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\delta)\big)$, which is given by $\delta^\star_T \in (0,1]$ in (25), where the value of zero cannot be attained because $d_{1T} > d_{2T}$. Indeed, we would have $d_{1T} = d_{2T}$ if and only if all correlations in $\boldsymbol{\Sigma}_{N_T}$ were equal to one, which cannot be the case because we assume $\boldsymbol{\Sigma}_{N_T}$ is positive definite. Finally, plugging $\delta^\star_T$ into (A68) yields the expression for $\text{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\delta^\star_T)\big)$ in (26).

Part ii). For a given $(\xi, \delta) \in \mathbb{R}^2$, the MSE of $\hat{\boldsymbol{\mu}}^{TL}(\xi, \delta)$ in (7) simplifies to

$$\text{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\xi,\delta)\big) = \frac{1}{N_T}\text{tr}\Big(\mathbb{V}\Big[(1-\delta)\overline{\mathbf{X}}^{(T)} + \delta\xi\overline{X}^{(T)}_g \mathbf{1}_{N_T}\Big]\Big) + \Big|\Big|\text{bias}\big(\hat{\boldsymbol{\mu}}^{TL}(\delta,\xi)\big)\Big|\Big|^2_{N_T}, \tag{A69}$$

where

$$\begin{aligned}
& \frac{1}{N_T}\text{tr}\Big(\mathbb{V}\Big[(1-\delta)\overline{\mathbf{X}}^{(T)} + \delta\xi\overline{X}^{(T)}_g \mathbf{1}_{N_T}\Big]\Big) \\
= \; & \frac{1}{N_T}\text{tr}\Big(\Big((1-\delta)\mathbf{I}_{N_T} + \frac{\delta\xi}{N_T}\mathbf{1}_{N_T}\mathbf{1}'_{N_T}\Big)'\mathbb{V}\big[\overline{\mathbf{X}}^{(T)}\big]\Big((1-\delta)\mathbf{I}_{N_T} + \frac{\delta\xi}{N_T}\mathbf{1}_{N_T}\mathbf{1}'_{N_T}\Big)\Big) \\
= \; & (1-\delta)^2 d_{1T} + \big(\delta^2\xi^2 + 2\delta(1-\delta)\xi\big)d_{2T},
\end{aligned} \tag{A70}$$

and

$$\begin{aligned}\Big|\Big|\text{bias}\big(\hat{\boldsymbol{\mu}}^{TL}(\delta,\xi)\big)\Big|\Big|_{N_T}^2 &= \Big|\Big|(1-\delta)\boldsymbol{\mu}_{N_T}+\delta\xi\sqrt{d_{4T}}\mathbf{1}_{N_T}-\boldsymbol{\mu}_{N_T}\Big|\Big|_{N_T}^2 \\ &= \frac{\delta^2}{N_T}\boldsymbol{\mu}'_{N_T}\Big(\mathbf{I}_{N_T}-\frac{\xi}{N_T}\mathbf{1}_{N_T}\mathbf{1}'_{N_T}\Big)^2\boldsymbol{\mu}_{N_T} \\ &= \delta^2\big(d_{3T}+d_{4T}+d_{4T}\xi(\xi-2)\big). \end{aligned} \tag{A71}$$

Combining (A69) and (A70), we obtain

$$\text{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\xi,\delta)\big) = (1-\delta)^2 d_{1T}+\delta\xi(\delta\xi+2(1-\delta))d_{2T}+\delta^2 d_{3T}+\delta^2(1-\xi)^2 d_{4T}, \tag{A72}$$

which is convex in ξ for any $\delta\in\mathbb{R}$ because $\min(d_{2T},d_{4T})\geq 0$. Therefore, for a given $\delta\in\mathbb{R}$, the optimal ξ is given by

$$\xi(\delta) := \underset{\xi\in\mathbb{R}}{\text{argmin}}\,\text{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\xi,\delta)\big) = 1-\frac{d_{2T}/\delta}{d_{2T}+d_{4T}}. \tag{A73}$$

Moreover,

$$\text{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\xi(\delta),\delta)\big) = (1-\delta)^2 d_{1T}+\delta(2-\delta)d_{2T}+\delta^2 d_{3T}-\frac{d_{2T}^2}{d_{2T}+d_{4T}}, \tag{A74}$$

which is convex in δ because $\min(d_{1T}-d_{2T},d_{3T})\geq 0$, and thus, is minimized at $\delta=\delta_T^\star$ in (25) and $\xi=\xi(\delta_T^\star)=\xi_T^\star$ in (27). Finally, plugging $\delta_T^\star$ into (A74) gives the expression for $\text{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star,\delta_T^\star)\big)$ in (28).

Part iii). The inequalities $\text{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star,\delta_T^\star)\big)\leq\text{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)\big)\leq\text{MSE}\big(\overline{\mathbf{X}}^{(T)}\big)$ follow because the underlying feasibility sets are increasingly restrictive. Moreover, $\text{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)\big)<\text{MSE}\big(\overline{\mathbf{X}}^{(T)}\big)$ because $d_{1T}=\text{MSE}\big(\overline{\mathbf{X}}^{(T)}\big)>d_{2T}$ and $w<1$ in (26). Finally, $\text{MSE}\big(\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star,\delta_T^\star)\big)<\text{MSE}\big(\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)\big)$ because $d_{2T}>0$ and $d_{4T}\geq 0$ in (28). This completes the proof.

F.5 Proof of Theorem 5

We start by writing $\delta_T^\star$ in (25) and $\hat{\delta}_T^\star$ in (30) as

$$\delta_T^\star = \frac{\tilde{a}_{1T}}{\tilde{a}_{2T}} \quad \text{and} \quad \hat{\delta}_T^\star = \frac{\hat{\tilde{a}}_{1T}}{\hat{\tilde{a}}_{2T}}, \tag{A75}$$

where $\tilde{a}_{1T}$, $\tilde{a}_{2T}$, $\hat{\tilde{a}}_{1T}$, and $\hat{\tilde{a}}_{2T}$ are defined in (A17)–(A18).

Part i). We have

$$\begin{aligned}
\left|\left|\hat{\boldsymbol{\mu}}^{OL}(\hat{\delta}_T^\star) - \hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)\right|\right|_{N_T}^2 &= \left(\hat{\delta}_T^\star - \delta_T^\star\right)^2 \hat{d}_{3T} \\
&= \frac{\left(\hat{\tilde{a}}_{1T}\tilde{a}_{2T} - \tilde{a}_{1T}\hat{\tilde{a}}_{2T}\right)^2}{\left(\hat{\tilde{a}}_{2T}\right)^2 \tilde{a}_{2T}^2}\left(\hat{\tilde{a}}_{2T} - \hat{\tilde{a}}_{1T}\right) \\
&\leq \frac{\left(\hat{\tilde{a}}_{1T}\tilde{a}_{2T} - \tilde{a}_{1T}\hat{\tilde{a}}_{2T}\right)^2}{\hat{\tilde{a}}_{2T}\tilde{a}_{2T}^2} \\
&=: \frac{\Delta_T}{\hat{\tilde{a}}_{2T}\tilde{a}_{2T}^2},
\end{aligned} \tag{A76}$$

where the inequality holds almost surely for any (T, N_T) due to Lemma 5iii. We can now apply Lemma 2 to the last line in (A76) with $\hat{a}_T = \hat{\tilde{a}}_{2T}$, $a_T = \tilde{a}_{2T}$, $\tau_1 = 1$, $\tau_2 = 2$, $M_1 = 1$ and $M_2 = 0$, which implies (31). Note that the three conditions in Lemma 2 are satisfied due to Lemma 5ii and Lemma 5iii, the fact that $\hat{\delta}_T^\star \in [0,1]$ holds almost surely, $\delta_T^\star \in [0,1]$ for all (T, N_T), and similar derivations to those in (A47).

Part ii). We show (32) using similar arguments to those used in the proof of Theorem 2ii. The Cauchy-Schwarz inequality implies that

$$\begin{aligned}
&\mathbb{E}\left[\left|\left|\left|\hat{\boldsymbol{\mu}}^{OL}(\hat{\delta}_T^\star) - \boldsymbol{\mu}_{N_T}\right|\right|_{N_T}^2 - \left|\left|\hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star) - \boldsymbol{\mu}_{N_T}\right|\right|_{N_T}^2\right|\right] \\
&\leq \sqrt{\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{OL}(\hat{\delta}_T^\star) - \hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star)\right|\right|_{N_T}^2\right]}\sqrt{\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{OL}(\hat{\delta}_T^\star) + \hat{\boldsymbol{\mu}}^{OL}(\delta_T^\star) - 2\boldsymbol{\mu}_{N_T}\right|\right|_{N_T}^2\right]}.
\end{aligned} \tag{A77}$$

Now, the first term in the right-hand side of (A77) converges to zero due to (31), while the

second term in the right-hand side of (A77) is bounded for all $T \geq 2$ because

$$\begin{aligned}
\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{OL}(\hat{\delta}_T^{\star}) - \boldsymbol{\mu}_{N_T}\right|\right|_{N_T}^2\right] &\leq 2\,\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{OL}(\hat{\delta}_T^{\star})\right|\right|_{N_T}^2\right] + 2a_{1T} \\
&\leq 4\,\mathbb{E}\left[\left|\left|(1-\hat{\delta}_T^{\star})\overline{\mathbf{X}}^{(T)}\right|\right|_{N_T}^2\right] + 4\,\mathbb{E}\left[\left|\left|\hat{\delta}_T^{\star}\overline{X}_g^{(T)}\mathbf{1}_{N_T}\right|\right|_{N_T}^2\right] + 2a_{1T} \\
&\leq 4\,\mathbb{E}\left[\left|\left|\overline{\mathbf{X}}^{(T)}\right|\right|_{N_T}^2\right] + 4\,\mathbb{E}\left[\left|\left|\overline{X}_g^{(T)}\mathbf{1}_{N_T}\right|\right|_{N_T}^2\right] + 2a_{1T} \\
&= 4a_{2T} + \frac{4}{N_T}\mathbb{E}\big[\hat{a}_{1T} - \hat{d}_{3T}\big] + 2a_{1T} \\
&= 4a_{2T} + \frac{4}{N_T^3}\mathbb{E}\Big[\big(\mathbf{1}'_{N_T}\overline{\mathbf{X}}^{(T)}\big)^2\Big] + 2a_{1T} \\
&\leq 4a_{2T} + \frac{4}{N_T}a_{2T} + 2a_{1T} \\
&\leq 2a_{1T} + 8a_{2T}, \qquad \text{(A78)}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{OL}(\delta_T^{\star}) - \boldsymbol{\mu}_{N_T}\right|\right|_{N_T}^2\right] &\leq 2\,\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{OL}(\delta_T^{\star})\right|\right|_{N_T}^2\right] + 2a_{1T} \\
&\leq 4\,\mathbb{E}\left[\left|\left|(1-\delta_T^{\star})\overline{\mathbf{X}}^{(T)}\right|\right|_{N_T}^2\right] + 4\,\mathbb{E}\left[\left|\left|\delta_T^{\star}\overline{X}_g^{(T)}\mathbf{1}_{N_T}\right|\right|_{N_T}^2\right] + 2a_{1T} \\
&\leq 4\,\mathbb{E}\left[\left|\left|\overline{\mathbf{X}}^{(T)}\right|\right|_{N_T}^2\right] + 4\,\mathbb{E}\left[\left|\left|\overline{X}_g^{(T)}\mathbf{1}_{N_T}\right|\right|_{N_T}^2\right] + 2a_{1T} \\
&\leq 2a_{1T} + 8a_{2T}, \qquad \text{(A79)}
\end{aligned}$$

where Assumption 2 has been used multiple times in conjunction with (A31) to obtain (A78)–(A79). Finally, (A77)–(A79) implies that the left hand-side of (A77) converges to zero as $T \to \infty$, which proves (32).

Part iii). Using (26) and some further derivations, we conclude that we need to show that

$$\mathbb{E}\left[\left|\hat{a}_{2T} - \hat{a}_{1T} - \frac{\big(\hat{\tilde{a}}_{1T}\big)^2}{\hat{\tilde{a}}_{2T}} - \left(a_{2T} - a_{1T} - \frac{\big(\tilde{a}_{1T}\big)^2}{\tilde{a}_{2T}}\right)\right|\right] \to 0 \quad \text{as} \quad T \to \infty, \qquad \text{(A80)}$$

for which is sufficient to show that

$$\mathbb{E}\big[|\hat{a}_{2T} - a_{2T} - (\hat{a}_{1T} - a_{1T})|\big] \to 0 \quad \text{as} \quad T \to \infty \qquad \text{(A81)}$$

and

$$\mathbb{E}\left[\left|\frac{\left(\hat{\tilde{a}}_{1T}\right)^2}{\hat{\tilde{a}}_{2T}} - \frac{\left(\tilde{a}_{1T}\right)^2}{\tilde{a}_{2T}}\right|\right] \to 0 \quad \text{as} \quad T \to \infty. \tag{A82}$$

Lyapunov's inequality and the L_2-convergence results in Lemma 3iii imply (A81) as follows:

$$\begin{aligned}
\mathbb{E}\big[|\hat{a}_{2T} - a_{2T} - (\hat{a}_{1T} - a_{1T})|\big] &\leq \mathbb{E}\big[|\hat{a}_{2T} - a_{2T}|\big] + \mathbb{E}\big[|\hat{a}_{1T} - a_{1T}|\big] \\
&\leq \sqrt{\mathbb{E}\Big[\big(\hat{a}_{2T} - a_{2T}\big)^2\Big]} + \sqrt{\mathbb{E}\Big[\big(\hat{a}_{1T} - a_{1T}\big)^2\Big]} \\
&\to 0 \quad \text{as} \quad T \to \infty.
\end{aligned} \tag{A83}$$

Relation (A82) is due to Lemma 2 by taking

$$\begin{aligned}
\Delta_T &= \left|\big(\hat{\tilde{a}}_{1T}\big)^2 \tilde{a}_{2T} - \big(\tilde{a}_{1T}\big)^2 \hat{\tilde{a}}_{2T}\right| \\
&= \left|2\tilde{a}_{1T}\tilde{a}_{2T}\big(\hat{\tilde{a}}_{1T} - \tilde{a}_{1T}\big) + \tilde{a}_{2T}\big(\hat{\tilde{a}}_{1T} - \tilde{a}_{1T}\big)^2 - \big(\tilde{a}_{1T}\big)^2\big(\hat{\tilde{a}}_{2T} - \tilde{a}_{2T}\big)\right|
\end{aligned} \tag{A84}$$

as well as $\hat{a}_T = \hat{\tilde{a}}_{2T}$, $a_T = \tilde{a}_{2T}$, $\tau_1 = \tau_2 = M_1 = M_2 = 1$. Note that the three conditions in Lemma 2 are satisfied due to Lemma 5ii and Lemma 5iii,

$$\frac{\Delta_T}{\hat{\tilde{a}}_{2T}\tilde{a}_{2T}} \leq \frac{\left(\hat{\tilde{a}}_{1T}\right)^2 \tilde{a}_{2T} + \left(\tilde{a}_{1T}\right)^2 \hat{\tilde{a}}_{2T}}{\hat{\tilde{a}}_{2T}\tilde{a}_{2T}} \leq \hat{\tilde{a}}_{2T} + \tilde{a}_{2T}, \tag{A85}$$

and the fact that

$$\begin{aligned}
\mathbb{E}\big[\Delta_T\big] &\leq 2\tilde{a}_{1T}\tilde{a}_{2T}\sqrt{\mathbb{E}\Big[\big(\hat{\tilde{a}}_{1T} - \tilde{a}_{1T}\big)^2\Big]} + \tilde{a}_{2T}\mathbb{E}\Big[\big(\hat{\tilde{a}}_{1T} - \tilde{a}_{1T}\big)^2\Big] + \big(\tilde{a}_{1T}\big)^2\sqrt{\mathbb{E}\Big[\big(\hat{\tilde{a}}_{2T} - \tilde{a}_{2T}\big)^2\Big]} \\
&\to 0 \quad \text{as} \quad T \to \infty,
\end{aligned} \tag{A86}$$

which is a consequence of Lyapunov's inequality and the L_2-convergence results in Lemma 5iii. This yields (33) and completes the proof.

F.6 Proof of Theorem 6

Part i). We have that

$$\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{TL}(\hat{\xi}_T^\star, \hat{\delta}_T^\star) - \hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star, \delta_T^\star)\right|\right|_{N_T}^2\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\left|\left|\left(\frac{\tilde{a}_{1T}}{\tilde{a}_{2T}} - \frac{\hat{\tilde{a}}_{1T}}{\hat{\tilde{a}}_{2T}}\right)\left(\overline{X}_g^{(T)}\mathbf{1}_{N_T} - \overline{\mathbf{X}}^{(T)}\right) - \left(\frac{\hat{d}_{2T}}{\hat{d}_{2T}+\hat{d}_{4T}} - \frac{d_{2T}}{d_{2T}+d_{4T}}\right)\overline{X}_g^{(T)}\mathbf{1}_{N_T}\right|\right|_{N_T}^2\right] \\
&\leq 2\,\mathbb{E}\left[\left(\frac{\hat{\tilde{a}}_{1T}}{\hat{\tilde{a}}_{2T}} - \frac{\tilde{a}_{1T}}{\tilde{a}_{2T}}\right)^2 \hat{d}_{3T}\right] + 2\,\mathbb{E}\left[\left(\frac{\hat{d}_{2T}}{\hat{d}_{2T}+\hat{d}_{4T}} - \frac{d_{2T}}{d_{2T}+d_{4T}}\right)^2 \hat{d}_{4T}\right]. \qquad \text{(A87)}
\end{aligned}
$$

Now, note that

$$
\mathbb{E}\left[\left(\frac{\hat{\tilde{a}}_{1T}}{\hat{\tilde{a}}_{2T}} - \frac{\tilde{a}_{1T}}{\tilde{a}_{2T}}\right)^2 \hat{d}_{3T}\right] \leq \mathbb{E}\left[\frac{\left(\hat{\tilde{a}}_{1T}\tilde{a}_{2T} - \hat{\tilde{a}}_{2T}\tilde{a}_{1T}\right)^2}{\hat{\tilde{a}}_{2T}\left(\tilde{a}_{2T}\right)^2}\right] \to 0 \quad \text{as} \quad T \to \infty, \qquad \text{(A88)}
$$

where the convergence follows as in (A76). Further,

$$
\mathbb{E}\left[\left(\frac{\hat{d}_{2T}}{\hat{d}_{2T}+\hat{d}_{4T}} - \frac{d_{2T}}{d_{2T}+d_{4T}}\right)^2 \hat{d}_{4T}\right] \leq \mathbb{E}\left[\frac{\left(\hat{d}_{2T}\left(d_{2T}+d_{4T}\right) - d_{2T}\left(\hat{d}_{2T}+\hat{d}_{4T}\right)\right)^2}{\left(\hat{d}_{2T}+\hat{d}_{4T}\right)\left(d_{2T}+d_{4T}\right)^2}\right] \to 0 \qquad \text{(A89)}
$$

as $T \to \infty$, where the convergence is a consequence of Lemma 2 with $\hat{a}_T = \hat{d}_{2T} + \hat{d}_{4T}$, $a_T = d_{2T} + d_{4T}$, $\tau_1 = 1$, $\tau_2 = 2$, $M_1 = 1$, $M_2 = 0$, and the fact that

$$
\begin{aligned}
\hat{d}_{2T} - d_{2T} &= \hat{a}_{2T} - a_{2T} - \left(\hat{a}_{1T} - a_{1T}\right) - \left(\hat{\tilde{a}}_{1T} - \tilde{a}_{1T}\right) \xrightarrow{L_2} 0 \quad \text{as} \quad T \to \infty, \\
\hat{d}_{4T} - d_{4T} &= \hat{a}_{1T} - a_{1T} + \left(\hat{\tilde{a}}_{1T} - \tilde{a}_{1T}\right) - \left(\hat{\tilde{a}}_{2T} - \tilde{a}_{2T}\right) \xrightarrow{L_2} 0 \quad \text{as} \quad T \to \infty, \qquad \text{(A90)}
\end{aligned}
$$

which is due to Lemma 3ii and Lemma 5ii. Note that applying Lemma 2 in the above also requires the following result:

$$
\mathbb{E}\Big[\left(\hat{d}_{2T}d_{4T} - d_{2T}\hat{d}_{4T}\right)^2\Big] \leq 2d_{4T}^2\mathbb{E}\Big[\left(\hat{d}_{2T} - d_{2T}\right)^2\Big] + 2d_{2T}^2\mathbb{E}\Big[\left(\hat{d}_{4T} - d_{4T}\right)^2\Big] \quad \text{as} \quad T \to \infty, \qquad \text{(A91)}
$$

which is due to (A90), the boundedness of $\max\left(d_{2T}, d_{4T}\right)$ guaranteed by (A36), Lemma 3i, and the fact that $0 \leq d_{4T} \leq a_{1T}$. By combining (A87)–(A89), we obtain (34).

Part ii). As in the previous proofs, e.g., that of Theorem 5ii, it is sufficient to conclude that

$$
\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{TL}(\hat{\xi}_T^\star, \hat{\delta}_T^\star) + \hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star, \delta_T^\star) - 2\boldsymbol{\mu}_{N_T}\right|\right|_{N_T}^2\right] = O(1) \quad \text{as} \quad T \to \infty. \qquad \text{(A92)}
$$

We have

$$\begin{aligned}
&\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{TL}(\hat{\xi}_T^\star,\hat{\delta}_T^\star)+\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star,\delta_T^\star)-2\boldsymbol{\mu}_{N_T}\right|\right|_{N_T}^2\right]\\
\leq\;& 3\,\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{TL}(\hat{\xi}_T^\star,\hat{\delta}_T^\star)\right|\right|_{N_T}^2\right]+3\,\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star,\delta_T^\star)\right|\right|_{N_T}^2\right]+12a_{1T}\\
\leq\;& 6\,\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{TL}(\hat{\xi}_T^\star,\hat{\delta}_T^\star)-\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star,\delta_T^\star)\right|\right|_{N_T}^2\right]+9\,\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star,\delta_T^\star)\right|\right|_{N_T}^2\right]+12a_{1T}. \qquad \text{(A93)}
\end{aligned}$$

The second term in the last line of (A93) satisfies

$$\begin{aligned}
\mathbb{E}\left[\left|\left|\hat{\boldsymbol{\mu}}^{TL}(\xi_T^\star,\delta_T^\star)\right|\right|_{N_T}^2\right] &\leq 2\big(1-\delta_T^{\star\star}\big)^2\mathbb{E}\big[\hat{a}_{1T}\big]+2\big(1-\delta_T^{\star\star}\xi_T^\star\big)^2\mathbb{E}\big[\hat{d}_{4T}\big]\\
&\leq 2\mathbb{E}\big[\hat{a}_{1T}\big]+2\mathbb{E}\big[\hat{d}_{4T}\big]\\
&\leq 4\mathbb{E}\big[\hat{a}_{1T}\big]\\
&= 4\left(a_{1T}+\frac{1}{TN_T}\mathrm{tr}\big(\boldsymbol{\Sigma}_{N_T}\big)\right)=O(1)\quad\text{as}\;\;T\to\infty, \qquad \text{(A94)}
\end{aligned}$$

where the last step is due to (A31) and Lemma 3i. Equations (A93)–(A94), Lemma 3i, and part i) of this theorem imply (A92), and thus, (35).

Part iii). Using (26) and (28) and some further derivations, we conclude that we need to show that

$$\begin{aligned}
\mathbb{E}\Bigg[\Bigg|&\frac{\hat{d}_{1T}\,\hat{d}_{3T}}{\hat{d}_{1T}-\hat{d}_{2T}+\hat{d}_{3T}}\times\frac{\hat{d}_{4T}}{\hat{d}_{2T}+\hat{d}_{4T}}+\big(\hat{d}_{3T}+\hat{d}_{4T}\big)\frac{\hat{d}_{2T}}{\hat{d}_{2T}+\hat{d}_{4T}}\times\frac{\hat{d}_{1T}-\hat{d}_{2T}}{\hat{d}_{1T}-\hat{d}_{2T}+\hat{d}_{3T}} \qquad \text{(A95)}\\
&-\frac{d_{1T}\,d_{3T}}{d_{1T}-d_{2T}+d_{3T}}\times\frac{d_{4T}}{d_{2T}+d_{4T}}-\big(d_{3T}+d_{4T}\big)\frac{d_{2T}}{d_{2T}+d_{4T}}\times\frac{d_{1T}-d_{2T}}{d_{1T}-d_{2T}+d_{3T}}\Bigg|\Bigg]\to 0
\end{aligned}$$

as $T\to\infty$, for which it is sufficient to show that

$$\mathbb{E}\left[\left|\frac{\hat{d}_{1T}\,\hat{d}_{3T}}{\hat{d}_{1T}-\hat{d}_{2T}+\hat{d}_{3T}}\times\frac{\hat{d}_{4T}}{\hat{d}_{2T}+\hat{d}_{4T}}-\frac{d_{1T}\,d_{3T}}{d_{1T}-d_{2T}+d_{3T}}\times\frac{d_{4T}}{d_{2T}+d_{4T}}\right|\right]\to 0 \qquad \text{(A96)}$$

and

$$\mathbb{E}\left[\left|\big(\hat{d}_{3T}+\hat{d}_{4T}\big)\frac{\hat{d}_{2T}}{\hat{d}_{2T}+\hat{d}_{4T}}\times\frac{\hat{d}_{1T}-\hat{d}_{2T}}{\hat{d}_{1T}-\hat{d}_{2T}+\hat{d}_{3T}}-\big(d_{3T}+d_{4T}\big)\frac{d_{2T}}{d_{2T}+d_{4T}}\times\frac{d_{1T}-d_{2T}}{d_{1T}-d_{2T}+d_{3T}}\right|\right]\to 0 \qquad \text{(A97)}$$

as $T \to \infty$. We first prove (A96). We have that

$$\begin{aligned}
& \mathbb{E}\left[\left| \frac{\hat{d}_{1T}\,\hat{d}_{3T}}{\hat{d}_{1T}-\hat{d}_{2T}+\hat{d}_{3T}} \times \frac{\hat{d}_{4T}}{\hat{d}_{2T}+\hat{d}_{4T}} - \frac{d_{1T}\,d_{3T}}{d_{1T}-d_{2T}+d_{3T}} \times \frac{d_{4T}}{d_{2T}+d_{4T}} \right|\right] \\
\leq\ & \mathbb{E}\left[\left| \left(\hat{d}_{1T}-d_{1T}\right) \frac{\hat{d}_{3T}}{\hat{d}_{1T}-\hat{d}_{2T}+\hat{d}_{3T}} \times \frac{\hat{d}_{4T}}{\hat{d}_{2T}+\hat{d}_{4T}} \right|\right] \\
& + \mathbb{E}\left[\left| d_{1T} \left(\frac{\hat{d}_{3T}}{\hat{d}_{1T}-\hat{d}_{2T}+\hat{d}_{3T}} - \frac{d_{3T}}{d_{1T}-d_{2T}+d_{3T}} \right) \times \frac{\hat{d}_{4T}}{\hat{d}_{2T}+\hat{d}_{4T}} \right|\right] \\
& + \mathbb{E}\left[\left| \frac{d_{1T}\,d_{3T}}{d_{1T}-d_{2T}+d_{3T}} \left(\frac{\hat{d}_{4T}}{\hat{d}_{2T}+\hat{d}_{4T}} - \frac{d_{4T}}{d_{2T}+d_{4T}} \right) \right|\right] \\
=:\ & I_{1T} + I_{2T} + I_{3T}. \qquad \text{(A98)}
\end{aligned}$$

We can show that

$$\max\left(I_{2T}, I_{3T}\right) \leq d_{1T} \to 0 \quad \text{as} \quad T \to \infty \qquad \text{(A99)}$$

by recalling Assumptions 1 and 2 as well as (A31). Moreover,

$$I_{1T} \leq \mathbb{E}\Big[\Big|\left(\hat{d}_{1T}-d_{1T}\right)\Big|\Big] \leq \sqrt{\mathbb{E}\Big[\left(\hat{d}_{1T}-d_{1T}\right)\Big]^2} \to 0 \quad \text{as} \quad T \to \infty, \qquad \text{(A100)}$$

which is due to Lyapunov's inequality and the following direct implication of Lemma 3ii:

$$\hat{d}_{1T} - d_{1T} = \hat{a}_{2T} - a_{2T} - \left(\hat{a}_{1T} - a_{1T}\right) \xrightarrow{L_2} 0 \quad \text{as} \quad T \to \infty. \qquad \text{(A101)}$$

We obtain (A96) by combining (A98)–(A100).

Next, we prove (A97). We have that

$$\begin{aligned}
& \mathbb{E}\left[\left| \left(\hat{d}_{3T}+\hat{d}_{4T}\right) \frac{\hat{d}_{2T}}{\hat{d}_{2T}+\hat{d}_{4T}} \times \frac{\hat{d}_{1T}-\hat{d}_{2T}}{\hat{d}_{1T}-\hat{d}_{2T}+\hat{d}_{3T}} - \left(d_{3T}+d_{4T}\right) \frac{d_{2T}}{d_{2T}+d_{4T}} \times \frac{d_{1T}-d_{2T}}{d_{1T}-d_{2T}+d_{3T}} \right|\right] \\
\leq\ & \mathbb{E}\left[\left| \left(\hat{d}_{3T}+\hat{d}_{4T}-d_{3T}-d_{4T}\right) \frac{\hat{d}_{1T}-\hat{d}_{2T}}{\hat{d}_{1T}-\hat{d}_{2T}+\hat{d}_{3T}} \times \frac{\hat{d}_{2T}}{\hat{d}_{2T}+\hat{d}_{4T}} \right|\right] \\
& + \mathbb{E}\left[\left| \left(d_{3T}+d_{4T}\right) \left(\frac{\hat{d}_{1T}-\hat{d}_{2T}}{\hat{d}_{1T}-\hat{d}_{2T}+\hat{d}_{3T}} - \frac{d_{1T}-d_{2T}}{d_{1T}-d_{2T}+d_{3T}} \right) \times \frac{\hat{d}_{2T}}{\hat{d}_{2T}+\hat{d}_{4T}} \right|\right] \\
& + \mathbb{E}\left[\left| \left(d_{3T}+d_{4T}\right) \frac{d_{1T}-d_{2T}}{d_{1T}-d_{2T}+d_{3T}} \times \left(\frac{\hat{d}_{2T}}{\hat{d}_{2T}+\hat{d}_{4T}} - \frac{d_{2T}}{d_{2T}+d_{4T}} \right) \right|\right]
\end{aligned}$$

$$=: \quad I_{4T} + I_{5T} + I_{6T}. \tag{A102}$$

We can show that

$$\begin{aligned} I_{4T} &\leq \mathbb{E}\Big[\Big|\hat{d}_{3T} + \hat{d}_{4T} - d_{3T} - d_{4T}\Big|\Big] \\ &\leq \mathbb{E}\Big[\Big|\hat{d}_{3T} - d_{3T}\Big|\Big] + \mathbb{E}\Big[\Big|\hat{d}_{4T} - d_{4T}\Big|\Big] \\ &\leq \sqrt{\mathbb{E}\Big[\big(\hat{d}_{3T} - d_{3T}\big)\Big]^2} + \sqrt{\mathbb{E}\Big[\big(\hat{d}_{4T} - d_{4T}\big)\Big]^2} \to 0 \quad \text{as} \;\; T \to \infty, \end{aligned} \tag{A103}$$

which is due to Lyapunov's inequality, (A90), and the fact that $\hat{d}_{3T} - d_{3T} \overset{L_2}{\to} 0$ as $T \to \infty$ because

$$\hat{d}_{3T} - d_{3T} = \hat{a}_{1T} - a_{1T} - \big(\hat{d}_{4T} - d_{4T}\big) \overset{L_2}{\to} 0 \quad \text{as} \;\; T \to \infty, \tag{A104}$$

where the latter is due to Lemma 3ii and (A90).

Assume now that Condition (b) is true. Then, it is not difficult to find that

$$\max\big(I_{5T}, I_{6T}\big) \leq d_{3T} + d_{4T} = o(1) \quad \text{as} \;\; T \to \infty, \tag{A105}$$

which together with (A102)–(A103) implies (A97) when Condition (b) holds. Therefore, (A96)–(A97) are true, and in turn, (36) is true if Condition (b) holds.

Assume now that Condition (a) is true. The latter and Lemma 3i imply that there exists a constant $\widetilde{M}_1 > 0$ independent of (T, N_T) for which

$$\frac{\big(d_{3T} + d_{4T}\big)^2}{d_{1T} - d_{2T} + d_{3T}} \leq \frac{\big(d_{3T} + d_{4T}\big)^2}{d_{3T}} \leq \widetilde{M}_1 \quad \text{for all } T \geq 1. \tag{A106}$$

The above and Lyapunov's inequality then yield

$$\begin{aligned} I_{5T} &\leq \sqrt{\widetilde{M}_1 \mathbb{E}\left[\frac{\Big(\big(\hat{d}_{1T} - \hat{d}_{2T}\big)\big(d_{1T} - d_{2T} + d_{3T}\big) - \big(d_{1T} - d_{2T}\big)\big(\hat{d}_{1T} - \hat{d}_{2T} + \hat{d}_{3T}\big)\Big)^2}{\big(\hat{d}_{1T} - \hat{d}_{2T} + \hat{d}_{3T}\big)^2\big(d_{1T} - d_{2T} + d_{3T}\big)}\right]} \\ &= \sqrt{\widetilde{M}_1 \mathbb{E}\left[\frac{\big(\hat{\tilde{a}}_{1T}\tilde{a}_{2T} - \tilde{a}_{1T}\hat{\tilde{a}}_{2T}\big)^2}{\big(\hat{\tilde{a}}_{2T}\big)^2\tilde{a}_{2T}}\right]} \to 0 \quad \text{as} \;\; T \to \infty, \end{aligned} \tag{A107}$$

where the latter is due to Lemma 2 by taking $\hat{a}_T = \hat{\tilde{a}}_{2T}$, $a_T = \tilde{a}_{2T}$, $\tau_1 = 2$, $\tau_2 = 1$, $M_1 = 0$, $M_2 = 1$, and $\Delta_T = \big(\hat{\tilde{a}}_{1T}\tilde{a}_{2T} - \tilde{a}_{1T}\hat{\tilde{a}}_{2T}\big)^2$. We can apply Lemma 2 because

$$\begin{aligned}\mathbb{E}[\Delta_T] &= \mathbb{E}\Big[\Big(\tilde{a}_{2T}\big(\hat{\tilde{a}}_{1T} - \tilde{a}_{1T}\big) - \tilde{a}_{1T}\big(\hat{\tilde{a}}_{2T} - \tilde{a}_{2T}\big)\Big)^2\Big] \\ &\leq 2\big(\tilde{a}_{2T}\big)^2\mathbb{E}\Big[\big(\hat{\tilde{a}}_{1T} - \tilde{a}_{1T}\big)^2\Big] + 2\big(\tilde{a}_{1T}\big)^2\mathbb{E}\Big[\big(\hat{\tilde{a}}_{2T} - \tilde{a}_{2T}\big)^2\Big] \to 0 \quad \text{as} \quad T \to \infty, \text{(A108)}\end{aligned}$$

which is a consequence of Lemma 5i and Lemma 5ii. Similarly, there exists a constant $\widetilde{M}_2 > 0$ independent of (T, N_T) such that

$$\frac{\big(d_{3T} + d_{4T}\big)^2}{d_{2T} + d_{4T}} \leq \frac{\big(d_{3T} + d_{4T}\big)^2}{d_{4T}} \leq \widetilde{M}_2 \quad \text{for all } T \geq 1. \tag{A109}$$

The above and Lyapunov's inequality yield

$$\begin{aligned}I_{6T} &\leq \mathbb{E}\left[\left|\big(d_{3T} + d_{4T}\big)\left(\frac{\hat{d}_{2T}}{\hat{d}_{2T} + \hat{d}_{4T}} - \frac{d_{2T}}{d_{2T} + d_{4T}}\right)\right|\right] \\ &\leq \sqrt{\widetilde{M}_2\mathbb{E}\left[\frac{\Big(\hat{d}_{2T}\big(d_{2T} + d_{4T}\big) - d_{2T}\big(\hat{d}_{2T} + \hat{d}_{4T}\big)\Big)^2}{\big(\hat{d}_{2T} + \hat{d}_{4T}\big)^2\big(d_{2T} + d_{4T}\big)}\right]} \to 0 \quad \text{as} \quad T \to \infty, \text{(A110)}\end{aligned}$$

where the latter is due to Lemma 2 by taking $\hat{a}_T = \hat{d}_{2T} + \hat{d}_{4T}$, $a_T = d_{2T} + d_{4T}$, $\tau_1 = 2$, $\tau_2 = 1$, $M_1 = 0$, $M_2 = 1$, and $\Delta_T = \big(\hat{d}_{2T}d_{4T} - d_{2T}\hat{d}_{4T}\big)^2$. We can apply Lemma 2 because

$$\begin{aligned}\mathbb{E}[\Delta_T] &= \mathbb{E}\Big[\Big(d_{4T}\big(\hat{d}_{2T} - d_{2T}\big) - d_{2T}\big(\hat{d}_{4T} - d_{4T}\big)\Big)^2\Big] \\ &\leq 2d_{4T}^2\mathbb{E}\Big[\big(\hat{d}_{2T} - d_{2T}\big)^2\Big] + 2d_{2T}^2\mathbb{E}\Big[\big(\hat{d}_{4T} - d_{4T}\big)^2\Big] \to 0 \quad \text{as} \quad T \to \infty, \text{(A111)}\end{aligned}$$

which is a consequence of the boundedness of d_{2T} and d_{4T} as well as (A90). Finally, we obtain (A97) when Condition (a) holds by combining (A102), (A103), (A107), and (A110). Therefore, (A96)–(A97) are true in this case as well and, in turn, (36) is true if Condition (a) holds. This completes the proof.

References in Supplementary Material

Asimit, V., Cidota, M. A., Chen, Z. & Asimit, J. (2025), 'Slab and shrinkage linear regression estimation', `https://openaccess.city.ac.uk/id/eprint/35005/`.

Bai, J. & Perron, P. (2003), 'Computation and analysis of multiple structural change models', *Journal of Applied Econometrics* **18**(1), 1–22.

Ledoit, O. & Wolf, M. (2004), 'A well-conditioned estimator for large-dimensional covariance matrices', *Journal of Multivariate Analysis* **88**(2), 365–411.

Severini, T. A. (2005), *Elements of Distribution Theory*, Cambridge University Press.