Mean-Variance Efficiency of a Portfolio in Ultra-High Dimensions and Cases of Maximum Out-of-Sample and Constrained Maximum Sharpe Ratios

Mehmet Caner* Marcelo Medeiros† Gabriel F. R. Vasconcelos‡

June 30, 2020

Abstract

In this paper, we analyze the maximum Sharpe ratio when the number of assets in a portfolio is larger than its time span. One obstacle in this high-dimensional setup is the singularity of the sample covariance matrix of the excess asset returns. To resolve this issue, we benefit from a technique called nodewise regression, which was developed by Meinshausen and Bühlmann (2006). It provides a sparse/weakly sparse and consistent estimate of the precision matrix using the lasso method. One of the key results in our paper is the mean-variance efficiency of the portfolios in high dimensions. Tied to that result, we also show that the maximum out-of-sample Sharpe ratio can be consistently estimated in this large portfolio of assets. Furthermore, we provide convergence rates and show that the number of assets slows the convergence up to a logarithmic factor. We also provide consistency of the maximum Sharpe ratio when the portfolio weights sum to one and a new formula for the constrained maximum Sharpe ratio. Finally, we obtain consistent estimates of the Sharpe ratios of the global minimum-variance portfolio and Markowitz's (1952) mean-variance portfolio. In terms of assumptions, we allow for dependent data. Simulations and out-of-sample forecasting exercises show that our new method performs well compared to factor- and shrinkage-based techniques.

*North Carolina State University, Nelson Hall, Department of Economics, NC 27695. Email: mcaner@ncsu.edu.

†Department of Economics, Pontifical Catholic University of Rio de Janeiro - Brazil. Email: mcm@econ.puc-rio.br

‡Department of Economics, University of California, Irvine. SSPB 3201, Irvine, CA. 92697. Email: gabriel.vasconcelos@uci.edu We thank Vanderbilt Economics Department seminar guests for comments. We are grateful for the comments by Harold Chiang, Srini Krishnamurthy, and Michael Wolf.

1 Introduction

One of the key issues in finance is the trade-off between the return and the risk of a portfolio. To obtain a better risk-adjusted return, we maximize the Sharpe ratio. In essence, the weights of the portfolio are chosen in such a way that the return-to-risk ratio is maximized. We contribute to this literature by studying the case of a large number of assets p, which may be greater than the time span of the portfolio n. Our analysis also involves time-series data for excess asset returns. To obtain the maximum Sharpe ratio, we make use of the asset's precision matrix. However, the sample covariance matrix is not invertible when $p > n$. Therefore, we need another way to estimate the precision matrix. To do so, we use a concept promoted by Meinshausen and Bühlmann (2006), which is called *nodewise regression*. To obtain the Sharpe ratio, we estimate the inverse of the precision matrix by a nodewise regression-based inverse as in van de Geer (2016). This method consists of running lasso regression of a given excess asset return on the remaining assets to form the rows of the precision matrix. This type of method assumes sparsity, or weak sparsity on the rows of the precision matrix, when $p \geq n$. Weak sparsity allows a non-sparse precision matrix, as long as the absolute ℓth power $(0 < \ell < 1)$ sum of absolute value of coefficients in each row does not diverge too fast; for this issue, see section 2.10 of van de Geer (2016). Note that we do not assume the sample covariance matrix to be sparse.

This assumption of the sparsity of the precision matrix can be interpreted as an asset being potentially correlated with a number, but not all, of the assets in a portfolio. Asset A may be linked to Asset B, and asset B may be linked to asset C, but there is no direct link between asset A and asset C. This is not a strong assumption, as we show in our empirical out-of-sample exercise in Section 7. Figure 2 shows that there are not too many large correlations for US assets in the two subsamples that we use in our study.

The related literature on nodewise regression is as follows. Chang et al. (2019) extend nodewise regression to time-series data and build confidence intervals for the cells in the precision matrix. Callot et al. (2019) provide the variance, risk, and weight estimation of the portfolio via nodewise regression. Caner and Kock (2018) establish uniform confidence intervals in the case of high-dimensional parameters in heteroskedastic setups using nodewise regression. Meinshausen and Bühlmann (2006) already provide an optimality result for nodewise regression in terms of predicting a certain excess asset return with other excess asset returns when the returns are normally distributed.

In this paper, we analyze three important aspects of the maximum Sharpe ratio when $p \geq n$.

First, we analyze the maximum out-of-sample Sharpe ratio and the mean-variance efficiency of a large portfolio. Our technique, and hence its contribution, will be complementary to the existing papers. One difference is that we analyze $p \geq n$ when both the number of assets and the time span go to infinity in a time-series framework. Recently, important contributions have been in this area by using shrinkage and factor models. Ledoit and Wolf (2017) propose a nonlinear shrinkage estimator in which small eigenvalues of the sample covariance matrix are increased and large eigenvalues are decreased by a shrinkage formula. Their main contribution is the optimal shrinkage function, which they find by minimizing a loss function. The maximum out-of-sample Sharpe ratio is an inverse function of this loss. Their results cover the iid case and when $p/n \to (0,1) \cup (1, +\infty)$. For the analysis of mean-variance efficiency, Ao et al. (2019) make a novel contribution in which they take a constrained optimization, maximize returns subject to risk of the portfolio, and show that it is equivalent to an unconstrained objective function, where they minimize a scaled return of the portfolio error by choosing optimal weights. To obtain these weights, they use lasso regression and hence assume a sparse number of nonzero weights of the portfolio, and they analyze $p/n \to (0,1)$. They show that their method maximizes the expected return of the portfolio and satisfies the risk constraint. This is an important result on its own.

Our main contribution is that we are able to obtain mean-variance efficiency for large portfolios even when $p > n$ when both dimensions are growing. Relatedly, the consistency of our nodewise-based maximum-out-of-sample Sharpe ratio estimate is established. We also provide the rate of convergence and see that the number of assets slows the rate of convergence up to a logarithmic factor in p; hence, consistent estimation of the Sharpe ratio of large portfolios is possible.

Second, we consider the rate of convergence and consistency of the maximum Sharpe ratio when the weights of the portfolio are normalized to one and $p > n$. Recently, Maller and Turkington (2002) and Maller et al. (2016) analyze the limit with a fixed number of assets and extend that approach to a large number of assets but a number less than the time span of the portfolio. Their papers make a key discovery: in the case of weight constraints (summing to one), the formula for the maximum Sharpe ratio depends on a technical term, unlike the unconstrained maximum Sharpe ratio case. Practitioners could obtain the minimum Sharpe ratio instead of the maximum if they are using the unconstrained formula. Our paper extends their paper by analyzing two issues, first the case of $p > n$, with both quantities growing to infinity, and second by handling the uncertainty created by this technical term, which we can estimate and use to obtain a new constrained and consistent Sharpe ratio.

Our third contribution is that we consider the Sharpe ratios in the global minimum-variance portfolio and Markowitz mean-variance portfolio. Our analysis uncovers consistent estimators even when $p > n$. We show that our method performs well in simulations and empirical applications. The reason for the good performance is due to the correlation structure of the excess asset returns. The test (out-of-sample) periods that we analyze have a small number of large correlations and are hence more in line with our sparsity assumptions, which can be seen in Figure 1. In Figure 1, Subsample 1 and Subsample 2 correspond to two out-of-sample data periods in Section 7, where we cover January 2005-December 2017 and January 2000-December 2017, respectively. Additionally, in the same figure, we superimpose a simulation design that comes from a widely used factor model design in Section 6. The factor design does not conform with the two subperiods that we analyze via real-life data. The factor design misses all negatively correlated assets and concentrates heavily on the mean, so in that sense, it reflects a highly restricted sparse model.

Regarding other papers, Ledoit and Wolf (2003) and Ledoit and Wolf (2004) propose a linear shrinkage estimator to estimate the covariance matrix and apply it to portfolio optimization. Ledoit and Wolf (2017) shows that nonlinear shrinkage performs better in out-of-sample forecasts. Lai et al. (2011) and Garlappi et al. (2007) approach the same problem from a Bayesian perspective by aiming to maximize a utility function tied to portfolio optimization. Another avenue of the literature improves the performance of the portfolios by introducing constraints on the weights. This is in the case of the global minimum-variance portfolio. Examples of works investigating this problem include Jagannathan and Ma (2003) and Fan et al. (2012). We also see a combination of different portfolios proposed by Kan and Zhou (2007) and Tu and Zhou (2011).

This paper is organized as follows. Section 2 considers our assumptions and precision matrix estimation. Section 3 addresses the maximum out-of-sample Sharpe ratio and the mean-variance efficiency. Section 4 handles the case of the maximum Sharpe ratio when the weights are normalized to one. Section 5 concerns the global minimum-variance and Markowitz mean-variance portfolio Sharpe ratios. Section 6 provides simulations that compare several methods. Section 7 presents an out-of-sample forecasting exercise. The main proofs are in the Appendix, and the Supplementary Appendix has some benchmark results used in the main proofs section. Let $\|\nu\|_{l_1}, \|\nu\|_{l_2}, \|\nu\|_\infty$ be the l_1, l_2, l_∞ norms of a generic vector ν. For matrices, we have $\|A\|_\infty$,, which is the sup norm.

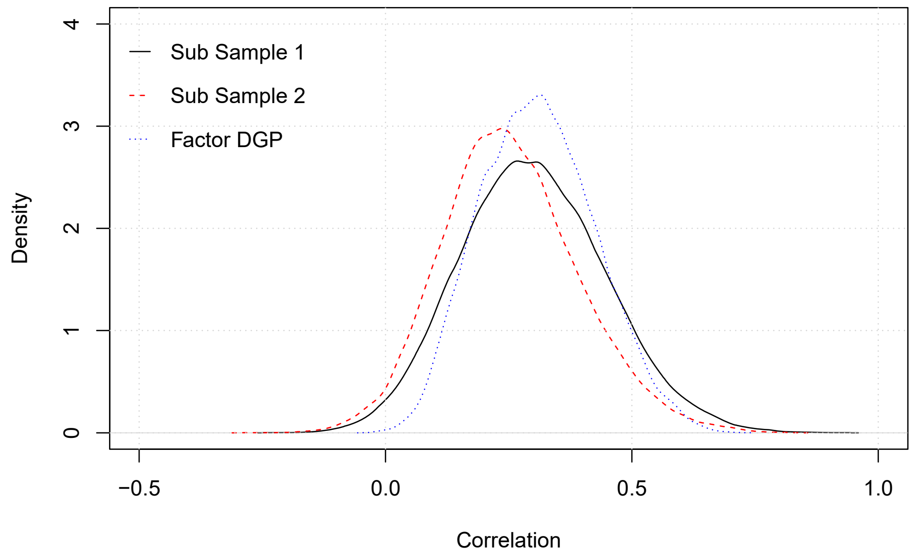

Figure 1: Correlation Densities

2 Precision Matrix and Its Estimate

2.1 Assumptions

Define $r_t := (r_{t,1}, r_{t,2}, \cdots, r_{t,p})'$ as the excess asset returns for a p asset portfolio, which is a $p \times 1$ vector. Define μ as the target excess asset return of a portfolio, $\mu := (\mu_1, \cdots, \mu_p)'$, which is a $p \times 1$ vector. The covariance matrix of excess asset returns is $\Sigma := E(r_t - \mu)(r_t - \mu)'$, and we define the sample covariance matrix of excess asset returns

$$\hat{\Sigma} := \frac{1}{n} \sum_{t=1}^{n} (r_t - \bar{r})(r_t - \bar{r})'.$$

Denote $\bar{r} := \frac{1}{n} \sum_{t=1}^{n} r_t$, which is a $p \times 1$ vector of mean excess asset returns. The matrix of excess asset returns (demeaned) is r^*, which is an $n \times p$ matrix. To make things clearer, set $r^*_{t,j} := r_{t,j} - \bar{r}_j$, which is the demeaned tth period, jth asset's excess return, and $\bar{r}_j := \frac{1}{n} \sum_{t=1}^{n} r_{t,j}$. Moreover, set r^*_j as the jth asset's demeaned excess return ($n \times 1$ vector), $j = 1, 2, \cdots, p$. Set r^*_{-j} as the matrix of demeaned excess returns ($n \times p - 1$ matrix), except the jth one. Let $r^*_{t,-j}$ represent the $p - 1$ vector of excess returns for all except the jth one. Furthermore, set $\hat{\mu} := \bar{r}$.

To understand the assumptions, we define a model that will link us to the nodewise regression concept in the next section. For $t = 1, \cdots, j, \cdots, n$

$$r^*_{t,j} = \gamma'_j r^*_{t,-j} + \eta_{t,j}, \tag{1}$$

where $\eta_{t,j}$ is the unobserved error. This is equation (5) in Chang et al. (2019). For the iid case, see equation (B.30) of Caner and Kock (2018). Here, we provide the assumptions.

Assumption 1. *There exist constants that are independent of* p *and* n*, such that* $K_1 > 0, K_2 > 1$, $0 < \alpha_1 \leq 2$*, and* $0 < \alpha_2 \leq 2$ *for* $t = 1, \cdots, n$

$$\max_{1 \leq j \leq p} Eexp(K_1 |r^*_{t,j}|^{\alpha_1}) \leq K_2, \quad \max_{1 \leq j \leq p} Eexp(K_1 |\eta_{t,j}|^{\alpha_2}) \leq K_2.$$

Assumption 2. *(i). The minimum eigenvalue of* Σ^{-1} *is denoted as* $Eigmin(\Sigma^{-1}) \geq c > 0,$*, where* c *is a positive constant, and the maximum eigenvalue of* Σ^{-1} *is denoted as* $Eigmax(\Sigma^{-1}) \leq K < \infty$*, where* K *is a positive constant. (ii). Moreover, for all* $j = 1, \cdots, p$*:* $0 < c_l \leq |\mu_j|$*, and for all* $j = 1, \cdots, p$ $|\mu_j| \leq c_u < \infty$*, where* c_l, c_u *are positive constants.*

Assumption 3. *The matrix of excess asset returns (demeaned)* r^* *has strictly stationary* β *mixing rows with* β *mixing coefficient satisfying* $\beta_k \leq exp(-K_3 k^{\alpha_3})$ *for any positive constant* k*, with* *constants* $K_3 > 0, \alpha_3 > 0$ *that are independent of* p *and* n*. Set* $\rho = min([\frac{1}{\alpha_1} + \frac{1}{\alpha_2} + \frac{1}{\alpha_3}]^{-1}, [\frac{1}{2\alpha_1} + \frac{1}{\alpha_3}]^{-1})$*. Additionally,* $lnp = o(n^{\rho/(2-\rho)})$*. With* $\rho \leq 1$*, we have that* $lnp = o(n)$*.*

Assumptions 1-2(i)-3 are from Chang et al. (2019). Assumption 1 allows us to apply the exponential tail inequalities used by Chang et al. (2019). Assumption 2(ii) does not allow a zero return for all assets, and all returns should also be finite. For technical and practical reasons, we also do not allow local to zero returns. Assumption 2 prevents the case of a zero maximum Sharpe ratio. Assumption 3 allows for weak dependence in the data. Chang et al. (2019) shows that causal ARMA processes with continuous error distributions are β mixing with exponentially decaying β_k. Stationary GARCH models with finite second moments and continuous error distributions satisfy Assumption 3. Some stationary Markov chains also satisfy Assumption 3. Note that we benefit from the first and fourth results of Lemma 1 on pp.70-71 Chang et al. (2019), so our ρ condition is a subset of theirs.

2.2 Precision Matrix Formula

In this subsection, we provide a precision matrix formula. This subsection is taken from Callot et al. (2019), and we repeat so that it will become clear how the precision matrix estimate is derived in the next subsection. The next subsection shows how this is related to the concept of the nodewise regression. We show how a formula for $\Theta := \Sigma^{-1}$ can be obtained under a strictly stationary time-series excess asset return. This is an extension of the iid case in Caner and Kock (2018). Let $\Sigma_{-j,-j}$ represent the $p-1 \times p-1$ submatrix of Σ, where the jth row and column have been removed. Additionally, $\Sigma_{j,-j}$ is the jth row of Σ with the jth element removed. Then, $\Sigma_{-j,j}$ represents the jth column of Σ with its jth element removed. From the inverse formula for the block matrices, we have the following for the jth main diagonal term:

$$\Theta_{j,j} = (\Sigma_{j,j} - \Sigma_{j,-j}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j})^{-1}, \tag{2}$$

and for the jth row of Θ with jth element removed

$$\Theta_{j,-j} = -(\Sigma_{j,j} - \Sigma_{j,-j}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j})^{-1}\Sigma_{j,-j}\Sigma_{-j,-j}^{-1} = -\Theta_{j,j}\Sigma_{j,-j}\Sigma_{-j,-j}^{-1}. \tag{3}$$

We now try to relate (2)(3) to a linear regression that we describe below in (7). Define γ_j ($p-1 \times 1$ vector) as the value of γ that minimizes

$$E[r_{t,j}^* - (r_{t,-j}^*)'\gamma]^2,$$

for all $t = 1, \cdots, n$. We can obtain a solution as

$$\gamma_j = \Sigma_{-j,-j}^{-1} \Sigma_{-j,j}, \tag{4}$$

by using strict stationarity of the data. Using symmetry of Σ and (4), we can write (3) as

$$\Theta_{j,-j} = -\Theta_{j,j} \gamma_j'. \tag{5}$$

Define the following $\Sigma_{-j,j} := E r_{t,-j}^* r_{t,j}^*$, $\Sigma_{-j,-j} := E r_{t,-j}^* r_{t,-j}^{*'}$. By (1), $\eta_{t,j} := r_{t,j}^* - (r_{t,-j}^*)' \gamma_j$. Then, it is easy to see by (4) that

$$\begin{aligned} E r_{t,-j}^* \eta_{t,j} &= E r_{t,-j}^* r_{t,j}^* - [E r_{t,-j}^* (r_{t,-j}^*)'] \gamma_j \\ &= \Sigma_{-j,j} - \Sigma_{-j,-j} \Sigma_{-j,-j}^{-1} \Sigma_{-j,j} = 0. \end{aligned} \tag{6}$$

This means that we can formulate (1) as a regression model with covariates orthogonal to errors

$$r_{t,j}^* = (r_{t,-j}^*)' \gamma_j + \eta_{t,j}, \tag{7}$$

for $t = 1, \cdots, n$. We can see that $\Theta_{j,-j}$ and hence all of the row Θ_j is sparse if and only if γ_j is sparse by comparing (5) and (7).

To derive a formula for Θ, we see that given (6)(7)

$$\begin{aligned} \Sigma_{j,j} := E[r_{t,j}^*]^2 &= \gamma_j' \Sigma_{-j,-j} \gamma_j + E \eta_{t,j}^2 \\ &= \Sigma_{j,-j} \Sigma_{-j,-j}^{-1} \Sigma_{-j,j} + E \eta_{t,j}^2, \end{aligned} \tag{8}$$

where we use (4) in the last equality in (8). Now, define $\tau_j^2 := E \eta_{t,j}^2$ for $t = 1, \cdots, n$, $j = 1, \cdots, p$. By (8)

$$\tau_j^2 = \Sigma_{j,j} - \Sigma_{j,-j} \Sigma_{-j,-j}^{-1} \Sigma_{-j,j} = \frac{1}{\Theta_{jj}}, \tag{9}$$

where we use (2) for the second equality. Next, define a $p \times p$ matrix

$$C_p := \begin{bmatrix} 1 & -\gamma_{1,2} & \cdots & -\gamma_{1,p} \\ -\gamma_{2,1} & 1 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ -\gamma_{p,1} & -\gamma_{p,2} & \cdots & 1 \end{bmatrix},$$

and $T^{-2} := diag(\tau_1^{-2}, \cdots, \tau_p^{-2})$, which is a diagonal matrix ($p \times p$ dimension). We can write

$$\Theta = T^{-2} C_p, \tag{10}$$

and to obtain (10), we use (2) and (9)

$$\Theta_{j,j} = \frac{1}{\tau_j^2}, \tag{11}$$

and by (5) with (11)

$$\Theta_{j,-j} = -\Theta_{j,j}\gamma_j' = \frac{-\gamma_j'}{\tau_j^2}.$$

2.3 Optimality of Nodewise Regression

As previously mentioned, the idea of nodewise regression was developed by Meinshausen and Bühlmann (2006). Nodewise regression stems from the idea of neighborhood selection. In a portfolio, neighborhood selection (nodewise regression) will select a "neighborhood" of a jth asset return (excess) in such a way that the smallest subset of returns of other assets in a portfolio will be conditionally dependent on this jth asset return. All the conditionally independent assets will receive a zero in the precision matrix. This method carries an optimality property when the asset returns are normally distributed. The normality assumption will be used only in this subsection. The best predictor for an excess asset return, $r_{t,j}^*$, in the portfolio of p assets is its neighborhood. Denote this neighborhood by $\mathcal{A}$. Then,

$$\gamma_j^* = argmin_{\gamma_j:\gamma_{j,k}=0\,\forall k\notin\mathcal{A}} E[r_{t,j}^* - \sum_{k\in\Gamma_{-j}} \gamma_{j,k} r_{t,k}^*]^2,$$

where $\mathcal{A} \subseteq \Gamma_{-j}$, $\Gamma_{-j} = \Gamma - \{j\}$, and $\Gamma = \{1, 2, \cdots, j, \cdots, p\}$. This is equation (2) in Meinshausen and Bühlmann (2006), where they have a detailed explanation for this result.

2.4 Estimate

A possible way of estimating the precision matrix when the number of assets is larger than the sample size is by nodewise regression. In the time series, this is developed by Chang et al. (2019). Callot et al. (2019) also use these results in portfolio risk. Here, we summarize the concept as in Callot et al. (2019). This is a concept based on the exact formula for the precision matrix. We borrow the main concepts from Bühlmann and van de Geer (2011). The precision matrix estimate follows the steps below.

1. Lasso nodewise regression is defined, for each $j = 1, 2, \cdots, p$, as

$$\hat{\gamma}_j = \underset{\gamma\in R^{p-1}}{\text{argmin}}[\|r_j^* - r_{-j}^*\gamma_j\|_2^2/n + 2\lambda_j\|\gamma\|_1], \tag{12}$$

where λ_j is a positive tuning parameter (sequence) and its choice, which will be discussed in the simulation section. Let S_j be the set of coefficients that are nonzero in row j of Σ^{-1}, and let $s_j = |S_j|$

be their cardinality. The maximum number of nonzero coefficients is set at $\bar{s} = \max_{1 \le j \le p} s_j$. Therefore, we make a sparsity assumption. Alternatively, but costly in notation, is weak sparsity, where we allow for the absolute lth power sum of coefficients in each row of the precision matrix to be diverging but not at a faster rate than the sample size. This of course demands a larger tuning parameter than does the sparsity assumption in practice. It is easy to incorporate weak sparsity into the proofs, as seen in Lemma 2.3 of van de Geer (2016). To avoid prolonging the paper, we have not pursued this track and required sparsity.

2. Setup the following matrix, which will be a key input in the precision matrix estimate:

$$\hat{C}_p = \begin{pmatrix} 1 & -\hat{\gamma}_{12} & \cdots & -\hat{\gamma}_{1p} \\ -\hat{\gamma}_{21} & 1 & \cdots & -\hat{\gamma}_{2p} \\ \cdots & \cdots & \ddots & \cdots \\ -\hat{\gamma}_{p1} & -\hat{\gamma}_{p2} & \cdots & 1 \end{pmatrix}.$$

3. Another key input is the following diagonal matrix with each scalar element $\hat{\tau}_j^2$, $j = 1, \cdots, p$

$$\hat{\tau}_j^2 = \frac{\|r_j^* - r_{-j}^* \hat{\gamma}_j\|_2^2}{n} + \lambda_j \|\hat{\gamma}_j\|_1.$$

From $\hat{T}^2 = diag(\hat{\tau}_1^2, \cdots, \hat{\tau}_p^2)$, which is the $p \times p$ matrix.

4. Set the precision matrix estimate (nodewise) as $\hat{\Theta} = \hat{T}^{-2} \hat{C}_p$.

We provide the first result in Lemma 1 of Chang et al. (2019) in the following Theorem. The iid data case with bounded moments is established in Caner and Kock (2018)

Theorem 1. *Under Assumptions 1-3,*

(i).

$$\max_{1 \le j \le p} \|\hat{\Theta}_j - \Theta_j\|_1 = O_p(\bar{s}\sqrt{\frac{lnp}{n}}).$$

(ii).

$$\|\hat{\mu} - \mu\|_\infty = O_p(\frac{\sqrt{lnp}}{\sqrt{n}}).$$

Note that Lemma 1 of Chang et al. (2019) applies to the estimation of sample covariance, whereas our theorem also shows the estimation of sample mean. From the proof of Lemma 1 for sample covariance in Chang et al. (2019), sample mean estimation can also be shown.

We provide the following assumption for the sparsity of coefficients in the nodewise regression estimate.

Assumption 4. *We have the following sparsity condition:*

$$\bar{s}\frac{\sqrt{lnp}}{\sqrt{n}} = o(1).$$

This is standard in the high-dimensional econometrics literature. By Assumption 4, it is easy to see that, via Theorem 1, the rows of the precision matrix are estimated consistently. The sparsity of the precision matrix does not imply that the covariance matrix is also sparse. It is possible to have, for example, a Toeplitz structure in the covariance matrix that is non-sparse but sparsity in the precision matrix.

2.5 Why use Nodewise Regression?

In finance, our method considers more complicated cases of $p > n$ and $p/n \to \infty$ when both $p, n \to \infty$. We also allow the $p = n$ case, while it is a hindrance to technical analysis in some shrinkage papers such as in the illuminating and very useful Ledoit and Wolf (2017). Our theorems also allow for non-iid data. Our technique should be seen as a complement to existing factor and shrinkage models and as carrying a certain optimality property, as outlined in subsection 2.3. Additionally, with our technique, one can obtain the mean-variance efficiency even when $p > n$ in the case of the maximum out-of-sample Sharpe ratio.

3 Maximum Out-of-Sample Sharpe Ratio

This section analyzes the maximum out of Sharpe ratio that is considered in Ao et al. (2019). To obtain that formula, we need the optimal calculation of the weights of the portfolio. The optimization of the portfolio weights is formulated as

$$argmax_w w'\mu \quad subject\, to \quad w'\Sigma w \leq \sigma^2, \tag{13}$$

where we maximize the return subject to a specified positive and finite risk constraint, $\sigma > 0$. After solving for the optimal weight, which will be shown in the next subsection, we can obtain the maximum out-of-sample Sharpe ratio. Equation (A.2) of Ao et al. (2019) defines the estimated maximum out-of-sample ratio when $p < n$, with the inverse of the sample covariance matrix used as an estimator for the precision matrix estimate, as:

$$\widehat{SR}_{moscov} := \frac{\mu'\hat{\Sigma}^{-1}\hat{\mu}}{\sqrt{\hat{\mu}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu}}},$$

and the theoretical version as

$$SR^* := \sqrt{\mu'\Sigma^{-1}\mu}.$$

Then, equation (1.1) of Ao et al. (2019) shows that when $p/n \to r_1 \in (0,1)$, the above plug-in maximum out-of-sample ratio cannot consistently estimate the theoretical version. We provide a nodewise version of the plug-in estimate that can estimate the theoretical Sharpe ratio even when $p > n$. Our maximum out-of-sample Sharpe ratio estimate using the nodewise estimate $\hat{\Theta}$ is:

$$\widehat{SR}_{mosnw} := \frac{\mu'\hat{\Theta}\hat{\mu}}{\sqrt{\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu}}}.$$

Theorem 2. *Under Assumptions 1-4,*

$$\left| \left[\frac{\widehat{SR}_{mosnw}}{SR^*} \right]^2 - 1 \right| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1).$$

Remarks. 1. Note that p.4353 of Ledoit and Wolf (2017) shows that the maximum out-of-sample Sharpe ratio is equivalent to minimizing a certain loss function of the portfolio. The limit of the loss function is derived under an optimal shrinkage function in Theorem 1. After that, they provide a shrinkage function even in the cases of $p/n \to r_1 \in (0,1) \cup (1,+\infty)$. Their proofs allow for iid data.

3.1 Mean-Variance Efficiency When $p > n$

This subsection formally shows that we can obtain mean-variance efficiency in an out-of-sample context when the number of assets in the portfolio is larger than the sample size, a novel result in the literature. Ao et al. (2019) show that this is possible when $p \leq n$, when both p, and n are large. That article is a very important contribution since they also demonstrate that other methods before theirs could not obtain that result, and it is a difficult issue to address. Given a risk level of $\sigma > 0$ and finite, the optimal weights of a portfolio are given in (2.3) of Ao et al. (2019) in an out-of-sample context. This comes from maximizing the expected portfolio return subject to its variance being constrained by the square of the risk, where this is shown in (13). Since $\Theta := \Sigma^{-1}$, the formula for weights is

$$w_{oos} = \frac{\sigma\Theta\mu}{\sqrt{\mu'\Theta\mu}}.$$

The estimates that we will use

$$\hat{w}_{oos} = \frac{\sigma \hat{\Theta} \hat{\mu}}{\sqrt{\hat{\mu}' \hat{\Theta} \hat{\mu}}}.$$

We are interested in maximized out-of-sample expected return $\mu' w_{oos}$ and its estimate $\mu' \hat{w}_{oos}$. Additionally, we are interested in the out-of-sample variance of the portfolio returns $w_{oos}' \Sigma w_{oos}$ and its estimate $\hat{w}_{oos}' \Sigma \hat{w}_{oos}$. Note also that by the formula for weights $w_{oos}' \Sigma w_{oos} = \sigma^2$, given $\Theta := \Sigma^{-1}$.

Below, we show that our estimates based on nodewise regression are consistent, and furthermore, we also provide the rate of convergence results.

Theorem 3. *(i). Under Assumptions 1-4,*

$$\left| \frac{\mu' \hat{w}_{oos}}{\mu' w_{oos}} - 1 \right| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1).$$

(ii). Under Assumptions 1-4,

$$\left| \hat{w}_{oos}' \Sigma \hat{w}_{oos} - \sigma^2 \right| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1).$$

Remarks. 1. From the results, we allow $p > n$, and still there is consistency. Additionally, the sparsity of the maximum number of nonzero elements in a row of the precision matrix $\bar{s}$ can grow to infinity but at a rate not larger than $\bar{s} = o((n/lnp)^{1/2})$ for the case in (i).

2. Therefore, we can allow $p = exp(n^{\kappa})$, with $0 < \kappa < 1$, and $\bar{s}$ can be a slowly varying function in n. This clearly shows that it is possible to have $p/n \to \infty$ in that scenario. In Theorem 3(i), we can have $p = n^2$, and $\bar{s} = o((n/lnn)^{1/2})$, and $p/n \to \infty$. The case of $p = 2 * n$ is also possible with $\bar{s} = o((n/lnn)^{1/2})$, with $p/n = 2$.

3. From the convergence rates, it is clear that we are penalized by the number of assets but in a logarithmic fashion; hence, our method is feasible to use in large-portfolio cases.

4. Ao et al. (2019) provide new results of the mean-variance efficiency of a large portfolio when $p \leq n$ and the returns of the assets are normally distributed. They provide a novel way of estimating return and risk. This involves lasso-sparse estimation of the weights of the portfolio.

4 Maximum Sharpe Ratio: Portfolio Weights Normalized to One

In this section, we define the maximum Sharpe ratio when the weights of the portfolio are normalized to one. This in turn will depend on a critical term that will determine the formula below.

The maximum Sharpe ratio is defined as follows, with w as the $p \times 1$ vector of portfolio weights:

$$max_w \frac{w'\mu}{\sqrt{w'\Sigma w}}, \text{s.}to \quad 1_p' w = 1,$$

where 1_p is a vector of ones. This maximum Sharpe ratio is constrained to have portfolio weights that sum to one. Maller et al. (2016) shows that depending on a scalar, it has two solutions. When $1_p'\Sigma^{-1}\mu \geq 0$, we have the square of the maximum Sharpe ratio:

$$MSR^2 = \mu'\Sigma^{-1}\mu. \tag{14}$$

When $1_p'\Sigma^{-1}\mu < 0$, we have

$$MSR_c^2 = \mu'\Sigma^{-1}\mu - (1_p'\Sigma^{-1}\mu)^2/(1_p'\Sigma^{-1}1_p). \tag{15}$$

This is equation (6.1) of Maller et al. (2016). Equation (14) is used in the literature, and this is the formula when the weights do not necessarily sum to one given a return constraint as in Ao et al. (2019).

These equations can be estimated by their sample counterparts, but in the case of $p > n$, $\hat{\Sigma}$ is not invertible, so we need to use new tools from high-dimensional statistics. We analyze the nodewise regression precision matrix estimate of Meinshausen and Bühlmann (2006). This is denoted by $\hat{\Theta}$. Therefore, we analyze the asymptotic behavior of the estimate of the maximum Sharpe ratio squared via nodewise regression. We will also introduce the maximum Sharpe ratio, which addresses the uncertainty regarding whether we should analyze MSR or MSR_c. This is

$$(MSR^*)^2 = MSR^2 1_{\{1_p'\Sigma^{-1}\mu \geq 0\}} + MSR_c^2 1_{\{1_p'\Sigma^{-1}\mu < 0\}}.$$

The estimators of MSR, MSR_c, MSR^* will be introduced in the next subsection.

4.1 Consistency and Rate of Convergence of Constrained Maximum Sharpe Ratio Estimators

First, when $1_p'\Sigma^{-1}\mu \geq 0$, we have the square of the maximum Sharpe ratio as in (14). To obtain an estimate by using nodewise regression, we replace Σ^{-1} with $\hat{\Theta}$. Namely, the estimate of the square of the maximum Sharpe ratio is:

$$\widehat{MSR}^2 = \hat{\mu}'\hat{\Theta}\hat{\mu}. \tag{16}$$

Using the result in Theorem 1, we can obtain the consistency of the maximum Sharpe ratio (squared).

Theorem 4. *Under Assumptions 1-4 with* $1_p' \Sigma^{-1} \mu \geq 0$,

$$\left| \frac{\widehat{MSR}^2}{MSR^2} - 1 \right| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1).$$

Remark. To the best of our knowledge, no existing result deals with MSR when $p > n$ and p can grow exponentially in n. We also allow for time-series data and establish a rate of convergence. The rate shows that precision matrix non-sparsity can badly affect the estimation error. The number of assets, on the other hand, can also increase the error by on a logarithmic scale.

Note that the maximum Sharpe ratio above relies on $1_p' \Sigma^{-1} \mu \geq 0$, where 1_p is a column vector of ones. This was recently pointed out in equation (6.1) Maller et al. (2016). If $1_p' \Sigma^{-1} \mu < 0$, the Sharpe ratio is minimized, as shown on p.503 of Maller and Turkington (2002). The new maximal Sharpe ratio in the case when $1_p' \Sigma^{-1} \mu < 0$ is in Theorem 2.1 of Maller and Turkington (2002). The square of the maximum Sharpe ratio when $1_p' \Sigma^{-1} \mu < 0$ is given in (15).

An estimator in this case is

$$\widehat{MSR}_c^2 = \hat{\mu}' \hat{\Theta} \hat{\mu} - (1_p' \hat{\Theta} \hat{\mu})^2 / (1_p' \hat{\Theta} 1_p). \tag{17}$$

The optimal portfolio allocation for such a case is given in (2.10) of Maller and Turkington (2002). The limit for such estimators when the number of assets is fixed (p fixed) is given in Theorems 3.1b-c of Maller et al. (2016).

We set up some notation for the next theorem. Set $1_p' \Sigma^{-1} 1_p / p = A$, $1_p' \Sigma^{-1} \mu / p = B$, $\mu' \Sigma^{-1} \mu / p = D$.

Theorem 5. *If* $1_p' \Sigma^{-1} \mu < 0$, *and under Assumptions 1-4 with* $AD - B^2 \geq C_1 > 0$, *where* C_1 *is a positive constant,*

$$\left| \frac{\widehat{MSR}_c^2}{MSR_c^2} - 1 \right| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1).$$

Remarks. 1. Condition $AD - B^2 \geq C_1 > 0$ is not restrictive, and it is used in Callot et al. (2019) as a condition that helps us to obtain a finite optimal portfolio variance in the Markowitz (1952) mean-variance portfolio below.

2. In Theorem 4, we allow $p > n$, and time-series data are allowed, unlike the iid or normal return cases in the literature when dealing with large p, n. Theorem 5 is new and will help us establish a new MSR result in the following Theorem.

We provide an estimate that takes into account uncertainties about the term $1_p'\Sigma^{-1}\mu$. Note that the term can be consistently estimated, as shown in Lemma A.3 in the Supplementary Appendix. A practical estimate for a maximum Sharpe ratio that will be consistent is:

$$\widehat{MSR}^* = \widehat{MSR} 1_{\{1_p'\hat{\Theta}\hat{\mu}>0\}} + \widehat{MSR}_c 1_{\{1_p'\hat{\Theta}\hat{\mu}<0\}},$$

where we excluded the case of $1_p'\hat{\Theta}\hat{\mu} = 0$ in the estimator. That specific scenario is very restrictive in terms of returns and variance. Note that under a mild assumption on the term, we show that by (A.44)(A.45)(A.48)(A.49), when $1_p'\Sigma^{-1}\mu > 0$, we have $1_p'\hat{\Theta}\hat{\mu} > 0$, and when $1_p'\Sigma^{-1}\mu < 0$, we have $1_p'\hat{\Theta}\hat{\mu} < 0$ with probability approaching one.

Theorem 6. *Under Assumptions 1-4 with $AD - B^2 \geq C_1 > 0$, where C_1 is a positive constant, and assuming $|1_p'\Sigma^{-1}\mu|/p \geq C > 2\epsilon > 0$, with a sufficiently small positive $\epsilon > 0$, and C being a positive constant,*

$$\left| \frac{(\widehat{MSR}^*)^2}{(MSR^*)^2} - 1 \right| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1).$$

Remarks 1. Condition $|1_p'\Sigma^{-1}\mu|/p \geq C > 2\epsilon > 0$ shows that apart from a small region around 0, we include all cases. This is similar to the $\beta - min$ condition in high-dimensional statistics used to achieve model selection. Note further that since $\Theta = \Sigma^{-1}$,

$$|1_p'\Theta\mu/p| = |\sum_{j=1}^{p}\sum_{k=1}^{p}\Theta_{j,k}\mu_k/p|,$$

which is a sum measure of roughly theoretical mean divided by standard deviations. It is difficult to see how this double sum in p will be a small number, unless the terms in the sum cancel out one another. Therefore, we exclude that type of case with our assumption. Additionally, ϵ is not arbitrary, from the proof this is the upper bound on the $|\hat{B} - B|$ in Lemma A.3 in Supplementary Appendix, and it is of order

$$\epsilon = O(\bar{s}\sqrt{\frac{lnp}{n}}) = o(1),$$

where the asymptotically small term follows Assumption 4.

2. In the case of $p > n$, we only consider consistency since standard central limit theorems (apart from those in rectangles or sparse convex sets) do not apply, and ideas such as multiplier bootstrap and empirical bootstrap with self-normalized moderate deviation results do not extend to this specific Sharpe ratio formulation.

3. This is a new result under the assumption that all portfolio weights sum to one and the uncertainty about the term $1_p'\Sigma^{-1}\mu$. We allow $p > n$ and time-series data.

4. When the precision matrix is non-sparse, i.e., $\bar{s} = p$, we have the rate of convergence as $p\sqrt{lnp/n}$. To have the estimation error converge to zero, we need $p\sqrt{lnp} = o(n^{1/2})$. In the non-sparse precision matrix case, we clearly allow only $p << n$.

5 Commonly Used Portfolios with a Large Number of Assets

Here, we provide consistent estimates of the Sharpe ratio of the global minimum-variance and Markowitz mean-variance portfolios when $p > n$.

5.1 Global Minimum-Variance Portfolio

In this part, we analyze not the maximum Sharpe ratio under the constraints of portfolio weights adding up to one but the Sharpe ratio we can infer from the global minimum-variance portfolio. This is the portfolio in which weights are chosen to minimize the variance of the portfolio subject to the weights summing to one. Specifically,

$$w_u = argmin_{w \in R^p} w'\Sigma w, \quad \text{such that} \quad w'1_p = 1.$$

In the main, this is similar to the maximum Sharpe ratio problem, but we minimize the square of the denominator in the Sharpe ratio definition subject to the same constraint in the maximum Sharpe ratio case above. The solution to the above problem is well known and is given by

$$w_u = \frac{\Sigma^{-1}1_p}{1_p'\Sigma^{-1}1_p}.$$

Next, substitute these weights into the Sharpe ratio formula, normalized by the number of assets

$$SR = \frac{w_u'\mu}{\sqrt{w_u'\Sigma w_u}} = \sqrt{p}(\frac{1_p'\Sigma^{-1}\mu}{p})(\frac{1_p'\Sigma^{-1}1_p}{p})^{-1/2}. \tag{18}$$

We estimate (18) by nodewise regression

$$\widehat{SR}_{nw} = \sqrt{p}(\frac{1_p'\hat{\Theta}\hat{\mu}}{p})(\frac{1_p'\hat{\Theta}1_p}{p})^{-1/2}. \tag{19}$$

To the best of our knowledge, the following theorem is a novel result in the literature when $p > n$ and establishes both consistency and rate of convergence in the case of the Sharpe ratio in the global minimum-variance portfolio.

Theorem 7. *Under Assumptions 1-4 with* $|1_p'\Sigma^{-1}\mu|/p \geq C > 2\epsilon > 0$,

$$\left|\frac{\widehat{SR}_{nw}^2}{SR^2} - 1\right| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1).$$

Remark. We see that a large p only affects the error by a logarithmic factor. The estimation error increases with the non-sparsity of the precision matrix.

5.2 Markowitz Mean-Variance Portfolio

Markowitz (1952) portfolio selection is defined as finding the smallest variance given a desired expected return ρ_1. The decision problem is

$$w_{MV} = argmin_{w \in R^p}(w'\Sigma w) \quad \text{such that} \quad w'1_p = 1, \quad w'\mu = \rho_1.$$

The formula for optimal weight is

$$\begin{aligned} w_{MV} &= \frac{(\mu'\Sigma^{-1}\mu) - \rho_1(1_p'\Sigma^{-1}\mu)}{(1_p'\Sigma^{-1}1_p)(\mu'\Sigma^{-1}\mu) - (1_p'\Sigma^{-1}\mu)^2}(\Sigma^{-1}1_p) \\ &+ \frac{\rho_1(1_p'\Sigma^{-1}1_p) - (1_p'\Sigma^{-1}\mu)}{(1_p'\Sigma^{-1}1_p)(\mu'\Sigma^{-1}\mu) - (1_p'\Sigma^{-1}\mu)^2}(\Sigma^{-1}\mu), \end{aligned}$$

which can be rewritten as

$$w_{MV} = \left[\frac{D - \rho_1 B}{AD - B^2}\right](\Sigma^{-1}1_p/p) + \left[\frac{\rho_1 A - B}{AD - B^2}\right](\Sigma^{-1}\mu/p), \tag{20}$$

where we use A, B, D formulas $A := 1_p'\Sigma^{-1}1_p/p, B := 1_p'\Sigma^{-1}\mu/p, D := \mu'\Sigma^{-1}\mu/p$. We define the estimators of these terms as $\hat{A} := 1_p'\hat{\Theta}1_p/p, \hat{B} := 1_p'\hat{\Theta}\hat{\mu}/p, \hat{D} := \hat{\mu}'\hat{\Theta}\hat{\mu}/p$.

The optimal variance of the portfolio in this scenario is normalized by the number of assets

$$V = \frac{1}{p}\left[\frac{A\rho_1^2 - 2B\rho_1 + D}{AD - B^2}\right].$$

The estimate of that variance is

$$\hat{V} = \frac{1}{p}\left[\frac{\hat{A}\rho_1^2 - 2\hat{B}\rho_1 + \hat{D}}{\hat{A}\hat{D} - \hat{B}^2}\right].$$

By our constraint, we obtain

$$w_{MV}'\mu = \rho_1.$$

Using the variance V above

$$SR_{MV} = \rho_1\sqrt{p\left(\frac{AD - B^2}{A\rho_1^2 - 2B\rho_1 + D}\right)}.$$

The estimate of the Sharpe ratio under the Markowitz mean-variance portfolio is

$$\widehat{SR}_{MV} = \rho_1 \sqrt{p\left(\frac{\hat{A}\hat{D} - \hat{B}^2}{\hat{A}\rho_1^2 - 2\hat{B}\rho_1 + \hat{D}}\right)}.$$

We provide the consistency of the maximum Sharpe ratio (squared) in this framework when the number of assets is larger than the sample size. This is a novel result in the literature.

Theorem 8. *Under Assumptions 1-4 with condition* $|1_p'\Sigma^{-1}\mu/p| \geq C > 2\epsilon > 0$ *and* $AD - B^2 \geq C_1 > 0$*,* $A\rho_1^2 - 2B\rho_1 + D \geq C_1 > 0$*, with* ρ_1 *uniformly bounded away from zero and infinity, we have*

$$\left| \frac{\widehat{SR}_{MV}^2}{SR_{MV}^2} - 1 \right| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1).$$

Remarks. 1. Conditions $AD - B^2 \geq C_1 > 0$ show that the variance is bounded away from infinity, and $A\rho_1^2 - 2B\rho_1 - D \geq C_1 > 0$ restricts the variance to be positive and bounded away from zero.

2. We provide the rate of convergence of the estimators, which increases with p in a logarithmic way, and the non-sparsity of the precision matrix affects the error in a linear way.

6 Simulations

6.1 Models and Implementation Details

In this section, we compare the nodewise regression with several models in a simulation exercise. The two aims of the exercise are to determine whether our method achieves consistency under a sparse setup and to check under two different setups how our method performs compared to others in the estimation of the constrained maximum Sharpe ratio, out-of-sample maximum Sharpe ratio, and Sharpe ratio in global minimum-variance and Markowitz mean-variance portfolios.

The other methods that are used widely in the literature and benefit from high-dimensional techniques are the principal orthogonal complement thresholding (POET) from Fan et al. (2013), the nonlinear shrinkage (NL-LW) and the single factor nonlinear shrinkage (SF-NL-LW) from Ledoit and Wolf (2017) and the maximum Sharpe ratio estimation and sparse regression (MAXSER) from Ao et al. (2019). All models except for the MAXSER are plug-in estimators, where the first step is to estimate the precision/covariance matrix, and the second step is to plug-in the estimate in the desired equation.

The POET uses principal components to estimate the covariance matrix and allows some eigenvalues of Σ to be spiked and grow at a rate $O(p)$, which allows common and idiosyncratic components to be identified and principal components analysis and can consistently estimate the space spanned by the eigenvectors of Σ. However, Fan et al. (2013) point out that the absolute convergence rate of the model is not satisfactory for estimating Σ and consistency can only be achieved in terms of the relative error matrix.

Nonlinear shrinkage is a method that individually determines the amount of shrinkage of each eigenvalue in the covariance matrix with respect to a particular loss function. The main aim is to increase the value of the lowest eigenvalues and decrease the largest eigenvalues to stabilize the high-dimensional covariance matrix. This is a very novel and excellent idea. Ledoit and Wolf (2017) propose a function that captures the objective of an investor using portfolio selection. As a result, they have an optimal estimator of the covariance matrix for portfolio selection for a large number of assets. The SF-NL-LW method extracts a single factor structure from the data prior to the estimation of the covariance matrix, which is simply an equal-weighted portfolio with all assets.

Finally, the MAXSER starts with the estimation of the adjusted squared maximum Sharpe ratio that is used in a penalized regression to obtain the portfolio weights. Of all the discussed models, the MAXSER is the only one that does not use an estimate of the precision matrix in a plug-in estimator of the maximum Sharpe ratio.

Regarding implementation, the POET and both models from Ledoit and Wolf (2017) are available in the R packages POET Fan et al. (2016) and nlshrink Ramprasad (2016). The SF-NL-LW needed some minor adjustments following the procedures described in Ledoit and Wolf (2017). For the MAXSER, we followed the steps for the non-factor case in Ao et al. (2019), and we used the package lars (Hastie and Efron, 2013) for the penalized regression estimation. We estimated the nodewise regression following the steps in Section 2.4 using the glmnet package Friedman et al. (2010) for penalized regressions. We used two alternatives to select the regularization parameter λ, a 10-fold cross validation (CV) and the generalized information criterion (GIC) from Zhang et al. (2010).

The GIC procedure starts by fitting $\hat{\gamma}_j$ in subsection 2.4 for a range of λ_j that goes from the intercept-only model to the largest feasible model. This is automatically done by the glmnet package. Then, for the GIC procedure, we calculate the information criterion for a given λ_j among the ranges of all possible tuning parameters

$$GIC_j(\lambda_j) = \frac{SSR(\lambda_j)}{n} + q(\lambda_j)\log(p-1)\frac{\log(\log(n))}{n} \tag{21}$$

where $SSR(\lambda_j)$ is the sum squared error for a given λ_j, $q(\lambda_j)$ is the number of variables, given λ_j, in the model that is nonzero, and p is the number of assets. The last step is to select the model with the smallest GIC. Once this is done for all assets $j = 1, \ldots, p$, we can proceed to obtain $\hat{\Theta}_{GIC}$.

For the CV procedure, we split the sample into k subsamples and fit the model for a range of λ_j as in the GIC procedure. However, we will fit models in the subsamples. We always estimate the models in $k-1$ subsamples, leaving one subsample as a test sample, where we compute the mean squared error (MSE). After repeating the procedure using all k subsamples as a test, we finally compute the average MSE across all subsamples and select the λ_j for each asset j that yields the smallest average MSE. We can then use the estimated $\hat{\gamma}_j$ to obtain $\hat{\Theta}_{CV}$.

6.2 Data Generation Process and Results

We used two DGPs to test the nodewise regression. The first DGP consists of a Toeplitz covariance matrix of excess asset returns, where

$$\Sigma_{i,j} = \rho^{|i-j|},$$

with values ρ equal to 0.25, 0.5 and 0.75 and the vector μ sampled from a normal distribution $N(0.5, 1)$.

The second DGP is based on a simplified version of the factor DGP in Ao et al. (2019):

$$r_j = \alpha_j + \sum_{k=1}^{K} \beta_{j,k} f_k + e_j, \tag{22}$$

where f_j are the factor returns, $\beta_{j,k}$ are the individual stock sensitivities to the factors, and $\alpha_j + e_j$ represent the idiosyncratic component of each stock. We adopted the Fama & French three factors[1] (FF3) monthly returns as factors with $\boldsymbol{\mu}_f$ and $\boldsymbol{\Sigma}_f$ being the factors' sample mean and covariance matrix. The βs, αs and $\hat{\Sigma}_e$ were estimated using a simple least squares regression using returns from the S&P500 stocks that were part of the index in the entire period from 2008 to 2017. In each simulation, we randomly selected p stocks from the pool with replacement because our simulations require more than the total number of available stocks. We then used the selected stocks to generate individual returns with $\boldsymbol{\Sigma}_e = \gamma diag(e_j)$, where gamma is assumed to be 1, 2 and 4.

[1]The factors are book-to-market, market capitalization and the excess return of the market portfolio.

Tables 1 and 2 show the results. The values in each cell show the average absolute estimation error for estimating the square of the Sharpe ratio in the case of global minimum-variance and Markowitz mean-variance portfolios in Section 5, out-of-sample forecasting, and the maximum Sharpe ratio in the case of constrained portfolio optimization in Sections 3-4 across iterations. Each eight column block in the table shows the results for a different sample size. In each of these blocks, the first four columns are for $p = 0.5 * n$, and the last four columns are for $p = 1.5 * n$. MSR, MSR-OOS, GMV-SR and MKW-SR are the constrained maximum Sharpe ratio, the out-of-sample maximum Sharpe ratio, the Sharpe ratio from the global minimum-variance portfolio and the Sharpe ratio from the Markowitz portfolio with target returns set to 1%, respectively. Therefore, there are four categories to evaluate the different estimates. The MAXSER risk constraint was set to 0.04 following Ao et al. (2019). We ran 100 iterations in each simulation setup. All bold-face entries in tables show category champions.

Starting with Table 1, we clearly see that our method performs very well in a sparse Toeplitz scenario. When the correlation is 0.5 or 0.75, our method has the smallest error of all those tested for MSR and MSR-OOS. We also see that with the GMV-SR and MKW-SR scenarios, the SF-NL-LW method generally performs the best. To give a specific example, with $n = 400, p = 600$ and $\rho = 0.75$, our OOS-MSR error is 0.118 (GIC based nodewise), the second-best is our CV-based nodewise with 0.259, and the third is SF-NL-LW with a 0.868 error. On the other hand, in the GMV-SR category, the best is SF-NL-LW with 0.551 error, whereas our best method is GIC nodewise with 0.664 error as third among all methods.

Additionally, we see that consistency is achieved with our methods, as our theorems suggest under sparse scenarios, as in Table 1. To see this, with $n = 100, p = 150$, our error in the OOS-MSR category is 0.336 (GIC, nodewise) and declines to 0.118 at $n = 400, p = 600$ at $\rho = 0.75$. Similar results exist in all other categories for our method in Table 1.

Table 2 paints a different picture under a factor model scenario; both NL-LW and SF-NL-LW perform the best in minimizing the errors for the constrained Markowitz-Sharpe ratio and global minimum-variance and Markowitz mean-variance portfolios. We also note that the MAXSER generally obtains the best results in estimating the out-of-sample maximum Sharpe ratio when $p = n/2$.

Table 1: Simulation Results – Toeplitz DGP

	n=100								n=200								n=400							
	p=n/2				p=1.5n				p=n/2				p=1.5n				p=n/2				p=1.5n			
	MSR	OOS-MSR	GVM-SR	MKW-SR	MSR	OOS-MSR	GVM-SR	MKW-SR	MSR	OOS-MSR	GVM-SR	MKW-SR	MSR	OOS-MSR	GVM-SR	MKW-SR	MSR	OOS-MSR	GVM-SR	MKW-SR	MSR	OOS-MSR	GVM-SR	MKW-SR
Toeplitz DGP $\rho = 0.25$																								
NW-GIC	0.072	0.158	0.402	0.330	**0.041**	0.179	0.405	0.343	**0.042**	0.129	0.371	0.310	**0.027**	**0.150**	0.388	0.330	**0.021**	**0.074**	0.301	0.258	**0.014**	**0.094**	0.329	0.282
NW-CV	0.073	0.171	0.416	0.342	0.043	0.186	0.411	0.349	0.045	0.152	0.401	0.334	0.028	0.161	0.401	0.341	0.025	0.119	0.366	0.313	0.016	0.131	0.377	0.321
POET	0.087	0.190	0.402	0.348	0.046	0.193	0.423	0.361	0.049	0.180	0.415	0.347	0.029	0.179	0.414	0.352	0.030	0.170	0.406	0.344	0.022	0.171	0.404	0.344
NL-LW	**0.065**	**0.129**	**0.317**	0.273	0.048	**0.155**	0.368	0.308	0.042	**0.127**	0.333	0.277	0.032	0.152	0.373	0.316	0.033	0.123	0.333	0.283	0.031	0.150	0.375	0.316
SF-NL-LW	0.085	0.153	0.366	**0.141**	0.072	0.181	**0.198**	**0.068**	0.094	0.136	**0.265**	**0.144**	0.078	0.167	**0.143**	**0.053**	0.098	0.124	**0.197**	**0.130**	0.092	0.156	**0.093**	**0.034**
MAXSER		0.149								0.267								0.363						
Toeplitz DGP $\rho = 0.50$																								
NW-GIC	**0.094**	**0.206**	0.695	0.466	0.133	**0.293**	0.557	0.516	0.096	**0.133**	0.468	0.427	0.115	**0.168**	0.494	0.460	0.091	**0.078**	0.407	0.379	0.104	**0.094**	0.431	0.402
NW-CV	0.100	0.344	1.049	0.552	**0.113**	0.405	0.621	0.575	**0.077**	0.234	0.573	0.523	**0.095**	0.271	0.576	0.539	**0.070**	0.155	0.521	0.489	**0.085**	0.176	0.535	0.500
POET	0.302	0.847	0.706	0.540	0.334	0.959	0.667	0.603	0.340	0.930	0.646	0.580	0.346	0.969	0.669	0.605	0.374	0.939	0.662	0.600	0.388	0.960	0.665	0.604
NL-LW	0.177	0.306	**0.423**	0.372	0.304	0.577	0.544	0.502	0.189	0.296	0.398	0.370	0.292	0.579	0.549	0.506	0.182	0.288	0.401	0.380	0.305	0.572	0.556	0.507
SF-NL-LW	0.197	0.310	0.698	**0.195**	0.323	0.562	**0.371**	**0.347**	0.214	0.288	**0.344**	**0.189**	0.332	0.546	**0.363**	**0.342**	0.207	0.276	**0.256**	**0.205**	0.350	0.530	**0.375**	**0.338**
MAXSER		0.251								0.405								0.518						
Toeplitz DGP $\rho = 0.75$																								
NW-GIC	0.267	**0.233**	**0.730**	0.669	0.371	**0.336**	0.726	0.726	0.263	**0.159**	**0.655**	0.655	0.311	**0.211**	0.705	0.694	0.226	**0.097**	0.636	0.625	0.254	**0.118**	0.664	0.651
NW-CV	**0.204**	0.484	0.767	0.775	**0.272**	0.553	0.806	0.796	**0.197**	0.349	0.787	0.764	**0.243**	0.397	0.788	0.777	**0.173**	0.232	0.753	0.746	**0.206**	0.259	0.766	0.754
POET	1.564	3.320	1.263	0.658	1.725	4.709	0.833	0.800	1.761	4.253	0.772	0.736	1.755	5.131	0.848	0.815	1.835	4.792	0.823	0.798	1.860	5.196	0.852	0.818
NL-LW	0.283	0.341	1.142	0.374	0.688	0.898	**0.563**	0.608	0.292	0.324	650.076	0.375	0.680	0.884	0.582	0.604	0.283	0.320	**0.556**	0.391	0.685	0.882	0.614	0.606
SF-NL-LW	0.293	0.354	5.172	**0.322**	0.689	0.912	0.568	**0.553**	0.301	0.329	1.143	**0.302**	0.685	0.878	**0.503**	**0.543**	0.286	0.321	1.527	**0.315**	0.691	0.868	**0.551**	**0.543**
MAXSER		0.392								0.574								0.706						

The table shows the simulation results for the Toeplitz DGP. Each simulation was done with 100 iterations. We used sample sizes n of 100, 200 and 400, and the number of stocks was either $n/2$ or $1.5n$ for the low-dimensional and the high-dimensional case, respectively. Each block of rows shows the results for a different value of ρ in the Toeplitz DGP. The values in each cell show the average absolute estimation error for estimating the square of the Sharpe ratio in the case of the global minimum-variance and Markowitz mean-variance portfolios in Section 5 and maximum Sharpe ratio in the case of out-of-sample forecasting and constrained portfolio optimization in Sections 3-4 across iterations.

Table 2: Simulation Results – Factor DGP

	n=100								n=200								n=400							
	p=n/2				p=1.5n				p=n/2				p=1.5n				p=n/2				p=1.5n			
	MSR	OOS-MSR	GVM-SR	MKW-SR	MSR	OOS-MSR	GVM-SR	MKW-SR	MSR	OOS-MSR	GVM-SR	MKW-SR	MSR	OOS-MSR	GVM-SR	MKW-SR	MSR	OOS-MSR	GVM-SR	MKW-SR	MSR	OOS-MSR	GVM-SR	MKW-SR
Factor DGP $\gamma = 1$																								
NW-GIC	0.654	2.835	4.010	0.695	0.675	9.743	0.843	0.769	0.618	5.498	0.783	0.760	0.649	19.800	0.825	0.799	0.609	10.105	0.810	0.787	0.618	36.215	0.831	0.813
NW-CV	0.671	3.129	1.423	0.710	0.705	8.218	0.787	0.786	0.637	4.894	0.783	0.768	0.669	14.725	0.819	0.811	0.618	7.584	0.801	0.789	0.630	23.422	0.817	0.818
POET	0.554	1.140	0.968	0.300	0.586	1.651	0.458	0.384	0.413	0.857	0.367	0.339	0.440	2.095	0.451	0.404	0.332	1.289	0.427	0.363	0.345	3.714	0.474	0.403
NL-LW	**0.509**	1.736	0.900	**0.112**	**0.532**	2.264	0.486	**0.092**	**0.342**	0.984	0.295	**0.086**	**0.354**	1.375	0.266	**0.129**	**0.216**	0.645	0.205	0.098	**0.217**	0.954	0.217	0.179
SF-NL-LW	0.537	1.169	**0.722**	0.192	0.564	**1.299**	**0.344**	0.231	0.361	0.582	**0.234**	0.145	0.378	**0.682**	**0.194**	0.151	0.234	0.321	**0.151**	**0.082**	0.245	**0.360**	**0.116**	**0.087**
MAXSER		**0.372**								**0.162**								**0.083**						
Factor DGP $\gamma = 2$																								
NW-GIC	0.806	1.554	3.022	0.811	0.827	4.940	0.866	0.868	0.794	2.712	0.864	0.858	0.818	9.939	0.900	0.890	0.795	4.999	0.890	0.881	0.805	18.126	0.910	0.901
NW-CV	0.815	1.788	1.247	0.821	0.844	4.153	0.847	0.877	0.804	2.396	0.865	0.863	0.829	7.330	0.897	0.896	0.799	3.692	0.885	0.881	0.812	11.623	0.902	0.904
POET	0.750	1.084	0.924	0.568	0.780	1.203	0.635	0.647	0.683	0.607	0.603	0.609	0.710	1.178	0.690	0.674	0.649	0.691	0.666	0.643	0.666	1.877	0.720	0.685
NL-LW	**0.724**	1.863	0.720	**0.348**	**0.751**	2.201	**0.440**	**0.386**	**0.645**	1.023	**0.365**	**0.373**	**0.665**	1.334	**0.380**	**0.384**	**0.588**	0.659	**0.387**	**0.389**	**0.601**	0.923	**0.358**	**0.380**
SF-NL-LW	0.740	1.207	**0.696**	0.500	0.768	**1.183**	0.528	0.560	0.655	0.581	0.449	0.494	0.678	**0.613**	0.523	0.535	0.597	**0.316**	0.464	0.484	0.615	**0.317**	0.517	0.519
MAXSER		**0.645**								**0.527**								0.446						
Factor DGP $\gamma = 4$																								
NW-GIC	0.887	0.965	2.522	0.879	0.906	2.515	0.904	0.921	0.885	1.296	0.911	0.913	0.904	4.976	0.941	0.937	0.889	2.419	0.933	0.930	0.899	9.061	0.950	0.946
NW-CV	0.892	1.192	1.169	0.885	0.914	2.107	0.894	0.927	0.891	1.132	0.912	0.915	0.909	3.614	0.939	0.941	0.892	1.732	0.929	0.930	0.902	5.714	0.945	0.948
POET	0.854	1.248	0.943	0.722	0.880	**1.077**	0.763	0.789	0.824	**0.570**	0.741	0.759	0.846	0.764	0.817	0.813	0.810	0.441	0.795	0.790	0.827	0.980	0.843	0.828
NL-LW	**0.839**	2.223	**0.692**	**0.581**	**0.864**	2.360	**0.574**	**0.632**	**0.802**	1.189	**0.565**	**0.613**	**0.823**	1.408	**0.619**	**0.647**	**0.777**	0.756	**0.609**	**0.640**	**0.793**	0.957	**0.641**	**0.661**
SF-NL-LW	0.848	1.446	0.752	0.679	0.873	1.249	0.694	0.736	0.808	0.693	0.633	0.688	0.829	**0.641**	0.717	0.734	0.782	**0.383**	0.668	0.697	0.801	**0.328**	0.730	0.737
MAXSER		**0.794**								0.735								0.700						

The table shows the simulation results for the factor DGP. Each simulation was done with 100 iterations. We used sample sizes n of 100, 200 and 400, and the number of stocks was either $n/2$ or $1.5n$ for the low-dimensional and the high-dimensional case, respectively. Each block of rows shows the results for a different value of γ in the factor DGP. The values in each cell show the average absolute estimation error for estimating the square of the Sharpe ratio in the case of global minimum-variance and Markowitz mean-variance portfolios in Section 5 and maximum Sharpe ratio in out-of-sample forecasting and the case of constrained portfolio optimization in Sections 3-4 across iterations.

7 Empirical Application

For the empirical application, we use two subsamples. The first subsample uses all data from January 1995 to December 2017 with an out-of-sample period from January 2005 to December 2017. We selected all stocks that were in the S&P 500 index for at least one month in the out-of-sample period and have data for the entire 1995-2017 period, which resulted in 383 stocks. The second subsample starts in January 1990 and ends in December 2017 with an out-of-sample period from January 2000 to December 2017. Using the same criterion as the first subsample, the number of stocks was 323. Given that this is an out-of-sample competition between models, we only estimated GMV and Markowitz portfolios for the plug-in estimators. The first out-of-sample period includes only the recession of 2008. The second out-of-sample period includes the recessions of 2000 and 2008, and the out-of-sample periods reflect recent history.

The Markowitz return constraint ρ_1 is 0.8% per month, and the MAXSER risk constraint is 4%. In the low-dimensional experiment, we randomly select 50 stocks from the pool to estimate the models. In the high-dimensional case, we use all stocks.

We use a rolling window setup for the out-of-sample estimation of the Sharpe ratio following Callot et al. (2019). Specifically, samples of size n are divided into in-sample $(1 : n_I)$ and out-of-sample $(n_I + 1 : n)$. We start by estimating the portfolio $\hat{w}_{nI}$ in the in-sample period and the out-of-sample portfolio returns $\hat{w}'_{nI} r_{n_I+1}$. Then, we roll the window by one element $(2 : n_I + 1)$ and form a new in-sample portfolio $\hat{w}_{n_I+1}$ and out-of-sample portfolio returns $\hat{w}_{n_I+1} r_{n_I+2}$. This procedure is repeated until the end of the sample.

The out-of-sample average return and variance without transaction costs are

$$\hat{\mu}_{os} = \frac{1}{n - n_I} \sum_{t=n_I}^{n-1} \hat{w}'_t r_{t+1}, \qquad \hat{\sigma}^2_{os} = \frac{1}{n - n_I - 1} \sum_{t=n_I}^{n-1} (\hat{w}'_t r_{t+1} - \hat{\mu}_{os})^2.$$

We estimate the Sharpe ratios with and without transaction costs. The transaction cost, c, is defined as 50 basis points following DeMiguel et al. (2007). Let $r_{P,t+1} = \hat{w}'_t r_{t+1}$ be the return of the portfolio in period $t + 1$; in the presence of transaction costs, the returns will be defined as

$$r^{Net}_{P,t+1} = r_{P,t+1} - c(1 + r_{P,t+1}) \sum_{j=1}^{p} |\hat{w}_{t+1,j} - \hat{w}^+_{t,j}|,$$

where $\hat{w}^+_{t,j} = \hat{w}_{t,j}(1 + R_{t+1,j})/(1 + R_{t+1,P})$ and $R_{t,j}$ and $R_{t,P}$ are the excess returns of asset j and the portfolio P added to the risk-free rate. The adjustment made in $\hat{w}^+_{t,j}$ is because the portfolio at the end of the period has changed compared to the portfolio at the beginning of the period.

The Sharpe ratio is calculated from the average return and the variance of the portfolio in the out-of-sample period

$$SR = \frac{\hat{\mu}_{os}}{\hat{\sigma}_{os}}.$$

The portfolio returns are replaced by the returns with transaction costs when we calculate the Sharpe ratio with transaction costs.

We use the same test as Ao et al. (2019) to compare the models. Specifically,

$$H_0 \ : \ SR_{Best} \leq SR_0 \ \ vs \ \ H_a \ : \ SR_{Best} > SR_0, \tag{23}$$

where SR_{Best} is the model with the largest Sharpe ratio, which is tested against all remaining models. This is the Jobson and Korkie (1981) test with Memmel (2003) correction. We also considered the method of Ledoit and Wolf (2008) for testing the significance of the winner and using the equally weighted portfolio as a benchmark; the results were very similar and hence are not reported.

We also included equally weighted portfolio (EW). GMV-NW-GIC and GMV-NW-CV denote the nodewise method with GIC and cross validation tuning parameter choices, respectively, in the global minimum-variance portfolio (GMV). GMV-POET, GMV-NL-LW, and GMV-SF-NL-LW denote the POET, nonlinear shrinkage, and single factor nonlinear shrinkage methods, respectively, which are described in the simulation section and also used in the global minimum-variance portfolio. The MAXSER is also used and explained in the simulation section. MW denotes the Markowitz mean-variance portfolio, and MW-NW-GIC denotes the nodewise method with GIC tuning parameter selection in the Markowitz portfolio. All the other methods with MW headers are analogous and thus self-explanatory.

The results are presented in Tables 3 and 4. Table 3 shows the results for the 2005-2017 out-of-sample period. Nodewise methods dominate in terms of the Sharpe ratio in Table 3. For example, with transaction costs in the high-dimensional portfolio category, in terms of Sharpe ratio (SR) (averaged over the out-of-sample time period), GMV-NW-GIC is the best model. It has an SR of 0.208. GMV-POET, GMV-NL-LW, and GMV-SF-NL-LW have SRs of 0.175, 0.144, and 0.140, respectively. If we were to analyze only the Markowitz portfolio in Table 3, with transaction costs in high dimensions, MW-NW-GIC has the highest SR of 0.205. Therefore, even in subcategories, the nodewise method dominates. Although statistical significance is not established, it is not clear that these significance tests have high power in our high-dimensional cases.

Table 4 shows the results for the out-of-sample January 2000-2017 subsample. We see that nodewise methods dominate all scenarios except for the low-dimensional case with no transaction costs. In the case of high dimensionality with transaction costs, MW-NW-GIC (Markowitz-nodewise-GIC) has an SR of 0.224, and the closest is EW with 0.207.

Table 3: Empirical Results – Out-of-Sample Period from Jan. 2005 to Dec. 2017

	Without TC								With TC							
	Low Dim.				High Dim.				Low Dim.				High Dim			
Portfolio	SR	Avg.	SD	p-value	SR	Avg.	SD	p-value	SR	Avg.	SD	p-value	SR	Avg.	SD	p-value
EW	0.221	0.010	0.047	0.074	0.200	0.010	0.049	0.283	0.215	0.010	0.047	0.178	0.194	0.009	0.049	0.520
GMV-NW-GIC	0.260	0.009	0.036	0.200	0.215	0.009	0.040	0.215	0.247	0.009	0.036	0.454	**0.208**	0.008	0.040	
GMV-NW-CV	0.256	0.010	0.038	0.219	0.219	0.009	0.039	0.482	0.242	0.009	0.038	0.387	0.203	0.008	0.039	0.098
GMV-POET	0.256	0.008	0.032	0.648	0.192	0.006	0.031	0.652	0.242	0.008	0.032	0.756	0.175	0.005	0.031	0.658
GMV-NL-LW	0.245	0.007	0.031	0.620	0.201	0.006	0.031	0.793	0.216	0.007	0.031	0.522	0.144	0.005	0.032	0.488
GMV-SF-NL-LW	0.251	0.008	0.030	0.685	0.198	0.006	0.031	0.777	0.225	0.007	0.030	0.613	0.140	0.004	0.031	0.475
MW-NW-GIC	**0.277**	0.010	0.035		0.217	0.008	0.038	0.020	**0.257**	0.009	0.035		0.205	0.008	0.038	0.760
MW-NW-CV	0.271	0.010	0.036	0.556	**0.225**	0.008	0.037		0.251	0.009	0.036	0.547	0.205	0.008	0.037	0.724
MW-POET	0.270	0.008	0.030	0.881	0.198	0.006	0.030	0.726	0.250	0.008	0.030	0.882	0.177	0.005	0.031	0.702
MW-NL-LW	0.252	0.008	0.031	0.705	0.205	0.006	0.031	0.833	0.220	0.007	0.031	0.569	0.148	0.005	0.031	0.518
MW-SF-NL-LW	0.259	0.008	0.030	0.781	0.202	0.006	0.030	0.811	0.229	0.007	0.030	0.666	0.143	0.004	0.031	0.506
MAXSER	0.045	0.003	0.060	0.054					-0.082	-0.005	0.061	0.003				

The table shows the Sharpe ratio (SR), average returns (Avg), standard deviation (SD) and p-value of the Jobson and Korkie (1981) test with Memmel (2003) correction. We also applied the Ledoit and Wolf (2008) test with circular bootstrap, and the results were very similar; therefore we only report those of the first test. The tests were always performed using the equal-weighted portfolio as benchmark. The statistics were calculated from 156 rolling windows covering the period from Jan. 2005 to Dec. 2017, and the size of the estimation window was 120 observations.

In Table 5, we analyze turnover, leverage and maximum leverage (equations (24), (25) and (26), respectively) of the portfolios in Tables 3-4.

The definitions are as follows for turnover:

$$\text{turnover} = \sum_{j=1}^{p} |\hat{w}_{t+1,j} - \hat{w}_{t,j}^{+}|, \tag{24}$$

and leverage

$$\text{leverage} = \left| \sum_{j=1}^{p} \min\{\hat{w}_{t+1,j}, 0\} \right|, \tag{25}$$

and maximum leverage

$$\text{max leverage} = \max_{j} \{ \left| \min\{\hat{w}_{t+1,j}, 0\} \right| \}. \tag{26}$$

Our method performs very well compared to the others in terms of turnover, leverage and maximum leverage. Nodewise-based methods even come closer to the EW (equally weighted) portfolio. To provide some perspective, in Table 5, in high dimension from January 2005 to December 2017,

Table 4: Empirical Results – Out-of-Sample Period from Jan. 2000 to Dec. 2017

	Without TC								With TC							
	Low Dim.				High Dim.				Low Dim.				High Dim			
Portfolio	SR	Avg.	SD	p-value	SR	Avg.	SD	p-value	SR	Avg.	SD	p-value	SR	Avg.	SD	p-value
EW	0.203	0.010	0.052	0.430	0.214	0.010	0.047	0.274	0.197	0.010	0.052	0.178	0.207	0.010	0.047	0.405
GMV-NW-GIC	0.228	0.009	0.040	0.607	0.227	0.009	0.039	0.215	0.217	0.009	0.040	0.151	0.219	0.008	0.039	0.482
GMV-NW-CV	0.227	0.009	0.041	0.606	0.227	0.009	0.038	0.262	0.213	0.009	0.041	0.103	0.211	0.008	0.038	0.153
GMV-POET	0.201	0.007	0.035	0.337	0.191	0.006	0.033	0.513	0.187	0.006	0.035	0.344	0.174	0.006	0.033	0.462
GMV-NL-LW	0.248	0.008	0.033	0.708	0.201	0.006	0.030	0.617	0.218	0.007	0.033	0.742	0.146	0.004	0.030	0.265
GMV-SF-NL-LW	0.247	0.008	0.032	0.812	0.204	0.006	0.029	0.663	0.221	0.007	0.032	0.805	0.148	0.004	0.029	0.297
MW-NW-GIC	0.249	0.009	0.038	0.948	0.236	0.009	0.038	0.951	**0.234**	0.009	0.038		**0.224**	0.008	0.038	
MW-NW-CV	0.247	0.010	0.039	0.918	**0.236**	0.009	0.037		0.230	0.009	0.039	0.579	0.216	0.008	0.037	0.014
MW-POET	0.217	0.007	0.033	0.490	0.190	0.006	0.032	0.536	0.199	0.007	0.033	0.492	0.169	0.005	0.032	0.461
MW-NL-LW	**0.252**	0.008	0.033		0.207	0.006	0.030	0.678	0.220	0.007	0.033	0.756	0.151	0.005	0.030	0.299
MW-SF-NL-LW	0.247	0.008	0.032	0.793	0.207	0.006	0.029	0.707	0.220	0.007	0.032	0.771	0.150	0.004	0.029	0.329
MAXSER	0.078	0.007	0.083	0.142					-0.021	-0.002	0.085	0.024				

The table shows the Sharpe ratio (SR), average returns (Avg), standard deviation (SD) and p-value of the Jobson and Korkie (1981) test with Memmel (2003) correction. We also applied the Ledoit and Wolf (2008) test with circular bootstrap, and the results were very similar; therefore we only report those of the first test. The tests were always performed using the equal-weighted portfolio as benchmark. The statistics were calculated from 216 rolling windows covering the period from Jan. 2000 to Dec. 2017, and the size of the estimation window was 120 observations.

the nodewise GMV-NW-GIC has a turnover of 0.057, which is much smaller than those of GMV-POET, GMV-NL-LW, GMV-SF-NL-LW of 0.092, 0.328, and 0.323, respectively. Our leverage and maximum leverage are also very small compared to those of the other methods.

Note that to better understand why we perform well in the out-of-sample exercise, we show the correlation matrices for the two out-of-sample periods that we analyzed in figure 2. Subsample 1 corresponds to January 2005-December 2017, and Subsample 2 corresponds to January 2000-December 2017; in Figures 2a and 2b, we colored the correlation of assets. Blue (dark in black and white) is anything above a 0.3 positive correlation (which is the average), yellow is anything between a 0 and 0.3 positive correlation (light gray in black and white), and red (dark in black and white) represents the very few negative correlations. Figures 2a and 2b clearly show that dark blue areas do not predominate. This follows our assumptions, where large correlations between assets should not dominate the correlation matrix of assets.

8 Conclusion

We provided a nodewise regression method that can control for risk and obtain the maximum expected return of a large portfolio. Our result is novel and holds even when $p > n$. We also show that the maximum out-of-sample Sharpe ratio can be estimated consistently. Furthermore, we also develop a formula for the maximum Sharpe ratio when the weights of the portfolio sum to one. A

Table 5: Turnover and Leverage

	Low Dimension			High Dimension		
2005-2017 Subsample	Turnover	Leverage	Max Leverage	Turnover	Leverage	Max Leverage
EW	0.048	0.000	0.000	0.054	0.000	0.000
GMV-NW-GIC	0.087	0.009	0.005	0.057	0.001	0.001
GMV-NW-CV	0.100	0.001	0.001	0.122	0.003	0.002
GMV-POET	0.074	0.193	0.037	0.092	0.301	0.007
GMV-NL-LW	0.164	0.403	0.065	0.328	0.806	0.023
GMV-SF-NL-LW	0.146	0.369	0.051	0.323	0.872	0.026
MW-NW-GIC	0.132	0.032	0.012	0.079	0.006	0.001
MW-NW-CV	0.141	0.018	0.008	0.144	0.009	0.002
MW-POET	0.109	0.217	0.037	0.109	0.305	0.007
MW-NL-LW	0.184	0.425	0.068	0.330	0.809	0.023
MW-SF-NL-LW	0.166	0.392	0.053	0.325	0.874	0.025
MAXSER	1.529	0.313	0.153			
2000-2017 Subsample						
EW	0.060	0.000	0.000	0.057	0.000	0.000
GMV-NW-GIC	0.081	0.007	0.004	0.059	0.001	0.000
GMV-NW-CV	0.106	0.004	0.004	0.122	0.003	0.002
GMV-POET	0.088	0.216	0.042	0.105	0.326	0.008
GMV-NL-LW	0.188	0.403	0.073	0.308	0.784	0.028
GMV-SF-NL-LW	0.154	0.366	0.062	0.300	0.829	0.026
MW-NW-GIC	0.110	0.016	0.007	0.084	0.005	0.001
MW-NW-CV	0.130	0.010	0.006	0.147	0.010	0.003
MW-POET	0.105	0.221	0.042	0.122	0.334	0.009
MW-NL-LW	0.201	0.415	0.072	0.311	0.787	0.028
MW-SF-NL-LW	0.167	0.372	0.061	0.303	0.832	0.025
MAXSER	1.652	0.346	0.197			

The table shows the average turnover, average leverage and average max leverage for all portfolios across all out-of-sample windows. The top panel shows the results for the 2000-2017 out-of-sample period, and the second panel shows the results for the 2005-2017 out-of-sample period.

consistent estimate for the constrained case is also shown. Then, we extended our results to the consistent estimation of Sharpe ratios in two widely used portfolios in the literature. It will be important to extend our results to more restrictions on portfolios.

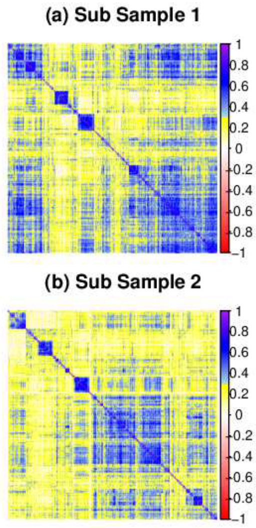

Figure 2: Data Correlation Matrices

Appendix

This appendix contains the proofs. The Supplementary Appendix contains additional proofs that are the building blocks of the proofs in and independent of the results in this appendix.

Proof of Theorem 2. (A.2) of Ao et al. (2019) shows that the squared ratio of the estimated maximum out-of-sample Sharpe ratio to the theoretical ratio can be written as

$$[\frac{\widehat{SR}_{mosnw}}{SR^*}]^2 = \frac{\frac{(\mu'\hat{\Theta}\hat{\mu})^2}{\hat{\mu}'\hat{\Theta}'\Sigma\hat{\Theta}\hat{\mu}}}{\mu'\Sigma^{-1}\mu} = \frac{\left[\frac{\mu'\hat{\Theta}\hat{\mu}}{\mu'\Sigma^{-1}\mu}\right]^2}{\left[\frac{\hat{\mu}'\hat{\Theta}'\Sigma\hat{\Theta}'\hat{\mu}}{\mu'\Sigma^{-1}\mu}\right]}. \tag{A.1}$$

The proof will consider the numerator and the denominator of the squared maximum out-of-sample Sharpe ratio. We start with the numerator using $\Theta := \Sigma^{-1}$

$$\frac{\mu'\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu} = \frac{\mu'\hat{\Theta}\hat{\mu} - \mu'\Theta\mu}{\mu'\Theta\mu} + 1. \tag{A.2}$$

Consider the fraction on the right-hand side. Start with the numerator in (A.2).

$$\begin{aligned}
|\mu'\hat{\Theta}\hat{\mu} - \mu'\Theta\mu|/p &= |\mu'\hat{\Theta}\hat{\mu} - \mu'\Theta\hat{\mu} + \mu'\Theta\hat{\mu} - \mu'\Theta\mu|/p \\
&\leq |\mu'(\hat{\Theta} - \Theta)\hat{\mu}|/p + |\mu'\Theta(\hat{\mu} - \mu)|/p \\
&\leq |\mu'(\hat{\Theta} - \Theta)(\hat{\mu} - \mu)|/p + |\mu'(\hat{\Theta} - \Theta)\mu|/p + |\mu'\Theta(\hat{\mu} - \mu)|/p \\
&= [O_p(\bar{s}\frac{lnp}{n}) + O_p(\bar{s}\sqrt{\frac{lnp}{n}}) + O_p(\bar{s}^{1/2}\sqrt{\frac{lnp}{n}}) \\
&= O_p(\bar{s}\sqrt{\frac{lnp}{n}}), \qquad \text{(A.3)}
\end{aligned}$$

where we use (A.87)-(A.89) for the rates and the dominant rate in the last equality is by Assumption 4. Next, we analyze the denominator in (A.2). Then, by Assumption 2, seeing that $\Sigma^{-1} = \Theta$, by definition

$$\mu'\Sigma^{-1}\mu/p \geq Eigmin(\Sigma^{-1})\|\mu\|_2^2/p \geq cc_l^2 > 0 \tag{A.4}$$

since for all j: $0 < c_l \leq |\mu_j|$ by Assumption 2, and $Eigmin(\Sigma^{-1}) \geq c > 0$, where c is a positive constant

Then, by (A.3)(A.4)

$$\frac{\mu'\hat{\Theta}\hat{\mu}/p}{\mu'\Theta\mu/p} \leq \frac{|\mu'\hat{\Theta}\hat{\mu} - \mu'\Theta\mu|/p}{\mu'\Theta\mu/p} + 1 = O_p(\bar{s}\sqrt{lnp/n}) + 1. \tag{A.5}$$

We now attempt to show that the denominator

$$\frac{\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu}}{\mu'\Sigma^{-1}\mu} \xrightarrow{p} 1. \tag{A.6}$$

In that respect, bearing in mind that $\Theta = \Sigma^{-1}$ is symmetric

$$\frac{\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu}}{\mu'\Sigma^{-1}\mu} = \frac{\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu} - \mu'\Theta\Sigma\Theta\mu}{\mu'\Theta\Sigma\Theta\mu} + 1 \geq 1 - \left|\frac{\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu} - \mu'\Theta\Sigma\Theta\mu}{\mu'\Theta\Sigma\Theta\mu}\right|. \tag{A.7}$$

We can write

$$\hat{\Theta}\hat{\mu} - \Theta\mu = (\hat{\Theta} - \Theta)\hat{\mu} + \Theta(\hat{\mu} - \mu). \tag{A.8}$$

Using (A.8)

$$\begin{aligned}
|\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu} - \mu'\Theta\Sigma\Theta\mu| &= |[(\hat{\mu}'\hat{\Theta} - \mu\Theta) + \mu\Theta]'\Sigma[(\hat{\mu}'\hat{\Theta} - \mu\Theta) + \mu\Theta] - \mu'\Theta\Sigma\Theta\mu| \\
&\leq |[(\hat{\Theta} - \Theta)\hat{\mu}]'\Sigma[(\hat{\Theta} - \Theta)\hat{\mu}]| && \text{(A.9)} \\
&+ 2|[(\hat{\Theta} - \Theta)\hat{\mu}]'\Sigma\Theta(\hat{\mu} - \mu)| && \text{(A.10)} \\
&+ 2|[(\hat{\Theta} - \Theta)\hat{\mu}]'\Sigma\Theta\mu| && \text{(A.11)} \\
&+ |[\Theta(\hat{\mu} - \mu)]'\Sigma[\Theta(\hat{\mu} - \mu)]| && \text{(A.12)} \\
&+ 2|[\Theta(\hat{\mu} - \mu)]'\Sigma\Theta\mu| && \text{(A.13)}
\end{aligned}$$

First, we consider (A.9)

$$\begin{aligned}
|\hat{\mu}'(\hat{\Theta} - \Theta)'\Sigma(\hat{\Theta} - \Theta)\hat{\mu}| &\leq Eigmax(\Sigma)\|(\hat{\Theta} - \Theta)\hat{\mu}\|_2^2 \\
&= Eigmax(\Sigma)[\sum_{j=1}^{p}\{(\hat{\Theta}_j - \Theta_j)'\hat{\mu}\}^2] \\
&\leq Eigmax(\Sigma)p \max_{1 \leq j \leq p}[(\hat{\Theta}_j - \Theta_j)'\hat{\mu}]^2 \\
&\leq Eigmax(\Sigma)p(\max_{1 \leq j \leq p}\|\hat{\Theta}_j - \Theta_j\|_1)^2\|\hat{\mu}\|_\infty^2 \\
&= O(1)pO_p(\bar{s}^2\frac{lnp}{n})O_p(1), && \text{(A.14)}
\end{aligned}$$

where we use Holder's inequality for the third inequality and Theorem 1(i), (ii) and Assumption 2 for the rate. Now, consider (A.10), and by definition $\Theta := \Sigma^{-1}$.

$$
\begin{aligned}
|[(\hat{\Theta}-\Theta)\hat{\mu}]'\Sigma\Theta(\hat{\mu}-\mu)| &= |\hat{\mu}'(\hat{\Theta}-\Theta)'(\hat{\mu}-\mu)| \\
&\leq |(\hat{\mu}-\mu)'(\hat{\Theta}-\Theta)'(\hat{\mu}-\mu)| + |\mu'(\hat{\Theta}-\Theta)'(\hat{\mu}-\mu)| \\
&= p[O_p(\bar{s}(\frac{lnp}{n})^{3/2}) + O_p(\bar{s}(\frac{lnp}{n}))] \\
&= pO_p(\bar{s}(\frac{lnp}{n})), \qquad \text{(A.15)}
\end{aligned}
$$

by (A.85)(A.88) for the second equality, and the dominant rate in third equality can be seen from Assumption 4. Next, consider (A.11), and recall that $\Theta := \Sigma^{-1}$

$$
\begin{aligned}
|[(\hat{\Theta}-\Theta)\hat{\mu}]'\Sigma\Theta\mu| &= |\hat{\mu}'(\hat{\Theta}-\Theta)\mu| \\
&\leq |(\hat{\mu}-\mu)'(\hat{\Theta}-\Theta)\mu| + |\mu'(\hat{\Theta}-\Theta)'\mu| \\
&= p[O_p(\bar{s}\frac{lnp}{n}) + O_p(\bar{s}\sqrt{\frac{lnp}{n}})] \\
&= pO_p(\bar{s}\sqrt{\frac{lnp}{n}}), \qquad \text{(A.16)}
\end{aligned}
$$

where we use (A.88)(A.89) for the second equality, and the dominant rate in the third equality can be seen from Assumption 4. Consider now (A.12) by the symmetry of $\Theta = \Sigma^{-1}$

$$
\begin{aligned}
|[\Theta(\hat{\mu}-\mu)]'\Sigma\Theta(\hat{\mu}-\mu)| &= |(\hat{\mu}-\mu)'\Theta(\hat{\mu}-\mu)| \\
&= pO_p(\bar{s}^{1/2}\frac{lnp}{n}) \qquad \text{(A.17)}
\end{aligned}
$$

by (A.86). Next, analyze (A.13) by the symmetricity of $\Theta = \Sigma^{-1}$

$$
\begin{aligned}
|[\Theta(\hat{\mu}-\mu)]'\Sigma\Theta\mu| &= |(\hat{\mu}-\mu)'\Theta\mu| \\
&= pO_p(\bar{s}^{1/2}\sqrt{\frac{lnp}{n}}), \qquad \text{(A.18)}
\end{aligned}
$$

by (A.87). Combine the rates and terms (A.14)-(A.18) in (A.9)-(A.13) to obtain

$$
|\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu} - \mu'\Theta\Sigma\Theta\mu| = pO_p(\bar{s}\sqrt{\frac{lnp}{n}}), \qquad \text{(A.19)}
$$

by the dominant rate in (A.16), as seen in Assumption 4.

See that by $\Theta = \Sigma^{-1}$, by (A.4),

$$
\mu'\Theta\Sigma\Theta\mu = \mu'\Sigma^{-1}\mu \geq Eigmin(\Sigma^{-1})\|\mu\|_2^2 \geq c\|\mu\|_2^2 \geq (c)(c_l^2)p, \qquad \text{(A.20)}
$$

by Assumption 2.

Combine (A.19)(A.20) in the second term on the right-hand side of (A.7) to have from Assumption 2 and Assumption 4

$$
\frac{|\hat{\mu}'\hat{\Theta}'\Sigma\hat{\Theta}\hat{\mu} - \mu'\Theta\Sigma\Theta\mu|/p}{\mu'\Theta\Sigma\Theta\mu/p} \leq \frac{cO_p(\bar{s}\sqrt{\frac{lnp}{n}})}{c(c_l^2)} = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \qquad \text{(A.21)}
$$

Therefore, we show (A.6). Then, combine (A.5)(A.6) in (A.1) to obtain the desired result. **Q.E.D.**

Proof of Theorem 3. (i). Start with definition of weights, and its estimates

$$\frac{\left(\frac{\sigma \mu'\hat{\Theta}\hat{\mu}}{\sqrt{\hat{\mu}'\hat{\Theta}\hat{\mu}}}\right)}{\left(\frac{\sigma \mu'\Theta\mu}{\sqrt{\mu'\Theta\mu}}\right)} - 1 = \left[\frac{\mu'\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu}\left(\frac{\mu'\Theta\mu}{\hat{\mu}'\hat{\Theta}\hat{\mu}}\right)^{1/2}\right] - 1 \tag{A.22}$$

See that

$$\begin{aligned}\left(\frac{\mu'\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu}\right) &\times \left(\frac{\mu'\Theta\mu}{\hat{\mu}'\hat{\Theta}\hat{\mu}}\right)^{1/2} - 1 \\ &\le \left[\left|\frac{\mu'\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu} - 1\right| + 1\right]\left[\left|\left(\frac{\mu'\Theta\mu}{\hat{\mu}'\hat{\Theta}\hat{\mu}}\right)^{1/2} - 1\right| + 1\right] - 1 \\ &= \left|\frac{\mu'\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu} - 1\right|\left|\left(\frac{\mu'\Theta\mu}{\hat{\mu}'\hat{\Theta}\hat{\mu}}\right)^{1/2} - 1\right| \\ &+ \left|\frac{\mu'\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu} - 1\right| + \left|\left(\frac{\mu'\Theta\mu}{\hat{\mu}'\hat{\Theta}\hat{\mu}}\right)^{1/2} - 1\right| \end{aligned} \tag{A.23}$$

By (A.5)

$$\left|\frac{\mu'\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu} - 1\right| = O_p(\bar{s}\sqrt{lnp/n}). \tag{A.24}$$

Next, we have

$$\begin{aligned}\frac{\mu'\Theta\mu}{\hat{\mu}\hat{\Theta}\hat{\mu}} &= \frac{\mu'\Theta\mu - \hat{\mu}'\hat{\Theta}\hat{\mu}}{\hat{\mu}'\hat{\Theta}\hat{\mu}} + 1 \\ &\le \frac{|\mu'\Theta\mu/p - \hat{\mu}'\hat{\Theta}\hat{\mu}/p|}{\mu'\Theta\mu/p - |\hat{\mu}'\hat{\Theta}\hat{\mu}/p - \mu'\Theta\mu/p|} + 1, \end{aligned} \tag{A.25}$$

where we divided both the numerator and denominator by p, and

$$\hat{\mu}'\hat{\Theta}\hat{\mu}/p \ge \mu'\Theta\mu/p - |\hat{\mu}'\hat{\Theta}\hat{\mu}/p - \mu'\Theta\mu/p|.$$

By (A.4),(A.25), Lemma A.4 in the Supplementary Appendix, and Assumption 2, with $\mu'\Theta\mu/p \ge cc_l^2$,

$$\frac{\mu'\Theta\mu}{\hat{\mu}'\hat{\Theta}\hat{\mu}} \le \frac{O_p(\bar{s}\sqrt{lnp/n})}{cc_l^2 - O_p(\bar{s}\sqrt{lnp/n})} + 1 = O_p(\bar{s}\sqrt{lnp/n}) + 1. \tag{A.26}$$

Then,

$$\left|\left(\frac{\mu'\Theta\mu}{\hat{\mu}'\hat{\Theta}\hat{\mu}}\right)^{1/2} - 1\right| = \{[1 + O_p(\bar{s}\sqrt{lnp/n})]^{1/2} - 1\} \tag{A.27}$$

Now, use Assumption 4 in (A.24)(A.27) and (A.23) to obtain the desired result.

. **Q.E. D**

(ii). Now, we analyze the risk. See that

$$\hat{w}_{oos}'\Sigma\hat{w}_{oos} - \sigma^2 = \sigma^2 \left(\frac{\hat{\mu}'\hat{\Theta}'\Sigma\hat{\Theta}\hat{\mu}}{\hat{\mu}'\hat{\Theta}\hat{\mu}} - 1 \right) = \sigma^2 \left(\frac{\frac{\hat{\mu}'\hat{\Theta}'\Sigma\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu}}{\frac{\hat{\mu}'\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu}} - 1 \right),$$

where we multiplied and divided by $\mu'\Theta\mu$, which is positive by Assumption 2. By (A.6)(A.21),

$$\left| \frac{\hat{\mu}'\hat{\Theta}'\Sigma\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu} - 1 \right| = O_p(\bar{s}\sqrt{lnp/n}). \tag{A.28}$$

Additionally, by Lemma A.4 in the Supplementary Appendix and Assumptions 2 and 4,

$$|\frac{\hat{\mu}'\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu} - 1| = o_p(1). \tag{A.29}$$

By (A.28)(A.29) and Assumption 4,

$$|\hat{w}_{oos}\Sigma\hat{w}_{oos} - \sigma^2| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1).$$

Q.E.D.

Proof of Theorem 4. See that by Assumption 2,

$$\left| \frac{\widehat{MSR}^2/p}{MSR^2/p} - 1 \right| = \left| \frac{\hat{\mu}'\hat{\Theta}\hat{\mu}/p}{\mu'\Sigma^{-1}\mu/p} - 1 \right| = \frac{|\hat{\mu}'\hat{\Theta}\hat{\mu}/p - \mu'\Sigma^{-1}\mu/p|}{\mu'\Sigma^{-1}\mu/p}.$$

Lemma A.4 in the Supplementary Appendix shows that under Assumptions 1-3,

$$|\hat{\mu}'\hat{\Theta}\hat{\mu}/p - \mu'\Sigma^{-1}\mu/p| = O(\bar{s}\sqrt{lnp/n}). \tag{A.30}$$

Combining (A.4),(A.30) with Assumption 4,

$$\left| \frac{\hat{\mu}'\hat{\Theta}\hat{\mu}/p}{\mu'\Sigma^{-1}\mu/p} - 1 \right| = O(\bar{s}\sqrt{lnp/n}) = o_p(1).$$

Q.E.D.

Proof of Theorem 5. Note that by the definition of MSR_c in (15) and A, B, D terms,

$$\frac{MSR_c^2}{p} = D - (B^2/A),$$

and the estimate is

$$\frac{\widehat{MSR}_c^2}{p} = \hat{D} - (\hat{B}^2/\hat{A}),$$

where $\hat{A} = 1_p'\hat{\Theta}1_p/p$, $\hat{B} = 1_p'\hat{\Theta}\hat{\mu}/p$, $\hat{D} = \hat{\mu}'\hat{\Theta}\hat{\mu}/p$.

Then, clearly

$$\frac{\frac{\widehat{MSR}_c^2}{p}}{\frac{MSR_c^2}{p}} = \left[\frac{\hat{A}\hat{D} - \hat{B}^2}{AD - B^2} \right] \left[\frac{A}{\hat{A}} \right]. \tag{A.31}$$

We start with

$$|\hat{A} - A| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1), \tag{A.32}$$

by Assumption 4 and Lemma A.2 in the Supplementary Appendix. Then, $A \geq Eigmin(\Sigma^{-1}) \geq c > 0$ with c a positive constant by Assumption 2. Thus, clearly we obtain, since $|\hat{A}| \geq A - |\hat{A} - A|$,

$$\left|\frac{A}{\hat{A}} - 1\right| = \left|\frac{A - \hat{A}}{\hat{A}}\right| \leq \frac{|\hat{A} - A|}{A - |\hat{A} - A|}$$

which implies

$$|\frac{A}{\hat{A}} - 1| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \tag{A.33}$$

Next, Lemma A.6 in the Supplementary Appendix establishes that under our Assumptions 1-4,

$$|(\hat{A}\hat{D} - \hat{B}^2) - (AD - B^2)| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1).$$

We can use the condition that $AD - B^2 \geq C_1 > 0$, and thus we combine the results above to obtain

$$\left|\frac{\hat{A}\hat{D} - \hat{B}^2}{AD - B^2} - 1\right| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \tag{A.34}$$

Since

$$\frac{\frac{\widehat{MSR}_c^2}{p}}{\frac{MSR_c^2}{p}} = \left[\left(\frac{\hat{A}\hat{D} - \hat{B}^2}{AD - B^2} - 1\right) + 1\right]\left[\left(\frac{A}{\hat{A}} - 1\right) + 1\right]$$

Combine (A.33)(A.34) in (A.31) to obtain

$$\begin{aligned}
\left|\frac{\widehat{MSR}_c^2/p}{MSR_c^2/p} - 1\right| &\leq \left|\frac{\hat{A}\hat{D} - \hat{B}^2}{AD - B^2} - 1\right| |\frac{A}{\hat{A}} - 1| \\
&+ |\frac{A}{\hat{A}} - 1| + \left|\frac{\hat{A}\hat{D} - \hat{B}^2}{AD - B^2} - 1\right| &\text{(A.35)} \\
&= O_p(\bar{s}\sqrt{lnp/n}) = o_p(1), &\text{(A.36)}
\end{aligned}$$

where the rate is the slowest among the three right-hand-side terms. **Q.E. D**

Proof of Theorem 6. We need to start with

$$\left|\frac{(\widehat{MSR}^*)^2/p}{(MSR^*)^2/p} - 1\right| = \frac{\left|(\widehat{MSR}^*)^2/p - (MSR^*)^2/p\right|}{(MSR^*)^2/p} \tag{A.37}$$

As a first step, analyze the denominator in (A.37). Note that $1_p'\Sigma^{-1}\mu/p \geq 0$ is implied by $1_p'\Sigma^{-1}\mu/p \geq C > 2\epsilon > 0$, and thus

$$MSR^2/p = \mu'\Sigma^{-1}\mu/p \geq Eigmin(\Sigma^{-1})\|\mu\|_2^2/p \geq cc_l^2 > 0,$$

by Assumption 2. Note that $1_p'\Sigma^{-1}\mu/p \leq -C < -2\epsilon < 0$ implies $1_p'\Sigma^{-1}\mu/p < 0$. Thus,

$$MSR_c^2/p = D - (B^2/A) = (AD - B^2)/A \geq C_1/K > 0,$$

since by Assumption $AD - B^2 \geq C_1 > 0$ and $A = 1_p'\Sigma^{-1}1_p/p \leq Eigmax(\Sigma^{-1}) \leq K < \infty$ and K is a positive constant by Assumption 2. Then, clearly by combining the results,

$$(MSR^*)^2/p = (MSR^2)1_{\{1_p'\Sigma^{-1}\mu \geq 0\}} + (MSR_c)^2 1_{\{1_p'\Sigma^{-1}\mu < 0\}} \geq min(cc_l^2, C_1/K) > 0. \tag{A.38}$$

Next, we consider the numerator. We need to show that

$$\begin{aligned} p^{-1}|(\widehat{MSR}^*)^2 - (MSR^*)^2| &= p^{-1}|(\widehat{MSR})^2 1_{\{1_p'\hat{\Theta}\hat{\mu} > 0\}} - (MSR)^2 1_{\{1_p'\Sigma^{-1}\mu \geq 0\}} \\ &+ [(\widehat{MSR}_c)^2 1_{\{1_p'\hat{\Theta}\hat{\mu} < 0\}} - (MSR_c)^2 1_{\{1_p'\Sigma^{-1}\mu < 0\}}]| \\ &= O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \end{aligned} \tag{A.39}$$

First, see that on the right-hand side of (A.39)

$$\begin{aligned} p^{-1}|(\widehat{MSR})^2 1_{\{1_p'\hat{\Theta}\hat{\mu} > 0\}} &- (MSR)^2 1_{\{1_p'\Sigma^{-1}\mu \geq 0\}}| \leq p^{-1}|(\widehat{MSR})^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p > 0\}} - (MSR)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p > 0\}}| \\ &+ p^{-1}|(MSR)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p > 0\}} - (MSR)^2 1_{\{1_p'\Sigma^{-1}\mu/p \geq 0\}}|, \end{aligned} \tag{A.40}$$

where division by p in the indicator function does not change the results since the function operates when it is positive.

Then, in (A.40),

$$\begin{aligned} p^{-1}|(\widehat{MSR})^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p > 0\}} - (MSR)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p > 0\}}| &\leq p^{-1}|(\widehat{MSR})^2 - (MSR)^2||1_{\{1_p'\hat{\Theta}\hat{\mu}/p > 0\}}| \\ &\leq p^{-1}|(\widehat{MSR})^2 - (MSR)^2| \\ &= O_p(\bar{s}\sqrt{lnp/n}) = o_p(1), \end{aligned} \tag{A.41}$$

by (A.30) and Assumption 4 for the rate. In (A.40) above, consider

$$\begin{aligned} p^{-1}|(MSR)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p > 0\}} &- (MSR)^2 1_{\{1_p'\Sigma^{-1}\mu/p \geq 0\}}| \\ &\leq p^{-1}MSR^2|1_{\{1_p'\hat{\Theta}\hat{\mu}/p > 0\}} - 1_{\{1_p'\Sigma^{-1}\mu/p \geq 0\}}|. \end{aligned} \tag{A.42}$$

Note that by definition of MSR^2/p,

$$MSR^2/p = \mu'\Sigma^{-1}\mu/p \leq Eigmax(\Sigma^{-1})\|\mu\|_2^2/p \leq Kc_u^2 < \infty, \tag{A.43}$$

where we use Assumption 2. Define the event $E_1 = \{|1_p'\hat{\Theta}\hat{\mu}/p - 1_p'\Sigma^{-1}\mu/p| \leq \epsilon\}$, where $\epsilon > 0$. Start with the condition $1_p'\Sigma^{-1}\mu/p \geq C > 2\epsilon > 0$; then on the event E_1,

$$\begin{aligned} \frac{1_p'\hat{\Theta}\hat{\mu}}{p} &= \frac{1_p'\hat{\Theta}\hat{\mu}}{p} - \frac{1_p'\Sigma^{-1}\mu}{p} + \frac{1_p'\Sigma^{-1}\mu}{p} \\ &\geq \frac{1_p'\Sigma^{-1}\mu}{p} - |\frac{1_p'\hat{\Theta}\hat{\mu}}{p} - \frac{1_p'\Sigma^{-1}\mu}{p}| \\ &\geq \frac{1_p'\Sigma^{-1}\mu}{p} - \epsilon \\ &\geq C - \epsilon > 2\epsilon - \epsilon = \epsilon > 0, \end{aligned} \tag{A.44}$$

where we use E_1 in the second inequality and the condition for the third inequality. This clearly shows that at event E_1, when the condition $1_p'\Sigma^{-1}\mu/p \geq C > 2\epsilon > 0$ holds, we have $1_p'\hat{\Theta}\hat{\mu}/p > \epsilon > 0$.

By Lemma A.3 of the Supplementary Appendix, E_1 occurs with a probability approaching one under our Assumptions 1-4

$$|1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}} - 1_{\{1_p'\Sigma^{-1}\mu/p\geq 0\}}| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1), \tag{A.45}$$

where we use (A.44) and $1_p'\Sigma^{-1}\mu/p \geq C > 2\epsilon > 0$, implying $1_p'\Sigma^{-1}\mu/p \geq 0$.

Next, combine (A.43)-(A.45) into (A.42)

$$p^{-1}|(MSR)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}} - (MSR)^2 1_{\{1_p'\Sigma^{-1}\mu/p\geq 0\}}| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \tag{A.46}$$

By (A.41)(A.46), we have in (A.40)

$$p^{-1}|(\widehat{MSR}^2) 1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}} - (MSR^2) 1_{\{1_p'\Sigma^{-1}\mu/p\geq 0\}} = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \tag{A.47}$$

The proof for $p^{-1}|(\widehat{MSR}_c)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p<0\}} - (MSR_c)^2 1_{\{1_p'\Sigma^{-1}\mu/p<0\}}|$ is identical to that in (A.47) given Theorem 5, except that we have to show that

$$|1_{\{1_p'\hat{\Theta}\hat{\mu}/p<0\}} - 1_{\{1_p'\Sigma^{-1}\mu/p<0\}}| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1), \tag{A.48}$$

instead of (A.45). Assume that we use event E_1:

$$\begin{aligned} \frac{1_p'\Sigma^{-1}\mu}{p} &= \frac{1_p'\Sigma^{-1}\mu}{p} - \frac{1_p'\hat{\Theta}\hat{\mu}}{p} + \frac{1_p'\hat{\Theta}\hat{\mu}}{p} \\ &\geq \frac{1_p'\hat{\Theta}\hat{\mu}}{p} - |\frac{1_p'\Sigma^{-1}\mu}{p} - \frac{1_p'\hat{\Theta}\hat{\mu}}{p}| \\ &\geq \frac{1_p'\hat{\Theta}\hat{\mu}}{p} - \epsilon. \end{aligned} \tag{A.49}$$

Then, in (A.49), using the condition $1_p'\Sigma^{-1}\mu/p \leq -C < -2\epsilon < 0$ (note that this also implies $1_p'\Sigma^{-1}\mu/p < 0$)

$$0 > -2\epsilon > -C \geq 1_p'\Sigma^{-1}\mu/p \geq 1_p'\hat{\Theta}\hat{\mu}/p - \epsilon,$$

which implies that, with $C > 2\epsilon$, adding ϵ to all sides above yields

$$0 > -\epsilon > -(C-\epsilon) \geq 1_p'\hat{\Theta}\hat{\mu}/p,$$

which clearly shows that when $1_p'\Sigma^{-1}\mu/p < 0$, we will have $1_p'\hat{\Theta}\hat{\mu}/p < 0$. Note that event E_1 occurs with probability approaching one by Lemma A.3 in the Supplementary Appendix, so we have proven (A.48). This implies with the result of Theorem 5 that

$$p^{-1}|(\widehat{MSR}_c)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p<0\}} - (MSR_c)^2 1_{\{1_p'\Sigma^{-1}\mu/p<0\}}| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \tag{A.50}$$

By now combining (A.47)(A.50), we proved (A.39) via the triangle inequality. With (A.38) and (A.39), the desired result follows (A.37).

Q.E.D.

Proof of Theorem 7. First, we start with definitions of $\hat{A} := 1_p'\hat{\Theta}1_p/p$, $\hat{B} := 1_p'\hat{\Theta}\hat{\mu}/p$, $A := 1_p'\Sigma^{-1}1_p/p$, $B := 1_p'\Sigma^{-1}\mu/p$.

$$
\begin{aligned}
\left|\frac{\widehat{SR}_{nw}^2}{SR^2} - 1\right| &= \left|\frac{p(1_p'\hat{\Theta}\hat{\mu}/p)^2(1_p'\hat{\Theta}1_p/p)^{-1}}{p(1_p'\Sigma^{-1}\mu/p)^2(1_p'\Sigma^{-1}1_p/p)^{-1}} - 1\right| \\
&= \left|\frac{\hat{B}^2 A}{B^2\hat{A}} - 1\right| \\
&= \left|\frac{\hat{B}^2 A - B^2\hat{A}}{B^2\hat{A}}\right| \qquad (A.51)
\end{aligned}
$$

We analyze the denominator in (A.51). To that effect, by Assumption 2,

$$
A = 1_p'\Sigma^{-1}1_p/p \geq Eigmin(\Sigma^{-1}) \geq c > 0.
$$

By the condition in the statement of Theorem 7,

$$
|B| = |\frac{1_p'\Sigma^{-1}\mu}{p}| \geq C > 2\epsilon > 0.
$$

Then, by Lemma A.2 and Lemma A.5 in the Supplementary Appendix

$$
|B^2\hat{A}| = |B^2(\hat{A} - A) + B^2 A| \geq B^2 A - B^2|\hat{A} - A| \geq C^2 c + o_p(1) > 0. \qquad (A.52)
$$

Now consider the numerator in (A.51):

$$
\begin{aligned}
|\hat{B}^2 A - B^2\hat{A}| &= |\hat{B}^2 A - \hat{B}^2\hat{A} + \hat{B}^2\hat{A} - B^2\hat{A}| \\
&\leq |\hat{B}^2(\hat{A} - A)| + |(\hat{B}^2 - B^2)\hat{A}| \\
&\leq |\hat{B}^2(\hat{A} - A)| + |\hat{B} - B||\hat{B} + B||\hat{A}|. \qquad (A.53)
\end{aligned}
$$

Analyze the first term on the right side of (A.53):

$$
\begin{aligned}
\hat{B}^2 &= |\hat{B}^2 - B^2 + B^2| \\
&\leq |\hat{B}^2 - B^2| + B^2 \\
&\leq |\hat{B} - B||\hat{B} + B| + B^2. \qquad (A.54)
\end{aligned}
$$

Then, by Lemma A.3 in the Supplementary Appendix,

$$
|\hat{B} - B| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \qquad (A.55)
$$

Then,

$$
\begin{aligned}
|\hat{B} + B| &\leq |\hat{B}| + |B| \\
&\leq |\hat{B} - B| + 2|B| \\
&= o_p(1) + 2|B| \\
&= O_p(1), \qquad (A.56)
\end{aligned}
$$

where we use (A.55) and Lemma A.5 in the Supplementary Appendix.

By (A.55)(A.56) in (A.54), we have

$$\hat{B}^2 = O_p(1). \tag{A.57}$$

Then, by Lemma A.2 in the Supplementary Appendix and (A.57),

$$|\hat{B}^2(\hat{A} - A)| \leq \hat{B}^2|\hat{A} - A| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \tag{A.58}$$

Then, the second term on the right side of (A.53) is

$$|\hat{B} - B||\hat{B} + B||\hat{A}| = O_p(\bar{s}\sqrt{\frac{lnp}{n}})O_p(1)O_p(1) = o_p(1), \tag{A.59}$$

by (A.55)(A.56) and Lemma A.2, Lemma A.5 in the Supplementary Appendix. Use (A.58)(A.59) in (A.53)

$$|\hat{B}^2 A - B^2 \hat{A}| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \tag{A.60}$$

Combine (A.52) with (A.60) in (A.51) to obtain the desired result. **Q.E.D.**

Proof of Theorem 8. To ease the notation in the proofs, set $AD - B^2 = z$, $A\rho_1^2 - 2B\rho_1 + D = v$. The estimates will be $\hat{z} = \hat{A}\hat{D} - \hat{B}^2$, $\hat{v} = \hat{A}\rho_1^2 - 2\hat{B}\rho_1 + \hat{D}$. Then,

$$\begin{aligned}
\left|\frac{\widehat{SR}^2_{MV}}{SR^2_{MV}} - 1\right| &= \left|\frac{\hat{z}/\hat{v}}{z/v} - 1\right| \\
&= \left|\frac{\hat{z}}{\hat{v}}\frac{v}{z} - 1\right| \\
&= \left|\frac{\hat{z}v - \hat{v}z}{\hat{v}z}\right|.
\end{aligned} \tag{A.61}$$

First, analyze the denominator of (A.61).

$$\begin{aligned}
|\hat{v}z| &= |(\hat{v} - v)z + vz|. \\
&\geq |vz| - |(\hat{v} - v)z| \\
&\geq |vz| - |\hat{v} - v||z|.
\end{aligned} \tag{A.62}$$

Then, by Lemma A.2-A.4 in the Supplementary Appendix, triangle inequality and ρ_1 being bounded away from zero and finite, by Assumption 4,

$$|\hat{v} - v| = |(\hat{A} - A)\rho_1^2 - 2(\hat{B} - B)\rho_1 + (\hat{D} - D)| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \tag{A.63}$$

We also know that by the conditions in theorem statement $z = AD - B^2 \geq C_1 > 0$, and $v = A\rho_1^2 - 2B\rho_1 + D \geq C_1 > 0$. Then, see that by Lemma A.5 in the Supplementary Appendix

$$|z| = |AD - B^2| \leq AD = O(1). \tag{A.64}$$

Thus, by (A.63)(A.64) and $z \geq C_1 > 0, v \geq C_1 > 0$ in (A.62), we have

$$|\hat{v}z| = o_p(1) + C_1^2 > 0. \tag{A.65}$$

Consider the numerator in (A.61):

$$|\hat{z}v - \hat{v}z| = |\hat{z}v - vz + vz - \hat{v}z| \leq |\hat{z} - z||v| + |z||\hat{v} - v|. \tag{A.66}$$

By Lemma A.6 in the Supplementary Appendix,

$$|\hat{z} - z| = |(\hat{A}\hat{D} - \hat{B}^2) - (AD - B^2)| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \tag{A.67}$$

Clearly, by Lemma A.5 in the Supplementary Appendix and triangle inequality with ρ_1 being finite,

$$|v| = |A\rho_1 - 2B\rho_1 + D| = O(1). \tag{A.68}$$

Then, use (A.63)(A.64)(A.67)(A.68) in (A.66) by Assumption 4

$$|\hat{z}v - \hat{v}z| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \tag{A.69}$$

Use (A.65)(A.69) in (A.61) to obtain the desired result. **Q.E.D.**

Supplementary Appendix

Here, we provide supplemental results. We provide a matrix norm inequality. Let x be a generic vector, which is $p \times 1$. M is a square matrix of dimension p, where M_j' is the jth row of dimension $1 \times p$, and M_j is the transpose of this row vector.

Lemma A.1.

$$\|Mx\|_1 \leq p \max_{1 \leq j \leq p} \|M_j\|_1 \|x\|_\infty.$$

Proof of Lemma A.1.

$$\begin{aligned}
\|Mx\|_1 &= |M_1'x| + |M_2'x| + \cdots + |M_p'x| \\
&\leq \|M_1\|_1\|x\|_\infty + \|M_2\|_1\|x\|_\infty + \cdots + \|M_p\|_1\|x\|_\infty \\
&= [\sum_{j=1}^{p} \|M_j\|_1]\|x\|_\infty \\
&\leq p \max_j \|M_j\|_1\|x\|_\infty, \qquad \text{(A.70)}
\end{aligned}$$

where we use Holder's inequality to obtain each inequality. **Q.E.D.**

The following lemmata are all from Callot et al. (2019) and repeated for the benefit of readers. Recall the definition of $A := 1_p'\Sigma^{-1}1_p/p$ and $\hat{A} := 1_p'\hat{\Theta}1_p/p$.

Lemma A.2. *Under Assumptions 1-4 uniformly in $j \in \{1, \cdots, p\}$, for $\lambda_j = O(\sqrt{lnp/n})$,*

$$|\hat{A} - A| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1).$$

Proof of Lemma A.2. First, see that

$$\hat{A} - A = (1_p'\hat{\Theta}1_p - 1_p'\Theta 1_p)/p = (1_p'(\hat{\Theta} - \Theta)1_p)/p. \qquad \text{(A.71)}$$

Now, consider the right-hand side of (A.71)

$$\begin{aligned}
|1_p'(\hat{\Theta} - \Theta)1_p|/p &\leq \|(\hat{\Theta} - \Theta)1_p\|_1\|1_p\|_\infty/p \\
&\leq \max_{1 \leq j \leq p} \|\hat{\Theta}_j - \Theta_j\|_1 \\
&= O_p(\bar{s}\sqrt{\ln p/n}) = o_p(1), \qquad \text{(A.72)}
\end{aligned}$$

where Holder's inequality is used in the first inequality, Lemma A.1 is used for the second inequality, and the last equality is obtained by using Theorem 1 and imposing Assumption 4. **Q.E.D.**

Before the next Lemma, we define $\hat{B} := 1_p'\hat{\Theta}\hat{\mu}/p$, and $B := 1_p'\Theta\mu/p$.

Lemma A.3. *Under Assumptions 1-4 uniformly in $j \in \{1, \cdots, p\}$, for $\lambda_j = O(\sqrt{lnp/n})$,*

$$|\hat{B} - B| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1).$$

Proof of Lemma A.3. We can decompose $\hat{B}$ by simple addition and subtraction into

$$\hat{B} - B = [1_p'(\hat{\Theta} - \Theta)(\hat{\mu} - \mu)]/p \quad (A.73)$$
$$+ [1_p'(\hat{\Theta} - \Theta)\mu]/p \quad (A.74)$$
$$+ [1_p'\Theta(\hat{\mu} - \mu)]/p \quad (A.75)$$

Now, we analyze each of the terms above. Since $\hat{\mu} = n^{-1}\sum_{t=1}^{n} r_t$,

$$\begin{aligned} |1_p'(\hat{\Theta} - \Theta)(\hat{\mu} - \mu)|/p &\leq \|(\hat{\Theta} - \Theta)1_p\|_1 \|\hat{\mu} - \mu\|_\infty / p \\ &\leq [\max_{1\leq j\leq p} \|\hat{\Theta}_j - \Theta_j\|_1]\|\hat{\mu} - \mu\|_\infty \\ &= O_p(\bar{s}\sqrt{\ln p/n})O_p(\sqrt{lnp/n}), \end{aligned} \quad (A.76)$$

where we use Holder's inequality in the first inequality and Lemma A.1 with $M = \hat{\Theta} - \Theta$, $x = 1_p$ in the second inequality above, and the rate is from Theorem 1.

Therefore, we consider (A.74) above. Since we have Assumption 2, $\|\mu\|_\infty < c_u < \infty$, where c_u is a positive constant,

$$\begin{aligned} |1_p'(\hat{\Theta} - \Theta)\mu|/p &\leq \|(\hat{\Theta} - \Theta)1_p\|_1 \|\mu\|_\infty / p \\ &\leq c_u[\max_{1\leq j\leq p} \|\hat{\Theta}_j - \Theta_j\|_1] \\ &= c_u O_p(\bar{s}\sqrt{\ln p/n}), \end{aligned} \quad (A.77)$$

where we use Holder's inequality in the first inequality and Lemma A.1 with $M = \hat{\Theta} - \Theta$, $x = 1_p$ in the second inequality above, and the rate is from Theorem 1.

Now consider (A.75).

$$\begin{aligned} |1_p'\Theta(\hat{\mu} - \mu)|/p &\leq \|\Theta 1_p\|_1 \|\hat{\mu} - \mu\|_\infty / p \\ &\leq [\max_{1\leq j\leq p} \|\Theta_j\|_1]\|\hat{\mu} - \mu\|_\infty \\ &= O(\sqrt{\bar{s}})O_p(\sqrt{\ln p/n}), \end{aligned} \quad (A.78)$$

where we use Holder's inequality in the first inequality and Lemma A.1 with $M = \Theta$, $x = 1_p$ in the second inequality above, and the rate is from Theorem 1 and (B.55) from Caner and Kock (2018) $[\max_{1\leq j\leq p} \|\Theta_j\|_1] = O(\sqrt{\bar{s}})$.

Combine (A.76)(A.77)(A.78) in (A.73)-(A.75), and note that the largest rate is coming from (A.77). Therefore, use Assumption 4, $\bar{s}\sqrt{lnp/n} = o(1)$ to obtain

$$|\hat{B} - B| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \quad (A.79)$$

. **Q.E.D.**

Note that $D := \mu'\Theta\mu/p$, and its estimator is $\hat{D} := \hat{\mu}'\hat{\Theta}\hat{\mu}/p$.

Lemma A.4. *Under Assumptions 1-4 uniformly in* $j \in \{1, \cdots, p\}$, *for* $\lambda_j = O(\sqrt{lnp/n})$,

$$|\hat{D} - D| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1).$$

Proof of Lemma A.4. By simple addition and subtraction,

$$\hat{D} - D = [(\hat{\mu} - \mu)'(\hat{\Theta} - \Theta)(\hat{\mu} - \mu)]/p \quad (A.80)$$
$$+ [(\hat{\mu} - \mu)'\Theta(\hat{\mu} - \mu)]/p \quad (A.81)$$
$$+ [2(\hat{\mu} - \mu)'\Theta\mu]/p \quad (A.82)$$
$$+ [2\mu'(\hat{\Theta} - \Theta)(\hat{\mu} - \mu)]/p \quad (A.83)$$
$$+ [\mu'(\hat{\Theta} - \Theta)\mu]/p. \quad (A.84)$$

We start with (A.80).

$$\begin{aligned} |(\hat{\mu} - \mu)'(\hat{\Theta} - \Theta)(\hat{\mu} - \mu)|/p &\leq \|(\hat{\Theta} - \Theta)(\hat{\mu} - \mu)\|_1 \|\hat{\mu} - \mu\|_\infty / p \\ &\leq [\|\hat{\mu} - \mu\|_\infty]^2 [\max_j \|\hat{\Theta}_j - \Theta_j\|_1] \\ &= O_p(\ln p/n) O_p(\bar{s}\sqrt{\ln p/n}) \\ &= O_p(\bar{s}(\ln p/n)^{3/2}), \end{aligned} \quad (A.85)$$

where Holder's inequality is used for the first inequality above, and the inequality Lemma A.1, with $M = \hat{\Theta} - \Theta$, and $x = \hat{\mu} - \mu$ for the second inequality above, and for the rates we use Theorem 1.

We continue with (A.81).

$$\begin{aligned} |(\hat{\mu} - \mu)'(\Theta)(\hat{\mu} - \mu)|/p &\leq \|(\Theta)(\hat{\mu} - \mu)\|_1 \|\hat{\mu} - \mu\|_\infty / p \\ &\leq [\|\hat{\mu} - \mu\|_\infty]^2 [\max_j \|\Theta_j\|_1] \\ &= O_p(\ln p/n) O(\sqrt{\bar{s}}) \\ &= O_p(\sqrt{\bar{s}}(\ln p/n)), \end{aligned} \quad (A.86)$$

where Holder's inequality is used for the first inequality above, and the inequality Lemma A.1, with $M = \Theta$, and $x = \hat{\mu} - \mu$ for the second inequality above, and for the rates, we use Theorem 1 and (B.55) of Caner and Kock (2018).

Then, we consider (A.82), using $\|\mu\|_\infty \leq c_u$,

$$\begin{aligned} |(\hat{\mu} - \mu)'(\Theta)(\mu)|/p &\leq \|(\Theta)(\hat{\mu} - \mu)\|_1 \|\mu\|_\infty / p \\ &\leq c_u [\|\hat{\mu} - \mu\|_\infty][\max_j \|\Theta_j\|_1] \\ &= O_p(\sqrt{\ln p/n}) O(\sqrt{\bar{s}}) \\ &= O_p(\sqrt{\bar{s}}\sqrt{lnp/n}), \end{aligned} \quad (A.87)$$

where Holder's inequality is used for the first inequality above, and the inequality Lemma A.1, with $M = \Theta$, and $x = \hat{\mu} - \mu$ for the second inequality above, and for the rates, we use Theorem 1 and (B.55) of Caner and Kock (2018).

Then, we consider (A.83).

$$\begin{aligned} |(\mu)'(\hat{\Theta} - \Theta)(\hat{\mu} - \mu)|/p &\leq \|(\hat{\Theta} - \Theta)(\mu)\|_1 \|\hat{\mu} - \mu\|_\infty / p \\ &\leq \|\mu\|_\infty \max_j \|\hat{\Theta}_j - \Theta_j\|_1 \|\hat{\mu} - \mu\|_\infty \\ &\leq c_u [\max_j \|\hat{\Theta}_j - \Theta_j\|_1] \|(\hat{\mu} - \mu)\|_\infty \\ &= O_p(\bar{s}\sqrt{\ln p/n}) O_p(\sqrt{\ln p/n}) \\ &= O_p(\bar{s} \ln p/n), \end{aligned} \quad (A.88)$$

where Holder's inequality is used for the first inequality above, and the inequality Lemma A.1, with $M = \hat{\Theta} - \Theta$, $x = \mu$ for the second inequality above, for the third inequality above, we use $\|\mu\|_\infty \leq c_u$, and for the rates, we use Theorem 1.

Then, we consider (A.84):

$$\begin{aligned} |(\mu)'(\hat{\Theta} - \Theta)(\mu)|/p &\leq \|(\hat{\Theta} - \Theta)(\mu)\|_1 \|\mu\|_\infty / p \\ &\leq [\|\mu\|_\infty]^2 \max_j \|\hat{\Theta}_j - \Theta_j\|_1 \\ &\leq c_u^2 [\max_j \|\hat{\Theta}_j - \Theta_j\|_1] \\ &= O_p(\bar{s}\sqrt{lnp/n}), \end{aligned} \tag{A.89}$$

where Holder's inequality is used for the first inequality above, and the inequality Lemma A.1, with $M = \hat{\Theta} - \Theta$, $x = \mu$ for the second inequality above, for the third inequality above, we use $\|\mu\|_\infty \leq c_u$, and for the rate, we use Theorem 1. Note that the last rate above in (A.89) derives our result, since it is the largest rate by Assumption 4.

Combine (A.85)-(A.89) in (A.80)-(A.84) and use the rate (A.89) to obtain

$$|\hat{D} - D| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \tag{A.90}$$

Q.E.D.

The following lemma establishes orders for the terms in the optimal weight, A, B, D. Note that both A, D are positive by Assumption 2 and uniformly bounded away from zero.

Lemma A.5. *Under Assumption 2*

$$A = O(1).$$

$$|B| = O(1).$$

$$D = O(1).$$

Proof of Lemma A.5. We complete the proof for the term $D = \mu'\Theta\mu/p$. The proof for $A = 1_p'\Theta 1_p/p$ is the same.

$$D = \mu'\Theta\mu/p \leq Eigmax(\Theta)\|\mu\|_2^2/p = O(1),$$

where we use the fact that each μ_j is a constant as in Assumption 2, and the maximal eigenvalue of $\Theta = \Sigma^{-1}$ is finite by Assumption 2. For the term B, the proof can be obtained by using the Cauchy-Schwartz inequality first and the same analysis as for terms A and D. **Q.E.D.**

Next, we need the following technical lemma, which provides the limit and the rate for the denominator in the optimal portfolio.

Lemma A.6. *Under Assumptions 1-4 uniformly over j in $\lambda_j = O(\sqrt{lnp/n})$,*

$$|(\hat{A}\hat{D} - \hat{B}^2) - (AD - B^2)| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1).$$

Proof of Lemma A.6. Note that by simple addition and subtraction,

$$\hat{A}\hat{D} - \hat{B}^2 = [(\hat{A} - A) + A][(\hat{D} - D) + D] - [(\hat{B} - B) + B]^2.$$

Then, using this last expression and simplifying, A, D being both positive,

$$\begin{aligned} |(\hat{A}\hat{D} - \hat{B}^2) - (AD - B^2)| &\leq \{|\hat{A} - A||\hat{D} - D| + |\hat{A} - A|D \\ &+ A|\hat{D} - D| + (\hat{B} - B)^2 + 2|B||\hat{B} - B|\} \\ &= O_p(\bar{s}\sqrt{lnp/n}) = o_p(1), \end{aligned} \quad \text{(A.91)}$$

where we use (A.72)(A.79)(A.90), Lemma A.5, and Assumption 4: $\bar{s}\sqrt{lnp/n} = o(1)$. **Q.E.D.**

References

Ao, M., Y. Li, and X. Zheng (2019). Approaching mean-variance efficiency for large portfolios. *Review of Financial Studies Forthcoming*.

Bühlmann, P. and S. van de Geer (2011). *Statistics for High Dimensional Data*. Springer Verlag.

Callot, L., M. Caner, O. Onder, and E. Ulasan (2019). A nodewise regression approach to estimating large portfolios. *Journal of Business and Economic Statistics Forthcoming*.

Caner, M. and A. Kock (2018). Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso. *Journal of Econometrics 203*, 143–168.

Chang, J., Y. Qiu, Q. Yao, and T. Zou (2019). Confidence regions for entries of a large precision matrix. *Journal of Econometrics Forthcoming*.

DeMiguel, V., L. Garlappi, and R. Uppal (2007). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies 22*(5), 1915–1953.

Fan, J., Y. Li, and K. Yu (2012). Vast volatility matrix estimation using high frequency data for portfolio selection. *Journal of the American Statistical Association 107*, 412–428.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(4), 603–680.

Fan, J., Y. Liao, and M. Mincheva (2016). *POET: Principal Orthogonal Complement Thresholding (POET) Method*. R package version 2.0.

Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software 33*(1), 1.

Garlappi, L., R. Uppal, and T. Wang (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies 20*, 41–81.

Hastie, T. and B. Efron (2013). *lars: Least Angle Regression, Lasso and Forward Stagewise*. R package version 1.2.

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

Jobson, J. D. and B. M. Korkie (1981). Performance hypothesis testing with the sharpe and treynor measures. *The Journal of Finance 36*(4), 889–908.

Kan, R. and G. Zhou (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis 42*.

Lai, T., H. Xing, and Z. Chen (2011). Mean-variance portfolio optimization when means and covariances are unknown. *The Annals of Applied Statistics 5*, 798–823.

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

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

Ledoit, O, M. and M. Wolf (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Review of Financial Studies 30*, 4349–4388.

Ledoit, O. and M. Wolf (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance 15*(5), 850–859.

Maller, R., S. Roberts, and R. Tourky (2016). The large sample distribution of the maximum sharpe ratio with and without short sales. *Journal of Econometrics 194*, 138–152.

Maller, R. and D. Turkington (2002). New light on portfolio allocation problem. *Mathematical Methods of Operations Research 56*, 501–511.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance 7*, 77–91.

Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 1436–1462.

Memmel, C. (2003). Performance hypothesis testing with the sharpe ratio. *Finance Letters 1*(1).

Ramprasad, P. (2016). *nlshrink: Non-Linear Shrinkage Estimation of Population Eigenvalues and Covariance Matrices*. R package version 1.0.1.

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

van de Geer, S. (2016). *Estimation and testing under sparsity*. Springer-Verlag.

Zhang, Y., R. Li, and C.-L. Tsai (2010). Regularization parameter selections via generalized information criterion. *Journal of the American Statistical Association 105*(489), 312–323.