Sharpe Ratio in High Dimensions: Cases of Maximum Out of Sample, Constrained Maximum, and Optimal Portfolio Choice

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

February 3, 2020

Preliminary

Abstract

In this paper, we analyze maximum Sharpe ratio when the number of assets in a portfolio is larger than its time span. One obstacle in this large dimensional setup is the singularity of the sample covariance matrix of the excess asset returns. To solve 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. We analyze three issues. One of the key results in our paper is that mean-variance efficiency for the portfolios in large dimensions is established. Then 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 see that the number of assets slow down the convergence up to a logarithmic factor. Then, we provide consistency of maximum Sharpe Ratio when the portfolio weights add up to one, and also provide a new formula and an estimate for constrained maximum Sharpe ratio. Finally, we provide consistent estimates of the Sharpe ratios of global minimum variance portfolio and Markowitz's (1952) mean variance portfolio. In terms of assumptions, we allow for time series data. Simulation and out-of-sample forecasting exercise shows 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.
Caner thanks Vanderbilt Economics Department seminar guests for comments.
†Department of Economics, Pontifical Catholic University of Rio de Janeiro - Brazil.

‡Department of Economics, University of California, Irvine. SSPB 3201, Irvine, CA. 92697. Email: gabriel.vasconcelos@uci.edu

1 Introduction

One of the key issues in finance is the tradeoff between the return and the risk of the portfolio. To get a better risk-return, we maximize the Sharpe ratio. Basically the weights of the portfolio is chosen in such a way that return to risk ratio is maximized. We contribute to the literature by the case of a large number of assets p, which may be greater than the time span of the portfolio n. Our analysis also will involve time-series data for excess asset returns. To get the maximum Sharpe ratio we benefit from precision matrix, but the sample covariance matrix is not invertible when $p > n$. Therefore, we need another concept to estimate the precision matrix. To that effect we benefit from a concept promoted by Meinshausen and Bühlmann (2006) which is called *nodewise regression*. To get the Sharpe ratio we estimate the inverse of the precision matrix by a nodewise regression based inverse as in van de Geer (2016). This is based on running lasso regression on a given excess asset return on the other excess asset returns. Then the vector estimate is used in forming the rows of the estimate of the precision matrix. This type of method assumes sparsity, or weak sparsity on the rows of the precision matrix, when $p \geq n$. This sparsity restriction amounts to an asset having largely correlated with certain assets but not all of the assets in the portfolio. Also, weak sparsity allows non-sparse precision matrix, as long as absolute l th power $(0 < l < 1)$ sum of absolute value of coefficients in each row does not diverge too fast, see section 2.10 of van de Geer (2016).

In terms of asset correlations this sparsity assumption can be clearly interpreted. An 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 and in an empirical out-of sample exercise, in Section 7, Figure 2 we show that in US correlation matrix of assets as as well as the density of correlation between assets show that large correlations are not many. Note that we do not assume the sample covariance matrix to be sparse.

Chang et al. (2019) extend nodewise regression to time series context and build confidence intervals for the cells in the precision matrix. Callot et al. (2019) provide variance, risk, and the weight estimation of the portfolio via nodewise regression. Caner and Kock (2018) establish uniform confidence intervals in 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 others excess asset 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. Our technique, and hence contribution, will be complementary to the existing papers. One difference will be analyzing $p \geq n$, when both number of assets and time span goes to infinity in time-series data. Recently, there are important contributions in this area by using shrinkage and factor models. Recently, Ledoit and Wolf (2017) propose a nonlinear shrinkage estimator in which small eigenvalues of the sample covariance matrix are increased, and the large ones are decreased by a shrinkage formula. The main contribution is the optimal shrinkage function, where they find by minimizing a loss function and estimating this optimal shrinkage. 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) made a novel contribution. Ao et al. (2019) take a constrained optimization, maximize returns subject to risk of the portfolio, and show that its equivalent to an unconstrained objective function, where they minimize a scaled return of the portfolio error by choosing optimal weights. In order to get these weights they use lasso regression and hence assume sparse number of nonzero weights of the portfolio, and they analyze $p/n \to (0,1)$. They show that their method maximizes expected return of the portfolio and satisfy the risk constraint. This is an important result on its own.

Our main contribution is that we are able to get mean-variance efficiency for large portfolios even when $p > n$. Related to that consistency of our nodewise based maximum-out-of-sample Sharpe ratio estimate is established. We also provide rate of convergence, and see that number of assets slow down the rate of convergence up to a logarithmic factor in p, hence estimation of large portfolios are 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), Maller et al. (2016) analyze the limit with fixed number of assets, and also extend that approach to large number of assets but less than the time span of the portfolio. Their papers made a key discovery since 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. If practitioners could have used the unconstrained maximum Sharpe formula, instead of the constrained one, they may be getting a minimum Sharpe ratio instead. Our paper extends their paper, by analyzing two issues, first the case of $p \geq n$, with both quantities growing to infinity, and then also analysis of handling the uncertainty created by a technical term, estimating that term,

and using in a new constrained maximum Sharpe ratio which will be estimated consistently.

Third, we consider the Sharpe ratios in Global Minimum Variance portfolio and Markowitz mean-variance portfolio. Our analysis uncovers consistent estimators even when $p > n$. We also conduct simulations and out-of-sample forecasting exercises. Our method performs well. The reason of the good performance is due to the correlation structure of the excess asset returns, the test (out-of-sample, empirics) periods that we analyze have small number of large correlations, hence more in line with our sparsity assumptions which can be seen in Figure 1. In Figure 1, Sub Sample 1 and Sub Sample 2 correspond to two out-of-sample data periods in Section 7. Also in the same figure we super-impose a simulation design that comes from a much used factor model design in Section 6 . There the factor design does not confirm with two sub-periods that we analyze via real life data.

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

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

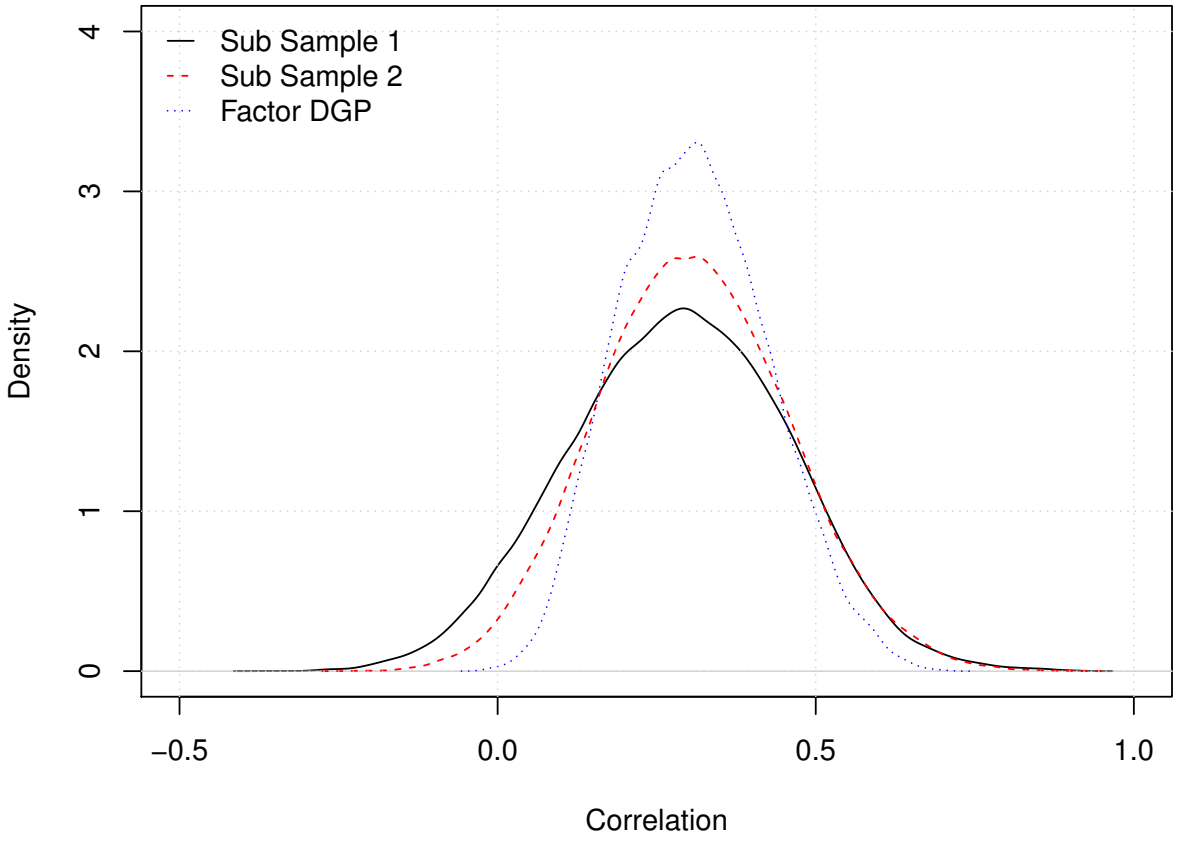

Figure 1: Correlation Densities

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. Covariance matrix of excess asset returns is: $\Sigma := E(r_t - \mu)(r_t - \mu)'$, 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 $n \times p$ matrix. To make things more clear set $r^*_{t,j} := r_{t,j} - \bar{r}_j$ which is the demeaned t th period, j th asset's excess return, and $\bar{r}_j := \frac{1}{n} \sum_{t=1}^{n} r_{t,j}$. Also set r^*_j as the j th 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 j th one. Let $r^*_{t,-j}$ represent the $p - 1$ vector of excess returns for all except the j th one. Also set $\hat{\mu} := \bar{r}$.

To understand the assumptions we define a model that will link us to 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 $K_1 > 0, K_2 > 1$, $0 < \alpha_1 \leq 2$, $0 < \alpha_2 \leq 2$ 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 $Eigmax(\Sigma^{-1}) \leq K < \infty$, wher K is a positive constant. (ii). Also 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 k, with constants*

$K_3 > 0, \alpha_3 > 0$ that are independent of p and n. Set $\rho = min([\frac{1}{\alpha_1} + \frac{1}{\alpha_2} + \frac{1}{\alpha_3}]^{-1}, [\frac{1}{2\alpha_1} + \frac{1}{\alpha_3}]^{-1})$. Also $lnp = o(n^{\rho/(2-\rho)})$. 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 use exponential tail inequalities used by Chang et al. (2019). Assumption 2(ii) does not allow zero return for all assets, also all returns should be finite. For technical and practical reasons, we do not allow local to zero returns too. Assumption 2 prevents the case of zero maximum Sharpe ratio. Assumption 3 allows for weak dependence in 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 first and fourth result of Lemma 1 in p.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 precision matrix formula. This subsection is taken from Callot et al. (2019), we repeat so that it will become clear how precision matrix estimate is derived in the next subsection. 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 strictly stationary time series excess asset return. This is an extension of 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. Also $\Sigma_{j,-j}$ is the j th row of Σ with j th element removed. Then $\Sigma_{-j,j}$ represent the j th column of Σ with its j th element removed. From the inverse formula for the block matrices, we have the following for the j th 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 j th row of Θ with j th 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 get a solution as

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

by using strict stationary 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 $\Sigma_{-j,j} := Er^*_{t,-j}r^*_{t,j}$, $\Sigma_{-j,-j} := Er^*_{t,-j}r^{*'}_{t,-j}$. By (1), $\eta_{t,j} := r^*_{t,j} - (r^*_{t,-j})'\gamma_j$. Then it is easy to see by (4),

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

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 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)

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

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

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

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 get (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 seen in the abstract, the idea of nodewise regression is developed by Meinshausen and Bühlmann (2006). Nodewise regression stems from the idea of the neighborhood selection. In a portfolio, neighborhood selection (nodewise regression) will select a "neighborhood" of a j th asset return (excess) in a way that the smallest subset of return of other assets in a portfolio will be conditionally dependent with this j th asset return. All the conditionally independent assets will receive a zero in 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. 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 $\mathcal{A} \subseteq \Gamma_{-j}$, and $\Gamma_{-j} = \Gamma - \{j\}$, and $\Gamma = \{1, 2, \cdots, j, \cdots, p\}$. This is equation (2) in Meinshausen and Bühlmann (2006), and there is detailed explanation for this result.

2.4 Estimate

A possible way of estimating precision matrix when the number of assets is larger than the sample size is by nodewise regression. In 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 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 as, for each $j = 1, 2, \cdots, p$

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

where λ_j is a positive tuning parameter (sequence) and its choice which will be discussed in simulation section. Let S_j be the set of coefficients which 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 $\bar{s} = \max_{1\le j\le p} s_j$. So we set a sparsity assumption. Alternative, but costly in notation, is weak sparsity where we allow for absolute l th power sum of coefficients in each row of the precision matrix to be diverging, but not a faster rate than the sample size. This of course demands a larger tuning parameter compared with sparsity assumption in practice. It is easy to incorporate weak sparsity to the proofs as can be

seen in Lemma 2.3 of van de Geer (2016). In order not to prolong the paper we have not pursued this track, and required sparsity.

2. Setup the following matrix, that will be 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 $\hat{\tau}_j^2$, $j = 1, \cdots, p$

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

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

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

(ii).

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

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

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. Sparsity of precision matrix does not imply covariance matrix to be sparse as well. It is possible to have Toeplitz structure in covariance matrix which is non-sparse but sparsity in precision matrix.

2.5 Why use Nodewise Regression?

In finance, our method leads to considering more complicated cases of $p > n$, and $p/n \to \infty$, when both $p, n \to \infty$. We also allow $p = n$ case where this is a hindrance to technical analysis is 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 model and shrinkage ones, and carry a certain optimality property as outlined in subsection 2.3. Also with our technique, we can get mean-variance efficiency even when $p > n$ in case of the maximum out-of-sample maximum 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). There are no constraints in portfolio weights unlike section 4.1. Equation (A.2) of Ao et al. (2019) defines the estimated maximum out of sample ratio when $p < n$, with inverse of sample covariance matrix used as an estimator for 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 which can even estimate the theoretical Sharpe ratio when $p > n$. Our maximum out of sample Sharpe ratio estimate by using 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 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 their Theorem 1. After that they provide an estimable shrinkage function even in the cases of $p/n \to r_1 \in (0,1) \cup (1,+\infty)$. Their proofs allow for iid data.

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

3.1 Mean Variance Efficiency When $p > n$

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

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

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 $\mu' \hat{w}_{oos}$. Also 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 from 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 rate of convergence results.

Theorem 3. *Under Assumptions 1-4*

(i).

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

(ii).

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

Remarks. 1. From the results clearly we allow $p > n$, and still there is consistency. Also 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 $\bar{s} = o(\sqrt{n/lnp})$.

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

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

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, 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 maximal Sharpe Ratio is constrained to have weights of the portfolio adding up 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:

$$MSR^2 = \mu'\Sigma^{-1}\mu. \tag{13}$$

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

$$MSR_c^2 = \mu'\Sigma^{-1}\mu - (1'_p\Sigma^{-1}\mu)^2/(1'_p\Sigma^{-1}1_p). \tag{14}$$

These are the equations (6.1) of Maller et al. (2016). Equation (13) is used in the literature, and this is the formula when the weights do not necessarily add up to one given a return constraint as in Ao et al. (2019).

These can be estimated by their sample counterparts, but in 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 was denoted by $\hat{\Theta}$. So we analyze the asymptotic behaviour of estimate of the maximal Sharpe Ratio squared via nodewise regression. We will also introduce maximum Sharpe ratio which takes care of the uncertainty about 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 of all, when $1_p'\Sigma^{-1}\mu \geq 0$, we have the square of maximum Sharpe ratio as in (13). To get an estimate by using nodewise regression we replace Σ^{-1} with $\hat{\Theta}$. Namely, estimate of the square of maximum Sharpe ratio is:

$$\widehat{MSR}^2 = \hat{\mu}'\hat{\Theta}\hat{\mu}. \tag{15}$$

Using the result in Theorem 1 we can obtain the consistency of 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. As far as we can search we are not aware of a result that deals with MSR when $p > n$, and p can grow exponentially in n. We also allow for time series, and establish a rate of convergence. The rate shows that precision matrix non-sparsity can affect estimation error badly. The number of assets on the other hand can also increase the error by in a logarithmic scale.

Note that maximum Sharpe ratio above rely on $1_p'\Sigma^{-1}\mu \geq 0$, where 1_p is a column vector of ones. This has been pointed out recently in equation (6.1) Maller et al. (2016). If $1_p'\Sigma^{-1}\mu < 0$ the Sharpe ratio is minimized as shown in p.503 of Maller and Turkington (2002). The new maximal Sharpe ratio in the case when $1_p'\Sigma^{-1}\mu < 0$ are 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 (14).

An estimator in this case is,

$$\widehat{MSR}_c^2 = \hat{\mu}'\hat{\Theta}\hat{\mu} - (1_p'\hat{\Theta}\hat{\mu})^2/(1_p'\hat{\Theta}1_p). \tag{16}$$

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 are 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 used in Callot et al. (2019) and is a condition that helps us to get a finite optimal portfolio variance in Markowitz (1952) mean variance portfolio below.

2. Exactly in Theorem 4, we allow $p > n$, and also time series data is allowed unlike 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 term can be consistently estimable, and this is shown in Lemma A.3 in the Supplement 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 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, $|1_p'\Sigma^{-1}\mu|$ below in

Theorem 6, , we show that by (A.56)(A.57)(A.60)(A.61) 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 is 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 $\beta - min$ condition in high dimensional statistics to achieve model selection. Also see 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 that this double sum in p will be a small number, unless the terms in the sum cancel each other. So we exclude that type of case with our assumption. Also ϵ is not arbitrary, from the proof this is the upper bound on the $|\hat{B} - B|$ in Lemma A.3 in Supplement Appendix, and it is of order

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

where the asymptotically small term is by Assumption 4.

2. In the case of $p > n$ we only consider consistency since standard Central Limit Theorems (apart from the ones in rectangles, or sparse convex sets) do not apply, and ideas such as multiplier bootstrap and empirical bootstrap with self normalized moderate deviations result do not extend to this specific Sharpe ratio formulation.

3. This is a new result taking into account 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 as well.

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 get the estimation error to converge to zero we need $p\sqrt{lnp} = o(n^{1/2})$. In non-sparse precision matrix case case clearly we allow only $p << n$.

5 Commonly Used Portfolios with Large Number of Assets

Here we provide consistent estimates of Sharpe ratio of Global Minimum Variance portfolio, 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 Global Minimum Variance Portfolio. This is the portfolio that weights are chosen to minimize the variance of the portfolio subject to weights adding up to one. Specifically

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

Mainly this is similar to maximum Sharpe ratio problem but we minimize the square of denominator in the Sharpe ratio definition subject to the same constraint in the maximum Sharpe ratio case above. 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 in the Sharpe ratio formula, normalized by number of assets

$$SR = \frac{w_u' \mu}{\sqrt{w_u' \Sigma w_u}} = \sqrt{p} (\frac{1_p' \Sigma^{-1} \mu}{p}) (\frac{1_p' \Sigma^{-1} 1_p}{p})^{-1/2}. \tag{17}$$

We estimate (17) by nodewise regression

$$\widehat{SR}_{nw} = \sqrt{p} (\frac{1_p' \hat{\Theta} \hat{\mu}}{p}) (\frac{1_p' \hat{\Theta} 1_p}{p})^{-1/2}. \tag{18}$$

As far as we know the following theorem is a new result in the literature when $p > n$, and establishes both consistency and rate of convergence in case of Sharpe ratio in 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 large p only affects the error by a logarithmic factor. Estimation error increases with non-sparsity of 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

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

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 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:

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

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 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 are larger than the sample size. This is a new 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 *is 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 $AD - B^2 \geq C_1 > 0$ shows 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 rate of convergence of estimators and it 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 see whether our method achieves consistency under a sparse setup, and also, check under two different setups, how our method performs compared to others in estimation of 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 also benefiting from high dimensional techniques are: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 Estimated and Sparse Regression (MAXSER) from Ao et al. (2019). All models except by the MAXSER are plug-in estimators where the first step is to estimate the precision 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 allowing 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 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 of that determines the amount of shrinkage of each eigenvalue in the covariance matrix individually 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 great idea. Ledoit and Wolf (2017) proposed 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 method SF-NL-LW 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 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.

As for implementations, 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) package for the 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 biggest 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 range of all possible tuning parameters

$$GIC_j(\lambda_j) = \frac{SSR(\lambda_j)}{n} + q(\lambda_j)\log(p-1)\frac{\log(\log(n))}{n} \tag{19}$$

where $SSR(\lambda_j)$ is the sum squared errors for a given λ_j, $q(\lambda_j)$ is the number of variables, given a λ_j, in the model that are 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 in k subsamples and fit the model for a range of λ_j

just like in the GIC procedure. However we will fit models in the subsamples. We estimate the models always 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 test we finally compute the average MSE across all subsamples and select the λ_j for each asset j that yield 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):

$$r_j = \alpha_j + \sum_{k=1}^{K} \beta_{j,k} f_k + e_j, \tag{20}$$

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 on 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 β's, α's 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 of 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.

Tables 1 and 2 show the results. The values in each cell show average absolute estimation error for estimating square of Sharpe ratio in case of Global Minimum Variance and Markowitz mean-variance portfolios in Section 5, and Maximum Sharpe ratio in case of constrained portfolio optimization, and in out-of-sample forecasting in Sections 3-4 respectively, 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 respectively the max Sharpe Ratio, the out-of-sample max

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

Sharpe Ratio, the Sharpe Ratio from the Global Minimum Variance and the Sharpe Ratio from the Markowitz portfolio with target returns set to 1%. So 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. All bold face entries in Tables show category champions.

To start with Table 1, we clearly see our method performs very well under a sparse Toeplitz scenario. When the correlation is 0.5, and 0.75, our method has the smallest error among all others with MSR and MSR-OOS. We also see that with GMV-SR and MKW-SR scenarios, SF-NL-LW method does the best generally. To give a specific example, with $N = 400, p = 600$ and $\rho = 0.75$, our OOS-MSR error is 0.118 (GIC based nodewise) and the second best is our CV based nodewise with 0.259 and third one is: SF-NL-LW has 0.868 error. On the other hand, in 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 methods.

Also we see that consistency is achieved with our methods as our theorems suggest under sparse scenario as in Table 1. To see this, with $N = 100, p = 150$, our error in 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 does the best in minimizing the errors for constrained Markowitz-Sharpe ratio, and Global Minimum Variance and Markowitz mean-variance portfolio. We also note that MAXSER gets the best results generally in estimating out-of-sample Maximum Sharpe ratio when $p = N/2$.

Table 1: Simulations Results - Toeplitz DGP

Toeplitz DGP $\rho = 0.25$

	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
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 average absolute estimation error for estimating square of Sharpe ratio in case of Global Minimum Variance and Markowitz mean-variance portfolios in Section 5, and Maximum Sharpe ratio in case of constrained portfolio optimization, and in out-of-sample forecasting in Sections 3-4 respectively, across iterations.

Table 2: Simulations Results - Factor DGP

Factor DGP $\gamma = 1$

	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
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 average absolute estimation error for estimating square of Sharpe ratio in case of Global Minimum Variance and Markowitz mean-variance portfolios in Section 5, and Maximum Sharpe ratio in case of constrained portfolio optimization, and in out-of-sample forecasting in Sections 3-4 respectively, across iterations.

7 Empirical Application

For the empirical application, we used data from stocks covering the monthly period between January 1995 to December 2017. Our data has 395 stocks that were in the S&P 500 during the period. Given that this is an out-of-sample competition between models, we only estimated GMV and Markowitz portfolios for the plug-in estimators. The MAXSER has its own way of recovering weights for the Max Sharpe Ratio. Our out-of-sample periods are from January 2010 to December 2017 and from January 2006 to December 2017. These periods are chosen to represent first, a non-recession period, and then a period including the great recession of 2008-2009. This out-of-sample periods reflect the recent history.

The Markowitz return constraint ρ_1 is 2% and the MAXSER risk constraint is 4%. In the low dimension experiment, we randomly selected 100 stock from the pool to estimate the models. In the high dimensional case, we use all 395 stocks.

We used 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 in 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 was 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 estimated 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 at period $t + 1$, on 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 due to the fact that the portfolio in the end of the period has changed compared to the portfolio in 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 being 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,

$$H_0 \ : \ SR_{Best} \leq SR_0 \ \ vs \ \ H_a \ : \ SR_{Best} > SR_0, \tag{21}$$

where SR_{Best} is the model with the biggest Sharpe Ratio, which is tested against all remaining models. This is the Jobson and Korkie (1981) test with Memmel (2003) correction.

In empirics section we also included Equally Weighted Portfolio (EW). GMV-NW-GIC, GMV-NW-CV denote the nodewise method with GIC and Cross validation tuning parameter choices respectively in Global Minimum Variance Portfolio(GMV). GMV-POET, GMV-NL-LW, GMV-SF-NL-LW denote the POET, Nonlinear Shrinkage, Single Factor Nonlinear Shrinkage methods are described in the simulation section and used in Global Minimum Variance Portfolio as well. 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 Markowitz portfolio. All the other methods with MW headers are self-explanatory in the same way.

The results are presented in Tables 3 and 4. Table 3 shows the results for the 2010-2017 subsample, which is basically a expansion period. Nodewise using the GIC to select the tuning parameter in a GMV portfolio had the biggest Sharpe-Ratios in all cases except the low-dimensional case with no transaction costs, where MAXSER performed very well. However, the subpool procedure required by the MAXSER makes it highly affected by transaction costs because the portfolio changes a lot between periods.

To give an example, with Transaction Costs in the high dimensional portfolio category, Sharpe-Ratio (SR) (averaged over out-of-sample time period), GMV-NW-GIC is the best. It has the SR of 0.375. GMV-POET, GMV-NL-LW, GMV-SF-NL-LW have SR of 0.263, 0.265, 0.266 respectively. These are all statistically different from our nodewise method at 10% level. 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 0.314. So even in sub-category wise nodewise method dominates.

Table 4 shows the results for the 2006-2017 sub-sample. In this case we have the entire sub-prime crisis in the test sample and the Sharpe Ratios are smaller for all models. We see again in all four categories, low dimension without transaction costs, high dimension without transaction costs, and low dimension with transaction costs, high dimension with transaction costs, Nodewise methods dominate. Specifically, GMV-NW-GIC, GMV-NW-CV, GMV-NW-GIC, GMV-NW-GIC are the leaders in each of the four categories explained above, respectively. However, some of the alternatives are not statistically significant in this Table 4. In this Table with Transaction Costs in the low dimensional category, GMV-NW-GIC has the highest SR with 0.220, but this is not stastistically significantly different from its alternatives at 10% level.

Table 3: Empirical Results - Subperiod from Jan. 2010 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	0.309	0.013	0.041	0.241	0.318	0.013	0.040	0.008	0.302	0.012	0.041	0.023	0.310	0.012	0.040	0.010
GMV-NW-GIC	0.383	0.012	0.032	0.446	**0.385**	0.012	0.032		**0.370**	0.012	0.032		**0.375**	0.012	0.032	
GMV-NW-CV	0.377	0.012	0.032	0.426	0.384	0.012	0.032	0.370	0.358	0.012	0.032	0.014	0.365	0.012	0.032	0.000
GMV-POET	0.249	0.013	0.051	0.131	0.270	0.013	0.049	0.002	0.242	0.012	0.051	0.006	0.263	0.013	0.049	0.003
GMV-NL-LW	0.251	0.013	0.050	0.135	0.272	0.013	0.048	0.002	0.245	0.012	0.050	0.006	0.265	0.013	0.048	0.003
GMV-SF-NL-LW	0.254	0.013	0.050	0.139	0.273	0.013	0.048	0.002	0.247	0.012	0.050	0.006	0.266	0.013	0.048	0.003
MW-NW-GIC	0.374	0.012	0.033	0.402	0.359	0.011	0.032	0.213	0.333	0.011	0.033	0.180	0.314	0.010	0.032	0.034
MW-NW-CV	0.360	0.012	0.034	0.353	0.367	0.011	0.030	0.304	0.309	0.011	0.034	0.064	0.294	0.009	0.030	0.015
MW-POET	0.329	0.017	0.051	0.282	0.309	0.015	0.048	0.059	0.321	0.016	0.051	0.186	0.296	0.014	0.048	0.055
MW-NL-LW	0.297	0.016	0.054	0.200	0.203	0.012	0.058	0.003	0.285	0.015	0.054	0.063	0.179	0.010	0.058	0.002
MW-SF-NL-LW	0.297	0.016	0.053	0.199	0.204	0.012	0.057	0.003	0.286	0.015	0.053	0.062	0.181	0.010	0.058	0.001
MAXSER	**0.398**	0.015	0.038						0.183	0.007	0.038	0.046				

The table shows the Sharpe Ratio (SR), Average Returns (Avg), Standard Deviation(SD) and the p-value of the Jobson and Korkie (1981) test with the Memmel (2003) correction for all portfolios. The test was always performed using the model with the biggest Sharpe Ratio against all other models. The statistics were calculated from 96 rolling windows covering Jan. 2010 to Dec. 2017 and the estimation window size was of 180 observations.

Note that to understand better why we perform well in the out-of-sample exercise we put correlation matrices for the two periods that we analyzed. Sub Sample 1 corresponds to January 2010-December 2017, and the Sub Sample 2 corresponds to January 2006-December 2017. in Figures 2a, 2b we colored the correlation of assets. Blue (dark in black and white printer) is anything above 0.3 positive correlation (which is the average), and yellow is anything between 0 and 0.3 positive (light gray in black and white printer), and for very few negative correlations red color (dark for large negative correlations in black and white printer). Figures 2a-2b clearly show that dark blue areas are not dominant in the picture. This is close to our assumptions where the large correlations between assets should not dominate the correlation matrix of assets.

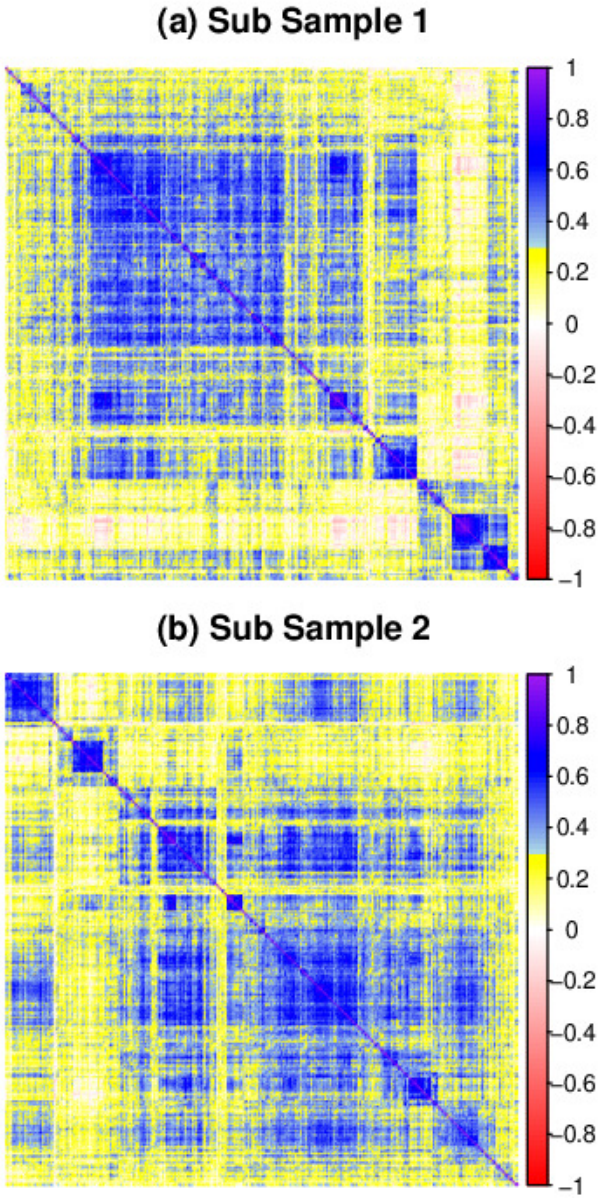

Figure 2: Data Correlation Matrices

Table 4: Empirical Results - Subperiod from Jan. 2006 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	0.196	0.010	0.049	0.085	0.197	0.010	0.051	0.162	0.190	0.009	0.049	0.109	0.191	0.010	0.051	0.203
GMV-NW-GIC	**0.230**	0.009	0.039		0.216	0.009	0.041	0.147	**0.220**	0.009	0.039		**0.208**	0.009	0.041	
GMV-NW-CV	0.226	0.009	0.040	0.162	**0.218**	0.009	0.041		0.210	0.008	0.040	0.015	0.203	0.008	0.041	0.005
GMV-POET	0.170	0.011	0.062	0.065	0.178	0.011	0.061	0.114	0.164	0.010	0.062	0.078	0.172	0.011	0.061	0.136
GMV-NL-LW	0.170	0.010	0.061	0.061	0.179	0.011	0.060	0.117	0.164	0.010	0.061	0.074	0.174	0.010	0.060	0.139
GMV-SF-NL-LW	0.172	0.010	0.060	0.065	0.180	0.011	0.060	0.118	0.166	0.010	0.060	0.077	0.174	0.010	0.060	0.140
MW-NW-GIC	0.211	0.008	0.037	0.362	0.193	0.007	0.036	0.309	0.170	0.006	0.037	0.176	0.152	0.005	0.036	0.130
MW-NW-CV	0.201	0.008	0.038	0.285	0.192	0.007	0.034	0.327	0.146	0.006	0.038	0.079	0.120	0.004	0.035	0.066
MW-POET	0.204	0.012	0.057	0.277	0.167	0.009	0.053	0.115	0.196	0.011	0.057	0.290	0.156	0.008	0.053	0.108
MW-NL-LW	0.170	0.010	0.060	0.100	0.081	0.005	0.064	0.012	0.159	0.009	0.060	0.093	0.059	0.004	0.064	0.007
MW-SF-NL-LW	0.174	0.010	0.059	0.115	0.084	0.005	0.064	0.013	0.163	0.010	0.059	0.109	0.063	0.004	0.064	0.008
MAXSER	0.180	0.009	0.051	0.328					-0.008	-0.000	0.052	0.014				

The table shows the Sharpe Ratio (SR), Average Returns (Avg), Standard Deviation (SD) and the p-value of the Jobson and Korkie (1981) test with the Memmel (2003) correction for all portfolios. The test was always performed using the model with the biggest Sharpe Ratio against all other models. The statistics were calculated from 144 rolling windows covering Jan. 2006 to Dec. 2017 and the estimation window size was of 132 observations.

8 Conclusion

In this paper we provide a nodewise regression method that can control the risk and get the maximum expected return of a large portfolio. Our result is new and holds even when $p > n$. We also show maximum-out of sample Sharpe ratio can be estimated consistently. Furthermore, we also develop a formula for maximum Sharpe ratio when weights of the portfolio add up to one. A consistent estimate for the constrained case is also shown. Then we extend our results to 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 portfolio.

Appendix

This appendix has the proofs.

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

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

The proof will consider the numerator and the denominator of the squared maximum out of sample Sharpe ratio. We start with the numerator.

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

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

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

where we use (A.99)-(A.101) for the rates and the dominant rate at last equality is by Assumption 4. Next we analyze the denominator in (A.2). By definition $\Theta := \Sigma^{-1}$, and using (A.41) the denominator is bounded away from zero.

Then by Assumptions 2, 4, (A.3)(A.41)

$$\frac{\mu'\hat{\Theta}\hat{\mu}/p}{\mu'\Theta\mu/p} \leq \frac{|\mu'\hat{\Theta}\hat{\mu} - \mu'\Theta\mu|/p}{\mu'\Theta\mu/p} + 1 = o_p(1) + 1. \tag{A.4}$$

We now try to show that the denominator

$$\frac{\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu}}{\mu'\Sigma^{-1}\mu} \xrightarrow{p} 1. \tag{A.5}$$

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

$$\frac{\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu}}{\mu'\Sigma^{-1}\mu} = \frac{\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu} - \mu'\Theta\Sigma\Theta\mu}{\mu'\Theta\Sigma\Theta\mu} + 1 \geq 1 - \left|\frac{\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu} - \mu'\Theta\Sigma\Theta\mu}{\mu'\Theta\Sigma\Theta\mu}\right|. \tag{A.6}$$

We can write

$$\hat{\Theta}\hat{\mu} - \Theta\mu = (\hat{\Theta} - \Theta)\hat{\mu} + \Theta(\hat{\mu} - \mu). \tag{A.7}$$

Using (A.7)

$$
\begin{aligned}
|\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu} - \mu'\Theta\Sigma\Theta\mu| &\leq |[(\hat{\Theta}-\Theta)\hat{\mu}]'\Sigma[(\hat{\Theta}-\Theta)\hat{\mu}]|. && \text{(A.8)}\\
&+ 2|[(\hat{\Theta}-\Theta)\hat{\mu}]'\Sigma\Theta(\hat{\mu}-\mu)| && \text{(A.9)}\\
&+ 2|[(\hat{\Theta}-\Theta)\hat{\mu}]'\Sigma\Theta\mu| && \text{(A.10)}\\
&+ |[\Theta(\hat{\mu}-\mu)]'\Sigma[\Theta(\hat{\mu}-\mu)]| && \text{(A.11)}\\
&+ 2|[\Theta(\hat{\mu}-\mu)]'\Sigma\Theta\mu| && \text{(A.12)}
\end{aligned}
$$

First we consider (A.8)

$$
\begin{aligned}
|\hat{\mu}'(\hat{\Theta}-\Theta)'\Sigma(\hat{\Theta}-\Theta)\hat{\mu}| &\leq Eigmax(\Sigma)\|(\hat{\Theta}-\Theta)\hat{\mu}\|_2^2\\
&= Eigmax(\Sigma)[\sum_{j=1}^{p}\{(\hat{\Theta}_j-\Theta_j)'\hat{\mu}\}^2]\\
&\leq Eigmax(\Sigma)p\max_{1\leq j\leq p}[(\hat{\Theta}_j-\Theta_j)'\hat{\mu}]^2\\
&\leq Eigmax(\Sigma)p[\max_{1\leq j\leq p}\|\hat{\Theta}_j-\Theta_j\|_1^2]\|\hat{\mu}\|_\infty^2\\
&= O(1)pO_p(\bar{s}^2\frac{lnp}{n})O_p(1), && \text{(A.13)}
\end{aligned}
$$

where we use Holder's inequality for the third inequality, and Theorem 1(i), (ii) and Assumption 2 for the rate. Now consider (A.9), and by definition $\Theta := \Sigma^{-1}$.

$$
\begin{aligned}
|[(\hat{\Theta}-\Theta)\hat{\mu}]'\Sigma\Theta(\hat{\mu}-\mu)| &= |\hat{\mu}'(\hat{\Theta}-\Theta)'(\hat{\mu}-\mu)|\\
&\leq |(\hat{\mu}-\mu)'(\hat{\Theta}-\Theta)'(\hat{\mu}-\mu)| + |\mu'(\hat{\Theta}-\Theta)'(\hat{\mu}-\mu)|\\
&= p[O_p(\bar{s}(\frac{lnp}{n})^{3/2}) + O_p(\bar{s}(\frac{lnp}{n}))]\\
&= pO_p(\bar{s}(\frac{lnp}{n})), && \text{(A.14)}
\end{aligned}
$$

by (A.97)(A.100) for the second equality, and the dominant rate in third equality can be seen by Assumption 4. Next, consider (A.10), and remembering $\Theta := \Sigma^{-1}$

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

where we use (A.100)(A.101) for the second equality, and the dominant rate in the third equality can be seen by Assumption 4. Consider now (A.11) by symmetry of $\Theta = \Sigma^{-1}$

$$
\begin{aligned}
|[\Theta(\hat{\mu}-\mu)]'\Sigma\Theta(\hat{\mu}-\mu)| &= |(\hat{\mu}-\mu)'\Theta(\hat{\mu}-\mu)|\\
&= pO_p(\bar{s}^{1/2}\frac{lnp}{n}) && \text{(A.16)}
\end{aligned}
$$

by (A.98). Next analyze (A.12) by symmetricity of $\Theta = \Sigma^{-1}$

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

by (A.99). Combine the rates and terms (A.13)-(A.17) in (A.8)-(A.12) to have

$$|\hat{\mu}'\hat{\Theta}\Sigma\hat{\Theta}\hat{\mu} - \mu'\Theta\Sigma\Theta\mu| = pO_p(\bar{s}\sqrt{\frac{lnp}{n}}), \tag{A.18}$$

by the dominant rate in (A.15), as seen in Assumption 4.

See that by $\Theta = \Sigma^{-1}$

$$\mu'\Theta\Sigma\Theta\mu = \mu'\Sigma^{-1}\mu \geq Eigmin(\Sigma^{-1})\|\mu\|_2^2 \geq c\|\mu\|_2^2, \tag{A.19}$$

by Assumption 2.

Combine (A.18)(A.19) in the second term on the right side of (A.6) to have by Assumption 2, Assumption 4

$$\frac{|\hat{\mu}'\hat{\Theta}'\Sigma\hat{\Theta}\hat{\mu} - \mu'\Theta\Sigma\Theta\mu|/p}{\mu'\Theta\Sigma\Theta\mu/p} \leq \frac{cO_p(\bar{s}\sqrt{\frac{lnp}{n}})}{c\|\mu\|_2^2/p} = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \tag{A.20}$$

So we showed (A.5). Then combine (A.4)(A.5) in (A.1) to have the desired result. **Q.E.D.**

Proof of Theorem 3. (i). First by Holder's inequality

$$\left|\frac{\mu'\hat{w}_{oos}}{\mu'w_{oos}} - 1\right| \leq \frac{\|\mu\|_\infty\|\hat{w}_{oos} - w_{oos}\|_1}{|\mu'w|}. \tag{A.21}$$

In (A.21) consider the denominator, by w_{oos} definition and $\Theta := \Sigma^{-1}$, by Assumption 2

$$|\mu'w_{oos}| = \sigma\left|\frac{\mu'\Theta\mu}{\sqrt{\mu'\Theta\mu}}\right| = \sigma|\sqrt{\mu'\Theta\mu}| \geq \sigma c^{1/2}c_l p^{1/2}, \tag{A.22}$$

by (A.41). We consider the numerator in (A.21). Next, set a $p \times 1$ vector as $\hat{y} = \sigma\hat{\Theta}\hat{\mu}$, and another $p \times 1$ vector $y = \sigma\Theta\mu$, a positive scalar $x = (\mu'\Theta\mu)^{1/2}$, and its estimate as $\hat{x} = (\hat{\mu}'\hat{\Theta}\hat{\mu})^{1/2}$. This is done to clarify the algebraic steps without going through burdensome notation. Now analyze the weights in the numerator in (A.21).

$$\|\hat{w}_{oos} - w_{oos}\|_1 = \left\|\frac{\sigma\hat{\Theta}\hat{\mu}}{\sqrt{\hat{\mu}'\hat{\Theta}\hat{\mu}}} - \frac{\sigma\Theta\mu}{\sqrt{\mu'\Theta\mu}}\right\|_1 = \|\frac{\hat{y}}{\hat{x}} - \frac{y}{x}\|_1.$$

We can further simplify the above expression as

$$\begin{aligned} \|\frac{\hat{y}}{\hat{x}} - \frac{y}{x}\|_1 &= \|\frac{\hat{y}x - yx + yx - \hat{x}y}{\hat{x}x}\|_1 \\ &= \|\frac{(\hat{y}-y)x + (x-\hat{x})y}{(\hat{x}-x)x + x^2}\|_1 \\ &\leq \frac{[\|\hat{y}-y\|_1]x}{x^2 - |\hat{x}-x|x} + \frac{[|\hat{x}-x|]\|y\|_1}{x^2 - |\hat{x}-x|x}, \end{aligned} \tag{A.23}$$

where we use add and subtract yx to the numerator for the first equality, and triangle inequality for the inequality. Since we simplified a bit on the expected return estimation, we turn to all the elements in (A.23). First

$$\begin{aligned} \|\hat{y} - y\|_1 &= \sigma \|\hat{\Theta}\hat{\mu} - \Theta\mu\|_1 \\ &= \sigma \|(\hat{\Theta} - \Theta + \Theta)(\hat{\mu} - \mu + \mu) - \Theta\mu\|_1 \\ &\leq \sigma \left(\|(\hat{\Theta} - \Theta)(\hat{\mu} - \mu)\|_1 + \|\Theta(\hat{\mu} - \mu)\|_1 + \|(\hat{\Theta} - \Theta)\mu\|_1 \right), \end{aligned} \tag{A.24}$$

where we use simple add and subtract and then triangle inequality. In (A.24) consider each of the right side terms. By Lemma A.1 and Theorem 1

$$\|(\hat{\Theta} - \Theta)(\hat{\mu} - \mu)\|_1 \leq p \max_{1 \leq j \leq p} \|\hat{\Theta}_j - \Theta_j\|_1 \|\hat{\mu} - \mu\|_\infty = pO_p(\bar{s}\sqrt{lnp/n})O_p(\sqrt{lnp/n}). \tag{A.25}$$

By Lemma A.1, (B.55) of Caner and Kock (2018) and Theorem 1

$$\|\Theta(\hat{\mu} - \mu)\|_1 \leq p \max_{1 \leq j \leq p} \|\Theta_j\|_1 \|\hat{\mu} - \mu\|_\infty = pO(\sqrt{\bar{s}})O_p(\sqrt{lnp/n}). \tag{A.26}$$

Next analyze the third term in (A.7)

$$\|(\hat{\Theta} - \Theta)\mu\|_1 \leq p \max_{1 \leq j \leq p} \|\hat{\Theta}_j - \Theta_j\|_1 \|\mu\|_\infty = pO_p(\bar{s}\sqrt{lnp/n})O(1), \tag{A.27}$$

by Lemma A.1, Theorem 1 and Assumption 2. By Assumption 4, the slowest rate in (A.24) is by (A.27), so

$$\|\hat{y} - y\|_1 = \sigma \|\hat{\Theta}\hat{\mu} - \Theta\mu\|_1 = pO_p(\bar{s}\sqrt{lnp/n}). \tag{A.28}$$

Now analyze

$$x = (\mu'\Theta\mu)^{1/2} \leq K^{1/2} c_u [p]^{1/2}, \tag{A.29}$$

where we use $\Theta := \Sigma^{-1}$ definition, Lemma A.5 and Assumption 2. Also by Assumption 2

$$x^2 \geq c c_l^2 p. \tag{A.30}$$

Next consider

$$|\hat{x} - x| = |(\hat{\mu}'\hat{\Theta}\hat{\mu})^{1/2} - (\mu'\Theta\mu)^{1/2}| = \left| \frac{\hat{\mu}'\hat{\Theta}\hat{\mu} - \mu'\Theta\mu}{(\hat{\mu}'\hat{\Theta}\hat{\mu})^{1/2} + (\mu'\Theta\mu)^{1/2}} \right|. \tag{A.31}$$

Analyze (A.31) by Lemma A.4

$$|\hat{\mu}'\hat{\Theta}\hat{\mu} - \mu'\Theta\mu| = O_p(\bar{s}\sqrt{lnp/n}).$$

Next by Assumption 4 and (A.29)(A.30) via Assumption 2

$$(\hat{\mu}'\hat{\Theta}\hat{\mu})^{1/2} + (\mu'\Theta\mu)^{1/2} \geq [c c_l^2 p - o_p(1)]^{1/2} + [c c_l^2 p]^{1/2} > 0.$$

Combine the last two results in (A.31) to have

$$|\hat{x} - x| = |(\hat{\mu}'\hat{\Theta}\hat{\mu})^{1/2} - (\mu'\Theta\mu)^{1/2}| = O_p(\frac{\bar{s}}{p^{1/2}}\sqrt{lnp/n}). \tag{A.32}$$

Consider by Lemma A.1, Assumption 2, and (B.55) of Caner and Kock (2018)

$$\begin{aligned}\|y\|_1 &= \frac{\sigma\|\Theta\mu\|_1}{(\mu'\Theta\mu)^{1/2}} \\ &\leq \frac{\sigma p \max_{1\leq j\leq p}\|\Theta_j\|_1\|\mu\|_\infty}{c^{1/2}c_l p^{1/2}} \\ &= \frac{\sigma p O(\bar{s}^{1/2})O(1)}{c^{1/2}c_l p^{1/2}} \\ &= O(\bar{s}^{1/2}p^{1/2}). \end{aligned} \tag{A.33}$$

Now we have all the terms in (A.23). We start with the first term on the right side of (A.23). To analyze the denominator see that by Assumption 4, we have $|\hat{x}-x| = o_p(1)$ in (A.32). Then by (A.29)(A.30) and $|\hat{x}-x| = o_p(1)$ as shown above

$$\frac{1}{x^2 - |\hat{x}-x|x} \leq \frac{1}{cc_l^2 p - o_p(1)K^{1/2}c_u p^{1/2}}. \tag{A.34}$$

To analyze the numerator for the first right side term in (A.23), we use (A.28)(A.29)

$$\|\hat{y}-y\|_1 x \leq pO_p(\bar{s}\sqrt{lnp/n})K^{1/2}c_u p^{1/2} = O_p(p^{3/2}\bar{s}\sqrt{lnp/n}). \tag{A.35}$$

Combine (A.34)(A.35) to have

$$\frac{\|\hat{y}-y\|_1 x}{x^2 - |\hat{x}-x|x} = O_p(p^{1/2}\bar{s}\sqrt{lnp/n}). \tag{A.36}$$

Then analyze the second right side term in (A.23) by (A.32)(A.33)(A.34)

$$\frac{|\hat{x}-x|\|y\|_1}{x^2 - |\hat{x}-x|x} = \frac{O_p(\bar{s}/p^{1/2}\sqrt{lnp/n})O(\bar{s}^{1/2}p^{1/2})}{cc_l^2 p - o_p(1)K^{1/2}c_u p^{1/2}} = O_p(\bar{s}\sqrt{lnp/n})O(\bar{s}^{1/2}/p). \tag{A.37}$$

Since $\bar{s}^{1/2} \leq p^{1/2}$, the rate in (A.36) is slower than than the one in (A.37), hence by (A.23)

$$\|\hat{w}_{oos} - w_{oos}\|_1 = O_p(p^{1/2}\bar{s}\sqrt{lnp/n}). \tag{A.38}$$

Use (A.22) with (A.38) to have (A.21) as

$$\begin{aligned}\left|\frac{\mu'\hat{w}_{oos}}{\mu' w_{oos}} - 1\right| &= O_p(\frac{p^{1/2}\bar{s}\sqrt{lnp/n}}{p^{1/2}}) \\ &= O_p(\bar{s}\sqrt{lnp/n}) \\ &= o_p(1), \end{aligned}$$

where the last step is by Assumption 4.**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.5)(A.20)

$$\left|\frac{\hat{\mu}'\hat{\Theta}'\Sigma\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu} - 1\right| = O_p(\bar{s}\sqrt{lnp/n}). \tag{A.39}$$

Also by Lemma A.4, Assumptions 2 and 4

$$|\frac{\hat{\mu}'\hat{\Theta}\hat{\mu}}{\mu'\Theta\mu} - 1| = o_p(1). \tag{A.40}$$

By (A.39)(A.40) 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

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

Then by Assumption 2, seeing that $\Sigma^{-1} = \Theta$ by definition

$$\mu'\Sigma^{-1}\mu/p \geq Eigmin(\Sigma^{-1})\|\mu\|_2^2/p \geq cc_l^2 > 0 \tag{A.41}$$

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

$$|\hat{\mu}'\hat{\Theta}\hat{\mu}/p - \mu'\Sigma^{-1}\mu/p| = O(\bar{s}\sqrt{lnp/n}). \tag{A.42}$$

Combining (A.41)-(A.42) 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 definition of MSR_c in (14) 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

$$\frac{\frac{\widehat{MSR}_c^2}{p}}{\frac{MSR_c^2}{p}} = \left[\frac{\hat{A}\hat{D} - \hat{B}^2}{AD - B^2}\right]\left[\frac{A}{\hat{A}}\right]. \tag{A.43}$$

We start with

$$|\hat{A} - A| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1), \tag{A.44}$$

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

$$|\frac{A}{\hat{A}} - 1| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \tag{A.45}$$

Then Lemma A.6 in Supplement 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 $AD - B^2 \geq C_1 > 0$, so we combine the results above to have

$$\left|\frac{\hat{A}\hat{D} - \hat{B}^2}{AD - B^2} - 1\right| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \tag{A.46}$$

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.45)(A.46) in (A.43) to have

$$\begin{aligned}\left|\frac{\widehat{MSR}_c^2/p}{MSR_c^2/p} - 1\right| &\leq \left|\frac{\hat{A}\hat{D} - \hat{B}^2}{AD - B^2} - 1\right||\frac{A}{\hat{A}} - 1| \\ &+ |\frac{A}{\hat{A}} - 1| + \left|\frac{\hat{A}\hat{D} - \hat{B}^2}{AD - B^2} - 1\right| \qquad \text{(A.47)} \\ &= O_p(\bar{s}\sqrt{lnp/n}) = o_p(1), \qquad \text{(A.48)}\end{aligned}$$

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

Proof of Theorem 6. We need to start with

$$\left|\frac{(\widehat{MSR}^*)^2/p}{(MSR^*)^2/p} - 1\right| = \frac{\left|(\widehat{MSR}^*)^2/p - (MSR^*)^2/p\right|}{(MSR^*)^2/p} \tag{A.49}$$

As a first step analyze the denominator in (A.49). Note that $1_p'\Sigma^{-1}\mu/p \geq 0$ is implied by $1_p'\Sigma^{-1}\mu/p \geq C > 2\epsilon > 0$, so

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

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

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

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 combining the results

$$(MSR^*)^2/p = (MSR^2)1_{\{1_p'\Sigma^{-1}\mu \geq 0\}} + (MSR_c)^2 1_{\{1_p'\Sigma^{-1}\mu < 0\}} \geq C_1/K > 0. \tag{A.50}$$

Next we consider the numerator. We need to show

$$\begin{aligned} p^{-1}|(\widehat{MSR}^*)^2 - (MSR^*)^2| &= p^{-1}|(\widehat{MSR})^2 1_{\{1_p'\hat{\Theta}\hat{\mu}>0\}} - (MSR)^2 1_{\{1_p'\Sigma^{-1}\mu\geq 0\}} \\ &+ [(\widehat{MSR}_c)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}<0\}} - (MSR_c)^2 1_{\{1_p'\Sigma^{-1}\mu<0\}}]| \\ &= O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \end{aligned} \tag{A.51}$$

First see that on the right side of (A.51)

$$\begin{aligned} p^{-1}|(\widehat{MSR})^2 1_{\{1_p'\hat{\Theta}\hat{\mu}>0\}} &- (MSR)^2 1_{\{1_p'\Sigma^{-1}\mu\geq 0\}}| \leq p^{-1}|(\widehat{MSR})^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}} - (MSR)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}}| \\ &+ p^{-1}|(MSR)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}} - (MSR)^2 1_{\{1_p'\Sigma^{-1}\mu/p\geq 0\}}|, \end{aligned} \tag{A.52}$$

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

Then in (A.52)

$$\begin{aligned} p^{-1}|(\widehat{MSR})^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}} - (MSR)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}}| &\leq p^{-1}|(\widehat{MSR})^2 - (MSR)^2||1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}}| \\ &\leq p^{-1}|(\widehat{MSR})^2 - (MSR)^2| \\ &= O_p(\bar{s}\sqrt{lnp/n}) = o_p(1), \end{aligned} \tag{A.53}$$

by (A.42) and Assumption 4 for the rate. In (A.52) above consider

$$\begin{aligned} p^{-1}|(MSR)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}} &- (MSR)^2 1_{\{1_p'\Sigma^{-1}\mu/p\geq 0\}}| \\ &\leq p^{-1}MSR^2|1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}} - 1_{\{1_p'\Sigma^{-1}\mu/p\geq 0\}}|. \end{aligned} \tag{A.54}$$

Note that by definition of MSR^2/p

$$MSR^2/p = \mu'\Sigma^{-1}\mu/p \leq Eigmax(\Sigma^{-1})\|\mu\|_2^2/p \leq Kc_u^2 < \infty, \tag{A.55}$$

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

$$\begin{aligned} \frac{1_p'\hat{\Theta}\hat{\mu}}{p} &= \frac{1_p'\hat{\Theta}\hat{\mu}}{p} - \frac{1_p'\Sigma^{-1}\mu}{p} + \frac{1_p'\Sigma^{-1}\mu}{p} \\ &\geq \frac{1_p'\Sigma^{-1}\mu}{p} - |\frac{1_p'\hat{\Theta}\hat{\mu}}{p} - \frac{1_p'\Sigma^{-1}\mu}{p}| \\ &\geq \frac{1_p'\Sigma^{-1}\mu}{p} - \epsilon \\ &\geq C - \epsilon > 2\epsilon - \epsilon = \epsilon > 0, \end{aligned} \tag{A.56}$$

where we use E_1 in the second inequality, and the condition for the third inequality. This clearly shows that on the Event E_1, when the condition $1_p'\Sigma^{-1}\mu/p \geq C > 2\epsilon > 0$ holds, we have $1_p'\hat{\Theta}\hat{\mu}/p \geq \epsilon > 0$. By Lemma A.3 of Supplement Appendix, E_1 happens with probability approaching one under our Assumptions 1-4

$$|1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}} - 1_{\{1_p'\Sigma^{-1}\mu/p\geq 0\}}| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1), \tag{A.57}$$

where we use (A.56) and $1_p'\Sigma^{-1}\mu/p \geq C > 2\epsilon > 0$ implying $1_p'\Sigma^{-1}\mu/p \geq 0$.

Next combine (A.55)-(A.57) into (A.54)

$$p^{-1}|(MSR)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}} - (MSR)^2 1_{\{1_p'\Sigma^{-1}\mu/p\geq 0\}}| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \tag{A.58}$$

By (A.53)(A.58) we have in (A.52)

$$p^{-1}|(\widehat{MSR}^2)1_{\{1_p'\hat{\Theta}\hat{\mu}/p>0\}} - (MSR^2)1_{\{1_p'\Sigma^{-1}\mu/p\geq 0\}} = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \tag{A.59}$$

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 the one in (A.59) given Theorem 5, except that we have to show

$$|1_{\{1_p'\hat{\Theta}\hat{\mu}/p<0\}} - 1_{\{1_p'\Sigma^{-1}\mu/p<0\}}| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1), \tag{A.60}$$

instead of (A.57). Assume that we use event E_1

$$\begin{aligned} \frac{1_p'\Sigma^{-1}\mu}{p} &= \frac{1_p'\Sigma^{-1}\mu}{p} - \frac{1_p'\hat{\Theta}\hat{\mu}}{p} + \frac{1_p'\hat{\Theta}\hat{\mu}}{p} \\ &\geq \frac{1_p'\hat{\Theta}\hat{\mu}}{p} - |\frac{1_p'\Sigma^{-1}\mu}{p} - \frac{1_p'\hat{\Theta}\hat{\mu}}{p}| \\ &\geq \frac{1_p'\hat{\Theta}\hat{\mu}}{p} - \epsilon. \end{aligned} \tag{A.61}$$

Then in (A.61) using the condition $1_p'\Sigma^{-1}\mu/p \leq -C < -2\epsilon < 0$ (note that this implies also $1_p'\Sigma^{-1}\mu/p < 0$)

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

which implies that, with $C > 2\epsilon$

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

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 happens with probability approaching one by Lemma A.3 of Supplement Appendix, so we proved (A.60). This implies with the result of Theorem 5

$$p^{-1}|(\widehat{MSR}_c)^2 1_{\{1_p'\hat{\Theta}\hat{\mu}/p<0\}} - (MSR_c)^2 1_{\{1_p'\Sigma^{-1}\mu/p<0\}}| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \tag{A.62}$$

Combining now (A.59)(A.62) we proved (A.51) via triangle inequality. With (A.50) and (A.51) the desired result follows by (A.49).

Q.E.D.

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

$$\begin{aligned}\left|\frac{\widehat{SR}_{nw}^2}{SR^2}-1\right| &= \left|\frac{p(1_p'\hat{\Theta}\hat{\mu}/p)^2(1_p'\hat{\Theta}1_p/p)^{-1}}{p(1_p'\Sigma^{-1}\mu/p)^2(1_p'\Sigma^{-1}1_p/p)^{-1}}-1\right| \\ &= \left|\frac{\hat{B}^2A}{B^2\hat{A}}-1\right| \\ &= \left|\frac{\hat{B}^2A-B^2\hat{A}}{B^2\hat{A}}\right| \end{aligned} \tag{A.63}$$

We analyze the denominator in (A.63). 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

$$|B^2\hat{A}| = |B^2(\hat{A}-A)+B^2A| \geq B^2A - B^2|\hat{A}-A| \geq C^2c + o_p(1) > 0. \tag{A.64}$$

Now consider the numerator in (A.63)

$$\begin{aligned}|\hat{B}^2A-B^2\hat{A}| &= |\hat{B}^2A-\hat{B}^2\hat{A}+\hat{B}^2\hat{A}-B^2\hat{A}| \\ &\leq |\hat{B}^2(\hat{A}-A)|+|(\hat{B}^2-B^2)\hat{A}| \\ &\leq |\hat{B}^2(\hat{A}-A)|+|\hat{B}-B||\hat{B}+B||\hat{A}|. \end{aligned} \tag{A.65}$$

Analyze the first term on the right side of (A.65)

$$\begin{aligned}\hat{B}^2 &= |\hat{B}^2-B^2+B^2| \\ &\leq |\hat{B}^2-B^2|+B^2 \\ &\leq |\hat{B}-B||\hat{B}+B|+B^2. \end{aligned} \tag{A.66}$$

Then by Lemma A.3 in Supplement Appendix

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

Then

$$\begin{aligned}|\hat{B}+B| &\leq |\hat{B}|+|B| \\ &\leq |\hat{B}-B|+2|B| \\ &= o_p(1)+2|B| \\ &= O_p(1), \end{aligned} \tag{A.68}$$

where we use (A.67) and Lemma A.5 in Supplement Appendix.

By (A.67)(A.68) in (A.66) we have

$$\hat{B}^2 = O_p(1). \tag{A.69}$$

Then by Lemma A.2 in Supplement Appendix and (A.69)

$$|\hat{B}^2(\hat{A} - A)| \leq \hat{B}^2|\hat{A} - A| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \tag{A.70}$$

Then the second term on the right side of (A.65)

$$|\hat{B} - B||\hat{B} + B||\hat{A}| = O_p(\bar{s}\sqrt{\frac{lnp}{n}})O_p(1)O_p(1) = o_p(1), \tag{A.71}$$

by (A.67)(A.68) and Lemma A.2, Lemma A.5. Use (A.70)(A.71) in (A.65)

$$|\hat{B}^2 A - B^2\hat{A}| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \tag{A.72}$$

Combine (A.64) with (A.72) in (A.63) to have the desired result.**Q.E.D.**

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

$$\begin{aligned}
\left|\frac{\widehat{SR}_{MV}^2}{SR_{MV}^2} - 1\right| &= \left|\frac{\hat{x}/\hat{y}}{x/y} - 1\right| \\
&= \left|\frac{\hat{x}}{\hat{y}}\frac{y}{x} - 1\right| \\
&= \left|\frac{\hat{x}y - \hat{y}x}{\hat{y}x}\right|.
\end{aligned} \tag{A.73}$$

Analyze the denominator of (A.73) first.

$$\begin{aligned}
|\hat{y}x| &= |(\hat{y} - y)x + yx|. \\
&\geq |yx| - |(\hat{y} - y)x| \\
&\geq |yx| - |\hat{y} - y||x|.
\end{aligned} \tag{A.74}$$

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

$$|\hat{y} - y| = |(\hat{A} - A)\rho_1^2 - 2(\hat{B} - B)\rho_1 + (\hat{D} - D)| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \tag{A.75}$$

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

$$|x| = |AD - B^2| \leq AD = O(1). \tag{A.76}$$

So by (A.75)(A.76) and $x \geq C_1 > 0, y \geq C_1 > 0$ in (A.74) to have

$$|\hat{y}x| = o_p(1) + C_1^2 > 0. \tag{A.77}$$

Consider the numerator in (A.73)

$$|\hat{x}y - \hat{y}x| = |\hat{x}y - xy + xy - \hat{y}x| \leq |\hat{x} - x||y| + |x||\hat{y} - y|. \tag{A.78}$$

By Lemma A.6

$$|\hat{x} - x| = |(\hat{A}\hat{D} - \hat{B}^2) - (AD - B^2)| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \tag{A.79}$$

Clearly by Lemma A.5 and triangle inequality with ρ_1 being finite

$$|y| = |A\rho_1 - 2B\rho_1 + D| = O(1). \tag{A.80}$$

Then use (A.75)(A.76)(A.79)(A.80) in (A.78) by Assumption 4

$$|\hat{x}y - \hat{y}x| = O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1). \tag{A.81}$$

Use (A.77)(A.81) in (A.73) to have the desired result.**Q.E.D.**

Supplement Appendix

Here we provide supplemental results. We provide matrix norm inequalities. Let x be a generic vector, which is of p dimension. M being a square matrix of dimension p, with M_j' is the j the 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.

$$\begin{aligned} \|Mx\|_1 &= |M_1'x| + |M_2'x| + \cdots + |M_p'x| \\ &\leq \|M_1\|_1\|x\|_\infty + \|M_2\|_1\|x\|_\infty + \cdots + \|M_p\|_1\|x\|_\infty \\ &= [\sum_{j=1}^{p} \|M_j\|_1]\|x\|_\infty \\ &\leq p \max_j \|M_j\|_1\|x\|_\infty, \end{aligned} \tag{A.82}$$

where we use Holders inequality to get each inequality. **Q.E.D.**

The following Lemmata are all from Callot et al. (2019), and repeated for the benefit of readers. Remembering 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

$$\hat{A} - A = (1_p'\hat{\Theta}1_p - 1_p'\Theta 1_p)/p = (1_p'(\hat{\Theta} - \Theta)1_p)/p. \tag{A.83}$$

Now consider the the right side of (A.83)

$$\begin{aligned} |1_p'(\hat{\Theta} - \Theta)1_p|/p &\leq \|(\hat{\Theta} - \Theta)1_p\|_1\|1_p\|_\infty/p \\ &\leq \max_{1 \leq j \leq p} \|\hat{\Theta}_j - \Theta_j\|_1 \\ &= O_p(\bar{s}\sqrt{\ln p/n}) = o_p(1), \end{aligned} \tag{A.84}$$

where Holders inequality is used in the first inequality, and 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

$$\begin{aligned} \hat{B} - B &= [1_p'(\hat{\Theta} - \Theta)(\hat{\mu} - \mu)]/p & \text{(A.85)} \\ &+ [1_p'(\hat{\Theta} - \Theta)\mu]/p & \text{(A.86)} \\ &+ [1_p'\Theta(\hat{\mu} - \mu)]/p & \text{(A.87)} \end{aligned}$$

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

$$\begin{aligned} |1_p'(\hat{\Theta} - \Theta)(\hat{\mu} - \mu)|/p &\leq \|(\hat{\Theta} - \Theta)1_p\|_1 \|\hat{\mu} - \mu\|_\infty / p \\ &\leq [\max_{1 \leq j \leq p} \|\hat{\Theta}_j - \Theta_j\|_1] \|\hat{\mu} - \mu\|_\infty \\ &= O_p(\bar{s}\sqrt{\ln p/n}) O_p(\sqrt{lnp/n}), & \text{(A.88)} \end{aligned}$$

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 by Theorem 1.

So we consider (A.86) above. Since we have Assumption 2, $\|\mu\|_\infty < c_u < \infty$, where c_u is a positive constant.

$$\begin{aligned} |1_p'(\hat{\Theta} - \Theta)\mu|/p &\leq \|(\hat{\Theta} - \Theta)1_p\|_1 \|\mu\|_\infty / p \\ &\leq c_u [\max_{1 \leq j \leq p} \|\hat{\Theta}_j - \Theta_j\|_1] \\ &= c_u O_p(\bar{s}\sqrt{\ln p/n}), & \text{(A.89)} \end{aligned}$$

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 by Theorem 1.

Now consider (A.87).

$$\begin{aligned} |1_p'\Theta(\hat{\mu} - \mu)|/p &\leq \|\Theta 1_p\|_1 \|\hat{\mu} - \mu\|_\infty / p \\ &\leq [\max_{1 \leq j \leq p} \|\Theta_j\|_1] \|\hat{\mu} - \mu\|_\infty \\ &= O(\sqrt{\bar{s}}) O_p(\sqrt{\ln p/n}), & \text{(A.90)} \end{aligned}$$

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.88)(A.89)(A.90) in (A.85)-(A.87), and note that the largest rate is coming from (A.89). So use Assumption 4, $\bar{s}\sqrt{lnp/n} = o(1)$ to have

$$|\hat{B} - B| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \quad \text{(A.91)}$$

.**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

$$\begin{aligned}
\hat{D} - D &= [(\hat{\mu}-\mu)'(\hat{\Theta}-\Theta)(\hat{\mu}-\mu)]/p && \text{(A.92)}\\
&+ [(\hat{\mu}-\mu)'\Theta(\hat{\mu}-\mu)]/p && \text{(A.93)}\\
&+ [2(\hat{\mu}-\mu)'\Theta\mu]/p && \text{(A.94)}\\
&+ [2\mu'(\hat{\Theta}-\Theta)(\hat{\mu}-\mu)]/p && \text{(A.95)}\\
&+ [\mu'(\hat{\Theta}-\Theta)\mu]/p. && \text{(A.96)}
\end{aligned}$$

We start with (A.92).

$$\begin{aligned}
|(\hat{\mu}-\mu)'(\hat{\Theta}-\Theta)(\hat{\mu}-\mu)|/p &\leq \|(\hat{\Theta}-\Theta)(\hat{\mu}-\mu)\|_1\|\hat{\mu}-\mu\|_\infty/p\\
&\leq [\|\hat{\mu}-\mu\|_\infty]^2[\max_j \|\hat{\Theta}_j-\Theta_j\|_1]\\
&= O_p(\ln p/n)O_p(\bar{s}\sqrt{\ln p/n})\\
&= O_p(\bar{s}(\ln p/n)^{3/2}), && \text{(A.97)}
\end{aligned}$$

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.93).

$$\begin{aligned}
|(\hat{\mu}-\mu)'(\Theta)(\hat{\mu}-\mu)|/p &\leq \|(\Theta)(\hat{\mu}-\mu)\|_1\|\hat{\mu}-\mu\|_\infty/p\\
&\leq [\|\hat{\mu}-\mu\|_\infty]^2[\max_j \|\Theta_j\|_1]\\
&= O_p(\ln p/n)O(\sqrt{\bar{s}})\\
&= O_p(\sqrt{\bar{s}}(\ln p/n)), && \text{(A.98)}
\end{aligned}$$

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, for the rates we use Theorem 1 and (B.55) of Caner and Kock (2018).

Then we consider (A.94), with using $\|\mu\|_\infty \leq c_u$,

$$\begin{aligned}
|(\hat{\mu}-\mu)'(\Theta)(\mu)|/p &\leq \|(\Theta)(\hat{\mu}-\mu)\|_1\|\mu\|_\infty/p\\
&\leq c_u[\|\hat{\mu}-\mu\|_\infty][\max_j \|\Theta_j\|_1]\\
&= O_p(\sqrt{\ln p/n})O(\sqrt{\bar{s}})\\
&= O_p(\sqrt{\bar{s}}\sqrt{lnp/n}), && \text{(A.99)}
\end{aligned}$$

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, for the rates we use Theorem 1 and (B.55) of Caner and Kock (2018).

Then we consider (A.95).

$$\begin{aligned}
|(\mu)'(\hat{\Theta}-\Theta)(\hat{\mu}-\mu)|/p &\leq \|(\hat{\Theta}-\Theta)(\mu)\|_1\|\hat{\mu}-\mu\|_\infty/p\\
&\leq \|\mu\|_\infty \max_j \|\hat{\Theta}_j-\Theta_j\|_1\|\hat{\mu}-\mu\|_\infty\\
&\leq c_u[\max_j \|\hat{\Theta}_j-\Theta_j\|_1]\|(\hat{\mu}-\mu)\|_\infty\\
&= O_p(\bar{s}\sqrt{\ln p/n})O_p(\sqrt{\ln p/n})\\
&= O_p(\bar{s}\ln p/n), && \text{(A.100)}
\end{aligned}$$

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

Then we consider (A.96),

$$\begin{aligned} |(\mu)'(\hat{\Theta} - \Theta)(\mu)|/p &\leq \|(\hat{\Theta} - \Theta)(\mu)\|_1 \|\mu\|_\infty / p \\ &\leq [\|\mu\|_\infty]^2 \max_j \|\hat{\Theta}_j - \Theta_j\|_1 \\ &\leq c_u^2 [\max_j \|\hat{\Theta}_j - \Theta_j\|_1] \\ &= O_p(\bar{s}\sqrt{lnp/n}), \end{aligned} \tag{A.101}$$

where Holder's inequality is used for the first inequality above, and the inequality Lemma A.1, with $M = \hat{\Theta} - \Theta$ and $x = \mu$ for the second inequality above, and 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.101) derives our result, since it is the largest rate by Assumption 4.

Combine (A.97)-(A.101) in (A.92)-(A.96) and the rate in (A.101) to have

$$|\hat{D} - D| = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1). \tag{A.102}$$

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 do the proof for 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 term B, the proof can be obtained by using Cauchy-Schwartz inequality first and the using the same analysis for terms A and D.**Q.E.D.**

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

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

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

Proof of Lemma A.6. Note that by simple adding and subtracting

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

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

$$\begin{aligned} |(\hat{A}\hat{D} - \hat{B}^2) - (AD - B^2)| &\leq \{|\hat{A} - A||\hat{D} - D| + |\hat{A} - A|D \\ &+ A|\hat{D} - D| + (\hat{B} - B)^2 + 2|B||\hat{B} - B|\} \\ &= O_p(\bar{s}\sqrt{lnp/n}) = o_p(1), \end{aligned} \tag{A.103}$$

where we use (A.84)(A.91)(A.102), Lemma A.5, and Assumption 4: $\bar{s}\sqrt{logp/n} = o(1)$.**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.

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.