TEXTO PARA DISCUSSÃO

No. 684

Residual Based Nodewise Regression in Factor Models with Ultra-High Dimensions: Analysis of Mean-Variance Portfolio Efficiency and Estimation of Out-of-Sample and Constrained Maximum Sharpe Ratios

Mehmet Caner
Marcelo Medeiros
Gabriel F. R. Vasconcelos

DEPARTAMENTO DE ECONOMIA
www.econ.puc-rio.br

Residual Based Nodewise Regression in Factor Models with Ultra-High Dimensions: Analysis of Mean-Variance Portfolio Efficiency and Estimation of Out-of-Sample and Constrained Maximum Sharpe Ratios

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

June 22, 2021

Abstract

We provide a new theory for nodewise regression when the residuals from a fitted factor model are used to apply our results to the analysis of maximum Sharpe ratio when the number of assets in a portfolio is larger than its time span. We introduce a new hybrid model where factor models are combined with feasible nodewise regression. Returns are generated from increasing number of factors plus idiosyncratic components (errors). The precision matrix of the idiosyncratic terms is assumed to be sparse, but the respective covariance matrix can be non-sparse. Since the nodewise regression is not feasible due to unknown nature of errors, we provide a feasible-residual based nodewise regression to estimate the precision matrix of errors, as a new method. Next, we show that the residual-based nodewise regression provides a consistent estimate for the precision matrix of errors. In another new development, we also show that the precision matrix of returns can be estimated consistently, even with increasing number of factors. Benefiting from the consistency of the precision matrix estimate of returns, we show that: (1) the portfolios in high dimensions are mean-variance efficient; (2) maximum out-of-sample Sharpe ratio estimator is consistent and the number of assets slows the convergence up to a logarithmic factor; (3) the maximum Sharpe ratio estimator is consistent when the portfolio weights sum to one; and (4) the Sharpe ratio estimators are consistent in global minimum-variance and mean-variance portfolios.

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

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

‡Head of Quantitative Research, BOCOM BBM Bank. Av. Barão de Tefé, 34 - 20º e 21º floors - Rio de Janeiro - RJ, 20220-460. Email: gabrielvasconcelos@bocombbm.com.br. We are very grateful to the co-editor, Torben Andersen, the Associate Editor and an anonymous referee for very insightful comments and suggestions which led to a much improved version of the manuscript. We thank Vanderbilt Economics Department seminar guests and the participants of the World Congress of the Econometric Society for useful comments. Finally, we are thankful for the comments by Harold Chiang, Srini Krishnamurthy, and Michael Wolf.

1 Introduction

One of the key issues in Finance, specially in empirical asset pricing, is the trade-off between the return and the risk of a portfolio. To obtain a better risk-adjusted returns, we maximize the Sharpe ratio. In essence, the weights of the portfolio are chosen in such a way that the return-to-risk ratio is maximized. We contribute to this literature by studying the case of a large number of assets p, which may be greater than the time span of the portfolio n. Our analysis also involves time-series data for excess asset returns. To obtain the maximum Sharpe ratio, we make use of the asset return's precision matrix. In order to get an estimate of the precision matrix for asset returns in a large portfolio, we propose that excess returns of assets are governed by an approximate factor model. Hence, asset returns (excess returns over a risk free asset) can be explained by increasing, but known and common number of factors with unknown errors entering the linear relation in an additive way. One major difference with the previous literature is that, in our case, the covariance matrix of errors can be non-sparse, but precision matrix has to be sparse. So this is a hybrid method that combines factor models with high-dimensional econometrics.

The first step in getting the maximum Sharpe ratio involves the precision matrix estimate of the idiosyncratic terms (errors). Estimating the precision matrix of errors is not an easy task and the simple nodewise regression idea as in Meinshausen and Bühlmann (2006) is not feasible. Therefore, we provide a simple feasible residual-based nodewise regression estimate for the precision matrix of errors. This feasible residual based nodewise regression is a new idea and it is shown to be consistently estimating the precision matrix of the errors. Next, we obtain consistent estimators to the precision matrix of asset returns.

Our main contribution is that we are able to obtain mean-variance efficiency for large portfolios even when $p > n$ when both dimensions are growing. Relatedly, the consistency of our nodewise-based maximum-out-of-sample Sharpe ratio estimate is established. We also provide the rate of convergence and show that the number of assets slows the rate of convergence up to a logarithmic factor in p. Consequently, consistent estimation of the Sharpe ratio of large portfolios is possible. Increasing number of factors slows the rate of convergence. Second, we consider the rate of convergence and consistency of the maximum Sharpe ratio when the weights of the portfolio are normalized to one and $p > n$. Recently, Maller and Turkington (2002) and Maller et al. (2016) analyze the limit with a fixed number of assets and extend that approach to a large number of assets but a number less than the time span of the portfolio. Their papers make a key discovery: in the case of weight constraints (summing to one), the formula for the maximum Sharpe ratio depends on a technical term, unlike the unconstrained maximum Sharpe ratio case. Practitioners could obtain the minimum Sharpe ratio instead of the maximum if they are using the unconstrained formula. Our paper extends their paper by analyzing two issues, first the case of $p > n$, with both quantities growing to infinity, and second by handling the uncertainty created by this technical term, which we can estimate and use to obtain a new constrained and consistent Sharpe ratio. Our third contribution is that we consider the Sharpe ratios in the global minimum-variance portfolio and Markowitz mean-variance portfolio. Our analysis uncovers consistent estimators even when $p > n$

1.1 A Brief Review of the Literature and Main Takeaways

In terms of the literature on nodewise regression and related methods, the most relevant papers are as follows. Meinshausen and Bühlmann (2006) establish the nodewise regression approach and provide an optimality result when data are normally distributed. Chang et al. (2018) extend the nodewise regression method to time-series data and build confidence intervals for the elements in the precision matrix. However, the goal of Chang et al. (2018) only centers on the elements of the precision matrix, and there is no connection to factor models. Furthermore, their results are based on the precision matrix of observed data and not on the residuals of a first-stage estimator. Finally, the authors do not consider the case of maximum Sharpe-Ratio and it is not clear if their results are directly applicable to financial applications. Caner and Kock (2018) establish uniform confidence intervals in the case of high-dimensional parameters in heteroskedastic setups using nodewise regression, but, as in the previous paper, there is no connection to factor models in empirical finance. Callot et al. (2021) provide the variance, the risk, and the weight estimation of a portfolio via nodewise regression. They take the nodewise regression directly from Meinshausen and Bühlmann (2006) and apply it to returns. However, they assume that the precision matrix of returns is sparse. Hence, it is more restrictive and less realistic than the method we propose. We combine factor models with sparsity of the precision matrix of errors. As a consequence, our method is much more connected to typical empirical asset pricing models. Furthermore, we do not impose any sparsity on the precision matrix of returns. Callot et al. (2021) also have no proofs about the estimation of the Sharpe ratio.

In terms of recent contributions to the literature on factor models and sparse regression we highlight Fan et al. (2021). The authors consider the combination of factor models and sparse regression in a very general setting. More specifically, they analyze a panel data model with a factor structure and idiosyncratic terms that are sparsely related. They also provide an inference procedure designed to test hypotheses on the entries of the covariance matrix of the residuals of pre-estimated models, including principal component regressions. Our paper differs from theirs in several directions. First, Fan et al. (2021) considers only the covariance matrix and not the precision matrix. Second, their approach is not based on nodewise regressions. Finally, Sharpe ratio estimation and portfolio allocation are not considered. A seminal paper is by Gagliardini et al. (2016), where they analyze time-varying risk premia in large portfolios with factor models. They develop a structural model and can tie that to factor models and after that they can estimate time varying risk-premia. One of the main assumptions is that maximum eigenvalue of covariance matrix of errors in the factor structure can diverge. Also they assume sparsity of covariance matrix of errors, and observed factors in the factor model. We also use diverging eigenvalue assumption in Assumption 7(i) in our paper, as well as increasing number of factors here, but with assumption of sparsity on the precision matrix of errors.

Recently, important contributions have been in this area by using shrinkage and factor models. Ledoit and Wolf (2017) propose a nonlinear shrinkage estimator in which small eigenvalues of the sample covariance matrix are increased and large eigenvalues are decreased by a shrinkage formula. Their main contribution is the optimal shrinkage function, which they find by minimizing a loss function. The maximum out-of-sample Sharpe ratio is an inverse function of this loss. Their results cover the independent and identically distributed case and when $p/n \to (0,1) \cup (1,+\infty)$. For the analysis of mean-variance efficiency, Ao et al. (2019) make a novel contribution in which they take a constrained optimization, maximize returns subject to risk of the portfolio, and show that it is equivalent to an unconstrained objective function, where they minimize a scaled return of the portfolio error by choosing optimal weights. To obtain these weights, they use lasso regression and hence, assume a sparse number of nonzero weights of the portfolio, and they analyze $p/n \to (0,1)$. They show that their method maximizes the expected return of the portfolio and satisfies the risk constraint. This is an important result on its own. One key paper in the literature is by Fan et al. (2011) which assumes an approximate factor model, but, on the other hand, the authors assume conditional sparsity-diagonality of the covariance matrix of errors. Fan et al. (2011) show for the first time how to build a precision matrix of returns in a large portfolio via factor models. Therefore, it is a key paper in the high-dimensional econometrics literature.

Regarding other papers, Ledoit and Wolf (2003,2004) propose a linear shrinkage estimator of the covariance matrix and apply it to portfolio optimization. Ledoit and Wolf (2017) shows that nonlinear shrinkage performs better in out-of-sample forecasts. Lai et al. (2011) and Garlappi et al. (2007) approach the same problem from a Bayesian perspective by aiming to maximize a utility function tied to portfolio optimization. Another avenue of the literature improves the performance of the portfolios by introducing constraints on the weights. This is in the case of the global minimum-variance portfolio. Examples of works investigating this problem include Jagannathan and Ma (2003) and Fan et al. (2012). We also see a combination of different portfolios proposed by Kan and Zhou (2007) and Tu and Zhou (2011). Very recently, Ding et al. (2020) extended factor models to assumptions that are more consistent with principal components analysis. They provide consistent estimation of risk of the portfolio under sparsity of covariance of errors with fixed number of factors. Barras et al. (2018), Brodie et al. (2009), Chamberlain and Rothschild (1983), DeMiguel et al. (2009), Fan et al. (2015) analyze mutual fund industry, sparsely constructed Markowitz portfolio, arbitrage and factor models in large portfolios, sparsely constructed mean-variance portfolios, and risks of large portfolios, respectively.

.

1.2 Organization of the Paper

This paper is organized as follows. Section 2 considers our assumptions and feasible precision matrix estimation for errors. Section 3 provides the feasible precision matrix estimate for asset returns. Section 4 addresses the maximum out-of-sample Sharpe ratio and the mean-variance efficiency. Section 5 handles the case of the maximum Sharpe ratio when the weights are normalized to one. Section 6 concerns the global minimum-variance and Markowitz mean-variance portfolio Sharpe ratios. Section 7 provides simulations that compare several methods. Section 8 presents an out-of-sample forecasting exercise. The main proofs are in the Appendix, and the Supplementary Appendix has some benchmark results used in the main proofs section. Let $\|\nu\|_{l_1}, \|\nu\|_{l_2}, \|\nu\|_\infty$ be the l_1, l_2, l_∞, norms of a generic vector ν. Let $\|v\|_n^2 := n^{-1}\sum_{t=1}^n v_t^2$ which is the prediction norm for an $n \times 1$ vector v. Let Eigmin(A) represents the minimum eigenvalue of a matrix A, and Eigmax(A) represent the maximum eigenvalue of the matrix A. For a generic matrix A, let $\|A\|_{l_1}, \|A\|_{l_\infty}, \|A\|_{l_2}$, be the l_1 induced matrix norm (i.e. maximum absolute column sum norm), l_∞ induced matrix norm (i.e. maximum absolute row sum norm), spectral matrix norm, respectively. $\|A\|_\infty$ is maximum absolute value of element of a matrix, and also a norm (but not a matrix norm). Matrix norms have the additional desirable feature of submultiplicativity property. For further information on matrix norms, see p.341 of Horn and Johnson (2013).

2 Factor Model and Feasible Nodewise Regression

We start with the following model, for j th asset return (excess asset return) at time t, $y_{j,t}$, for $j = 1, \cdots, p$, and time periods $t = 1, \cdots, n$

$$y_{j,t} = b_j' f_t + u_{j,t}. \tag{1}$$

where b_j are $K \times 1$ vector of factor loadings, and $f_t : K \times 1$ vector of common factors to all assets' returns, and $u_{j,t}$ is the scalar error term for asset return j, at time t. All the factors are observed. This model is used at Fan et al. (2011). From this point on, when asset return is mentioned, it should be understood as excess asset return.

For the j th asset return we can rewrite (1) in the vector form, for $j = 1, \cdots, p$

$$y_j = X' b_j + u_j, \tag{2}$$

where $X = (f_1, \cdots, f_n) : K \times n$ matrix, and $y_j = (y_{j,1}, \cdots, y_{j,n})' : n \times 1$ vector of j th asset returns. We can also express the same relation in a matrix form.

$$Y = BX + U, \tag{3}$$

where $Y : p \times n$ matrix, $B : p \times K$ matrix, and $U : p \times n$ matrix [1]. Define the covariance matrix of errors, with the definition of $u_t := (u_{1,t}, \cdots, u_{j,t}, \cdots, u_{p,t})' : p \times 1$ error vector

$$\Sigma_n := E u_t u_t'.$$

In Assumption 7 below, we assume that maximum eigenvalue of Σ_n can grow with sample size, this is due to Σ_n being a $p \times p$ matrix where p may grow with n. We will assume sparsity for the precision matrix of errors $\Omega := \Sigma_n^{-1}$. We assume that at each row of Ω, which will be denoted as $\Omega_j' : 1 \times p$, and the s_j number

[1] We can also write the returns for each period in time, $t = 1, \cdots, n$

$$y_t = B f_t + u_t,$$

where $y_t = (y_{1,t}, \cdots, y_{j,t}, \cdots, y_{p,t})' : p \times 1$ vector.

of cells are nonzero, and the rest of the row is zero. We represent the indices of nonzero cells in Ω_j' as S_j, for $l = 1, \cdots, p$

$$S_j := \{j : \Omega_{j,l} \neq 0\},$$

where $\Omega_{j,l}$ represents the j th row and l element in that row. Let S_j^c represent the index set of all zero elements in j th row of Ω_j'. Denote the maximum number of nonzero elements across all rows $j = 1, \cdots, p$ of the precision matrix Ω as

$$\bar{s} := \max_{1 \leq j \leq p} s_j.$$

This last point is an assumption, and will be shown below in Assumption 1. We can also allow weak sparsity in the precision matrix of errors as in Theorem 2.2 in van de Geer (2016), but this increases the notation with no change in the final results. Note that when $n \to \infty$, we allow $p \to \infty$, and $K \to \infty$, $\bar{s} \to \infty$. As in the literature we do not subscript them by n even though $p, K, \bar{s}$ can increase with n. Specifically, we assume $p, K, \bar{s}$ are nondecreasing in n. Also we allow for $p > n$ in our analysis which can be considered ultra-high dimensional portfolio analysis.

For future references, we denote all of the asset returns except j th one as

$$Y_{-j} = B_{-j} X + U_{-j}, \tag{4}$$

where $Y_{-j} : p-1 \times n$ matrix, is matrix Y without j th row, $B_{-j} : p-1 \times K$ matrix is matrix B without j th row, and $U_{-j} : p-1 \times n$ matrix, which is U matrix without p th row. Let $\Sigma_{n,-j,-j}$ represent the $p-1 \times p-1$ subset of Σ_n when the j th row and j th column of Σ_n has been removed. Let $\Sigma_{n,-j,j}$ represent the j th column of Σ_n with j th element removed, and $\Sigma_{n,j,-j}$ represent the j th row of Σ_n with its j th element is removed. In the assumptions below we will assume, stationary (strict) and ergodic errors for the time series dimension. Using inverse formula for partitioned matrices, Yuan (2010) shows

$$\Omega_{j,j} := (\Sigma_{n,j,j} - \Sigma_{n,j,-j} \Sigma_{n,-j,-j}^{-1} \Sigma_{n,-j,j})^{-1}, \tag{5}$$

and

$$\Omega_{j,-j}' := -\Omega_{j,j} \Sigma_{n,j,-j} \Sigma_{n,-j,-j}^{-1}. \tag{6}$$

where $\Omega_{j,-j}'$ is the j th row of Ω, without the j th element in that row. Let γ_j be the minimizing value of γ

$$E[u_{j,t} - u_{-j,t}' \gamma]^2,$$

where $u_{j,t}$ is the j the element of $u_t : p \times 1$ vector, similarly, $u_{-j,t}$ is the $p-1 \times 1$ vector of errors in t th time period, except j th term in u_t. Then

$$\gamma_j = \Sigma_{n,-j,-j}^{-1} \Sigma_{n,-j,j}, \tag{7}$$

and

$$\Omega_{j,-j}' = -\Omega_{j,j} \gamma_j'. \tag{8}$$

Define $\eta_{j,t} := u_{j,t} - u'_{-j,t}\gamma_j$, where we can show that by $\Sigma_{n,-j,j}, \Sigma_{n,-j,-j}$ definitions, and (7)

$$\begin{aligned} E u_{-j,t}\eta_{t,j} &= E u_{-j,t} u_{j,t} - E u_{-j,t} u'_{-j,t}\gamma_j \\ &= \Sigma_{n,-j,j} - \Sigma_{n,-j,-j}\Sigma^{-1}_{n,-j,-j}\Sigma_{n,-j,j} = 0. \end{aligned} \tag{9}$$

This provides us with

$$u_{j,t} = u'_{-j,t}\gamma_j + \eta_{j,t}, \tag{10}$$

for all $j = 1, \cdots, p, t = 1, \cdots, n$, and vector $u_{-j,t}$ and $\eta_{j,t}$ are orthogonal in (10). The vector form is, for each $j = 1, \cdots, p$,

$$u_j = U'_{-j}\gamma_j + \eta_j, \tag{11}$$

where $u_j = (u_{j,1}, \cdots, u_{j,n})'$ ($n \times 1$ vector), and U'_{-j}: ($n \times p - 1$) matrix. Proceeding with (10) and using (9)

$$E u^2_{j,t} = \gamma'_j \Sigma_{n,-j,-j}\gamma_j + E\eta^2_{j,t},$$

which implies by (7), and (5) with definition of $\Sigma_{n,j,j} := E u^2_{j,t}$,

$$E\eta^2_{j,t} = \Sigma_{n,j,j} - \Sigma_{n,j,-j}\Sigma^{-1}_{n,-j,-j}\Sigma_{n,-j,j} = \frac{1}{\Omega_{j,j}}. \tag{12}$$

Define for each $j = 1, \cdots, p$, where we use (12)

$$\tau^2_j := E\eta^2_{j,t} = \frac{1}{\Omega_{j,j}}. \tag{13}$$

Then clearly, the main diagonal term in the precision matrix is:

$$\Omega_{j,j} = \frac{1}{\tau^2_j}. \tag{14}$$

and j th row of precision matrix, without j th element, by (8)(14) is:

$$\Omega'_{j,-j} = \frac{-\gamma'_j}{\tau^2_j}. \tag{15}$$

Define

$$\Omega := T^{-2} C, \tag{16}$$

with

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

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

Based on (11) and the sparsity of the precision matrix, with a sequence $\lambda_n > 0$, for all $j = 1, \cdots, p$,

$$\tilde{\gamma}_j = argmin_{\gamma_j \in R^{p-1}} [\|u_j - U'_{-j}\gamma_j\|^2_n + 2\lambda_n \|\gamma_j\|_1]. \tag{17}$$

The main issue with (17) is, unlike nodewise regression in Caner and Kock (2018), it is infeasible due to error terms regressed on each other. This is a major problem, and has not been tackled in this literature before. We now show how to turn this to feasible regression and still consistently estimate γ_j.

To get estimates for b_j and B'_{-j}, Fan et al. (2011) uses Ordinary Least Squares (OLS) and p.3347 of Fan et al. (2011) shows

$$\hat{b}_j - b_j = (XX')^{-1} X u_j. \tag{18}$$

Note that we can define OLS residual by (2)

$$\begin{aligned} \hat{u}_j &= y_j - X'\hat{b}_j \\ &= u_j - X'(XX')^{-1} X u_j \\ &= M_X u_j, \end{aligned} \tag{19}$$

where we define, ($X' : n \times K$ matrix)

$$M_X := I_n - X'(XX')^{-1} X. \tag{20}$$

Then by OLS, with $\hat{B}'_{-j} : K \times p-1$, $B'_{-j} : K \times p-1$ matrix,

$$\hat{B}'_{-j} - B'_{-j} = (XX')^{-1} X U'_{-j}.$$

Define the residuals with transposing (4)

$$\begin{aligned} \hat{U}'_{-j} &= Y'_{-j} - X'\hat{B}'_{-j} \\ &= U'_{-j} - X'(\hat{B}'_{-j} - B'_{-j}) \\ &= U'_{-j} - X'(XX')^{-1} X U'_{-j} \\ &= M_X U'_{-j}. \end{aligned} \tag{21}$$

Note that $\hat{U}'_{-j} : n \times p-1$, $M_X : n \times n$, $U'_{-j} : n \times p-1$. Next pre-multiply each side of (11) by M_X, and use (19)(21)

$$\hat{u}_j = \hat{U}'_{-j}\gamma_j + \eta_{xj}, \tag{22}$$

where we define

$$\eta_{xj} := M_X \eta_j, \tag{23}$$

$n \times 1$ vector. Of course, the key difficulties are how the new η_{xj}, and usage of the residuals affect the consistent estimation of γ_j ? To answer these questions we define a feasible nodewise estimator

$$\hat{\gamma}_j = argmin_{\gamma_j \in R^{p-1}} \left[\| \hat{u}_j - \hat{U}'_{-j}\gamma_j \|_n^2 + \lambda_n \|\gamma_j\|_1 \right]. \tag{24}$$

Then to define $\hat{\Omega}'_j$ which is the j th row of precision matrix estimate, we need

$$\hat{\tau}_j^2 := \hat{u}'_j(\hat{u}_j - \hat{U}'_{-j}\hat{\gamma}_j)/n. \tag{25}$$

Now to form the j th row of $\hat{\Omega}$, set the j th element in j th row as

$$\hat{\Omega}_{j,j} := 1/\hat{\tau}_j^2 \tag{26}$$

in (14), and the rest of j th row estimate as in (15)

$$\hat{\Omega}'_{j,-j} := \frac{-\hat{\gamma}'_j}{\hat{\tau}_j^2}. \tag{27}$$

We want to show that for each $j = 1, \cdots, p$, $\hat{\Omega}'_j$ is consistent. We can write $\hat{\Omega}'_j = \hat{C}'_j/\hat{\tau}_j^2$ with $\hat{C}'_j : 1 \times p$ representing 1 in j th cell and $-\hat{\gamma}'_j$ in the other cells.

2.1 Assumptions

In this part we provide the assumptions that will be needed for consistency for the j th row of precision matrix estimate.

Assumption 1. *(i).* $\{u_t\}_{t=1}^n, \{f_t\}_{t=1}^n$ *are stationary (strictly), and ergodic. Furthermore,* $\{u_t\}_{t=1}^n, \{f_t\}_{t=1}^n$ *are independent.* u_t *has zero mean* $(p \times 1)$*, with variance matrix* Σ_n $(p \times p)$*.* $Eigmin(\Sigma_n) \geq c > 0$*, with c a positive constant, and* $\max_{1 \leq j \leq p} E u_{j,t}^2 \leq C < \infty$*. (ii). Let the maximum number of nonzero cells across rows of the precision matrix of errors,* $\Omega := \Sigma_n^{-1}$*, be defined as* $\max_{1 \leq j \leq p} s_j = \bar{s}$*. (iii). Let* $\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty$ *be the* Σ_n *algebras generated by* $\{(f_t, u_t)\}$*, for* $-\infty < t \leq 0$*, and* $n \leq t < \infty$ *respectively. Denote the strong mixing coefficient as*

$$\alpha(n) := sup_{\mathcal{A} \in \mathcal{F}_{-\infty}^0, \mathcal{B} \in \mathcal{F}_n^\infty} |P(\mathcal{A})P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B}|$$

and $\alpha(t) \leq exp(-Ct^{r_0})$*, for a positive constant* $r_0 > 0$*.*

Assumption 2. *There exists positive constants* $r_1, r_2, r_3 > 0$ *and another set of positive constants* $b_1, b_2, b_3, s_1, s_2, s_3 > 0$*, and for* $t = 1, \cdots, n$*, and* $j = 1, \cdots, p$*, with* $k = 1, \cdots, K$

(i).

$$P[|u_{j,t}| > s_1] \leq exp(-(s_1/b_1)^{r_1}).$$

(ii).

$$P[|\eta_{j,t}| > s_2] \leq exp(-(s_2/b_2)^{r_2}).$$

(iii).

$$P[|f_{k,t}| > s_3] \leq exp(-(s_3/b_3)^{r_3}).$$

(iv). There exists $0 < \gamma_1 < 1$ *such that* $\gamma_1^{-1} = 3r_1^{-1} + r_0^{-1}$*, and we also assume* $3r_2^{-1} + r_0^{-1} > 1$*, and* $3r_3^{-1} + r_0^{-1} > 1$*.*

Define $\gamma_2^{-1} := 1.5r_1^{-1} + 1.5r_2^{-1} + r_0^{-1}$, and $\gamma_3^{-1} := 1.5r_1^{-1} + 1.5r_3^{-1} + r_0^{-1}$, let $\gamma_{min} := min(\gamma_1, \gamma_2, \gamma_3)$.

Assumption 3. *(i).* $(lnp)^{(2/\gamma_{min})-1} = o(n)$*, and (ii).* $K^2 = o(n)$*, (iii).* $K = o(p)$*.*

Assumption 4. *(i). $Eigmin(cov f_t) \geq c > 0$, with $cov f_t$ being the covariance matrix of the factors f_t, $t = 1, \cdots, n$.*

(ii). $\max_{1 \leq k \leq K} E f_{kt}^2 \leq C < \infty$, $\min_{1 \leq k \leq k} E f_{kt}^2 \geq c > 0$.

(iii). $\max_{1 \leq j \leq p} E \eta_{j,t}^2 \leq C < \infty$.

Assumption 5.

$$\bar{s} K^2 \sqrt{\frac{lnp}{n}} = o(1).$$

Note that Assumptions 1-3 are standard assumptions and are used in Fan et al. (2011). Also we get $0 < \gamma_2 < 1, 0 < \gamma_3 < 1$ given Assumption 2(iv). Also by Assumption 3, $\sqrt{\frac{lnp}{n}} = o(1)$. Stationary GARCH models with finite second moments, and continuous error distributions, as well as causal ARMA processes with continuous error distributions, and a certain class of stationary Markov chains satisfy our Assumptions 1-2 and are discussed in p.61 of Chang et al. (2018). Chang et al. (2018) also uses similar assumptions. We should note that main interest of Chang et al. (2018) is just building confidence intervals around entries for a large precision matrix, there is no finance related theorems.

Assumption 4(i)-(ii) is also used in Fan et al. (2011), and the nodewise error asumption 4(iii) is used in nodewise regression context in Caner and Kock (2018). Assumption 5 is new and shows the interaction of sparsity of the precision matrix with factors, they both contribute negatively to biases that our analysis will show below in theorems. Compared to a regular nodewise sparsity assumption in Callot et al. (2021), we have an extra K^2 factor.

We provide one of the main Theorems in the paper. This is new in the literature and shows that a feasible-residual based nodewise regression is possible, and it provides consistent estimates for the rows of the precision matrix of errors.

Theorem 1. *Under Assumptions 1-5*

$$\|\hat{\Omega} - \Omega\|_{l_\infty} := \max_{1 \leq j \leq p} \|\hat{\Omega}_j' - \Omega_j'\|_1 = \max_{1 \leq j \leq p} \|\hat{\Omega}_j - \Omega_j\|_1 = O_p(\bar{s} \sqrt{\frac{lnp}{n}}) = o_p(1).$$

Remarks.

1. Note that $\hat{\Omega}_j', \Omega_j'$ are $1 \times p$ row vectors for the precision matrix estimate and the precision matrix respectively. The second inequality above just shows that l_1 norm for a row vector and its column version (transpose of that same row to column) is the same. $\hat{\Omega}_j, \Omega_j$ are the column versions of the rows of the precision matrix estimate $(\hat{\Omega}_j')$, and the precision matrix, (Ω_j') respectively.

2. Note that by Theorem 1 statement and matrix l_1 norm definition (i.e. l_1 induced matrix norm, maximum column sum norm of a matrix). So

$$\|\hat{\Omega}' - \Omega\|_{l_1} = \max_{1 \leq j \leq p} \|\hat{\Omega}_j - \Omega_j\|_1, \tag{28}$$

where $\hat{\Omega}'$ is the transpose of $\hat{\Omega}$. Also note that

$$\|\hat{\Omega}' - \Omega\|_{l_1} = \|\hat{\Omega} - \Omega\|_{l_\infty}, \tag{29}$$

where we use l_1, l_∞ matrix norm definitions.

3. As long as Assumption 5 is maintained, the rate of approximation error in Theorem 1 matches the factor model free nodewise regression error in Callot et al. (2021).

3 Precision Matrix Estimate for The Returns

Note that the covariance matrix for asset returns is defined, using independence between factors and errors, $\Sigma_y : p \times p$ matrix

$$\Sigma_y = Bcov f_t B' + \Sigma_n. \tag{30}$$

We start with precision matrix formula for the asset returns, based on factor model that we used. Using Sherman-Morrison-Woodbury formula, as in p.13 of Horn and Johnson (2013), $\Gamma := \Sigma_y^{-1}$ is defined as precision matrix for the returns:

$$\Gamma := \Omega - \Omega B[\{cov f_t\}^{-1} + B'\Omega B]^{-1} B'\Omega, \tag{31}$$

and precision matrix estimate for the returns are

$$\hat{\Gamma} := \hat{\Omega} - \hat{\Omega}\hat{B}[\{\widehat{cov f_t}\}^{-1} + \hat{B}'\hat{\Omega}_{sym}\hat{B}]^{-1}\hat{B}'\hat{\Omega}, \tag{32}$$

where $\hat{\Omega}_{sym} := \frac{\hat{\Omega}+\hat{\Omega}'}{2}$, and it is the symmetrized version of our feasible nodewise regression estimator for the precision matrix for errors. $\widehat{cov f_t} = n^{-1}XX' - n^{-2}X1_n1_n'X'$ is the estimate for the covariance matrix returns, and it is given in p.3327 of Fan et al. (2011) with 1_n representing a vector $(n \times 1)$ of ones. Also $\hat{B} = (YX')(XX')^{-1}$ are the least-squares estimates for the factor model in (3). Also $\hat{B}$ is a $p \times K$ matrix, and $\widehat{cov f_t}$ is $K \times K$ matrix. Note that we use a symmetric version of our precision matrix estimate for errors in the term in square brackets in (32), there is a technical reason behind that. The proofs depend on symmetry of the matrix in the square brackets in (32), but the other parts in the proof do not need symmetry of the precision matrix estimate, hence we use both symmetrized, $\hat{\Omega}_{sym}$ and standard (non-symmetric version) of the precision matrix estimate, $\hat{\Omega}$. We want to rewrite the precision matrix and its estimate so that its convenient to analyze them technically. In that respect define

$$L := B[\{cov f_t\}^{-1} + B'\Omega B]^{-1}B',$$

and

$$\hat{L} := \hat{B}[\{\widehat{cov f_t}\}^{-1} + \hat{B}'\hat{\Omega}_{sym}\hat{B}]^{-1}\hat{B}'.$$

So we have

$$\Gamma = \Omega - \Omega L \Omega, \quad \hat{\Gamma} = \hat{\Omega} - \hat{\Omega}\hat{L}\hat{\Omega}. \tag{33}$$

We need to find $\max_{1 \le j \le p} \|\hat{\Gamma}_j - \Gamma_j\|_1$ where $\Gamma_j', \hat{\Gamma}_j'$ are the $1 \times p$ dimensional rows of the precision matrix of the returns, and its estimate respectively. $\Gamma_j, \hat{\Gamma}_j$ are simply transposes of these rows which are $p \times 1$. In this respect, using (33)

$$\max_{1\le j\le p}\|\hat{\Gamma}_j-\Gamma_j\|_1=\max_{1\le j\le p}\|\hat{\Gamma}_j'-\Gamma_j'\|_1=\max_{1\le j\le p}\|(\hat{\Omega}_j'-\Omega_j')-(\hat{\Omega}_j'\hat{L}\hat{\Omega}-\Omega_j'L\Omega)\|_1. \tag{34}$$

Our aim is to simplify and get rates for the right side term in (34). To get a consistency and rate of convergence result for the precision matrix for returns, rather than the errors as in Theorem 1 above we need an assumption on factor loadings.

Assumption 6. *(i).* $\max_{1\le j\le p}\max_{1\le k\le K}|b_{jk}|\le C<\infty$.

(ii). $\|p^{-1}B'B-\Delta\|_{l_2}=o(1)$ *for some* $K\times K$ *symmetric positive definite matrix* Δ *such that* $Eigmin(\Delta)$ *is bounded away from zero.*

Also a strengthened assumption on sparsity compared to Assumption 5 is provided. Note that when $n\to\infty, p\to\infty$. Define a rate of convergence, $l_n>0$ which will be a part of the rate in Theorem 2 below

$$l_n:=r_n^2K^{5/2}\max(\bar{s},\bar{s}^{1/2}K^{1/2})\sqrt{\frac{max(lnp,lnn)}{n}}. \tag{35}$$

Specifically the rate l_n is the rate of estimation error for $\|\hat{L}-L\|_{l_\infty}$ as in Lemma A.13.

Assumption 7. *(i).* $Eigmax(\Sigma_n)\le Cr_n$, *with* $C>0$ *a positive constant, and* $r_n\to\infty$ *as* $n\to\infty$, *and* $r_n/p\to 0$, *and* r_n *is a positive sequence.*

(ii). Assume that

$$\bar{s}l_n\to 0.$$

Note that Assumption 6 is used in Fan et al. (2011). Assumption 7(i) is used in Gagliardini et al. (2016). Assumption 7(i) allows for the maximal eigenvalue of Σ_n to grow with n. In special case of a diagonal Σ_n, due to Assumption 1(i), the maximum eigenvalue of a diagonal Σ_n matrix is finite. But that diagonal matrix of variance of errors case is empirically less relevant and less realistic. We expect the errors to be correlated across assets. For an example of where maximum eigenvalue of Σ_n may diverge, we show that this may be the case for block diagonal matrix structure for Σ_n in (37).

Assumption 7(ii) is a sparsity assumption with tradeoffs between maximal eigenvalue and the sparsity of the precision matrix. This assumption is needed to analyze the precision matrix for the asset returns. To give an example, ignoring constants, we can have $\bar{s}=lnn, K=lnn, p=2n, r_n=n^{1/5}$. Then Assumption 7(ii) is satisfied

$$\frac{(lnn)^{9/2}ln(2n)^{1/2}n^{2/5}}{n^{1/2}}\to 0.$$

Next we provide definition for sample mean of asset returns, and the population mean of asset returns. Define $\hat{\mu}:=\frac{1}{n}\sum_{t=1}^n y_t$, where y_t is a $p\times 1$ vector of asset returns. Let $\mu:=Ey_t$. Next theorem provides one of our main results, which is the consistent estimation of the precision matrix for asset returns. This theorem is a crucial input in Sharpe-ratio.

Theorem 2. *(i). Under Assumptions 1-4, 6-7*

$$\max_{1 \leq j \leq p} \|\hat{\Gamma}_j - \Gamma_j\|_1 = O_p\left(\bar{s} l_n\right) = o_p(1).$$

(ii). Under Assumptions 1-6

$$\|\hat{\mu} - \mu\|_\infty = O_p\left(max(K\sqrt{\frac{lnn}{n}}, \sqrt{\frac{lnp}{n}})\right) = o_p(1).$$

This theorem is a new result in the literature, and merges two key concepts: factor models and nodewise regression in high dimensional econometrics-finance. Theorem 2 clearly shows that there is a tradeoff between the maximal eigenvalue of the errors, number of factors and the sparsity of the precision matrix. Clearly increasing number of factors in our model badly affect the rate of estimation of the precision matrix of the returns. In the case of fixed number of factors the rate in Theorem 2(i) is simplified to

$$r_n^2 \bar{s}^2 \sqrt{\frac{max(lnp, lnn)}{n}},$$

by using l_n definition in Assumption 7(ii).

3.1 Two examples relating precision matrix restriction to covariance matrix

We also tie the sparsity assumption on the precision matrix of errors to the structure of a covariance matrix of errors. We provide two examples for errors, one block-diagonal covariance matrix for errors, and the other one is Toeplitz form for covariance matrix of errors. Then we provide how they affect precision matrix, and Assumption 7(i).

Block Diagonal Covariance Matrix for Errors: There are $m = 1, \cdots, M$ blocks in a $p \times p$ square matrix of covariance of errors.

$$\Sigma_n := \begin{bmatrix} \Sigma_{n,p_1} & & \\ & \Sigma_{n,p_m} & \\ & & \Sigma_{n,p_M} \end{bmatrix}.$$

Each $\Sigma_{n,p_m} : p_m \times p_m$ block and $\sum_{m=1}^{M} p_m = p$. Clearly the inverse is sparse as well

$$\Omega := \Sigma_n^{-1} = \begin{bmatrix} \Sigma_{n,p_1}^{-1} & & \\ & \Sigma_{n,p_m}^{-1} & \\ & & \Sigma_{n,p_M}^{-1} \end{bmatrix}.$$

The sparsity assumption-Assumption 1- for Ω can be translated into Σ_n as $\max_{1 \leq m \leq M} \max_{1 \leq j \leq p_m} s_{p_m} = \bar{s}$, where this is the maximum number of nonzero cells in a given row of a block, across all blocks. For

Assumption 7 we need the following inequality from Corollary 6.1.5 of Horn and Johnson (2013), by seeing that spectral radius of a matrix is larger than or equal to absolute value of any eigenvalue for any square matrix, A, hence

$$Eigmax(A) \leq min(\|A\|_{l_1}, \|A\|_{l_\infty}). \tag{36}$$

For the same inequality also see Theorem 5.6.9a of Horn and Johnson (2013). Relating to Assumption 7(i)

$$Eigmax(\Sigma_n) \leq \max_{1 \leq j_1 \leq p} \sum_{j_2=1}^{p} |\Sigma_{n,j_1,j_2}| = \max_{1 \leq j_1 \leq p} \sum_{j_2=1}^{p} |Eu_{j_1,t}u_{j_2,t}|,$$

where Σ_{n,j_1,j_2} is the j_1, j_2 element of covariance matrix of errors. By (36) this last inequality becomes

$$Eigmax(\Sigma_n) \leq \max_{1 \leq m \leq M} \max_{1 \leq j_1 \leq p_m} \sum_{j_2=1}^{p_m} |Eu_{j_1,t}u_{j_2,t}|.$$

It is easy to see that using Assumption 1(i), and under sufficient conditions for Assumption 7(i), with $p_m \to \infty$ as $n \to \infty$

$$\max_{1 \leq m \leq M} \max_{1 \leq j_1 \leq p_m} \max_{1 \leq j_2 \leq p_m} |Eu_{j_1,t}u_{j_2,t}| \leq C < \infty, \ \max_{1 \leq m \leq M} \frac{p_m}{p} \to 0, \ r_n := max_{1 \leq m \leq M}\, p_m, \tag{37}$$

we get $Eigmax(\Sigma_n) \leq Cr_n, r_n/p \to 0$. This allows size of the blocks may be increasing with p, but the ratio of the maximum block size to total number of parameters should be small.

Toeplitz Analysis: the correlation among errors are $Eu_{j,t}u_{i,t} = \rho^{|i-j|}$, with $|\rho| < 1$.Then

$$\Sigma_n = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{p-1} \\ \rho & 1 & \cdots & \cdots & \cdots \\ \vdots & \cdots & \cdots & \cdots & 1 \end{bmatrix}.$$

We have the tri-diagonal inverse, with all other cells being zero except the main and two adjacent diagonals.

$$\Sigma_n^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & 0 & 0 & \cdots \\ -\rho & 1+\rho^2 & -\rho & 0 & \cdots \\ 0 & -\rho & 1+\rho^2 & -\rho & \cdots \\ & & & & 1 \end{bmatrix}.$$

Clearly $\bar{s} = 3$, and the covariance matrix for errors is not sparse. For Assumption 7(i), using (36)

$$Eigmax(\Sigma_n) \leq \|\Sigma_n\|_{l_\infty} = \max_{1 \leq j_1 \leq p} \sum_{j_2=1}^{p} |\rho^{|j_2 - j_1|}|.$$

Clearly Assumption 7(i) is satisfied since the sum on the right side converges to a constant.

3.2 Algorithm For Asset Return Based Precision Matrix Estimate

Here we provide a practical algorithm to get the precision matrix estimate for asset returns: $\hat{\Gamma}$, and it will depend on the precision matrix estimate for the residual based nodewise regression estimate $\hat{\Omega}$, and its symmetric version $\hat{\Omega}_{sym}$.

1. Use (19) to set up the residual from a least squares based regression via known factors with y_j as the j the asset returns $(n \times 1)$

$$\hat{u}_j = y_j - X'\hat{b}_j,$$

with $\hat{b}_j = (XX')^{-1}Xy_j$, and $X = (f_1, \cdots, f_t, \cdots, f_n) : K \times n$ matrix with $f_t : K \times 1$ known factor vector.

2. Form the transpose matrix of residuals for all asset returns except j th one; $\hat{U}'_{-j}$ which is $n \times p-1$ matrix as in (21)

$$\hat{U}'_{-j} = Y'_{-j} - X'\hat{B}'_{-j},$$

where $\hat{B}'_{-j} = (XX')^{-1}XY'_{-j}$, $(K \times p-1$ matrix), which is the transpose of factor loading estimates, $Y'_{-j} : (n \times p - 1)$ is the transpose matrix of asset returns except the j the asset.

3. Run (24), nodewise regression of $\hat{u}_j$ on $\hat{U}'_{-j}$ via lasso, and get λ_n from Cross-validation or Generalized Information Criterion as in Section 7.1.

4. Use (25) to get $\hat{\tau}_j^2$.

5. Now form $\hat{\Omega}'_j$ which is a row in the precision matrix estimate for the errors with $1/\hat{\tau}_j^2$ as j th element of that j th row, and put all other elements of the j th row, as $-\hat{\gamma}'_j/\hat{\tau}_j^2$.

6. Run steps 1-5 for all $j = 1, \cdots, p$. Stack all rows $j = 1, \cdots, p$ to form $p \times p$ matrix: $\hat{\Omega}$. Form symmetric version by $\hat{\Omega}_{sym} := \frac{\hat{\Omega}+\hat{\Omega}'}{2}$.

7. Form

$$\hat{B} = (YX')(XX')^{-1}$$

which is $p \times K$ matrix of OLS estimates, where $Y : p \times n$ matrix of all asset returns, where $j = 1, \cdots, p$ represent all columns-assets, and rows $t = 1, \cdots, n$ time periods. Also form the covariance matrix estimate for factors

$$\widehat{cov f_t} = n^{-1}XX' - n^{-2}X1_n1_n'X'$$

where 1_n is $n \times 1$ column vector of ones.

8. Now form the precision matrix estimate for all asset returns by (32), by steps 6-7

$$\hat{\Gamma} = \hat{\Omega} - \hat{\Omega}\hat{B}[(\widehat{cov f_t})^{-1} + \hat{B}'\hat{\Omega}_{sym}\hat{B}]^{-1}\hat{B}'\hat{\Omega}.$$

We use $\hat{\Omega}_{sym}$ in the inverse in square brackets, so that we can use specific inequalities for the inverse in our proof. $\hat{\Omega}$ is the nodewise regression estimator, and $\hat{\Omega}_{sym}$ is the symmetrized version.

3.3 Why use Nodewise Regression?

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

4 Maximum Out-of-Sample Sharpe Ratio

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

$$argmax_w w'\mu \quad subject\,to \quad w'\Sigma_y w \leq \sigma^2, \tag{38}$$

where we maximize the return subject to a specified positive and finite risk constraint, $\sigma^2 > 0$. After solving for the optimal weight, which will be shown in the next subsection, we can obtain the maximum out-of-sample Sharpe ratio. Equation (A.2) of Ao et al. (2019) defines the estimated maximum out-of-sample ratio when $p < n$, with the inverse of the sample covariance matrix, $\hat{\Sigma}_y^{-1} = [\frac{1}{n}\sum_{t=1}^{n} y_t y_t']^{-1}$ used as an estimator for the precision matrix estimate, as:

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

and the theoretical version as, since $\Gamma := \Sigma_y^{-1}$

$$SR^* := \sqrt{\mu'\Gamma\mu}.$$

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

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

We need the following sparsity assumption in Assumption 8(i) below that replaces Assumption 7(ii). Assumption 8(ii) is an assumption on mainly the signal in the problem, combined with Assumption 7(i), this new assumption tells that signal will dominate noise (in terms of maximum eigenvalues). More sufficient conditions and explanations are relegated to at the end of Supplement Appendix about Assumption 8(ii). We allow maximum eigenvalue of the covariance matrix of returns to be increasing with n, but at a rate at most K. Note that number of factors K is nondecreasing in n.

Assumption 8. *(i).*

$$K^3 \bar{s} l_n = o(1).$$

(ii). For a positive constant $C > 0$

$$Eigmax(\Sigma_y) \leq CK.$$

Now we provide one of our main results in the next theorem.

Theorem 3. *Under Assumptions 1-4,6, 7(i), 8*

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

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

2. Also see that with fixed number of factors, the rate in Theorem 3 is, by the definition of l_n in (35)

$$r_n^2 \bar{s}^2 \sqrt{\frac{max(lnp, lnn)}{n}}.$$

This rate is the same as in the precision matrix estimate of the returns with fixed number of factors, as shown in the paragraph after Theorem 2.

4.1 Mean-Variance Efficiency When $p > n$

This subsection formally shows that we can obtain mean-variance efficiency in an out-of-sample context when the number of assets in the portfolio is larger than the sample size, a novel result in the literature. Ao et al. (2019) show that this is possible when $p \leq n$, when both p, and n are large. That article is a very important contribution since they also demonstrate that other methods before theirs could not obtain that result, and it is a difficult issue to address. Given a finite risk level of $\sigma^2 > 0$, the optimal weights of a portfolio are given in (2.3) of Ao et al. (2019) in an out-of-sample context. This comes from maximizing the expected portfolio return subject to its variance being constrained by the square of the risk, where this is shown in (38). Since $\Gamma := \Sigma_y^{-1}$, the formula for weights is

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

The estimates that we will use

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

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

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

Theorem 4. *(i). Under Assumptions 1-4, 6,7(i),8*

$$\left| \frac{\mu' \hat{w}_{oos}}{\mu' w_{oos}} - 1 \right| = O_p(K^2 \bar{s} l_n) = o_p(1).$$

(ii). Under Assumptions 1-4,6,7(i),8

$$\left|\hat{w}_{oos}'\Sigma_y\hat{w}_{oos} - \sigma^2\right| = O_p(K^2\bar{s}l_n) = o_p(1).$$

Remarks. 1. From the results, we allow $p > n$, and still there is consistency.

2. From the convergence rates, it is clear that we are penalized by the number of assets but in a logarithmic fashion; hence, our method is feasible to use in large-portfolio cases. This logarithmic rate in assets can be seen from the definition of l_n in (35).

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

5 Maximum Sharpe Ratio: Portfolio Weights Normalized to One

In this section, we define the maximum Sharpe ratio when the weights of the portfolio are normalized to one. This in turn will depend on a critical term that will determine the formula below.
The maximum Sharpe ratio is defined as follows, with w as the $p \times 1$ vector of portfolio weights:

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

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

$$MSR^2 = \mu'\Sigma_y^{-1}\mu. \tag{39}$$

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

$$MSR_c^2 = \mu'\Sigma_y^{-1}\mu - (1_p'\Sigma_y^{-1}\mu)^2/(1_p'\Sigma_y^{-1}1_p). \tag{40}$$

This is equation (6.1) of Maller et al. (2016). Equation (39) is used in the literature, and this is the formula when the weights do not necessarily sum to one given a return constraint as in Ao et al. (2019).

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

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

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

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

First, when $1_p'\Sigma_y^{-1}\mu \geq 0$, we have the square of the maximum Sharpe ratio as in (39). To obtain an estimate by using nodewise regression, we replace $\Gamma := \Sigma_y^{-1}$ with $\hat{\Gamma}$. Namely, the estimate of the square of the maximum Sharpe ratio is:

$$\widehat{MSR}^2 = \hat{\mu}'\hat{\Gamma}\hat{\mu}. \tag{41}$$

Using the result in Theorem 2, we can obtain the consistency of the maximum Sharpe ratio (squared).

Theorem 5. *Under Assumptions 1-4, 6,7(i), 8 with* $1_p'\Gamma\mu \geq 0$*,*

$$\left| \frac{\widehat{MSR}^2}{MSR^2} - 1 \right| = O_p(K^2 \bar{s} l_n) = o_p(1).$$

Remark. To the best of our knowledge, no existing result deals with MSR when $p > n$ and p can grow exponentially in n. We also allow for time-series data and establish a rate of convergence. The number of assets, on the other hand, can also increase the error by on a logarithmic scale as can be seen in (35).

Note that the maximum Sharpe ratio above relies on $1_p'\Sigma_y^{-1}\mu \geq 0$, where 1_p is a column vector of ones. This was recently pointed out in equation (6.1) Maller et al. (2016). If $1_p'\Sigma_y^{-1}\mu < 0$, the Sharpe ratio is minimized, as shown on p.503 of Maller and Turkington (2002). The new maximum Sharpe ratio in the case when $1_p'\Sigma_y^{-1}\mu < 0$ is in Theorem 2.1 of Maller and Turkington (2002). The square of the maximum Sharpe ratio when $1_p'\Sigma_y^{-1}\mu < 0$ is given in (40).

An estimator in this case is

$$\widehat{MSR}_c^2 = \hat{\mu}'\hat{\Gamma}\hat{\mu} - (1_p'\hat{\Gamma}\hat{\mu})^2/(1_p'\hat{\Gamma}1_p). \tag{42}$$

The optimal portfolio allocation for such a case is given in (2.10) of Maller and Turkington (2002). The limit for such estimators when the number of assets is fixed (p fixed) is given in Theorems 3.1b-c of Maller et al. (2016). We set up some notation for the next theorem. Set $1_p'\Sigma_y^{-1}1_p/p = A$, $1_p'\Sigma_y^{-1}\mu/p = F$, $\mu'\Sigma_y^{-1}\mu/p = D$.

Theorem 6. *If* $1_p'\Sigma_y^{-1}\mu < 0$*, and under Assumptions 1-4,6,7(i), 8 with* $AD - F^2 \geq C_1 > 0$*, where* C_1 *is a positive constant,*

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

Remarks. 1. Condition $AD - F^2 \geq C_1 > 0$ is not restrictive, and it is used in Callot et al. (2021) as a condition that helps us to obtain a finite optimal portfolio variance in the Markowitz (1952) mean-variance portfolio below.

2. In Theorem 5, we allow $p > n$, and time-series data are allowed, unlike the iid or normal return cases in the literature when dealing with large p, n. Theorem 6 is new and will help us establish a new result in the following Theorem.

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

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

where we excluded the case of $1_p' \hat{\Gamma} \hat{\mu} = 0$ in the estimator. That specific scenario is very restrictive in terms of returns and variance. Note that under a mild assumption, when $1_p' \Sigma_y^{-1} \mu > 0$, we have $1_p' \hat{\Gamma} \hat{\mu} > 0$, and when $1_p' \Sigma_y^{-1} \mu < 0$, we have $1_p' \hat{\Gamma} \hat{\mu} < 0$ with probability approaching one in the proof of Theorem 7.

Theorem 7. *Under Assumptions 1-4,6,7(i), 8, with $AD - F^2 \geq C_1 > 0$, where C_1 is a positive constant, and assuming $|1_p' \Sigma_y^{-1} \mu|/p \geq C > 2\epsilon > 0$, with a sufficiently small positive $\epsilon > 0$, and C being a positive constant,*

$$\left| \frac{(\widehat{MSR}^*)^2}{(MSR^*)^2} - 1 \right| = O_p(K^2 \bar{s} l_n) = o_p(1).$$

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

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

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

$$\epsilon = O(K \bar{s} l_n) = o(1),$$

where the asymptotically small term follows Assumption 8.

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

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

4. Also see that when the number of factors are fixed, $K = O(1)$, then we see that approximation error improves, given the definition of l_n in (35) to

$$O(r_n^2 \bar{s}^2 \sqrt{\frac{max(lnp, lnn)}{n}}) = o(1).$$

6 Commonly Used Portfolios with a Large Number of Assets

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

6.1 Global Minimum-Variance Portfolio

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

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

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

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

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

$$SR = \frac{w_u' \mu}{\sqrt{w_u' \Sigma_y w_u}} = \sqrt{p} (\frac{1_p' \Sigma_y^{-1} \mu}{p}) (\frac{1_p' \Sigma_y^{-1} 1_p}{p})^{-1/2}. \tag{43}$$

We estimate (43) by nodewise regression, noting that $\Gamma := \Sigma_y^{-1}$,

$$\widehat{SR}_{nw} = \sqrt{p} (\frac{1_p' \hat{\Gamma} \hat{\mu}}{p}) (\frac{1_p' \hat{\Gamma} 1_p}{p})^{-1/2}. \tag{44}$$

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

Theorem 8. *Under Assumptions 1-4, 6,7(i), 8 with* $|1_p' \Sigma_y^{-1} \mu|/p \geq C > 2\epsilon > 0$,

$$\left| \frac{\widehat{SR}_{nw}^2}{SR^2} - 1 \right| = O_p(K^{5/2} \bar{s} l_n) = o_p(1).$$

Remarks.

1. We see that a large p only affects the error by a logarithmic factor as in definition of l_n in (35). The estimation error increases with the non-sparsity of the precision matrix.

2. We see that factors affect global minimum variance portfolio worse than the constrained portfolio in Theorem 7. This is related to different optimization here.

3. Also with fixed number of factors, $K = O(1)$ the approximation rate in theorem here and Theorem 7, Remark 4, is the same

$$O(r_n^2 \bar{s}^2 \sqrt{\frac{max(lnp, lnn)}{n}}) = o(1).$$

6.2 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

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

The formula for optimal weight is

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

which can be rewritten as

$$w_{MV} = \left[\frac{D - \rho_1 F}{AD - F^2} \right] (\Sigma_y^{-1} 1_p / p) + \left[\frac{\rho_1 A - F}{AD - F^2} \right] (\Sigma_y^{-1} \mu / p), \tag{45}$$

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

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

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

The estimate of that variance is

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

By our constraint, we obtain

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

Using the variance V above

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

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

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

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

Theorem 9. *Under Assumptions 1-4, 6,7(i), 8 with condition* $|1_p' \Sigma_y^{-1} \mu / p| \geq C > 2\epsilon > 0$ *and* $AD - F^2 \geq C_1 > 0$, $A\rho_1^2 - 2F\rho_1 + D \geq C_1 > 0$, *with* ρ_1 *uniformly bounded away from zero and infinity, we have*

$$\left| \frac{\widehat{SR}_{MV}^2}{SR_{MV}^2} - 1 \right| = O_p(K^3 \bar{s} l_n) = o_p(1).$$

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

2. We provide the rate of convergence of the estimators, which increases with p in a logarithmic way as in l_n definition in (35), and the non-sparsity of the precision matrix affects the error in a linear way.

3. This is the scenario that factors affect the estimation of Sharpe-ratio strongest. With fixed number of factors, as in previous theorem we can match the rate of approximation as in Theorem 7, Remark 4 there.

7 Simulations

7.1 Models and Implementation Details

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

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

The POET uses principal components to estimate the covariance matrix and allows some eigenvalues of Σ_n to be spiked and grow at a rate $O(p)$, which allows common and idiosyncratic components to be identified via principal components analysis and can consistently estimate the space spanned by the eigenvectors of

Σ_n. However, Fan et al. (2013) point out that the absolute convergence rate of the model is not satisfactory for estimating Σ_n and consistency can only be achieved in terms of the relative error matrix.

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

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

Regarding implementation, the POET and both models from Ledoit and Wolf (2017) are available in the R packages POET Fan et al. (2016) and nlshrink Ramprasad (2016). The SF-NL-LW need some minor adjustments following the procedures described in Ledoit and Wolf (2017). For the MAXSER, we follow the steps for the non-factor case in Ao et al. (2019), and we use the package lars (Hastie and Efron (2013)) for the penalized regression estimation. We estimate the nodewise regression following the steps in Section 3.2 using the glmnet package Friedman et al. (2010) for penalized regressions. We used two alternatives to select the regularization parameter λ, a 10-fold cross validation (CV) and the generalized information criterion (GIC) from Zhang et al. (2010).

The GIC procedure starts by fitting $\hat{\gamma}_j$ in (24) for a range of λ_j that goes from the intercept-only model to the largest feasible model. This is automatically done by the glmnet package. Then, for the GIC procedure, we calculate the information criterion for a given λ_j among the ranges of all possible tuning parameters

$$GIC_j(\lambda_j) = \frac{SSR(\lambda_j)}{n} + q(\lambda_j)\log(p-1)\frac{\log(\log(n))}{n} \tag{46}$$

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

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

7.2 Data Generation Process and Results

The DGP is based on a simplified version of the factor DGP in Ao et al. (2019), for $j = 1, \cdots, p$:

$$y_j = \alpha_j + \sum_{k=1}^{K} \beta_{j,k} f_k + e_j, \tag{47}$$

where y_j, f_j are the monthly asset returns, factor returns respectively, $\beta_{j,k}$ are the individual stock sensitivities to the factors, and $\alpha_j + e_j$ represent the idiosyncratic component of each stock. We start with two specifications that correspond to two Tables. Table 1 corresponds to 1 factor, which is excess return of the market portfolio, hence $K = 1$, and Table 2 corresponds to 3 factors from the Fama & French three factors, $K = 3$. [2] Let $\boldsymbol{\mu}_f$ and $\boldsymbol{\Sigma}_f$ be the factors' sample mean and covariance matrix. The β, and α and covariance matrix of residuals: $\widetilde{\Sigma_n}$ are estimated using a simple least squares regression using returns from the S&P500 stocks that were part of the index in the entire period from 2008 to 2017. In each simulation, we randomly select 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 covariance matrix of errors: $\widehat{\boldsymbol{\Sigma_n}} = \widetilde{\Sigma_n} \odot Toeplitz(\rho)$, where $Toeplitz(\rho)$ is the $p \times p$ matrix of the form, for (i,j) th element

$$Toeplitz(\rho)_{i,j} := \rho^{|i-j|},$$

with $\rho = 0.25, 0.5, 0.75$. $A \odot B$ represent element by element multiplication (Hadamard product) of two square matrices A, B of the same dimensions.

Tables 1-2 show the results. The values in each cell show the average absolute estimation error for estimating the square of the Sharpe ratio in the case of global minimum-variance and Markowitz mean-variance portfolios in Section 6, out-of-sample forecasting, and the maximum Sharpe ratio in the case of constrained portfolio optimization in Sections 4-5 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 $p = n/2$, and the last four columns are for $p = 3n/2$. MSR, MSR-OOS, GMV-SR and MKW-SR are the constrained maximum Sharpe ratio, the out-of-sample maximum Sharpe ratio, the Sharpe ratio from the global minimum-variance portfolio and the Sharpe ratio from the Markowitz portfolio with target returns set to 1%, respectively. Therefore, there are four categories to evaluate the different estimates. The MAXSER risk constraint was set to 0.04 following Ao et al. (2019). We ran 100 iterations in each simulation setup. All bold-face entries in tables show category champions.

Both Tables show that our method achieves consistency as shown in Theorems. Analyzing $K = 3$, Table 2, with $\rho = 0.50$ OOS-MSR (Out Of Sample-Maximum Sharpe Ratio) and Generalized Information Criterion tuning parameter selection, the estimation error at $p = 3n/2$, with $n = 100$ is 1.273, and this error declines to 0.666 at $p = 3n/2, n = 200$, and then declines to 0.360 at $p = 3n/2, n = 400$. So with jointly increasing n, p we show that the error declines, as predicted by our theorems. The main reason is that errors

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

grow with $\sqrt{lnp}$ but decline with $n^{1/2}$ rate. So the number of assets in a large portfolio only affects the error logarithmically. To give another example from Table 2, with $\rho = 0.50$, GMV-SR (Global Minimum Variance-Sharpe Ratio) and Cross Validation tuning parameter selection with our method, the estimation error is 0.743 with $p = n/2, n = 100$, then this error declines to 0.239 with $p = n/2, n = 200$, and further declines to 0.160 with $p = n/2, n = 400$.

Next, we consider which method achieves the smallest estimation error. Table 1 clearly favors SF-NL-LW (Single Factor Non-Linear Shrinkage of Ledoit-Wolf) since it has single factor built into this subset of their technique. We get better results in Table 2 ($K = 3$) for our methods. We have 4 categories: MSR, OOS-MSR, GMV-SR, MKW-SR corresponding to our Theorems 3-9. There are 9 possibilities in each category (given we are either at $p = n/2$ or $p = 3n/2$), representing 3 choices of sample sizes paired with 3 choices of different Toeplitz structure.

We analyze each category. We start with Table 1. With $p = 3n/2$ in OOS-MSR our NW-GIC method has the smallest errors 8 out of 9 categories. When $p = n/2$, MAXSER method dominates all others since it is specifically factor model designed to handle OOS-MSR with $p < n$. In GMV-SR, with $p = n/2$, in 3 out of 9 cases, our NW-GIC dominates. In the other categories in Table 1, non-linear shrinkage method of Ledoit-Wolf (2017) does the best, but our methods come very close second.

In Table 2, with $K = 3$, our methods perform better than in Table 1. We start with $p = 3n/2$, and in OOS-MSR, our NW-GIC provides the best result with the smallest error at all $\rho = 0.25, 0.5, 0.75$. This is a very good result, since this measures the estimation error in Out-Of-Sample Maximum Sharpe Ratio in large portfolios which reflects out of sample performance compared to other three categories. To give a brief example, with $\rho = 0.75, p = 3n/2, n = 400$, our NW-GIC error is 0.404, and our NW-CV method comes second with 0.405. SF-NL-LW (Single Factor Non-Linear Shrinkage of Ledoit-Wolf) comes third with 0.417, the other methods fare poorly, with POET coming with 3.607 error. POET depends on sparsity (conditional) of covariance of errors, hence it suffers under a more realistic non-sparse covariance matrix error setup. In case of $p = n/2$ with OOS-MSR category MAXSER method dominates the others, our methods come second. In the category of GMV-SR, with $p = 3n/2$, out of 9 possible configurations our methods have the least error in 7 cases. In case of the same category but with $p = 0.5n$, 5 out of 9 possibilities, our methods dominate. In case of the category of MKW-SR (Markowitz-Sharpe Ratio), our theorems predict that our methods may suffer from number of factors. In this category, we see that non-linear shrinkage methods are the best, and our methods are the second best. In the case of constrained maximum sharpe ratio, (MSR) again non-linear shrinkage methods perform the best, but differences between the methods are small.

Table 1: Simulation Results – Single Factor Toeplitz DGP with Real Factors

n = 100

	p = n/2				p = 3n/2			
	MSR	OOS-MSR	GMV-SR	MKW-SR	MSR	OOS-MSR	GMV-SR	MKW-SR
Toeplitz $\rho = 0.25$								
NW-GIC	0.517	1.062	0.680	0.114	0.520	1.095	0.251	0.126
NW-CV	0.517	1.061	0.634	0.112	0.521	1.099	0.251	0.128
POET	0.526	1.055	0.636	0.167	0.522	1.095	0.260	0.144
NL-LW	**0.487**	1.705	**0.559**	0.172	**0.480**	2.249	0.377	0.333
SF-NL-LW	0.516	1.069	0.689	**0.110**	0.517	**1.094**	**0.249**	**0.120**
MAXSER		**0.359**						
Toeplitz $\rho = 0.5$								
NW-GIC	0.525	1.067	0.829	0.134	0.529	**1.095**	**0.266**	0.157
NW-CV	0.526	1.067	0.726	0.132	0.531	1.099	0.267	0.159
POET	0.535	1.061	0.721	0.190	0.531	1.096	0.279	0.175
NL-LW	**0.495**	1.694	**0.558**	0.151	**0.489**	2.231	0.360	0.290
SF-NL-LW	0.523	1.076	0.819	**0.130**	0.526	1.095	0.266	**0.151**
MAXSER		**0.363**						
Toeplitz $\rho = 0.75$								
NW-GIC	0.542	1.105	1.300	0.183	0.549	**1.131**	0.318	0.227
NW-CV	0.542	1.108	1.131	0.183	0.551	1.144	0.319	0.231
POET	0.553	1.104	1.086	0.248	0.552	1.136	0.333	0.249
NL-LW	**0.510**	1.703	**0.548**	**0.109**	**0.510**	2.215	0.334	**0.187**
SF-NL-LW	0.537	1.115	0.922	0.176	0.545	1.136	**0.315**	0.220
MAXSER		**0.371**						

n = 200

	p = n/2				p = 3n/2			
	MSR	OOS-MSR	GMV-SR	MKW-SR	MSR	OOS-MSR	GMV-SR	MKW-SR
Toeplitz $\rho = 0.25$								
NW-GIC	0.331	0.515	**0.211**	0.073	0.352	**0.545**	0.140	0.079
NW-CV	0.331	0.514	0.212	**0.072**	0.352	0.546	0.140	0.079
POET	0.336	0.511	0.212	0.102	0.354	0.548	0.145	0.089
NL-LW	**0.301**	0.961	0.322	0.216	**0.300**	1.350	0.329	0.391
SF-NL-LW	0.330	0.515	0.215	0.072	0.350	0.545	**0.139**	**0.076**
MAXSER		**0.152**						
Toeplitz $\rho = 0.5$								
NW-GIC	0.342	0.521	**0.220**	0.100	0.365	**0.552**	0.161	0.113
NW-CV	0.342	0.521	0.221	0.100	0.365	0.553	0.161	0.113
POET	0.348	0.518	0.226	0.133	0.367	0.556	0.167	0.124
NL-LW	**0.306**	0.954	0.316	0.192	**0.313**	1.340	0.289	0.342
SF-NL-LW	0.340	0.523	0.225	**0.099**	0.363	0.553	**0.160**	**0.109**
MAXSER		**0.158**						
Toeplitz $\rho = 0.75$								
NW-GIC	0.366	0.558	0.248	0.166	0.390	**0.593**	0.224	0.189
NW-CV	0.366	0.560	0.248	0.167	0.390	0.595	0.223	0.189
POET	0.373	0.561	0.261	0.204	0.393	0.601	0.235	0.203
NL-LW	**0.324**	0.956	0.298	**0.132**	**0.337**	1.339	0.226	0.232
SF-NL-LW	0.361	0.563	**0.245**	0.157	0.387	0.597	**0.218**	**0.183**
MAXSER		**0.169**						

n = 400

	p = n/2				p=3n/2			
	MSR	OOS-MSR	GMV-SR	MKW-SR	MSR	OOS-MSR	GMV-SR	MKW-SR
Toeplitz $\rho = 0.25$								
NW-GIC	0.208	0.262	**0.091**	**0.051**	0.219	**0.273**	0.077	0.053
NW-CV	0.208	0.262	0.091	0.051	0.219	0.273	0.078	0.053
POET	0.212	0.263	0.096	0.066	0.220	0.276	0.081	0.058
NL-LW	**0.169**	0.645	0.265	0.258	**0.163**	0.931	0.333	0.416
SF-NL-LW	0.207	0.263	0.091	0.051	0.217	0.273	**0.076**	**0.051**
MAXSER		**0.098**						
Toeplitz $\rho = 0.5$								
NW-GIC	0.222	0.271	0.108	0.083	0.233	**0.283**	0.108	0.089
NW-CV	0.222	0.271	0.108	0.084	0.233	0.283	0.108	0.089
POET	0.226	0.273	0.118	0.100	0.235	0.286	0.113	0.095
NL-LW	**0.175**	0.641	0.233	0.231	**0.177**	0.926	0.281	0.365
SF-NL-LW	0.220	0.273	**0.106**	**0.082**	0.232	0.284	**0.106**	**0.087**
MAXSER		**0.091**						
Toeplitz $\rho = 0.75$								
NW-GIC	0.250	0.309	0.174	0.156	0.264	**0.326**	0.193	0.170
NW-CV	0.250	0.311	0.174	0.157	0.264	0.327	0.193	0.171
POET	0.256	0.319	0.192	0.179	0.267	0.335	0.201	0.180
NL-LW	**0.196**	0.647	0.194	0.151	**0.202**	0.931	**0.186**	0.252
SF-NL-LW	0.244	0.313	**0.160**	**0.149**	0.261	0.330	0.187	**0.167**
MAXSER		**0.082**						

The table shows the simulation results for the Toeplitz DGP. Each simulation was done with 100 iterations. We used sample sizes n of 100, 200 and 400, and the number of stocks was either $n/2$ or $1.5n$ for the low-dimensional and the high-dimensional case, respectively. Each block of rows shows the results for a different value of ρ in the Toeplitz DGP. The values in each cell show the average absolute estimation error for estimating the square of the Sharpe ratio in the case of the global minimum-variance and Markowitz mean-variance portfolios in Section 6 and maximum Sharpe ratio in the case of out-of-sample forecasting and constrained portfolio optimization in Sections 4-5 across iterations.

Table 2: Simulation Results – 3 Factor Toeplitz DGP with Real Factors

n = 100

	p = n/2				p = 3n/2			
	MSR	OOS-MSR	GMV-SR	MKW-SR	MSR	OOS-MSR	GMV-SR	MKW-SR
Toeplitz $\rho = 0.25$								
NW-GIC	0.544	1.237	0.739	0.152	0.579	**1.269**	0.343	0.220
NW-CV	0.543	1.237	**0.724**	0.151	0.580	1.309	**0.341**	0.224
POET	0.558	1.137	1.041	0.308	0.590	1.641	0.463	0.391
NL-LW	**0.511**	1.726	0.906	**0.110**	**0.535**	2.251	0.481	**0.086**
SF-NL-LW	0.540	1.168	0.736	0.199	0.567	1.292	0.349	0.238
MAXSER		**0.375**						
Toeplitz $\rho = 0.50$								
NW-GIC	0.548	1.247	0.760	0.163	0.584	**1.273**	0.352	0.232
NW-CV	0.548	1.248	**0.743**	0.162	0.585	1.314	**0.351**	0.236
POET	0.563	1.148	1.203	0.320	0.595	1.642	0.471	0.402
NL-LW	**0.516**	1.732	0.903	**0.107**	**0.540**	2.244	0.475	**0.079**
SF-NL-LW	0.543	1.181	0.753	0.206	0.572	1.297	0.358	0.249
MAXSER		**0.379**						
Toeplitz $\rho = 0.75$								
NW-GIC	0.556	1.295	0.812	0.189	0.593	**1.318**	**0.377**	0.263
NW-CV	0.556	1.301	0.802	0.187	0.595	1.340	0.383	0.268
POET	0.571	1.200	2.071	0.347	0.605	1.679	0.490	0.428
NL-LW	**0.523**	1.727	0.947	**0.106**	**0.550**	2.257	0.468	**0.071**
SF-NL-LW	0.549	1.228	**0.783**	0.222	0.581	1.346	0.382	0.272
MAXSER		**0.385**						

n = 200

	p = n/2				p=3n/2			
	MSR	OOS-MSR	GMV-SR	MKW-SR	MSR	OOS-MSR	GMV-SR	MKW-SR
Toeplitz $\rho = 0.25$								
NW-GIC	0.369	0.578	**0.235**	0.117	0.391	**0.658**	0.200	0.158
NW-CV	0.369	0.578	0.235	0.117	0.391	0.658	**0.199**	0.158
POET	0.418	0.852	0.372	0.346	0.445	2.078	0.459	0.411
NL-LW	**0.348**	0.988	0.289	**0.084**	**0.358**	1.370	0.260	**0.118**
SF-NL-LW	0.366	0.581	0.235	0.152	0.383	0.679	0.203	0.160
MAXSER		**0.166**						
Toeplitz $\rho = 0.50$								
NW-GIC	0.376	0.585	0.239	0.132	0.398	**0.666**	0.213	0.172
NW-CV	0.376	0.585	0.239	0.132	0.398	0.666	**0.212**	0.173
POET	0.424	0.855	0.382	0.358	0.451	2.065	0.472	0.422
NL-LW	**0.349**	0.990	0.293	**0.077**	**0.366**	1.372	0.253	**0.101**
SF-NL-LW	0.371	0.589	**0.237**	0.160	0.390	0.687	0.214	0.173
MAXSER		**0.168**						
Toeplitz $\rho = 0.75$								
NW-GIC	0.387	0.623	0.260	0.164	0.411	**0.708**	0.246	0.206
NW-CV	0.388	0.624	0.260	0.165	0.412	0.709	0.246	0.207
POET	0.436	0.886	0.409	0.385	0.464	2.068	0.501	0.448
NL-LW	**0.358**	1.009	0.287	**0.066**	**0.377**	1.397	0.242	**0.068**
SF-NL-LW	0.381	0.631	**0.246**	0.179	0.402	0.731	**0.241**	0.200
MAXSER		**0.173**						

n = 400

	p = n/2				p = 3n/2			
	MSR	OOS-MSR	GMV-SR	MKW-SR	MSR	OOS-MSR	GMV-SR	MKW-SR
Toeplitz $\rho = 0.25$								
NW-GIC	0.242	0.311	**0.148**	**0.085**	0.254	**0.350**	**0.122**	0.100
NW-CV	0.242	0.311	0.148	0.085	0.254	0.350	0.122	0.100
POET	0.338	1.279	0.436	0.371	0.351	3.682	0.484	0.410
NL-LW	**0.216**	0.642	0.206	0.092	**0.223**	0.953	0.206	0.169
SF-NL-LW	0.240	0.321	0.157	0.090	0.251	0.361	0.126	**0.096**
MAXSER		**0.081**						
Toeplitz $\rho = 0.50$								
NW-GIC	0.251	0.321	**0.159**	0.100	0.263	**0.360**	**0.143**	0.115
NW-CV	0.251	0.321	0.160	0.100	0.263	0.360	0.143	0.115
POET	0.346	1.276	0.451	0.383	0.359	3.648	0.498	0.420
NL-LW	**0.226**	0.648	0.206	**0.083**	**0.229**	0.955	0.185	0.151
SF-NL-LW	0.248	0.331	0.167	0.102	0.260	0.371	0.146	**0.110**
MAXSER		**0.080**						
Toeplitz $\rho = 0.75$								
NW-GIC	0.267	0.360	0.198	0.136	0.281	**0.404**	0.192	0.150
NW-CV	0.268	0.362	0.198	0.137	0.281	0.405	0.192	0.150
POET	0.361	1.292	0.483	0.410	0.375	3.607	0.529	0.445
NL-LW	**0.234**	0.659	**0.195**	**0.062**	**0.242**	0.979	**0.155**	**0.109**
SF-NL-LW	0.260	0.371	0.195	0.127	0.275	0.417	0.188	0.141
MAXSER		**0.079**						

The table shows the simulation results for the Toeplitz DGP. Each simulation was done with 100 iterations. We used sample sizes n of 100, 200 and 400, and the number of stocks was either $n/2$ or $1.5n$ for the low-dimensional and the high-dimensional case, respectively. Each block of rows shows the results for a different value of ρ in the Toeplitz DGP. The values in each cell show the average absolute estimation error for estimating the square of the Sharpe ratio in the case of the global minimum-variance and Markowitz mean-variance portfolios in Section 6 and maximum Sharpe ratio in the case of out-of-sample forecasting and constrained portfolio optimization in Sections 4-5 across iterations.

8 Empirical Application

For the empirical application, we use two subsamples. The first subsample uses data from January 1995 to December 2019 with an out-of-sample period from January 2005 to December 2019. We selected all stocks that were in the S&P 500 index for at least one month in the out-of-sample period and have data for the entire 1995-2019 period, which resulted in 382 stocks. The second subsample starts in January 1990 and ends in December 2019 with an out-of-sample period from January 2000 to December 2019. Using the same criterion as the first subsample, the number of stocks was 321, which is aroud 15% less stocks than the first subsample. The objective is to have an out-of-sample competition between models, we only estimated GMV and Markowitz portfolios for the plug-in estimators. The first out-of-sample period includes only the recession of 2008. The second out-of-sample period includes the recessions of 2000 and 2008, and the out-of-sample periods reflect recent history.

The Markowitz return constraint ρ_1 is 0.8% per month, and the MAXSER risk constraint is 4%. In the low-dimensional experiment, we randomly select 50 stocks from the pool to estimate the models with the same stocks for all windows. We also experimented with 25 stocks, but did not report. This is available from the authors on demand. In the high-dimensional case, we use all available stocks.

We use a rolling window setup for the out-of-sample estimation of the Sharpe ratio following Callot et al. (2021). Specifically, samples of size n are divided into in-sample $(1 : n_I)$ and out-of-sample $(n_I + 1 : n)$. We start by estimating the portfolio $\hat{w}_{nI}$ in the in-sample period and the out-of-sample portfolio returns $\hat{w}'_{nI} y_{n_I+1}$. 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 $\hat{w}_{n_I+1} y_{n_I+2}$. This procedure is repeated until the end of the sample.

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

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

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

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

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

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

$$SR = \frac{\hat{\mu}_{os}}{\widehat{\Sigma}_{y,os}}.$$

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

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

$$H_0 \;:\; SR_{NW} \leq SR_0 \;\; vs \;\; H_a \;:\; SR_{NW} > SR_0, \tag{48}$$

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

We also included equally weighted portfolio (EW). GMV-NW-GIC and GMV-NW-CV denote the nodewise method with GIC and cross validation tuning parameter choices, respectively, in the global minimum-variance portfolio (GMV).

In each of our feasible nodewise models with GIC, CV, we either use a single factor model (market as the only factor), or three factor model. They are denoted GMV-NW-GIC-SF, GMV-NW-GIC-3F for the global minimum variance portfolio analyzed with feasible nodewise method and GIC criterion for tuning parameter choice and single and three factor models respectively. In the same way, we define GMV-NW-CV-SF, GMV-NW-CV-3F. We take GMV-NW-GIC-SF as the benchmark to test against all other methods since it generally does well in different preliminary forecasts.

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

The results are presented in Tables 3 and 4. Table 3 shows the results for the 2005-2019 out-of-sample period. Feasible nodewise methods do well in terms of the Sharpe ratio in Table 3. For example, with transaction costs in the low-dimensional portfolio category, in terms of Sharpe ratio (SR) (averaged over the out-of-sample time period), GMV-NW-GIC-SF is the best model. It has an SR of 0.210. In the case of high dimensional case with transaction costs in the same table, GMV-POET and our GMV-NW-GIC-SF virtually tie (difference in favor of POET in fourth decimal) at 0.214 for the Sharpe-ratio.

If we were to analyze only the Markowitz portfolio in Table 3, with transaction costs in high dimensions, MW-NW-GIC-SF has the highest SR of 0.211. Therefore, even in other subcategories of Markowitz portfolio, the feasible nodewise method dominates. Although statistical significance is not established, it is not clear that these significance tests have high power in our high-dimensional cases.

Table 4 shows the results for the out-of-sample January 2000-2019 subsample. We see that feasible nodewise methods dominate all scenarios except for the low-dimensional case with transaction costs. In the case of high dimensionality with transaction costs, GMV-NW-GIC-sf (Markowitz-nodewise-GIC) has an SR of 0.225, and the closest is GMV-POET with 0.204. Also we experimented with two other out-sample periods of 2005-2017, 2000-2017, and the results are slightly better for our methods, and these can be shared on demand. We also try 25 stocks in the low dimensional case and the results for MAXSER were very similar, hence they are not reported. These results are available from authors on demand.

Table 3: Empirical Results – Out-of-Sample Period from Jan. 2005 to Dec. 2019

	Without TC								With TC							
	Low Dim.				High Dim.				Low Dim.				High Dim.			
	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	0.052	0.730	0.197	0.010	0.048	0.644	0.191	0.010	0.052	0.802	0.191	0.009	0.048	0.792
GMV-NW-GIC-SF	**0.229**	0.008	0.036		0.236	0.008	0.032		**0.210**	0.008	0.036		0.214	0.007	0.032	
GMV-NW-CV-SF	0.226	0.008	0.036	0.590	0.240	0.008	0.032	0.398	0.203	0.007	0.036	0.132	0.192	0.006	0.032	0.002
GMV-NW-GIC-3F	0.215	0.007	0.034	0.576	0.214	0.007	0.033	0.520	0.191	0.007	0.034	0.424	0.183	0.006	0.033	0.398
GMV-NW-CV-3F	0.212	0.007	0.034	0.474	0.226	0.007	0.032	0.790	0.183	0.006	0.034	0.278	0.132	0.004	0.032	0.032
GMV-POET	0.218	0.007	0.034	0.682	0.232	0.007	0.030	0.914	0.203	0.007	0.034	0.822	**0.214**	0.006	0.030	0.996
GMV-NL-LW	0.236	0.008	0.034	0.834	0.236	0.007	0.030	0.998	0.205	0.007	0.034	0.908	0.179	0.005	0.031	0.490
GMV-SF-NL-LW	0.216	0.007	0.034	0.684	**0.245**	0.007	0.030	0.886	0.190	0.007	0.034	0.546	0.184	0.006	0.030	0.600
MW-NW-GIC-SF	0.229	0.008	0.034	0.970	0.236	0.008	0.032	0.966	0.205	0.007	0.034	0.786	0.211	0.007	0.032	0.706
MW-NW-CV-SF	0.228	0.008	0.034	0.942	0.242	0.008	0.032	0.620	0.197	0.007	0.034	0.482	0.190	0.006	0.032	0.056
MW-NW-GIC-3F	0.214	0.007	0.033	0.628	0.217	0.007	0.033	0.606	0.185	0.006	0.033	0.444	0.183	0.006	0.033	0.416
MW-NW-CV-3F	0.212	0.007	0.033	0.574	0.225	0.007	0.032	0.790	0.177	0.006	0.033	0.302	0.125	0.004	0.032	0.032
MW-POET	0.223	0.007	0.032	0.880	0.229	0.007	0.030	0.844	0.200	0.006	0.032	0.794	0.207	0.006	0.030	0.840
MW-NL-LW	0.220	0.008	0.034	0.860	0.235	0.007	0.030	0.980	0.186	0.006	0.034	0.636	0.177	0.005	0.030	0.540
MW-SF-NL-LW	0.204	0.007	0.034	0.574	0.241	0.007	0.030	0.920	0.175	0.006	0.034	0.482	0.180	0.005	0.030	0.554
MAXSER	0.161	0.010	0.065	0.510					0.024	0.002	0.066	0.116				

The table shows the Sharpe ratio (SR), average returns (Avg), standard deviation (SD) and p-value of the Jobson and Korkie (1981) test with Memmel (2003) correction. We also applied the Ledoit and Wolf (2008) test with circular bootstrap, and the results were very similar; therefore we only report those of the first test in this table. The tests were always performed using the equal-weighted portfolio as benchmark. The statistics were calculated from 180 rolling windows covering the period from Jan. 2005 to Dec. 2019, and the size of the estimation window was 120 observations.

In Table 5, we analyze turnover, leverage and maximum leverage (equations (49), (50) and (51), respectively) of the portfolios in Tables 3-4.

The definitions are as follows for turnover:

$$\text{turnover} = \sum_{j=1}^{p} |\hat{w}_{t+1,j} - \hat{w}_{t,j}^{+}|, \tag{49}$$

and leverage

$$\text{leverage} = \left| \sum_{j=1}^{p} \min\{\hat{w}_{t+1,j}, 0\} \right|, \tag{50}$$

and maximum leverage

$$\text{max leverage} = \max_{j} \{ \left| \min\{\hat{w}_{t+1,j}, 0\} \right| \}. \tag{51}$$

It is clear that in Table 5 in terms of turnover, leverage, maximum leverage, GMV-POET annd GMV-NW-GIC-SF do well, with the best and close to best respectively if we discount EW portfolios.

Table 4: Empirical Results – Out-of-Sample Period from Jan. 2000 to Dec. 2019

	Without TC								With TC							
	Low Dim.				High Dim.				Low Dim.				High Dim.			
	SR	AVG	SD	p-value	SR	AVG	SD	p-value	SR	AVG	SD	p-value	SR	AVG	SD	p-value
EW	0.201	0.010	0.047	0.874	0.210	0.010	0.047	0.546	**0.195**	0.009	0.047	0.998	0.203	0.010	0.047	0.758
GMV-NW-GIC-SF	**0.213**	0.008	0.035		0.245	0.008	0.034		0.195	0.007	0.035		**0.225**	0.008	0.034	
GMV-NW-CV-SF	0.212	0.008	0.036	0.940	0.249	0.008	0.034	0.374	0.191	0.007	0.036	0.454	0.206	0.007	0.033	0.006
GMV-NW-GIC-3F	0.193	0.007	0.034	0.424	0.224	0.007	0.031	0.498	0.171	0.006	0.034	0.382	0.192	0.006	0.032	0.260
GMV-NW-CV-3F	0.188	0.006	0.034	0.348	0.231	0.007	0.031	0.700	0.161	0.006	0.034	0.196	0.139	0.004	0.031	0.016
GMV-POET	0.185	0.006	0.033	0.282	0.222	0.007	0.032	0.416	0.169	0.006	0.033	0.316	0.204	0.007	0.032	0.430
GMV-NL-LW	0.160	0.006	0.035	0.172	0.232	0.007	0.029	0.838	0.131	0.005	0.035	0.120	0.175	0.005	0.029	0.398
GMV-SF-NL-LW	0.172	0.006	0.034	0.252	0.242	0.007	0.028	0.934	0.145	0.005	0.034	0.196	0.184	0.005	0.028	0.398
MW-NW-GIC-SF	0.211	0.007	0.034	0.872	0.243	0.008	0.032	0.868	0.189	0.006	0.034	0.644	0.219	0.007	0.032	0.602
MW-NW-CV-SF	0.210	0.007	0.034	0.834	**0.249**	0.008	0.032	0.656	0.185	0.006	0.034	0.504	0.202	0.006	0.032	0.028
MW-NW-GIC-3F	0.191	0.006	0.034	0.442	0.226	0.007	0.031	0.584	0.165	0.006	0.034	0.338	0.190	0.006	0.031	0.326
MW-NW-CV-3F	0.184	0.006	0.034	0.324	0.228	0.007	0.031	0.652	0.153	0.005	0.034	0.162	0.132	0.004	0.031	0.038
MW-POET	0.181	0.006	0.032	0.282	0.216	0.007	0.031	0.408	0.161	0.005	0.033	0.240	0.195	0.006	0.031	0.402
MW-NL-LW	0.151	0.005	0.036	0.172	0.229	0.007	0.029	0.782	0.120	0.004	0.036	0.092	0.172	0.005	0.029	0.352
MW-SF-NL-LW	0.161	0.006	0.035	0.248	0.237	0.007	0.028	0.886	0.131	0.005	0.035	0.152	0.178	0.005	0.028	0.398
MAXSER	0.040	0.004	0.088	0.294					-0.039	-0.004	0.099	0.364				

The table shows the Sharpe ratio (SR), average returns (Avg), standard deviation (SD) and p-value of the Jobson and Korkie (1981) test with Memmel (2003) correction. We also applied the Ledoit and Wolf (2008) test with circular bootstrap, and the results were very similar; therefore we only report those of the first test in this table. The tests were always performed using the equal-weighted portfolio as benchmark. The statistics were calculated from 240 rolling windows covering the period from Jan. 2005 to Dec. 2019, and the size of the estimation window was 120 observations.

Table 5: Turnover and Leverage

	2005-2019 Subsample					
	Low Dimension			High Dimension		
	Turnover	Leverage	Max Leverage	Turnover	Leverage	Max Leverage
EW	0.053	0.000	0.000	0.054	0.000	0.000
GMV-NW-GIC-SF	0.125	0.312	0.042	0.130	0.376	0.009
GMV-NW-CV-SF	0.160	0.311	0.040	0.302	0.395	0.014
GMV-NW-GIC-3F	0.148	0.380	0.048	0.186	0.528	0.013
GMV-NW-CV-3F	0.190	0.382	0.049	0.593	0.567	0.030
GMV-POET	0.096	0.288	0.043	0.096	0.299	0.007
GMV-NL-LW	0.198	0.420	0.057	0.325	0.807	0.024
GMV-SF-NL-LW	0.163	0.383	0.050	0.341	0.904	0.025
MW-NW-GIC-SF	0.154	0.331	0.046	0.150	0.382	0.009
MW-NW-CV-SF	0.191	0.329	0.044	0.322	0.402	0.014
MW-NW-GIC-3F	0.179	0.401	0.052	0.207	0.539	0.013
MW-NW-CV-3F	0.220	0.401	0.051	0.626	0.582	0.030
MW-POET	0.128	0.306	0.046	0.117	0.307	0.008
MW-NL-LW	0.220	0.440	0.064	0.327	0.814	0.024
MW-SF-NL-LW	0.184	0.400	0.052	0.344	0.912	0.025
MAXSER	1.766	0.421	0.200			
	2000-2019 Sub Sample					
EW	0.056	0.000	0.000	0.056	0.000	0.000
GMV-NW-GIC-SF	0.120	0.283	0.049	0.127	0.342	0.011
GMV-NW-CV-SF	0.142	0.279	0.048	0.278	0.361	0.014
GMV-NW-GIC-3F	0.144	0.355	0.053	0.192	0.541	0.016
GMV-NW-CV-3F	0.181	0.353	0.053	0.557	0.572	0.030
GMV-POET	0.097	0.290	0.038	0.107	0.322	0.009
GMV-NL-LW	0.196	0.396	0.068	0.311	0.782	0.027
GMV-SF-NL-LW	0.173	0.383	0.062	0.310	0.849	0.026
MW-NW-GIC-SF	0.142	0.296	0.050	0.148	0.351	0.011
MW-NW-CV-SF	0.165	0.292	0.048	0.299	0.369	0.014
MW-NW-GIC-3F	0.165	0.368	0.054	0.209	0.548	0.016
MW-NW-CV-3F	0.203	0.364	0.054	0.582	0.581	0.030
MW-POET	0.121	0.301	0.041	0.126	0.333	0.009
MW-NL-LW	0.214	0.409	0.071	0.313	0.787	0.027
MW-SF-NL-LW	0.197	0.395	0.067	0.314	0.855	0.025
MAXSER	1.860	0.371	0.201			

The table shows the average turnover, average leverage and average max leverage for all portfolios across all out-of-sample windows. The top panel shows the results for the 2000-2019 out-of-sample period, and the second panel shows the results for the 2005-2019 out-of-sample period.

9 Conclusion

We provide a hybrid factor model combined with nodewise regression method that can control for risk and obtain the maximum expected return of a large portfolio. Our result is novel and holds even when $p > n$. We allow for increasing number of factors, with possible unbounded largest eigenvalue of the covariance matrix of errors. Sparsity is assumed on the precision matrix of errors, rather than the covariance matrix of errors. We also show that the maximum out-of-sample Sharpe ratio can be estimated consistently. Furthermore, we also develop a formula for the maximum Sharpe ratio when the weights of the portfolio sum to one. A consistent estimate for the constrained case is also shown. Then, we extended our results to the consistent estimation of Sharpe ratios in two widely used portfolios in the literature. It will be important to extend our results to more restrictions on portfolios.

Main Appendix

The main appendix is divided into several parts. The first part has preliminary proofs, norm inequalities, definitions, and a maximal inequality that is extended in a very minor form from the existing literature. The second part has the proofs of lemmata that lead to proof of Theorem 1. The first two parts relate only to the proof of Theorem 1. The third part is only related to the proof of Theorem 2. Part 4 is related to all the remaining proofs of Theorems in this paper. Supplementary appendix contains material that are common to all proofs.

PART 1:

We start with a lemma that provides norm inequalities. Let $A_1 : p \times K$, $B_1 : K \times K$ matrices and $x : K \times 1$ vector.

Lemma A.1. *(i).*

$$\|A_1 B_1 x\|_\infty \leq K^2 \|A_1\|_\infty \|B_1\|_\infty \|x\|_\infty.$$

(ii).

$$\|A_1 B_1 A_1'\|_\infty \leq K^2 \|A_1\|_\infty^2 \|B_1\|_\infty.$$

Proof of Lemma A.1. (i). Set $B_1 x = x_1$, and let a_j' be the $1 \times K$ row vector of A_1

$$\begin{aligned} \|A_1 x_1\|_\infty &= \max_{1 \leq j \leq p} |a_j' x_1| \leq \left[\max_{1 \leq j \leq p} \|a_j\|_1 \right] \|x_1\|_\infty && \text{(A.1)} \\ &\leq \left[K \|A_1\|_\infty \right] \|x_1\|_\infty && \text{(A.2)} \\ &= \left[K \|A_1\|_\infty \right] \|B_1 x\|_\infty \\ &\leq K^2 \|A_1\|_\infty \|B_1\|_\infty \|x\|_\infty, && \text{(A.3)} \end{aligned}$$

where we use Hölder's inequality for the first inequality, and the relation between l_1, l_∞ norms for the second inequality, and to get the last inequality we repeat the first two inequalities.

(ii). Use Section 4.3 of van de Geer (2016)

$$\|A_1 B_1 A_1'\|_\infty \leq \|A_1\|_\infty [\|B_1 A_1'\|_{l_1}], \tag{A.4}$$

where $\|.\|_{l_1}$ is the maximum absolute column sum norm of $B_1 A_1'$ matrix (i.e. l_1 induced matrix norm). Let b_l' be $1 \times K$ row vector of B_1, and a_j is the j th column of A_1' matrix ($K \times p$).

$$\begin{aligned} \|B_1 A_1'\|_{l_1} &= \max_{1 \leq j \leq p} \sum_{l=1}^{K} |b_l' a_j| \\ &\leq [\max_{1 \leq j \leq p} \|a_j\|_\infty][\sum_{l=1}^{K} \|b_l\|_1] \\ &\leq \|A_1\|_\infty [K \max_{1 \leq l \leq K} \|b_l\|_1] \\ &\leq \|A_1\|_\infty [K^2 \|B_1\|_\infty] && \text{(A.5)} \end{aligned}$$

where we use Hölder's inequality for the first inequality, and l_1, l_∞ norm relation for the other inequalities. Next use (A.5) in (A.4) to get

$$\|A_1 B_1 A_1'\|_\infty \leq K^2 \|A_1\|_\infty^2 \|B_1\|_\infty.$$

Q.E.D

Next we provide a lemma that is directly from Lemma A.2 of Fan et al. (2011).

Lemma A.2. *(Fan et al. (2011)). Suppose that two random variables* Z_1, Z_2 *satisfy the following exponential type tail condition. There exist* $r_{z_1}, r_{z_2} \in (0,1)$ *and* $b_{z_1}, b_{z_2} > 0$ *constant such that for all* $s > 0$

$$P[|Z_l| > s] \leq exp(1 - (s/b_{z_l})^{r_{z_l}}), \quad l = 1, 2,$$

Then for some $r_{z_3} > 0$, *and* $b_{z_3} > 0$

$$P|Z_1 Z_2| > s] \leq exp(1 - (s/b_{z_3})^{r_{z_3}}).$$

We provide now the following maximal inequality due to Theorem 1 of Merlevede et al. (2011), and used in the proof of Lemma A.3(i) and proof of Lemma B.1(ii) in Fan et al. (2011). To that effect, we provide a general assumption on data, and then show the theorem and its proof.

Assumption L1.*(i).* X_t, Y_t *are vectors of dimension* d_x, d_y *respectively, for* $t = 1, \cdots, n$. *They are both stationary and ergodic. Also* $\{X_t, Y_t\}$ *are strong mixing with strong mixing coefficients are satisfying*

$$\alpha(t) \leq exp(-Ct^{r_{xy}}),$$

with t, a positive integer, and $r_{xy} > 0$ *a positive constant. (ii). We also let* X_t, Y_t *satisfy the exponential tail condition for* $j_1 = 1, \cdots, d_x, j_2 = 1, \cdots, d_y$

$$P[|X_{j_1,t}| > s] \leq exp(-(s/b_x)^{r_x}),$$

for positive constants $b_x, r_x > 0$, *and*

$$P[|Y_{j_2,t}| > s] \leq exp(-(s/b_y)^{r_y}),$$

with positive constants $b_y, r_y > 0$. *Also we need to assume* $3r_x^{-1} + r_{xy}^{-1} > 1, 3r_y^{-1} + r_{xy}^{-1} > 1$.

Theorem A.1. *Under Assumption L1, (Fan et al. (2011)).*

$$\begin{aligned} P\left[\max_{1 \leq j_1 \leq d_x} \max_{1 \leq j_2 \leq d_y} \frac{|\sum_{t=1}^n (X_{j_1,t} Y_{j_2,t} - EX_{j_1,t} Y_{j_2,t})|}{n} > s\right] &\leq d_x d_y \{n exp\left(\frac{-(ns)^\gamma}{C_1}\right) \\ &+ exp\left(\frac{-(n^2 s^2)}{C_2(1 + nC_3)}\right) \\ &+ exp\left(\frac{-(ns)^2}{C_4 n} exp\left(\frac{(ns)^{\gamma(1-\gamma)}}{C_5 (lnns)^\gamma}\right)\right)\} \end{aligned}$$

with $0 < \gamma < 1$, *and* γ *is defined as* $\gamma^{-1} := 1.5r_x^{-1} + 1.5r_y^{-1} + r_{xy}^{-1}$.

Proof of Theorem A.1. This is a simple application of Lemma A.2 above with Assumption L1 for Theorem 1 of Merlevede et al. (2011), and Bonferroni union bound. **Q.E.D.**

PART. 2:

We start with an important maximal inequality applied to factor models in nodewise regression setting. Some of the results are already in Lemma A.3, Lemma B.1 of Fan et al. (2011). We show them so that readers can see all results without referral to other literature. We also provide two new results Lemma A.3(ii), (v) due to nodewise regression interaction with factor models.

Lemma A.3. *Under Assumptions 1-3, for* $C > C_m > 0$, *with* $m = 1, 2, 3, 4, 5$ *with* C_m *that is used in Theorem A.1.*

(i).

$$P\left[\max_{1 \leq j \leq p} \max_{1 \leq l \leq p} |\frac{1}{n} \sum_{t=1}^n u_{l,t} u_{j,t} - Eu_{l,t} u_{j,t}| > C\sqrt{lnp/n}\right] = O(\frac{1}{p^2}).$$

(ii). Denote $U_{-j}: p-1 \times n$ matrix in (4), and let the l th row and t th column element $U_{-j,l,t}$ and η_j as $n \times 1$ vector, and the t th element as $\eta_{j,t}$

$$P\left[\max_{1\le j\le p}\max_{1\le l\le p}|\frac{1}{n}\sum_{t=1}^{n}U_{-j,l,t}\eta_{j,t}| > C\sqrt{lnp/n}\right] = O(\frac{1}{p^2}).$$

(iii).

$$P\left[\max_{1\le k\le K}\max_{1\le j\le p}|\frac{1}{n}\sum_{t=1}^{n}f_{k,t}u_{j,t}| > C\sqrt{lnp/n}\right] = O(\frac{1}{p^2}).$$

(iv). Let $f_{k_1,t}, f_{k_2,t}$ represent k_1, k_2 factors (elements) of the vector f_t

$$P\left[\max_{1\le k_1\le K}\max_{1\le k_2\le K}|\frac{1}{n}\sum_{t=1}^{n}f_{k_1,t}f_{k_2,t} - Ef_{k_1,t}f_{k_2,t}| > C\sqrt{lnn/n}\right] = O(\frac{1}{n^2}).$$

(v).

$$P\left[\max_{1\le k\le K}\max_{1\le j\le p}|\frac{1}{n}\sum_{t=1}^{n}f_{k,t}\eta_{j,t}| > C\sqrt{\bar{s}lnp/n}\right] = O(\frac{1}{p^2}).$$

Proof of Lemma A.3. (i). This is Lemma A.3(i) of Fan et al. (2011).

(ii). The proof follows from Theorem A.1 and Assumption 3 provides the tail probability through the same algebra as in p.3346 of Fan et al. (2011).

(iii). This is Lemma B1(ii) of Fan et al. (2011).

(iv). This is Lemma B1(i) of Fan et al. (2011).

(v). The proof will involve several steps and this is due to interaction of factor models ($f_{k,t}$) and nodewise error ($\eta_{j,t}$). Start with the definition of

$$\begin{aligned}\eta_{j,t} &= u_{j,t} - u'_{-j,t}\gamma_j = [u_{j,t}, u'_{-j,t}]\begin{bmatrix}1\\ -\gamma_j\end{bmatrix} \quad (A.6)\\ &= u'_t\Omega_j, \quad (A.7)\end{aligned}$$

where $\Omega_j := \begin{bmatrix}1\\ -\gamma_j\end{bmatrix}$ $p \times 1$ vector, and $u'_t := [u_{j,t}, u'_{-j,t}]$ which is $1 \times p$ row vector. Next

$$\begin{aligned}\max_{1\le j\le p}\max_{1\le k\le K}|\frac{1}{n}\sum_{t=1}^{n}f_{k,t}\eta_{j,t}| &= \max_{1\le j\le p}\max_{1\le k\le K}|\frac{1}{n}\sum_{t=1}^{n}f_{k,t}u'_t\Omega_j|\\ &\le \max_{1\le k\le K}\|\frac{1}{n}\sum_{t=1}^{n}f_{k,t}u'_t\|_\infty \max_{1\le j\le p}\|\Omega_j\|_1\\ &= \max_{1\le j\le p}\max_{1\le k\le K}|\frac{1}{n}\sum_{t=1}^{n}f_{k,t}u_{j,t}| \max_{1\le j\le p}\|\Omega_j\|_1 \quad (A.8)\end{aligned}$$

where we use (A.7) for the first equality and Hölder's inequality for the first inequality.
Consider

$$\max_{1\le j\le p}\|\Omega_j\|_1 \le 1 + \max_{1\le j\le p}\|\gamma_j\|_1, \quad (A.9)$$

where we use Ω_j definition. Noting that $\Sigma_{n,-j,-j}$ is $p-1 \times p-1$ submatrix of Σ_n consisting all rows and columns of Σ_n except the j th one. See that

$$\frac{\gamma_j' \Sigma_{n,-j,-j} \gamma_j}{\gamma_j' \gamma_j} \geq Eigmin(\Sigma_{n,-j,-j}) \geq Eigmin(\Sigma_n) \geq c > 0.$$

Then

$$\|\gamma_j\|_2^2 = \gamma_j' \gamma_j \leq \frac{\gamma_j' \Sigma_{n,-j,-j} \gamma_j}{Eigmin(\Sigma_n)} \leq C < \infty, \tag{A.10}$$

where we use the last inequality above for the first inequality in (A.10) and (B.48) of Caner and Kock (2018) for the second inequality in (A.10), given our Assumption $\max_{1 \leq j \leq p} E u_{j,t}^2 \leq C < \infty$. Hence

$$\max_{1 \leq j \leq p} \|\gamma_j\|_1 \leq \sqrt{\bar{s}} \max_{1 \leq j \leq p} \|\gamma_j\|_2 \leq \sqrt{\bar{s}} C = O(\sqrt{\bar{s}}), \tag{A.11}$$

by (A.10). So clearly by (A.9)(A.11)

$$\max_{1 \leq j \leq p} \|\Omega_j\|_1 = O(\sqrt{\bar{s}}). \tag{A.12}$$

Next, use Lemma A.3(iii) and (A.12) in (A.8) to have

$$P[\max_{1 \leq j \leq p} \max_{1 \leq k \leq K} |\frac{1}{n} \sum_{t=1}^{n} f_{k,t} \eta_{j,t}| \geq C \frac{\sqrt{\bar{s}} lnp}{\sqrt{n}}] = O(1/p^2).$$

This also implies that since $X := (f_1, \cdots, f_n) : K \times n$ matrix, and $\eta_j := (\eta_{j,1}, \cdots, \eta_{j,n})' : n \times 1$

$$\max_{1 \leq j \leq p} \|X \eta_j / n\|_\infty = O_p(\frac{\sqrt{\bar{s}} lnp}{\sqrt{n}}). \tag{A.13}$$

Q.E.D.

Now we start defining two events, and we condition the next lemma, which is l_1 bound on nodewise regression estimates, on these two events. Then we relax this restriction, and show that an unconditional result for l_1 norm of the nodewise regression estimates after finding that these two events converge in probability to one. Define

$$\mathcal{A}_1 := \{2 \max_{1 \leq j \leq p} \|\eta_{xj}' \hat{U}_{-j}' / n\|_\infty \leq \lambda_n\}, \tag{A.14}$$

and define the population adaptive restricted eigenvalue condition, as in Caner and Kock (2018), for $j = 1, \cdots, p$, and let δ_{S_j} represent the vector with S_j indices in δ_j, and all the other elements than S_j indices in δ_j set to zero

$$\phi^2(s_j) := \min_{\delta_j \in R^{p-1}} \left\{ \frac{\delta_j' \Sigma_{n,-j,-j} \delta_j}{\|\delta_{S_j}\|_1^2} : \delta \in R^p - \{0\}, \|\delta_{S_j^c}\|_1 \leq 3\sqrt{s_j} \|\delta_{S_j}\|_2 \right\}, \tag{A.15}$$

and the empirical version of the adaptive restricted eigenvalue condition is as follows, with $\hat{U}_{-j} : p-1 \times n$ matrix

$$\hat{\phi}^2(s_j) := \min_{\delta_j \in R^{p-1}} \left\{ \frac{\delta_j' (\hat{U}_{-j} \hat{U}_{-j}' / n) \delta_j}{\|\delta_{S_j}\|_1^2} : \delta \in R^p - \{0\}, \|\delta_{S_j^c}\|_1 \leq 3\sqrt{s_j} \|\delta_{S_j}\|_2 \right\}, \tag{A.16}$$

and the event is for each $j = 1, \cdots, p$

$$\mathcal{A}_{2j} := \{\hat{\phi}^2(s_j) \geq \phi^2(s_j)/2\}.$$

We have the following l_1 bound result.

Lemma A.4. *Under $A_1 \cap A_{2j}$, with $\lambda_n > 0$ in (A.14)*

$$\max_{1 \le j \le p} \|\hat{\gamma}_j - \gamma_j\|_1 \le \frac{24\lambda_n \bar{s}}{\phi^2(\bar{s})} = O_p(\lambda_n \bar{s}),$$

We specify the formula and the rate for λ_n in the next Lemma.

Proof of Lemma A.4. Start with $\hat{\gamma}_j$ definition in (24)

$$\|\hat{u}_j - \hat{U}'_{-j}\hat{\gamma}_j\|_n^2 + 2\lambda_n \sum_{j=1}^{n} |\hat{\gamma}_j| \le \|\hat{u}_j - \hat{U}'_{-j}\gamma_j\|_n^2 + 2\lambda_n \sum_{j=1}^{p} |\gamma_j|. \tag{A.17}$$

Use (22) to have $\hat{u}_j - \hat{U}'_{-j}\hat{\gamma}_j = \eta_{xj} - \hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)$ and this last equation can be substituted into first left side term and first right side term in (A.17) to have

$$\|\eta_{xj} - \hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n^2 + 2\lambda_n \sum_{j=1}^{p} |\hat{\gamma}_j| \le \|\eta_{xj}\|_n^2 + 2\lambda_n \sum_{j=1}^{p} |\gamma_j|. \tag{A.18}$$

Simplify the first term on the left and the first term on the right side of (A.18),

$$\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n^2 + 2\lambda_n \sum_{j=1}^{p} |\hat{\gamma}_j| \le 2|\frac{\eta'_{xj}\hat{U}'_{-j}}{n}(\hat{\gamma}_j - \gamma_j)| + 2\lambda_n \sum_{j=1}^{p} |\gamma_j|. \tag{A.19}$$

Since we use $\mathcal{A}_1$ and then Hölder's inequality

$$\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n^2 + 2\lambda_n \sum_{j=1}^{p} |\hat{\gamma}_j| \le \lambda_n \|\hat{\gamma}_j - \gamma_j\|_1 + 2\lambda_n \sum_{j=1}^{p} |\gamma_j|. \tag{A.20}$$

Use $\|\hat{\gamma}_j\|_1 = \|\hat{\gamma}_{S_j}\|_1 + \|\hat{\gamma}_{S_j^c}\|_1$, on the second term on the left side of (A.20) (S_j represents the indices of nonzero cells in row j of the precision matrix, and S_j^c represents the indices of zero cells in row j of the precision matrix).

$$\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n^2 + 2\lambda_n \sum_{j \in S_j^c} |\hat{\gamma}_j| \le \lambda_n \|\hat{\gamma}_j - \gamma_j\|_1 + 2\lambda_n \sum_{j=1}^{p} |\gamma_j| - 2\lambda_n \sum_{j \in S_j} |\hat{\gamma}_j|. \tag{A.21}$$

Now use $\sum_{j \in S_j^c} |\gamma_j| = 0$ in the second term on the right side of (A.21) and use reverse triangle inequality

$$\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n^2 + 2\lambda_n \sum_{j \in S_j^c} |\hat{\gamma}_j| \le \lambda_n \|\hat{\gamma}_j - \gamma_j\|_1 + 2\lambda_n \sum_{j \in S_j} |\hat{\gamma}_j - \gamma_j|. \tag{A.22}$$

Next by $\|\hat{\gamma}_j - \gamma_j\|_1 = \|\hat{\gamma}_{S_j} - \gamma_{S_j}\|_1 + \|\hat{\gamma}_{S_j^c}\|_1$ for the first term on the right side of (A.22)

$$\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n^2 + \lambda_n \sum_{j \in S_j^c} |\hat{\gamma}_j| \le 3\lambda_n \sum_{j \in S_j} |\hat{\gamma}_j - \gamma_j|. \tag{A.23}$$

Use the norm inequality $\|\hat{\gamma}_{S_j} - \gamma_{S_j}\|_1 \le \sqrt{s_j}\|\hat{\gamma}_{S_j} - \gamma_{S_j}\|_2$

$$\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n^2 + \lambda_n \sum_{j \in S_j^c} |\hat{\gamma}_j| \le 3\lambda_n \sqrt{s_j}\|\hat{\gamma}_{S_j} - \gamma_{S_j}\|_2. \tag{A.24}$$

Now ignoring the first term above and dividing the rest by $\lambda_n > 0$, provides the restricted set condition (cone condition) in adaptive restricted eigenvalue condition

$$\|\hat{\gamma}_{S_j^c}\|_1 \le 3\sqrt{s_j}\|\hat{\gamma}_{S_j} - \gamma_{S_j}\|_2. \tag{A.25}$$

Set $\delta_j = \hat{\gamma}_j - \gamma_j$ in the empirical adaptive restricted set condition in (A.16), then use the empirical adaptive restricted eigenvalue condition in (A.24)

$$\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n^2 + \lambda_n \sum_{j \in S_j^c} |\hat{\gamma}_j| \leq 3\lambda_n \sqrt{s_j} \frac{\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n}{\hat{\phi}(s_j)}.$$

Then use $3ab \leq a^2/2 + 9b^2/2$ with $b = \frac{\lambda_n \sqrt{s_j}}{\hat{\phi}(s_j)}$, $a = \|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n$.

$$\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n^2 + \lambda_n \sum_{j \in S_j^c} |\hat{\gamma}_j| \leq \frac{\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n^2}{2} + \frac{9\lambda_n^2 s_j}{2\hat{\phi}^2(s_j)}.$$

Use $\mathcal{A}_{2j}$ in the first term on the right side and simplify

$$\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n^2 + 2\lambda_n \sum_{j \in S_j^c} |\hat{\gamma}_j| \leq \frac{18\lambda_n^2 s_j}{\phi^2(s_j)}.$$

This implies

$$\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n^2 \leq \frac{18\lambda_n^2 s_j}{\phi^2(s_j)}. \tag{A.26}$$

Now to get l_1 bound, ignore the first term in (A.24) and add both sides $\lambda_n \|\hat{\gamma}_{S_j} - \gamma_{S_j}\|_1$

$$\lambda_n \sum_{j \in S_j^c} |\hat{\gamma}_j| + \lambda_n \sum_{j \in S_j} |\hat{\gamma}_j - \gamma_j| = \lambda_n \|\hat{\gamma}_j - \gamma_j\|_1 \leq \lambda_n \|\hat{\gamma}_{S_j} - \gamma_{S_j}\|_1 + 3\lambda_n \sqrt{s_j} \|\hat{\gamma}_{S_j} - \gamma_{S_j}\|_2. \tag{A.27}$$

Use the norm inequality $\|\hat{\gamma}_{S_j} - \gamma_{S_j}\|_1 \leq \sqrt{s_j} \|\hat{\gamma}_{S_j} - \gamma_{S_j}\|_2$ for the first term on the right side of (A.27)

$$\lambda_n \|\hat{\gamma}_j - \gamma_j\|_1 \leq 4\lambda_n \sqrt{s_j} \|\hat{\gamma}_{S_j} - \gamma_{S_j}\|_2.$$

This can be simplified as

$$\|\hat{\gamma}_j - \gamma_j\|_1 \leq 4\sqrt{s_j} \|\hat{\gamma}_{S_j} - \gamma_{S_j}\|_2.$$

and can use the empirical adaptive restricted eigenvalue condition in (A.16)

$$\|\hat{\gamma}_j - \gamma_j\|_1 \leq 4\sqrt{s_j} \frac{\|\hat{U}'_{-j}(\hat{\gamma}_j - \gamma_j)\|_n}{\hat{\phi}(s_j)}.$$

Next use (A.26) and $\mathcal{A}_{2j}$ to have

$$\|\hat{\gamma}_j - \gamma_j\|_1 \leq \frac{24\lambda_n s_j}{\phi^2(s_j)} \leq \frac{24\lambda_n \bar{s}}{\phi^2(\bar{s})}.$$

Last inequality above is true by noticing $s_j \leq \bar{s}$ by $\bar{s}$ definition, and then by definition of population adaptive restricted eigenvalue condition $\phi^2(s_j) \geq \phi^2(\bar{s})$.**Q.E.D.**

Now we evaluate two events, in the next two lemmata.

Lemma A.5. *Under Assumptions 1-4*

$$P(\mathcal{A}_1) \geq 1 - O(\frac{1}{p^2}) - O(\frac{1}{n^2}),$$

and

$$\lambda_n := C[K^2 \frac{\sqrt{\bar{s}} lnp}{n} + \sqrt{\frac{lnp}{n}}] = O\left(K^2 \frac{\sqrt{\bar{s}} lnp}{n} + \sqrt{\frac{lnp}{n}}\right),$$

with $C > 0$.

Remark. If we add Assumption 5 we can show as in (A.51), the rate for λ_n simplifies to

$$\lambda_n = O(\sqrt{lnp/n}).$$

Proof of Lemma A.5. Start with $\mathcal{A}_1$ definition in (A.14). Use (21)(23) and $M_X := I_n - X'(XX')^{-1}X$ is idempotent

$$\frac{\hat{U}_{-j}\eta_{xj}}{n} = \frac{U_{-j}\eta_j}{n} - \left(\frac{U_{-j}X'}{n}\right)\left(\frac{XX'}{n}\right)^{-1}\left(\frac{X\eta_j}{n}\right) \tag{A.28}$$

Next we use triangle inequality in l_∞ norm

$$\|\frac{\eta_{xj}'\hat{U}_{-j}'}{n}\|_\infty = \|\frac{\hat{U}_{-j}\eta_{xj}}{n}\|_\infty \leq \|\frac{U_{-j}\eta_j}{n}\|_\infty + \|\left(\frac{U_{-j}X'}{n}\right)\left(\frac{XX'}{n}\right)^{-1}\left(\frac{X\eta_j}{n}\right)\|_\infty \tag{A.29}$$

Note that U is $p \times n$ matrix and U_{-j} is the $p-1 \times n$ submatrix, which is U without j th row. So

$$\max_{1\leq j\leq p} \|\frac{U_{-j}\eta_j}{n}\|_\infty \leq C\sqrt{lnp/n}, \tag{A.30}$$

with probability at least $1 - O(1/p^2)$ by Lemma A.3(ii). Next for the second right side term in (A.29)

$$\max_{1\leq j\leq p} \|\left(\frac{U_{-j}X'}{n}\right)\left(\frac{XX'}{n}\right)^{-1}\left(\frac{X\eta_j}{n}\right)\|_\infty \leq K^2 \max_{1\leq j\leq p}\|\frac{U_{-j}X'}{n}\|_\infty \|\left(\frac{XX'}{n}\right)^{-1}\|_\infty \max_{1\leq j\leq p}\|\frac{X\eta_j}{n}\|_\infty. \tag{A.31}$$

by Lemma A.1(i). We evaluate each term in (A.31). Note that $X = (f_1, \cdots, f_n) : K \times n$, $U : p \times n$

$$\max_{1\leq j\leq p} \|\frac{U_{-j}X'}{n}\|_\infty \leq \|UX'/n\|_\infty = \max_{1\leq j\leq p}\max_{1\leq k\leq K} |\frac{\sum_{t=1}^n u_{j,t}f_{k,t}}{n}|.$$

Then by Lemma A.3(iii)

$$P[\max_{1\leq j\leq p}\max_{1\leq k\leq K} |\frac{1}{n}\sum_{t=1}^n u_{j,t}f_{k,t}| \leq C\sqrt{lnp/n}] \geq 1 - O(1/p^2). \tag{A.32}$$

Then by Assumption 4 and Lemma A.3(iv)

$$\|(XX'/n)^{-1}\|_\infty \leq C, \tag{A.33}$$

with probability at least $1 - O(1/n^2)$. Next, since $\eta_j : n \times 1$, and $\eta_{j,t}$ is the t th element

$$\max_{1\leq j\leq p} \|\frac{X\eta_j}{n}\|_\infty = \max_{1\leq j\leq p}\max_{1\leq k\leq K} |\frac{1}{n}\sum_{t=1}^n f_{k,t}\eta_{j,t}|. \tag{A.34}$$

Then by Lemma A.3(v)

$$P[\max_{1\leq j\leq p}\max_{1\leq k\leq K} |\frac{1}{n}\sum_{t=1}^n f_{k,t}\eta_{j,t}| \leq C\sqrt{\frac{\bar{s}lnp}{n}}] \geq 1 - O(1/p^2). \tag{A.35}$$

Combine (A.30)-(A.35) in (A.29)

$$\max_{1\leq j\leq p} \|\frac{\hat{U}_{-j}\eta_{xj}}{n}\|_\infty \leq C\left[K^2\frac{\sqrt{\bar{s}}lnp}{n} + \sqrt{\frac{lnp}{n}}\right], \tag{A.36}$$

with probability at least $1 - O(1/p^2) - O(1/n^2)$.**Q.E.D.**

Lemma A.6. *Under Assumptions 1-5, for* $j = 1, \cdots, p$

$$P(\mathcal{A}_{2j}) \geq 1 - O(1/p^2) - O(1/n^2).$$

Proof of Lemma A.6.

For each $j = 1, \cdots, p$, add and subtract $\delta_j'(U_{-j}U_{-j}'/n)\delta_j$

$$\begin{aligned} \frac{\delta_j'\hat{U}_{-j}\hat{U}_{-j}'\delta_j}{n} &= \frac{\delta_j'\hat{U}_{-j}\hat{U}_{-j}'\delta_j}{n} - \frac{\delta_j'U_{-j}U_{-j}'\delta_j}{n} + \frac{\delta_j'U_{-j}U_{-j}'\delta_j}{n} \\ &\geq \frac{\delta_j'U_{-j}U_{-j}'\delta_j}{n} - \left|\frac{\delta_j'\hat{U}_{-j}\hat{U}_{-j}'\delta_j}{n} - \frac{\delta_j'U_{-j}U_{-j}'\delta_j}{n}\right| \end{aligned}$$

Next add and subtract $\frac{\delta_j'\Sigma_{n,-j,-j}\delta_j}{n}$ from the right side of the above inequality

$$\begin{aligned} \frac{\delta_j'\hat{U}_{-j}\hat{U}_{-j}'\delta_j}{n} &\geq \frac{\delta_j'\Sigma_{n,-j,-j}\delta_j}{n} - \left|\frac{\delta_j'U_{-j}U_{-j}'\delta_j}{n} - \frac{\delta_j'\Sigma_{n,-j,-j}\delta_j}{n}\right| \\ &- \left|\frac{\delta_j'\hat{U}_{-j}\hat{U}_{-j}'\delta_j}{n} - \frac{\delta_j'U_{-j}U_{-j}'\delta_j}{n}\right| \end{aligned} \tag{A.37}$$

Note that second right side term with absolute value in (A.37) can be bounded by using Hölder's inequality twice

$$\left|\frac{\delta_j'\hat{U}_{-j}\hat{U}_{-j}'\delta_j}{n} - \frac{\delta_j'U_{-j}U_{-j}'\delta_j}{n}\right| \leq \|\delta_j\|_1^2 \left\|\frac{\hat{U}_{-j}\hat{U}_{-j}'}{n} - \frac{U_{-j}U_{-j}'}{n}\right\|_\infty.$$

By the same analysis applied to the first right side term with absolute value in (A.37) and simplifying

$$\begin{aligned} \frac{\delta_j'\hat{U}_{-j}\hat{U}_{-j}'\delta_j}{n} &\geq \frac{\delta_j'\Sigma_{n,-j,-j}\delta_j}{n} \\ &- \|\delta_j\|_1^2 \left[\|\frac{\hat{U}_{-j}\hat{U}_{-j}'}{n} - \frac{U_{-j}U_{-j}'}{n}\|_\infty + \|\frac{U_{-j}U_{-j}'}{n} - \Sigma_{n,-j,-j}\|_\infty\right] \end{aligned} \tag{A.38}$$

In (A.38) we start considering by (21)

$$\frac{\hat{U}_{-j}\hat{U}_{-j}'}{n} - \frac{U_{-j}U_{-j}'}{n} = -\frac{U_{-j}X'}{n}\left(\frac{XX'}{n}\right)^{-1}\frac{XU_{-j}'}{n}. \tag{A.39}$$

Using (A.39) Lemma A.1(ii), and (A.32)(A.33)

$$\begin{aligned} \|\frac{\hat{U}_{-j}\hat{U}_{-j}'}{n} - \frac{U_{-j}U_{-j}'}{n}\|_\infty &\leq K^2\|\frac{U_{-j}X'}{n}\|_\infty^2\|(\frac{XX'}{n})^{-1}\|_\infty \\ &\leq C\frac{K^2 lnp}{n}, \end{aligned} \tag{A.40}$$

with probability at least $1 - O(1/p^2) - O(1/n^2)$. Next in (A.38) see that $\Sigma_{n,-j,-j}$ is a submatrix of Σ_n, and U_{-j} is a submatrix of U as described above

$$\|\frac{U_{-j}U_{-j}'}{n} - \Sigma_{n,-j,-j}\|_\infty \leq \|\frac{UU'}{n} - \Sigma_n\|_\infty \leq C\sqrt{lnp/n}, \tag{A.41}$$

with probability at least $1 - O(1/p^2)$ by Lemma A.3(i). We need to provide some simplification for $\|\delta_j\|_1^2$ term in (A.38). Next since cone condition in adaptive restricted eigenvalue condition is satisfied in (A.25)

$$\|\delta_{S_j^c}\|_1 \leq 3\sqrt{s_j}\|\delta_{S_j}\|_2.$$

Then add $\|\delta_{S_j}\|_1$ to the left side and right side and use the norm inequality that puts an upper bound on l_1 in terms of l_2

$$\begin{aligned} \|\delta_{S_j^c}\|_1 + \|\delta_{S_j}\|_1 &= \|\delta_j\|_1 \leq \|\delta_{S_j}\|_1 + 3\sqrt{s_j}\|\delta_{S_j}\|_2 \\ &\leq 4\sqrt{s_j}\|\delta_{S_j}\|_2. \end{aligned}$$

So we can have

$$\frac{\|\delta_j\|_1^2}{\|\delta_{S_j}\|_2^2} \leq 16 s_j. \tag{A.42}$$

Now divide (A.38) by $\|\delta_{S_j}\|_2^2 > 0$ and use (A.40) and (A.42)

$$\begin{aligned} \frac{\delta_j' \hat{U}_{-j} \hat{U}_{-j}' \delta_j / n}{\|\delta_{S_j}\|_2^2} &\geq \frac{\delta_j' \Sigma_{n,-j,-j} \delta_j}{\|\delta_{S_j}\|_2^2} - 16 s_j K^2 \|\frac{U_{-j} X'}{n}\|_\infty^2 \|(\frac{XX'}{n})^{-1}\|_\infty \\ &- 16 s_j \|\frac{U_{-j} U_{-j}'}{n} - \Sigma_{n,-j,-j}\|_\infty. \end{aligned} \tag{A.43}$$

Next using empirical and population adaptive restricted eigenvalue definitions, by minimizing over δ_j

$$\begin{aligned} \hat{\phi}^2(s_j) &\geq \phi^2(s_j) - 16 s_j K^2 \|\frac{U_{-j} X'}{n}\|_\infty^2 \|(\frac{XX'}{n})^{-1}\|_\infty \\ &- 16 s_j \|\frac{U_{-j} U_{-j}'}{n} - \Sigma_{n,-j,-j}\|_\infty. \end{aligned} \tag{A.44}$$

Note that if we have with with probability approaching one (wpa1 from now on)

$$16 s_j \left[K^2 \|\frac{U_{-j} X'}{n}\|_\infty^2 \|(\frac{XX'}{n})^{-1}\|_\infty + 16 s_j \|\frac{U_{-j} U_{-j}'}{n} - \Sigma_{n,-j,-j}\|_\infty \right] \leq \phi^2(s_j)/2, \tag{A.45}$$

we have with wpa1

$$\hat{\phi}^2(s_j)/2 \geq \phi^2(s_j)/2.$$

Thus we need to show that following probability goes to zero

$$P[\hat{\phi}^2(s_j) < \phi^2(s_j)/2] \leq P[16 s_j (K^2 \|\frac{U_{-j} X'}{n}\|_\infty^2 \|(\frac{XX'}{n})^{-1}\|_\infty + \|\frac{U_{-j} U_{-j}'}{n} - \Sigma_{n,-j,-j}\|_\infty) > \phi^2(s_j)/2] \tag{A.46}$$

Set $\epsilon_n := 16 s_j [K^2 \ln p/n + \sqrt{lnp/n}]$. Clearly by (A.40)(A.41)

$$P[16 s_j (K^2 \|\frac{U_{-j} X'}{n}\|_\infty^2 \|(\frac{XX'}{n})^{-1}\|_\infty + \|\frac{U_{-j} U_{-j}'}{n} - \Sigma_{n,-j,-j}\|_\infty) > \epsilon_n] \leq O(1/p^2) + O(1/n^2). \tag{A.47}$$

Since $\epsilon_n \to 0$ by Assumption 5, by (A.46)(A.47) $P(\hat{\phi}^2(s_j) < \phi^2(s_j)/2) \to 0$.**Q.E.D.**
One crucial point is that we need to get a low bound for $\cap_{j=1}^p \mathcal{A}_{2j}$. In that respect from (A.45)

$$\left(16 s_j [K^2 \|\frac{U_{-j} X'}{n}\|_\infty^2 \|(\frac{XX'}{n})^{-1}\|_\infty + \|\frac{U_{-j} U_{-j}'}{n} - \Sigma_{n,-j,-j}\|_\infty] \leq \phi^2(s_j)/2 \right) \subseteq \{\hat{\phi}^2(s_j) \geq \phi^2(s_j)/2\} = \mathcal{A}_{2j}.$$

Then clearly by $\Sigma_n, \Sigma_{n,-j,-j}$ definitions and population adaptive restricted eigenvalue condition

$$\begin{aligned} 16 s_j [K^2 \|\frac{U_{-j} X'}{n}\|_\infty^2 \|(\frac{XX'}{n})^{-1}\|_\infty &+ \|\frac{U_{-j} U_{-j}'}{n} - \Sigma_{n,-j,-j}\|_\infty] \\ &\leq 16 \bar{s} [K^2 \|\frac{UX'}{n}\|_\infty^2 \|(\frac{XX'}{n})^{-1}\|_\infty + \|\frac{UU'}{n} - \Sigma_n\|_\infty] \\ &\leq \phi^2(\bar{s})/2 \leq \phi^2(s_j)/2. \end{aligned} \tag{A.48}$$

So (A.48) implies that for $j = 1, \cdots, p$

$$[16 \bar{s} [K^2 \|UX'/n\|_\infty^2 \|(XX'/n)^{-1}\|_\infty + \|UU'/n - \Sigma_n\|_\infty] \leq \phi^2(\bar{s})] \subseteq \mathcal{A}_{2j}.$$

This means

$$[16 \bar{s} [K^2 \|UX'/n\|_\infty^2 \|(XX'/n)^{-1}\|_\infty + \|UU'/n - \Sigma_n\|_\infty] \leq \phi^2(\bar{s})] \subseteq \cap_{j=1}^p \mathcal{A}_{2j}.$$

Next by (A.33),(A.41) via Lemma A.3(iii)

$$P\left((\cap_{j=1}^p \mathcal{A}_{2,j})^c\right) \leq P(16\bar{s}[K^2\|UX'/n\|_\infty^2\|(XX'/n)^{-1}\|_\infty + \|UU'/n - \Sigma_n\|_\infty] > \phi^2(\bar{s})) \leq O(1/p^2) + O(1/n^2). \quad (A.49)$$

We analyze the λ_n rate here. See that by Lemma A.5

$$\lambda_n = O(K^2\bar{s}^{1/2}\frac{lnp}{n} + \sqrt{\frac{lnp}{n}}). \quad (A.50)$$

But under Assumption 5

$$\begin{aligned} K^2\bar{s}^{1/2}\frac{lnp}{n} + \sqrt{\frac{lnp}{n}} &= (K^2\bar{s}^{1/2}\sqrt{lnp/n} + 1)\sqrt{lnp/n} \\ &= [o(1)+1]\sqrt{lnp/n} \\ &= O(\sqrt{lnp/n}). \quad (A.51) \end{aligned}$$

We provide the main consistency result for residual based nodewise regression result. This is new in the literature due to its factor model.

Lemma A.7. *Under Assumptions 1-5*

$$\max_{1\leq j\leq p} \|\hat{\gamma}_j - \gamma_j\|_1 = O_p(\lambda_n \bar{s}) = O_p(\bar{s}\sqrt{lnp/n}) = o_p(1).$$

Proof of Lemma A.7. Use Lemma A.5-A.6 and (A.49) to have

$$P(\mathcal{A}_1 \cap \{\cap_{j=1}^p \mathcal{A}_{2j}\}) \geq 1 - O(1/p^2) - O(1/n^2).$$

Then combine above with Lemma A.4 to have the desired result via Assumption 5 and (A.50)(A.51) to have $\lambda_n\bar{s} = o(1)$.**Q.E.D.**

Next, we provide proof of consistency for the estimates of the reciprocal of the main diagonal elements of the precision matrix.

Lemma A.8. *Under Assumptions 1-5*

$$\max_{1\leq j\leq p} |\hat{\tau}_j^2 - \tau_j^2| = O_p\left(\bar{s}^{1/2}\sqrt{\frac{lnp}{n}}\right) = o_p(1).$$

Proof of Lemma A.8. Start with $\hat{\tau}_j^2$ definition in (25). For all $j = 1, \cdots, p$

$$\hat{\tau}_j^2 := \hat{u}_j'(\hat{u}_j - \hat{U}_{-j}'\hat{\gamma}_j)/n$$

and $\tau_j^2 := E\eta_{j,t}^2$, with $\eta_{j,t} := u_{j,t} - u_{-j,t}'\gamma_j$, and $\eta_j := (\eta_{j,1}, \cdots, \eta_{j,n})' : n \times 1$ vector $\eta_{xj} = M_X\eta_j$. Using (22) for $\hat{u}_j$ in $\hat{\tau}_j^2$ definition we have

$$\begin{aligned} \hat{\tau}_j^2 &= (\hat{U}_{-j}'\gamma_j + \eta_{xj})'(\eta_{xj} - \hat{U}_{-j}'(\hat{\gamma}_j - \gamma_j))/n \\ &= \frac{\eta_{xj}'\eta_{xj}}{n} - \frac{\eta_{xj}'\hat{U}_{-j}'(\hat{\gamma}_j - \gamma_j)}{n} \\ &+ \frac{\gamma_j'\hat{U}_{-j}\eta_{xj}}{n} - \frac{\gamma_j'\hat{U}_{-j}\hat{U}_{-j}'(\hat{\gamma}_j - \gamma_j)}{n}. \end{aligned}$$

Use triangle inequality

$$\max_{1\le j\le p} |\hat{\tau}_j^2 - \tau_j^2| \le \max_{1\le j\le p} |\frac{\eta_{xj}'\eta_{xj}}{n} - \tau_j^2| + \max_{1\le j\le p} |\frac{\eta_{xj}'\hat{U}_{-j}'(\hat{\gamma}_j - \gamma_j)}{n}| + \max_{1\le j\le p} |\frac{\gamma_j'\hat{U}_{-j}\eta_{xj}}{n}| + \max_{1\le j\le p} |\frac{\gamma_j'\hat{U}_{-j}\hat{U}_{-j}'(\hat{\gamma}_j - \gamma_j)}{n}|. \quad (A.52)$$

Consider each term in (A.52) carefully. Start with definition; $M_X := I_n - X'(XX')^{-1}X$, and M_X being idempotent.

$$\max_{1\le j\le p} |\frac{\eta_{xj}'\eta_{xj}}{n} - \tau_j^2| \le \max_{1\le j\le p} |\frac{\eta_j'\eta_j}{n} - \tau_j^2| + \max_{1\le j\le p} |\frac{\eta_j'X'}{n}\left(\frac{XX'}{n}\right)^{-1}\frac{X\eta_j}{n}|. \quad (A.53)$$

First, exactly as in Lemma A.3(i) with Assumption 2(ii)(iv), $3r_2^{-1} + r_0^{-1} > 1$ we have by Theorem A.1

$$\max_{1\le j\le p} |\frac{\eta_j'\eta_j}{n} - \tau_j^2| = O_p(\sqrt{lnp/n}). \quad (A.54)$$

Then note that $X\eta_j : K \times 1$ vector, and $XX' : K \times K$ matrix

$$\begin{aligned}
\max_{1\le j\le p} |\frac{\eta_j'X'}{n}\left(\frac{XX'}{n}\right)^{-1}\frac{X\eta_j}{n}| &\le \max_{1\le j\le p} \|\frac{\eta_j'X'}{n}\|_1 \max_{1\le j\le p} \|\left(\frac{XX'}{n}\right)^{-1}\frac{X\eta_j}{n}\|_\infty \\
&\le \left[\max_{1\le j\le p} \|\frac{\eta_j'X'}{n}\|_1\right] K\|\left(\frac{XX'}{n}\right)^{-1}\|_\infty \max_{1\le j\le p} \|\frac{X\eta_j}{n}\|_\infty \\
&\le \left[K \max_{1\le j\le p} \|\frac{\eta_j'X'}{n}\|_\infty\right]\left[K\|\left(\frac{XX'}{n}\right)^{-1}\|_\infty \max_{1\le j\le p} \|\frac{X\eta_j}{n}\|_\infty\right] \\
&= \left[K \max_{1\le j\le p} \|\frac{\eta_j'X'}{n}\|_\infty\right]^2 \|\left(\frac{XX'}{n}\right)^{-1}\|_\infty, \quad (A.55)
\end{aligned}$$

where we use Hölder's inequality for the first inequality, and (A.1)(A.2) for the second inequality, and the norm inequality between l_1, l_∞ norms for the third inequality (i.e. $\|x\|_1 \le dim(x)\|x\|_\infty, dim(x)$: dimension of the vector x). Next by (A.33)(A.34)(A.35), we have by (A.55)

$$\max_{1\le j\le p} |\frac{\eta_j'X'}{n}\left(\frac{XX'}{n}\right)^{-1}\frac{X\eta_j}{n}| = O_p(K^2\bar{s}\frac{lnp}{n}). \quad (A.56)$$

Combine (A.54)(A.56) in (A.53) to have the first term on the right side of (A.52) by Assumption 5 to get the last equality in (A.57)

$$\max_{1\le j\le p} |\frac{\eta_{xj}'\eta_{xj}}{n} - \tau_j^2| = O_p(K^2\bar{s}\frac{lnp}{n}) + O_p(\sqrt{\frac{lnp}{n}}) = O_p(\sqrt{\frac{lnp}{n}}). \quad (A.57)$$

See that by Lemma A.5 (wpa1) and (A.14)

$$\|\eta_{xj}'\hat{U}_{-j}'/n\|_\infty \le \lambda_n/2. \quad (A.58)$$

Then we also have by (A.51) $\lambda_n = O(\sqrt{lnp/n})$. In (A.52) consider the second term on the right side by

(A.58), Lemma A.7, with rate for λ_n in (A.51)

$$\begin{aligned}
\max_{1\le j\le p} |\frac{\eta_{xj}'\hat{U}_{-j}'}{n}(\hat{\gamma}_j - \gamma_j)| &\le \max_{1\le j\le p} \|\frac{\eta_{xj}'\hat{U}_{-j}'}{n}\|_\infty \max_{1\le j\le p} \|\hat{\gamma}_j - \gamma_j\|_1 \\
&= O_p(\sqrt{\frac{lnp}{n}})O_p(\sqrt{\frac{lnp}{n}}\bar{s}) \\
&= O_p\left(\bar{s}\frac{lnp}{n}\right). \qquad (A.59)
\end{aligned}$$

Consider the third term on the right side of (A.52), where we use Hölder's inequality

$$\begin{aligned}
\max_{1\le j\le p} |\gamma_j' \frac{\hat{U}_{-j}\eta_{xj}}{n}| &\le |[\max_{1\le j\le p} \|\gamma_j\|_1][\max_{1\le j\le p} \|\frac{\hat{U}_{-j}\eta_{xj}}{n}\|_\infty] \\
&= O(\bar{s}^{1/2})O_p(\sqrt{\frac{lnp}{n}}) \\
&= O_p\left(\sqrt{\bar{s}}\sqrt{\frac{lnp}{n}}\right), \qquad (A.60)
\end{aligned}$$

for the rates we use (A.11)(A.58)(A.51). Last we consider the fourth term on the right side of (A.52). To get a better rate, we start with Karush-Kuhn-Tucker (KKT) conditions in (24). The following $p-1$ equations form the KKT

$$\lambda_n \hat{\kappa}_j + \frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}\hat{\gamma}_j - \frac{\hat{U}_{-j}\hat{u}_j}{n} = 0_{p-1},$$

where $\hat{\kappa}_j$ is the sub-differrential and explained in more detail in p.160 of Caner and Kock (2018) which replaces the gradient in non-differential penalties. Also for all $j = 1\cdots, p$

$$\|\hat{\kappa}_j\|_\infty \le 1. \qquad (A.61)$$

Use (22) for $\hat{u}_j$

$$\lambda_n \hat{\kappa}_j + \frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}\hat{\gamma}_j - \frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}\gamma_j - \frac{\hat{U}_{-j}\eta_{xj}}{n} = 0_{p-1}.$$

Rewrite above equations as

$$\frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}(\hat{\gamma}_j - \gamma_j) = \frac{\hat{U}_{-j}\eta_{xj}}{n} - \lambda_n\hat{\kappa}_j.$$

Then by triangle inequality and (A.51),(A.58)

$$\begin{aligned}
\|\frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}(\hat{\gamma}_j - \gamma_j)\|_\infty &\le \|\frac{\hat{U}_{-j}\eta_{xj}}{n}\|_\infty + \lambda_n\|\hat{\kappa}_j\|_\infty \\
&\le \|\frac{\hat{U}_{-j}\eta_{xj}}{n}\|_\infty + \lambda_n \\
&= O_p(\sqrt{\frac{lnp}{n}}) + O_p(\sqrt{\frac{lnp}{n}}) \\
&= O_p(\sqrt{\frac{lnp}{n}}). \qquad (A.62)
\end{aligned}$$

Then the fourth term on the right side of (A.52)

$$\begin{aligned}
\max_{1\le j\le p} |\gamma_j' \frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}(\hat{\gamma}_j - \gamma_j)| &\le \max_{1\le j\le p} \|\gamma_j\|_1 \max_{1\le j\le p} \|\frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}(\hat{\gamma}_j - \gamma_j)\|_\infty \\
&= O(\sqrt{\bar{s}})O_p(\sqrt{\frac{lnp}{n}}) \\
&= O_p(\sqrt{\frac{\bar{s}lnp}{n}}), \qquad (A.63)
\end{aligned}$$

where we use Hölder's inequality, (A.11)(A.62). Clearly, (A.60)(A.63) are the slowest among the four terms on the right side of (A.52), and we use Assumption 5 to get the desired result.**Q.E.D.**

Proof of Theorem 1. First, we derive some of the key results. By definition of τ_j^2, for $j = 1, \cdots, p$, and since $\Omega : \Sigma_n^{-1}$, with Assumption 1

$$\tau_j^2 := \frac{1}{\Omega_{j,j}} \geq \frac{1}{Eigmax(\Omega)} = Eigmin(\Sigma_n) \geq c > 0. \tag{A.64}$$

Note that $min_{1 \leq j \leq p} \tau_j^2$ is bounded away from zero. Next

$$\min_{1 \leq j \leq p} \hat{\tau}_j^2 = \min_{1 \leq j \leq p} |\hat{\tau}_j^2 - \tau_j^2 + \tau_j^2| \geq \min_{1 \leq j \leq p} \tau_j^2 - \max_{1 \leq j \leq p} |\hat{\tau}_j^2 - \tau_j^2|. \tag{A.65}$$

is bounded away from zero wpa1 by Lemma A.8. Then

$$\max_{1 \leq j \leq p} |\frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2}| = \max_{1 \leq j \leq p} \frac{|\hat{\tau}_j^2 - \tau_j^2|}{\hat{\tau}_j^2 \tau_j^2} = O_p(\sqrt{\frac{\bar{s} lnp}{n}}) = o_p(1). \tag{A.66}$$

by Lemma A.8, (A.64)(A.65). Now we complete the proof by using the formula for $\hat{\Omega}_j, \Omega_j$.

$$\begin{aligned}
\max_{1 \leq j \leq p} \|\hat{\Omega}_j - \Omega_j\|_1 &= \max_{1 \leq j \leq p} \|\frac{\hat{C}_j}{\hat{\tau}_j^2} - \frac{C_j}{\tau_j^2}\|_1 \\
&\leq \max_{1 \leq j \leq p} |\frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2}| + \max_{1 \leq j \leq p} \|\frac{\hat{\gamma}_j}{\hat{\tau}_j^2} - \frac{\gamma_j}{\tau_j^2}\|_1 \\
&= \max_{1 \leq j \leq p} |\frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2}| + \max_{1 \leq j \leq p} \|\frac{\hat{\gamma}_j}{\hat{\tau}_j^2} - \frac{\gamma_j}{\hat{\tau}_j^2} + \frac{\gamma_j}{\hat{\tau}_j^2} - \frac{\gamma_j}{\tau_j^2}\|_1 \\
&= \max_{1 \leq j \leq p} |\frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2}| + \max_{1 \leq j \leq p} \frac{\|\hat{\gamma}_j - \gamma_j\|_1}{\hat{\tau}_j^2} \\
&+ \max_{1 \leq j \leq p} \|\gamma_j\|_1 \max_{1 \leq j \leq p} |\frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2}| \\
&= O_p(\sqrt{\frac{\bar{s} lnp}{n}}) + O_p(\bar{s}\sqrt{\frac{lnp}{n}}) + O(\bar{s}^{1/2}) O_p(\sqrt{\frac{\bar{s} lnp}{n}}) \\
&= O_p(\bar{s}\sqrt{\frac{lnp}{n}}) = o_p(1),
\end{aligned}$$

where we use (A.66), Lemma A.7, (A.11) for the rates, and the last equality is by Assumption 5.**Q.E.D.**

Part 3:

After the proof of Theorem 1 we provide lemmata that lead to proof of Theorem 2. We start with a lemma that is related to norm inequalities. First define generic matrices, $A_1 : p \times K, A_2 : K \times p, D_1 : K \times K, D_2 : p \times p$, also define a row vector $x' : 1 \times p$, and also define $p \times p$ matrices A_3, D_3.

Lemma A.9. *(i).*

$$\|A_1 D_1 A_2\|_{l_\infty} \leq p K^{1/2} \|A_1\|_{l_\infty} \|A_2\|_\infty \|D_1\|_{l_2}.$$

(ii).

$$\|A_2 D_2 A_1\|_\infty \leq p \|A_2\|_\infty \|A_1\|_\infty \|D_2\|_{l_\infty}.$$

(iii).

$$\|x' A_3 D_3\|_1 \leq \|x\|_1 \|D_3\|_{l_\infty} \|A_3\|_{l_\infty}.$$

(iv).

$$\|A_2 D_2 A_1\|_\infty \leq p \|A_2\|_\infty \|A_1\|_\infty \|D_2\|_{l_1}.$$

Proof of Lemma A.9.

(i).

$$\begin{aligned}\|A_1 D_1 A_2\|_{l_\infty} &\leq \|A_1\|_{l_\infty}\|D_1 A_2\|_{l_\infty}\\ &\leq \|A_1\|_{l_\infty}[p\|A_2\|_\infty]\|D_1\|_{l_\infty}\\ &\leq p\|A_1\|_{l_\infty}\|A_2\|_\infty[K^{1/2}\|D_1\|_{l_2}]\end{aligned}$$

where we use submultiplicativity of matrix norms for the first inequality, and submultiplicativity of matrix norms and the following for the second inequality,

$$\|A_2\|_{l_\infty} := \max_{1\leq k\leq K}\|A'_{2,k}\|_1 \leq p \max_{1\leq k\leq K}\max_{1\leq j\leq p}|A_{2,kj}|,$$

where $A'_{2,k}, A_{2,kj}$ are the k th row of A_2, and k, j element of A_2 respectively. Then for the last inequality, we use a matrix norm inequality that provides an upper bound for l_∞ matrix norm in terms of spectral norm in p.365 of Horn and Johnson (2013).

(ii).

$$\begin{aligned}\|A_2 D_2 A_1\|_\infty &\leq \|A_2\|_\infty\|D_2 A_1\|_{l_1}\\ &\leq p\|A_2\|_\infty\|A_1\|_\infty\|D_2\|_{l_\infty}\end{aligned}$$

where we use section 4.3 of van de Geer (2016) for the first inequality, and the second inequality can be seen by defining $D'_{2,j}$as the row of D_2, and $A_{1,k}$ as the k th column of A_1 and using Hölder's inequality

$$\begin{aligned}\|D_2 A_1\|_{l_1} &:= \max_{1\leq k\leq K}\sum_{j=1}^{p}|D'_{2,j}A_{1,k}| \leq \max_{1\leq k\leq K}\sum_{j=1}^{p}\|D_{2,j}\|_1\|A_{1,k}\|_\infty\\ &\leq p\max_{1\leq j\leq p}\|D_{2,j}\|_1(\max_{1\leq k\leq K}\|A_{1,k}\|_\infty)\\ &= p\|D_2\|_{l_\infty}\|A_1\|_\infty.\end{aligned}$$

(iii).

$$\begin{aligned}\|x'A_3D_3\|_1 = \|D'_3A'_3x\|_1 &\leq \|x\|_1\|D'_3A'_3\|_{l_1}\\ &\leq \|x\|_1\|D'_3\|_{l_1}\|A'_3\|_{l_1}\\ &= \|x\|_1\|D_3\|_{l_\infty}\|A_3\|_{l_\infty},\end{aligned}$$

where we use p.345 of Horn and Johnson (2013) for the first inequality, and l_1 matrix norm submultiplicativity for the second inequality, and the last equality is by seeing that transpose of l_1 matrix norm is l_∞ matrix norm.

(iv).

$$\begin{aligned}\|A_2 D_2 A_1\|_\infty &\leq \|A_2\|_\infty\|D_2 A_1\|_{l_1}\\ &\leq \|A_2\|_\infty\|D_2\|_{l_1}\|A_1\|_{l_1}\\ &\leq p\|A_2\|_\infty\|D_2\|_{l_1}\|A_1\|_\infty,\end{aligned}$$

where we use p.44 van de Geer (2016) dual norm inequality for the first inequality, then for the second inequality we use submultiplicativity property of matrix norms,and for the last inequality we use $\|A_1\|_{l_1} := \max_{1\leq k\leq K}\sum_{j=1}^{p}|A_{1,jk}| \leq p\|A_1\|_\infty$, where $A_{1,jk}$ is the j, k th cell in A_1. **Q.E.D.**

Lemma A.10. *Under Assumptions 1-6*

(i).

$$\|\hat{B} - B\|_{l_\infty} = O_p(K^{3/2}\sqrt{lnp/n}),$$

(ii).

$$\|\hat{B}\|_{l_\infty} = O_p(K),$$

(iii).

$$\|B\|_{l_\infty} = O(K).$$

(iv).

$$\|B'\|_{l_\infty} = \|B\|_{l_1} = O(p).$$

Proof of Lemma A.10. We start with reminding $X : K \times n$, $X = [f_1, \cdots, f_t, \cdots, f_n]$, where $f_t : K \times 1$ vector, and $u_j : n \times 1$ vector, with least squares estimate $\hat{b}_j - b_j = (\frac{XX'}{n})^{-1}\frac{Xu_j}{n}$ for $j = 1, \cdots, p$ as in p.3347 of Fan et al. (2011).

(i).

$$\begin{aligned}
\|\hat{B} - B\|_{l_\infty} &:= \max_{1 \le j \le p} \|\hat{b}_j' - b_j'\|_1 = \max_{1 \le j \le p} \|b_j - b_j\|_1 \\
&= \max_{1 \le j \le p} \|(\frac{XX'}{n})^{-1}\frac{Xu_j}{n}\|_1 \\
&\le \|(\frac{XX'}{n})^{-1}\|_{l_1} \max_{1 \le j \le p} \|\frac{Xu_j}{n}\|_1 \\
&\le \|(\frac{XX'}{n})^{-1}\|_{l_1} \left[K \max_{1 \le j \le p} \|\frac{Xu_j}{n}\|_\infty\right] \\
&\le K\left[K^{1/2}\|\frac{XX'}{n})^{-1}\|_{l_2}\right] \max_{1 \le j \le p} \|\frac{Xu_j}{n}\|_\infty \\
&= O(K^{3/2})O_p(\sqrt{lnp/n}),
\end{aligned}$$

where we use l_∞ norm definition for the first equality, and for the first inequality we use p.345 of Horn and Johnson (2013), which is $\|Ax\|_1 \le \|A\|_{l_1}\|x\|_1$ for a generic matrix A, and generic vector x, for the third inequality we use the upper bound of l_1 induced matrix norm in terms of spectral norm, as in p.365 of Horn and Johnson (2013). The rates are from (B.3)(B.4) of Fan et al. (2011), and Lemma A.3(iii).

(ii). See that

$$\|B\|_{l_\infty} = \max_{1 \le j \le p} \|b_j'\|_1 = O(K), \tag{A.67}$$

by Assumption 6 that $|b_{jk}| \le C$ for a positive constant C and uniformly over $j = 1, \cdots, p$, $k = 1, \cdots, K$. Next using the results above with Assumption 5, we have

$$\|\hat{B}\|_{l_\infty} = O_p(K).$$

(iii). This is proved in (A.67).

(iv). The proof of (iv) is the same as in (iii) above except, with b_k as the k th column of matrix B.

$$\|B\|_{l_1} = \max_{1 \le k \le K} \|b_k\|_1 = O(p),$$

by Assumption 6. **Q.E.D.**

Before the next lemma, we extend two following results which is Lemma B.4 in Fan et al. (2011) to the case of increasing maximal eigenvalue of errors.

Lemma A.11. *Under Assumptions 4,6,7(i), with $c > 0, C > 0$, and positive finite constants (i).*

$$Eigmin(B'\Omega B) \geq \frac{cp}{cr_N}.$$

(ii).

$$\|([Covf_t]^{-1} + B'\Omega B)^{-1}\|_{l_2} = O(\frac{r_n}{p}).$$

Proof of Lemma A.11. We follow the proof of Lemma B.4 in Fan et al. (2011). (i). Since $\Omega := \Sigma_n^{-1}$,

$$\begin{aligned} Eigmin(B'\Omega B) &\geq Eigmin(\Omega)Eigmin(B'B) \\ &= [Eigmax(\Sigma_n)]^{-1}Eigmin(B'B) \\ &\geq \frac{cp}{Cr_n}, \end{aligned}$$

by Assumption 6, 7(i).

(ii). Using Assumption 4

$$Eigmin[([Covf_t]^{-1} + B'\Omega B)] \geq Eigmin(B'\Omega B) \geq \frac{cp}{Cr_n}. \tag{A.68}$$

We have the desired result by (A.68), and since for an invertible matrix A, $Eigmax(A^{-1}) = 1/Eigmin(A)$. **Q.E.D.**

As described above in the main text, we form the symmetrized version of our feasible nodewise regression estimator for this part of the paper: $\hat{\Omega}_{sym} := \frac{\hat{\Omega}+\hat{\Omega}'}{2}$.

Lemma A.12. *(i). Under Assumptions 1-6*

$$\|\hat{B}'\hat{\Omega}_{sym}\hat{B} - B'\Omega B\|_{l_2} = O_p\left(pK\max(\bar{s}, \bar{s}^{1/2}K^{1/2})\sqrt{\frac{lnp}{n}}\right).$$

(ii). Under Assumptions 1-4, 6-7

$$\|\hat{G}\|_{l_2} = O_p\left(\frac{r_n}{p}\right),$$

with $\hat{G} := [[\widehat{covf_t}]^{-1} + \hat{B}'\hat{\Omega}_{sym}\hat{B}]^{-1}$.

Proof of Lemma A.12. (i). We start with simple adding and subtracting($\hat{B} = (\hat{B} - B) + B$), $\hat{\Omega}_{sym} = (\hat{\Omega}_{sym} - \Omega) + \Omega$) and triangle inequality

$$\begin{aligned} \|\hat{B}'\hat{\Omega}_{sym}\hat{B} - B'\Omega B\|_\infty &\leq \|(\hat{B} - B)'(\hat{\Omega}_{sym} - \Omega)(\hat{B} - B)\|_\infty \\ &+ 2\|(\hat{B} - B)'(\hat{\Omega}_{sym} - \Omega)B\|_\infty + \|(\hat{B} - B)'\Omega(\hat{B} - B)\|_\infty \\ &+ \|B'(\hat{\Omega}_{sym} - \Omega)B\|_\infty + 2\|(\hat{B} - B)'\Omega B\|_\infty. \end{aligned} \tag{A.69}$$

Analyze each term in (A.69), and by Lemma A.9(ii)(iv)

$$\begin{aligned} \|(\hat{B} - B)'(\hat{\Omega}_{sym} - \Omega)(\hat{B} - B)\|_\infty &\leq \frac{1}{2}\|(\hat{B} - B)'(\hat{\Omega} - \Omega)(\hat{B} - B)\|_\infty \\ &+ \frac{1}{2}\|(\hat{B} - B)'(\hat{\Omega}' - \Omega)(\hat{B} - B)\|_\infty \\ &\leq \frac{p}{2}\|\hat{B} - B\|_\infty^2\|\hat{\Omega} - \Omega\|_{l_\infty} + \frac{p}{2}\|\hat{B} - B\|_\infty^2\|\hat{\Omega}' - \Omega\|_{l_1} \\ &= pO_p(\frac{Klnp}{n})O_p(\frac{\bar{s}\sqrt{lnp}}{\sqrt{n}}), \end{aligned} \tag{A.70}$$

where we use (B.14) of Fan et al (2011) which is: $\|\hat{B}-B\|_\infty = O_p(\sqrt{\frac{Klnp}{n}})$, and Theorem 1 with

$$\|\hat{\Omega}-\Omega\|_{l_\infty} = \|\hat{\Omega}'-\Omega\|_{l_1} := \max_{1\leq j\leq p} \|\hat{\Omega}_j - \Omega_j\|_1, \tag{A.71}$$

since l_1 norm of transpose of $\hat{\Omega}$ involves rows of $\hat{\Omega}$ (hence columns of $\hat{\Omega}'$).

For the second term in (A.69)

$$\begin{aligned}
\|(\hat{B}-B)'(\hat{\Omega}_{sym}-\Omega)B\|_\infty &\leq \|(\hat{B}-B)'(\hat{\Omega}-\Omega)B\|_\infty/2 \\
&+ \|(\hat{B}-B)'(\hat{\Omega}'-\Omega)B\|_\infty/2 \\
&\leq \frac{p}{2}\|\hat{B}-B\|_\infty\|B\|_\infty\|\hat{\Omega}-\Omega\|_{l_\infty} \\
&+ \frac{p}{2}\|\hat{B}-B\|_\infty\|B\|_\infty\|\hat{\Omega}'-\Omega\|_{l_1} \\
&= pO_p(\sqrt{\frac{Klnp}{n}})O(1)O_p(\bar{s}\sqrt{\frac{lnp}{n}}),
\end{aligned} \tag{A.72}$$

where we use, $\hat{\Omega}_{sym}$, Lemma A.9(ii)(iv) for the first-second inequalities, (B.14) of Fan et al. (2011), Assumption 6, and Theorem 1 and (A.71) for the rates. Now consider the third term in (A.69)

$$\begin{aligned}
\|(\hat{B}-B)'\Omega(\hat{B}-B)\|_\infty &\leq p\|\hat{B}-B\|_\infty^2\|\Omega\|_{l_\infty} \\
&= p[O_p(\sqrt{\frac{Klnp}{n}})]^2 O(\bar{s}^{1/2}),
\end{aligned} \tag{A.73}$$

where we use Lemma A.9(ii) for the first inequality, (B.14) of Fan et al. (2011), and $\|\Omega\|_{l_\infty} := \max_{1\leq j\leq p}\|\Omega_j'\|_1 = \max_{1\leq j\leq p}\|\Omega_j\|_1 = O(\bar{s}^{1/2})$ as in (A.12). We consider the fourth term in (A.69)

$$\begin{aligned}
\|B'(\hat{\Omega}_{sym}-\Omega)B\|_\infty &\leq \frac{1}{2}\|B'(\hat{\Omega}-\Omega)B\|_\infty + \frac{1}{2}\|B'(\hat{\Omega}'-\Omega)B\|_\infty \\
&\leq \frac{p}{2}\|B\|_\infty^2\|\hat{\Omega}-\Omega\|_{l_\infty} + \frac{p}{2}\|B\|_\infty^2\|\hat{\Omega}'-\Omega\|_{l_1} \\
&= p[O(1)]^2 O_p(\bar{s}\sqrt{\frac{lnp}{n}}),
\end{aligned} \tag{A.74}$$

where we use symmetry of $\hat{\Omega}_{sym}$, Lemma A.9(ii)(iv) for the first and second inequality, and Assumption 6, and Theorem 1 (A.71) for the rates. Also analyze the fifth term in (A.69)

$$\begin{aligned}
\|(\hat{B}-B)'\Omega B\|_\infty &\leq p\|\hat{B}-B\|_\infty\|B\|_\infty\|\Omega\|_{l_\infty} \\
&= pO_p(\sqrt{\frac{Klnp}{n}})O(1)O(\bar{s}^{1/2}),
\end{aligned} \tag{A.75}$$

where we use Lemma A.9(ii) for the inequality, and the rates are by (B.14) of Fan et al. (2011), Assumption 6, and $\|\Omega\|_{l_\infty} := \max_{1\leq j\leq p}\|\Omega_j'\|_1 = \max_{1\leq j\leq p}\|\Omega_j\|_1 = O(\bar{s}^{1/2})$ as in (A.12). Since by Assumption 5, $\sqrt{Klnp/n} = o(1)$, the slowest rate is the maximum of the rates (A.74)(A.75) above. So

$$\|\hat{B}'\hat{\Omega}_{sym}\hat{B} - B'\Omega B\|_\infty = O_p\left(p\max(\bar{s}, \bar{s}^{1/2}K^{1/2})\sqrt{\frac{lnp}{n}}\right). \tag{A.76}$$

Then by norm inequality tying spectral norm to $\|.\|_\infty$ norm in p.365 of Horn and Johnson (2013), and since $\hat{B}'\hat{\Omega}_{sym}\hat{B} - B'\Omega B$ is $K\times K$ matrix

$$\|\hat{B}'\hat{\Omega}_{sym}\hat{B} - B'\Omega B\|_{l_2} \leq K\|\hat{B}'\hat{\Omega}_{sym}\hat{B} - B'\Omega B\|_\infty = O_p\left(pK\max(\bar{s}, \bar{s}^{1/2}K^{1/2})\sqrt{\frac{lnp}{n}}\right). \tag{A.77}$$

Q.E.D.

(ii). Since $\widehat{covf_t}^{-1}, (covf_t)^{-1}$ does not involve precision matrix estimate, we proceed as in Fan et al. (2011), Lemma B5(ii). Specifically (B.20) of Fan et al. (2011) provide

$$\|[\widehat{covf_t}]^{-1} - [covf_t]^{-1}\|_{l_2} = O_p(\sqrt{lnn/n}).$$

Using (A.77) and the equation above we develop a larger bound

$$\|([\widehat{covf_t}]^{-1} + \hat{B}'\hat{\Omega}_{sym}\hat{B}) - ([covf_t]^{-1} + B'\Omega B)\|_{l_2} = O_p\left(pK\max(\bar{s}, \bar{s}^{1/2}K^{1/2})\sqrt{\frac{max(lnp, lnn)}{n}}\right). \quad (A.78)$$

We have by (A.68)

$$Eigmin[[covf_t]^{-1} + B'\Omega B] \geq \frac{cp}{Cr_n}, \quad (A.79)$$

by Assumption 6, and since $r_n/p \to 0$. Note that

$$pK\max(\bar{s}, \bar{s}^{1/2}K^{1/2})\sqrt{\frac{max(lnp, lnn)}{n}} = o(p/r_n),$$

since by Assumption 7, $r_nK\max(\bar{s}, \bar{s}^{1/2}K^{1/2})\sqrt{\frac{max(lnp,lnn)}{n}} = o(1)$. So (A.78) has the rate

$$\|([\widehat{covf_t}]^{-1} + \hat{B}'\hat{\Omega}_{sym}\hat{B}) - ([covf_t]^{-1} + B'\Omega B)\|_{l_2} = o_p(p/r_n). \quad (A.80)$$

Then using Lemma A.1(i) of Fan et al. (2011), with (A.79)(A.80)

$$Eigmin([\widehat{covf_t}]^{-1} + \hat{B}'\hat{\Omega}_{sym}\hat{B}) \geq \frac{cp}{Cr_n}, \quad (A.81)$$

wpa1 with $r_n << p$ as in Assumption 7. By (A.81), and seeing that for invertible matrix A, $Eigmax(A^{-1}) = 1/Eigmin(A)$,

$$\|([\widehat{covf_t}]^{-1} + \hat{B}'\hat{\Omega}_{sym}\hat{B})^{-1}\|_{l_2} = O_p(\frac{r_n}{p}).$$

Q.E.D.

We restate the definitions of major terms that are used.

$$\hat{G} := (\widehat{covf_t}^{-1} + \hat{B}'\hat{\Omega}_{sym}\hat{B})^{-1}. \quad (A.82)$$

$$G := (covf_t^{-1} + B'\Omega B)^{-1}. \quad (A.83)$$

Next, remembering

$$\hat{L} := \hat{B}\hat{G}\hat{B}' \quad L := BGB'. \quad (A.84)$$

and

$$l_n := r_n^2K^{5/2}\max(\bar{s}, \bar{s}^{1/2}K^{1/2})\sqrt{\frac{\max(lnp, lnn)}{n}}. \quad (A.85)$$

We have the next lemma which will be instrumental in proving Theorem 2.

Lemma A.13. *Under Assumptions 1-4, 6-7*

$$\|\hat{L} - L\|_{l_\infty} = O_p(l_n) = o_p(1).$$

Proof of Lemma A.13. Start with, by adding and subtracting and triangle inequality

$$\begin{aligned}\|\hat{L} - L\|_{l_\infty} &\leq \|(\hat{B} - B)\hat{G}(\hat{B} - B)'\|_{l_\infty} + \|(\hat{B} - B)\hat{G}B'\|_{l_\infty} \\ &+ \|B\hat{G}(\hat{B} - B)'\|_{l_\infty} + \|B\hat{G}B' - BGB'\|_{l_\infty}. \qquad \text{(A.86)}\end{aligned}$$

Consider the first term in (A.86)

$$\begin{aligned}\|(\hat{B} - B)\hat{G}(\hat{B} - B)'\|_{l_\infty} &\leq pK^{1/2}\|\hat{B} - B\|_{l_\infty}\|\hat{G}\|_{l_2}\|\hat{B} - B\|_\infty \\ &= pK^{1/2}O_p(K^{3/2}\sqrt{\frac{lnp}{n}})O_p(\frac{r_n}{p})O_p(\sqrt{\frac{Klnp}{n}}) \\ &= O_p(r_n K^{5/2}\frac{lnp}{n}), \qquad \text{(A.87)}\end{aligned}$$

where we use Lemma A.9(i) for the first inequality, Lemma A.10-A.12, and (B.14) of Fan et al. (2011): $\|\hat{B} - B\|_\infty = O_p(\sqrt{\frac{Klnp}{n}})$ for the rates. Next in (A.86), we consider the second term on the right side

$$\begin{aligned}\|(\hat{B} - B)\hat{G}B'\|_{l_\infty} &\leq pK^{1/2}\|\hat{B} - B\|_{l_\infty}\|\hat{G}\|_{l_2}\|B\|_\infty \\ &= pK^{1/2}O_p(K^{3/2}\sqrt{\frac{lnp}{n}})O_p(\frac{r_n}{p})O(1) = O_p(r_n K^2\sqrt{\frac{lnp}{n}}), \qquad \text{(A.88)}\end{aligned}$$

where we use Lemma A.9(i) for the first inequality, and for the rates use Lemma A.10-A.12, and Assumption 6 which shows that factor loadings are uniformly bounded away from infinity. Analyze the third term in (A.86).

$$\begin{aligned}\|B\hat{G}(\hat{B} - B)'\|_{l_\infty} &\leq pK^{1/2}\|B\|_{l_\infty}\|\hat{G}\|_{l_2}\|\hat{B} - B\|_\infty \\ &= pK^{1/2}O(K)O_p(\frac{r_n}{p})O_p(\sqrt{\frac{Klnp}{n}}) = O_p(r_n K^2\sqrt{\frac{lnp}{n}}), \qquad \text{(A.89)}\end{aligned}$$

where we use Lemma A.9(i) for the first inequality, Lemma A.10-A.12, and (B.14) of Fan et al. (2011): $\|\hat{B} - B\|_\infty = O_p(\sqrt{\frac{Klnp}{n}})$ for the rates. Now we analyze the fourth term on the right side of (A.86).

$$\|B(\hat{G} - G)B'\|_{l_\infty} \leq pK^{1/2}\|B\|_{l_\infty}\|\hat{G} - G\|_{l_2}\|B\|_\infty, \qquad \text{(A.90)}$$

where we use Lemma A.9(i). We have from (A.82)(A.83) and by submultiplicativity of l_2 matrix norm (spectral norm)

$$\begin{aligned}\|\hat{G} - G\|_{l_2} &\leq \|\hat{G}\|_{l_2}\|G\|_{l_2}\|(\widehat{covf_t}^{-1} + \hat{B}'\hat{\Omega}_{sym}\hat{B}) - (covf_t^{-1} + B'\Omega B)\|_{l_2} \\ &= O_p(\frac{r_n}{p})O(\frac{r_n}{p})O_p\left(pK\max(\bar{s}, \bar{s}^{1/2}K^{1/2})\sqrt{\frac{\max(lnp, lnn)}{n}}\right) \\ &= O_p\left(r_n^2 p^{-1}K\max(\bar{s}, \bar{s}^{1/2}K^{1/2})\sqrt{\frac{max(lnp, lnn)}{n}}\right), \qquad \text{(A.91)}\end{aligned}$$

where we use Lemma A.12, and $\|G\|_{l_2} = O(r_n/p)$ by Lemma A.11, (A.78). Substitute (A.91) into (A.90) via Lemma A.10

$$\|B(\hat{G} - G)B'\|_{l_\infty} = O_p\left(r_n^2 K^{5/2}\max(\bar{s}, \bar{s}^{1/2}K^{1/2})\sqrt{\frac{max(lnp, ln)}{n}}\right). \qquad \text{(A.92)}$$

Since the last rate is the slowest among all on the right side of (A.86) we have the desired result.**Q.E.D.**

Proof of Theorem 2. From (34), and using triangle inequality

$$\max_{1\le j\le p}\|\hat{\Gamma}_j-\Gamma_j\|_1=\max_{1\le j\le p}\|\hat{\Gamma}_j'-\Gamma_j'\|_1\le\max_{1\le j\le p}\|\hat{\Omega}_j'-\Omega_j'\|_1+\max_{1\le j\le p}\|\hat{\Omega}_j'\hat{L}\hat{\Omega}-\Omega_j'L\Omega\|_1. \tag{A.93}$$

We consider second right side term in (A.93). Add and subtract $\Omega_j'\hat{L}\hat{\Omega}$ via triangle inequality

$$\max_{1\le j\le p}\|\hat{\Omega}_j'\hat{L}\hat{\Omega}-\Omega_j'L\Omega\|_1\le\max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'\hat{L}\hat{\Omega}\|_1+\max_{1\le j\le p}\|\Omega_j'(\hat{L}\hat{\Omega}-L\Omega)\|_1. \tag{A.94}$$

We analyze the first term on the right side of (A.94) and try to simplify by adding and subtracting $(\hat{\Omega}_j-\Omega_j)'\hat{L}\Omega$, and triangle inequality

$$\max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'\hat{L}\hat{\Omega}\|_1\le\max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'\hat{L}(\hat{\Omega}-\Omega)\|_1+\max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'\hat{L}\Omega\|_1. \tag{A.95}$$

Then on the first right side term in (A.95) add and subtract $(\hat{\Omega}_j-\Omega_j)'L(\hat{\Omega}-\Omega)$ via triangle inequality

$$\max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'\hat{L}(\hat{\Omega}-\Omega)\|_1\le\max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'(\hat{L}-L)(\hat{\Omega}-\Omega)\|_1+\max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'L(\hat{\Omega}-\Omega)\|_1.$$

Now for the second right side term in (A.95) add and subtract $(\hat{\Omega}_j-\Omega_j)'L\Omega$ via triangle inequality

$$\max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'\hat{L}\Omega\|_1\le\max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'(\hat{L}-L)\Omega\|_1+\max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'L\Omega\|_1.$$

Substitute the last two inequalities into (A.95)

$$\begin{aligned}\max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'\hat{L}\hat{\Omega}\|_1 &\le \max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'(\hat{L}-L)(\hat{\Omega}-\Omega)\|_1\\ &+ \max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'L(\hat{\Omega}-\Omega)\|_1\\ &+ \max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'(\hat{L}-L)\Omega\|_1\\ &+ \max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'L\Omega\|_1. \end{aligned}\tag{A.96}$$

Now in (A.94) we consider the second term on the right side, add and subtract $\Omega_j'L\hat{\Omega}$ via triangle inequality

$$\max_{1\le j\le p}\|\Omega_j'\hat{L}\hat{\Omega}-\Omega_j'L\Omega\|_1\le\max_{1\le j\le p}\|\Omega_j'(\hat{L}-L)\hat{\Omega}\|_1+\max_{1\le j\le p}\|\Omega_j'L(\hat{\Omega}-\Omega)\|_1. \tag{A.97}$$

Also add and subtract $\Omega_j'(\hat{L}-L)\Omega$ to the first term on the right side of (A.97) above, to have

$$\begin{aligned}\max_{1\le j\le p}\|\Omega_j'\hat{L}\hat{\Omega}-\Omega_j'L\Omega\|_1 &\le \max_{1\le j\le p}\|\Omega_j'(\hat{L}-L)(\hat{\Omega}-\Omega)\|_1\\ &+ \max_{1\le j\le p}\|\Omega_j'(\hat{L}-L)\Omega\|_1\\ &+ \max_{1\le j\le p}\|\Omega_j'L(\hat{\Omega}-\Omega)\|_1. \end{aligned}\tag{A.98}$$

Combine (A.96)(A.98) into (A.94) right side to have

$$\begin{aligned}\max_{1\le j\le p}\|\hat{\Omega}_j'\hat{L}\hat{\Omega}-\Omega_j'L\Omega\|_1 &\le \max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'(\hat{L}-L)(\hat{\Omega}-\Omega)\|_1\\ &+ \max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'L(\hat{\Omega}-\Omega)\|_1\\ &+ \max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'(\hat{L}-L)\Omega\|_1\\ &+ \max_{1\le j\le p}\|(\hat{\Omega}_j-\Omega_j)'L\Omega\|_1\\ &+ \max_{1\le j\le p}\|\Omega_j'(\hat{L}-L)(\hat{\Omega}-\Omega)\|_1\\ &+ \max_{1\le j\le p}\|\Omega_j'(\hat{L}-L)\Omega\|_1\\ &+ \max_{1\le j\le p}\|\Omega_j'L(\hat{\Omega}-\Omega)\|_1. \end{aligned}\tag{A.99}$$

To consider all the terms in (A.99) we need to find some rates about terms. In that respect,

$$\begin{aligned} \|L\|_{l_\infty} &:= \|BGB'\|_{l_\infty} \\ &\leq \|B\|_{l_\infty}\|B\|_{l_1}\|G\|_{l_\infty} \\ &\leq \|B\|_{l_\infty}\|B\|_{l_1}[K^{1/2}\|G\|_{l_2}] \\ &= O(K)O(p)K^{1/2}O(\frac{r_n}{p}) = O(r_n K^{3/2}), \end{aligned} \tag{A.100}$$

where we use definition of L for the first equality in (A.84), G is defined in (A.83), and we use submultiplicativity of l_∞ norm for the first inequality, and the relation between spectral norm and l_∞ norm from p.365 of Horn and Johnson (2013) for the second inequality, and the rates are from (A.67), Lemma A.10, Lemma A.11 and G definition. Next we need the following, by using the same analysis in (B.55) of Caner and Kock (2018) via strict stationary of the data, or (A.12) here

$$\|\Omega\|_{l_\infty} := \max_{1\leq j\leq p}\|\Omega_j'\|_1 = \max_{1\leq j\leq p}\|\Omega_j\|_1 = O(\sqrt{\bar{s}}). \tag{A.101}$$

We consider each term on the right side of (A.99).

$$\begin{aligned} max_{1\leq j\leq p}\|(\hat{\Omega}_j-\Omega_j)'(\hat{L}-L)(\hat{\Omega}-\Omega)\|_1 &\leq [\max_{1\leq j\leq p}\|\hat{\Omega}_j-\Omega_j\|_1^2]\|\hat{L}-L\|_{l_\infty} & \text{(A.102)} \\ &= [O_p(\bar{s}\sqrt{lnp/n})]^2 \\ &\times O_p(l_n), & \text{(A.103)} \end{aligned}$$

where we use Lemma A.9(iii), and

$$\|\hat{\Omega}-\Omega\|_{l_\infty} := \max_{1\leq j\leq p}\|\hat{\Omega}_j'-\Omega_j'\|_1 = \max_{1\leq j\leq p}\|\hat{\Omega}_j-\Omega_j\|_1, \tag{A.104}$$

for the inequality in (A.102) and use Lemma A.13, and Theorem 1 for the rates.

We consider the second term on the right side of (A.99).

$$\begin{aligned} max_{1\leq j\leq p}\|(\hat{\Omega}_j-\Omega_j)'L(\hat{\Omega}-\Omega)\|_1 &\leq [\max_{1\leq j\leq p}\|\hat{\Omega}_j-\Omega_j\|_1^2]\|L\|_{l_\infty} \\ &= [O_p(\bar{s}\sqrt{lnp/n})]^2 \\ &\times O_p(r_n K^{3/2}), \end{aligned} \tag{A.105}$$

where we use Lemma A.9(iii), and (A.104) for the inequality in (A.105) and use (A.100), and Theorem 1 for the rates. We analyze the third term on the right side of (A.99)

$$\begin{aligned} \max_{1\leq j\leq p}\|(\hat{\Omega}_j-\Omega_j)'(\hat{L}-L)\Omega\|_1 &\leq [\max_{1\leq j\leq p}\|\hat{\Omega}_j-\Omega_j\|_1]\,\|\hat{L}-L\|_{l_\infty}\,\|\Omega\|_{l_\infty} \\ &= O_p(\bar{s}\sqrt{lnp/n}) \\ &\times O_p(l_n)O(\bar{s}^{1/2}), \end{aligned} \tag{A.106}$$

where we use Lemma A.9(iii) for the first inequality, and for the rates, first see that by definition and since l_1 norm of a row vector is the same as l_1 norm of transpose of the row vector, (which is the column version of the row vector, not the column of the same matrix) $\|\Omega\|_{l_\infty} := \max_{1\leq j\leq p}\|\Omega_j'\|_1 = \max_{1\leq j\leq p}\|\Omega_j\|_1$, and the rates are by (A.101), Lemma A.13, Theorem 1. Now consider the fourth term on the right side of (A.99)

$$\begin{aligned} \max_{1\leq j\leq p}\|(\hat{\Omega}_j-\Omega_j)'L\Omega\|_1 &\leq \max_{1\leq j\leq p}\|\hat{\Omega}_j-\Omega_j\|_1\|L\|_{l_\infty}\|\Omega\|_{l_\infty} \\ &= O_p(\bar{s}\sqrt{lnp/n})O_p(r_n K^{3/2})O(\bar{s}^{1/2}), \end{aligned} \tag{A.107}$$

where we use Lemma A.9(iii) for the inequality, and Theorem 1, (A.100)(A.101) for the rate. Now consider the fifth term on the right side of (A.99).

$$\begin{aligned}\max_{1\le j\le p}\|\Omega_j'(\hat{L}-L)(\hat{\Omega}-\Omega)\|_1 &\le [\max_{1\le j\le p}\|\Omega_j\|_1]\|\hat{L}-L\|_{l_\infty}\|\hat{\Omega}-\Omega\|_{l_\infty}\\ &= O(\bar{s}^{1/2})O_p(l_n)\\ &\times O_p(\bar{s}\sqrt{lnp/n}),\end{aligned} \tag{A.108}$$

where we use lemma A.9(iii) for the first inequality, and Theorem 1, Lemma A.13, (A.101) for the rates. Consider the sixth term on the right side of (A.99)

$$\begin{aligned}\max_{1\le j\le p}\|\Omega_j'(\hat{L}-L)\Omega\|_1 &\le [\max_{1\le j\le p}\|\Omega_j\|_1^2]\|\hat{L}-L\|_{l_\infty}\\ &= O(\bar{s})O_p(l_n),\end{aligned} \tag{A.109}$$

where we use Lemma A.9(iii) for the inequality, and use (A.101), and Lemma A.13 for the rates. Now analyze the seventh term on the right side of (A.99)

$$\begin{aligned}\max_{1\le j\le p}\|\Omega_j'L(\hat{\Omega}-\Omega)\|_1 &\le [\max_{1\le j\le p}\|\Omega_j\|_1]\|L\|_{l_\infty}\|\hat{\Omega}-\Omega\|_{l_\infty}\\ &= O(\bar{s}^{1/2})O_p(r_nK^{3/2})O_p(\bar{s}\sqrt{lnp/n}),\end{aligned} \tag{A.110}$$

where we use Lemma A.9(iii) for the inequality, and for the rates we use (A.100)(A.101) Theorem 1. Note that among all (A.103)-(A.110), the slowest rate is by (A.109) by the definition of l_n in (A.85) and by Assumption 7.

So we have by Assumption 7

$$\max_{1\le j\le p}\|\hat{\Omega}_j'\hat{L}\hat{\Omega}-\Omega_j'L\Omega\|_1 = O_p(\bar{s}l_n) = o_p(1). \tag{A.111}$$

This ends the proof of (i) with using Theorem 1 and (A.111) in (A.93).

(ii). Since

$$y_t = Bf_t + u_t,$$

as in Fan et al. (2011) with y_t, u_t being $p\times 1$ vector of asset returns, and errors respectively at time $t = 1,\cdots,n$.

$$\hat{\mu}-\mu = B[\frac{1}{n}\sum_{t=1}^{n}(f_t - Ef_t)] + \frac{1}{n}\sum_{t=1}^{n}u_t,$$

by Assumption 1. Consider

$$\begin{aligned}\|\hat{\mu}-\mu\|_\infty &\le \|B(\frac{1}{n}\sum_{t=1}^{n}(f_t-Ef_t))\|_\infty + \|\frac{1}{n}\sum_{t=1}^{n}u_t\|_\infty\\ &\le \|B\|_{l_\infty}\|\frac{1}{n}\sum_{t=1}^{n}f_t - Ef_t\|_\infty + \|\frac{1}{n}\sum_{t=1}^{n}u_t\|_\infty\\ &= O(K)O_p(\sqrt{lnn/n}) + O_p(\sqrt{lnp/n}),\end{aligned}$$

Clearly by the proof of Lemma A.1(i) here we have $\|Ax\|_\infty \le \|A\|_{l_\infty}\|x\|_\infty$ for a generic vector x, and a matrix A. Then by Lemma A.10(iii) and Theorem A.1 we get the rate.

Q.E.D.

PART 4:

Proof of Theorem 3. (A.2) of Ao et al. (2019) shows that the squared ratio of the estimated maximum out-of-sample Sharpe ratio to the theoretical ratio can be written as

$$[\frac{\widehat{SR}_{mosnw}}{SR^*}]^2 = \frac{\frac{(\mu'\hat{\Gamma}\hat{\mu})^2}{\hat{\mu}'\hat{\Gamma}'\Sigma_y\hat{\Gamma}\hat{\mu}}}{\mu'\Gamma\mu} = \frac{\left[\frac{\mu'\hat{\Gamma}\hat{\mu}}{\mu'\Gamma\mu}\right]^2}{\left[\frac{\hat{\mu}'\hat{\Gamma}'\Sigma_y\hat{\Gamma}\hat{\mu}}{\mu'\Gamma\mu}\right]}. \tag{A.112}$$

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

$$\frac{\mu'\hat{\Gamma}\hat{\mu}}{\mu'\Gamma\mu} = \frac{\mu'\hat{\Gamma}\hat{\mu} - \mu'\Gamma\mu}{\mu'\Gamma\mu} + 1. \tag{A.113}$$

Consider the fraction on the right-hand side. Start with the numerator in (A.113).

$$\begin{aligned}
|\mu'\hat{\Gamma}\hat{\mu} - \mu'\Gamma\mu|/p &= |\mu'\hat{\Gamma}\hat{\mu} - \mu'\Gamma\hat{\mu} + \mu'\Gamma\hat{\mu} - \mu'\Gamma\mu|/p \\
&\leq |\mu'(\hat{\Gamma} - \Gamma)\hat{\mu}|/p + |\mu'\Gamma(\hat{\mu} - \mu)|/p \\
&\leq |\mu'(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu)|/p + |\mu'(\hat{\Gamma} - \Gamma)\mu|/p + |\mu'\Gamma(\hat{\mu} - \mu)|/p \\
&= O_p(K\bar{s}l_n \max(K\sqrt{ln/n}, \sqrt{lnp/n})) + O_p(K^2\bar{s}l_n) \\
&+ O_p(K^{5/2}\bar{s}r_n max(K\sqrt{ln/n}, \sqrt{lnp/n})) \\
&= O_p(K^2\bar{s}l_n),
\end{aligned} \tag{A.114}$$

where we use (SA.18)(SA.19)(SA.20) for the rates and the dominant rate in the last equality is by Assumption 8 and l_n definition (35).
Next, we analyze the denominator in (A.113). First see that

$$\frac{\mu'\Gamma\mu}{p} \geq \frac{Eigmin(\Gamma)\|\mu\|_2^2}{p}. \tag{A.115}$$

Then by Assumption 6

$$\|\mu\|_2^2 = \|BEf_t\|_2^2 \geq \|Ef_t\|_2^2 Eigmin(B'B) \geq \|Ef_t\|_2^2 cp. \tag{A.116}$$

By Assumption 4

$$\|Ef_t\|_2^2 = \sum_{k=1}^{K} Ef_{kt}^2 \geq Kc > 0. \tag{A.117}$$

Combine (A.116)(A.117)

$$\|\mu\|_2^2/p \geq Kc > 0. \tag{A.118}$$

Since $\Gamma := \Sigma_y^{-1}$ by definition, and via Assumption 8

$$Eigmin(\Gamma) := Eigmin(\Sigma_y^{-1}) = \frac{1}{Eigmax(\Sigma_y)} \geq \frac{1}{CK}. \tag{A.119}$$

By combining (A.118) with (A.119) in (A.115)(A.116)

$$\frac{\mu'\Gamma\mu}{p} \geq \frac{c}{C} > 0. \tag{A.120}$$

Then, by (A.114)(A.120) in (A.113)

$$\frac{\mu'\hat{\Gamma}\hat{\mu}/p}{\mu'\Gamma\mu/p} \leq \frac{|\mu'\hat{\Gamma}\hat{\mu} - \mu'\Gamma\mu|/p}{\mu'\Gamma\mu/p} + 1 = O_p(K^2\bar{s}l_n) + 1. \tag{A.121}$$

We now attempt to show that the denominator in (A.112)

$$\frac{\hat{\mu}'\hat{\Gamma}\Sigma_y\hat{\Gamma}\hat{\mu}}{\mu'\Sigma_y^{-1}\mu} \xrightarrow{p} 1. \tag{A.122}$$

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

$$\frac{\hat{\mu}'\hat{\Gamma}'\Sigma_y\hat{\Gamma}\hat{\mu}}{\mu'\Sigma_y^{-1}\mu} = \frac{\hat{\mu}'\hat{\Gamma}'\Sigma_y\hat{\Gamma}\hat{\mu} - \mu'\Gamma\Sigma_y\Gamma\mu}{\mu'\Gamma\Sigma_y\Gamma\mu} + 1 \geq 1 - \left|\frac{\hat{\mu}'\hat{\Gamma}'\Sigma_y\hat{\Gamma}\hat{\mu} - \mu'\Gamma\Sigma_y\Gamma\mu}{\mu'\Gamma\Sigma_y\Gamma\mu}\right|. \tag{A.123}$$

We can write

$$\hat{\Gamma}\hat{\mu} - \Gamma\mu = (\hat{\Gamma} - \Gamma)\hat{\mu} + \Gamma(\hat{\mu} - \mu). \tag{A.124}$$

Using (A.124)

$$\begin{aligned} |\hat{\mu}'\hat{\Gamma}'\Sigma_y\hat{\Gamma}\hat{\mu} - \mu'\Gamma\Sigma_y\Gamma\mu| &= |[(\hat{\mu}\hat{\Gamma} - \mu\Gamma) + \mu\Gamma]'\Sigma_y[(\hat{\mu}\hat{\Gamma} - \mu\Gamma) + \mu\Gamma] - \mu'\Gamma\Sigma_y\Gamma\mu| \\ &\leq |[(\hat{\Gamma} - \Gamma)\hat{\mu}]'\Sigma_y[(\hat{\Gamma} - \Gamma)\hat{\mu}]| && \text{(A.125)} \\ &+ 2|[(\hat{\Gamma} - \Gamma)\hat{\mu}]'\Sigma_y\Gamma(\hat{\mu} - \mu)| && \text{(A.126)} \\ &+ 2|[(\hat{\Gamma} - \Gamma)\hat{\mu}]'\Sigma_y\Gamma\mu| && \text{(A.127)} \\ &+ |[\Gamma(\hat{\mu} - \mu)]'\Sigma_y[\Gamma(\hat{\mu} - \mu)]| && \text{(A.128)} \\ &+ 2|[\Gamma(\hat{\mu} - \mu)]'\Sigma_y\Gamma\mu| && \text{(A.129)} \end{aligned}$$

First, we consider (A.125).

$$\begin{aligned} |\hat{\mu}'(\hat{\Gamma} - \Gamma)'\Sigma_y(\hat{\Gamma} - \Gamma)\hat{\mu}| &\leq Eigmax(\Sigma_y)\|(\hat{\Gamma} - \Gamma)\hat{\mu}\|_2^2 \\ &= Eigmax(\Sigma_y)[\sum_{j=1}^{p}\{(\hat{\Gamma}_j - \Gamma_j)'\hat{\mu}\}^2] \\ &\leq Eigmax(\Sigma_y)p\max_{1\leq j\leq p}[(\hat{\Gamma}_j - \Gamma_j)'\hat{\mu}]^2 \\ &\leq Eigmax(\Sigma_y)p(\max_{1\leq j\leq p}\|\hat{\Gamma}_j - \Gamma_j\|_1)^2\|\hat{\mu}\|_\infty^2 \\ &= pO(K)O_p(\bar{s}^2l_n^2)O_p(K^2), && \text{(A.130)} \end{aligned}$$

where we use Hölder's inequality for the third inequality and Theorem 2 and Assumption 8, (SA.8) for the rate. Now, consider (A.126), and by definition $\Gamma := \Sigma_y^{-1}$.

$$\begin{aligned} |[(\hat{\Gamma} - \Gamma)\hat{\mu}]'\Sigma_y\Gamma(\hat{\mu} - \mu)| &= |(\hat{\mu} - \mu)'(\hat{\Gamma} - \Gamma)\hat{\mu}| \\ &\leq |(\hat{\mu} - \mu)'(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu)| + |(\hat{\mu} - \mu)(\hat{\Gamma} - \Gamma)\mu| \\ &= p[O_p(max(K\sqrt{lnn/n}, \sqrt{lnp/n}))]^2O_p(\bar{s}l_n)] \\ &+ [pO(K)O_p(\bar{s}l_n)O_p(max(K\sqrt{lnn/n}, \sqrt{lnp/n}))] \\ &= pO_p(K\bar{s}l_n max(K\sqrt{lnn/n}, \sqrt{lnp/n})), && \text{(A.131)} \end{aligned}$$

by (SA.16)(SA.19) for the second equality, and the dominant rate in third equality can be seen from Assumption 8. Next, consider (A.127), and recall that $\Gamma := \Sigma_y^{-1}$

$$\begin{aligned} |[(\hat{\Gamma} - \Gamma)\hat{\mu}]'\Sigma_y \Gamma \mu| &= |\mu'(\hat{\Gamma} - \Gamma)\hat{\mu}| \\ &\leq |\mu'(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu)| + |\mu'(\hat{\Gamma} - \Gamma)\mu| \\ &= p[O(K)O_p(\bar{s}l_n)O_p(max(K\sqrt{lnn/n}, \sqrt{lnp/n})) + O_p(K^2\bar{s}l_n)] \\ &= pO_p(K^2\bar{s}l_n), \end{aligned} \quad \text{(A.132)}$$

where we use (SA.19)(SA.20) for the second equality, and the dominant rate in the third equality can be seen from Assumption 8. Consider now (A.128) by the symmetry of $\Gamma = \Sigma_y^{-1}$

$$\begin{aligned} |[\Gamma(\hat{\mu} - \mu)]'\Sigma_y \Gamma(\hat{\mu} - \mu)| &= |(\hat{\mu} - \mu)'\Gamma(\hat{\mu} - \mu)| \\ &= p[O_p(\max(K\sqrt{ln/n}, \sqrt{lnp/n}))]^2 O(\bar{s}r_n K^{3/2}) \end{aligned} \quad \text{(A.133)}$$

by (SA.17). Next, analyze (A.129) by the symmetricity of $\Gamma = \Sigma_y^{-1}$

$$\begin{aligned} |[\Gamma(\hat{\mu} - \mu)]'\Sigma_y \Gamma \mu| &= |(\hat{\mu} - \mu)'\Gamma\mu| \\ &= pO_p(\bar{s}r_n K^{5/2} \max(K\sqrt{lnn/n}, \sqrt{lnp/n})), \end{aligned} \quad \text{(A.134)}$$

by (SA.18). Combine the rates and terms (A.130)-(A.134) in (A.125)-(A.129) to obtain

$$|\hat{\mu}'\hat{\Gamma}'\Sigma_y\hat{\Gamma}\hat{\mu} - \mu'\Gamma\Sigma_y\Gamma\mu| = pO_p(K^2\bar{s}l_n), \quad \text{(A.135)}$$

by the dominant rate in (A.132), as seen in Assumption 8 $K^2\bar{s}l_n \to 0$, and l_n definition in Assumption 7.

See that by $\Gamma = \Sigma_y^{-1}$, by (A.120)

$$\frac{\mu'\Gamma\Sigma_y\Gamma\mu}{p} = \frac{\mu'\Gamma\mu}{p} \geq \frac{c}{C} > 0. \quad \text{(A.136)}$$

Combine (A.135)(A.136), in the second right side term in (A.123) via Assumption 8

$$\frac{|\hat{\mu}'\hat{\Gamma}'\Sigma_y\hat{\Gamma}\hat{\mu} - \mu'\Gamma\Sigma_y\Gamma\mu|/p}{\mu'\Gamma\Sigma_y\Gamma\mu/p} \leq O_p(K^2\bar{s}l_n) = o_p(1). \quad \text{(A.137)}$$

Therefore, we show (A.122) via (A.123). Then, combine (A.121)(A.122) in (A.112) to obtain the desired result. **Q.E.D.**

Proof of Theorem 4. (i). Start with definition of weights, and its estimates

$$\frac{\left(\frac{\sigma\mu'\hat{\Gamma}\hat{\mu}}{\sqrt{\hat{\mu}'\hat{\Gamma}\hat{\mu}}}\right)}{\left(\frac{\sigma\mu'\Gamma\mu}{\sqrt{\mu'\Gamma\mu}}\right)} - 1 = \left[\frac{\mu'\hat{\Gamma}\hat{\mu}}{\mu'\Gamma\mu}\left(\frac{\mu'\Gamma\mu}{\hat{\mu}'\hat{\Gamma}\hat{\mu}}\right)^{1/2}\right] - 1 \quad \text{(A.138)}$$

See that

$$
\begin{aligned}
\left(\frac{\mu'\hat{\Gamma}\hat{\mu}}{\mu'\Gamma\mu}\right) & \times \left(\frac{\mu'\Gamma\mu}{\hat{\mu}'\hat{\Gamma}\hat{\mu}}\right)^{1/2} - 1 \\
& \leq \left[\left|\frac{\mu'\hat{\Gamma}\hat{\mu}}{\mu'\Gamma\mu} - 1\right| + 1\right]\left[\left|\left(\frac{\mu'\Gamma\mu}{\hat{\mu}'\hat{\Gamma}\hat{\mu}}\right)^{1/2} - 1\right| + 1\right] - 1 \\
& = \left|\frac{\mu'\hat{\Gamma}\hat{\mu}}{\mu'\Gamma\mu} - 1\right|\left|\left(\frac{\mu'\Gamma\mu}{\hat{\mu}'\hat{\Gamma}\hat{\mu}}\right)^{1/2} - 1\right| \\
& + \left|\frac{\mu'\hat{\Gamma}\hat{\mu}}{\mu'\Gamma\mu} - 1\right| + \left|\left(\frac{\mu'\Gamma\mu}{\hat{\mu}'\hat{\Gamma}\hat{\mu}}\right)^{1/2} - 1\right| \qquad (A.139)
\end{aligned}
$$

By (A.121)

$$
\left|\frac{\mu'\hat{\Gamma}\hat{\mu}}{\mu'\Gamma\mu} - 1\right| = O_p(K^2\bar{s}l_n). \qquad (A.140)
$$

Next, we have

$$
\begin{aligned}
\frac{\mu'\Gamma\mu}{\hat{\mu}\hat{\Gamma}\hat{\mu}} & = \frac{\mu'\Gamma\mu - \hat{\mu}'\hat{\Gamma}\hat{\mu}}{\hat{\mu}'\hat{\Gamma}\hat{\mu}} + 1 \\
& \leq \frac{|\mu'\Gamma\mu/p - \hat{\mu}'\hat{\Gamma}\hat{\mu}/p|}{\mu'\Gamma\mu/p - |\hat{\mu}'\hat{\Gamma}\hat{\mu}/p - \mu'\Gamma\mu/p|} + 1, \qquad (A.141)
\end{aligned}
$$

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

$$
\hat{\mu}'\hat{\Gamma}\hat{\mu}/p \geq \mu'\Gamma\mu/p - |\hat{\mu}'\hat{\Gamma}\hat{\mu}/p - \mu'\Gamma\mu/p|.
$$

By (A.120),(A.141), Lemma SA.4 in the Supplementary Appendix, simplifying the ratio of two positive constants as $c/C = c$ and $K^2\bar{s}l_n = o(1)$ via Assumption 8 in the denominator below in (A.142)

$$
\frac{\mu'\Gamma\mu}{\hat{\mu}'\hat{\Gamma}\hat{\mu}} \leq \frac{O_p(K^2\bar{s}l_n)}{c - O_p(K^2\bar{s}l_n)} + 1 = O_p(K^2\bar{s}l_n) + 1. \qquad (A.142)
$$

Then,

$$
\left|\left(\frac{\mu'\Gamma\mu}{\hat{\mu}'\hat{\Gamma}\hat{\mu}}\right)^{1/2} - 1\right| = \{[1 + O_p(K^2\bar{s}l_n)]^{1/2} - 1\} \qquad (A.143)
$$

Now, use Assumption 8 in (A.140)(A.143) and (A.139) to obtain the desired result.

. **Q.E. D**

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

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

where we multiplied and divided by $\mu'\Gamma\mu$, which is positive by (A.120). By (A.122)(A.137), since $\Gamma := \Sigma_y^{-1}$

$$
\left|\frac{\hat{\mu}'\hat{\Gamma}'\Sigma_y\hat{\Gamma}\hat{\mu}}{\mu'\Gamma\mu} - 1\right| = O_p(K^2\bar{s}l_n). \qquad (A.144)
$$

Additionally, by Lemma SA.4 in the Supplementary Appendix and (A.120)

$$|\frac{\hat{\mu}'\hat{\Gamma}\hat{\mu}/p}{\mu'\Gamma\mu/p} - 1| = o_p(1). \tag{A.145}$$

By (A.144)(A.145) and Assumption 8,

$$|\hat{w}_{oos}\Sigma_y\hat{w}_{oos} - \sigma^2| = O_p(K^2\bar{s}l_n) = o_p(1).$$

Q.E.D.

Proof of Theorem 5. See that, since $\Gamma := \Sigma_y^{-1}$,

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

Lemma SA.4 in the Supplementary Appendix shows that

$$|\hat{\mu}'\hat{\Gamma}\hat{\mu}/p - \mu'\Sigma_y^{-1}\mu/p| = O_p(K^2\bar{s}l_n). \tag{A.146}$$

Combining (A.120),(A.146) with Assumption 8,

$$\left|\frac{\hat{\mu}'\hat{\Gamma}\hat{\mu}/p}{\mu'\Sigma_y^{-1}\mu/p} - 1\right| = O_p(K^2\bar{s}l_n) = o_p(1).$$

Q.E.D.

Proof of Theorem 6. Note that by the definition of MSR_c in (40) and A, F, D terms,

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

and the estimate is

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

where $\hat{A} = 1_p'\hat{\Gamma}1_p/p$, $\hat{F} = 1_p'\hat{\Gamma}\hat{\mu}/p$, $\hat{D} = \hat{\mu}'\hat{\Gamma}\hat{\mu}/p$.
Then, clearly

$$\frac{\frac{\widehat{MSR}_c^2}{p}}{\frac{MSR_c^2}{p}} = \left[\frac{\hat{A}\hat{D} - \hat{F}^2}{AD - F^2}\right]\left[\frac{A}{\hat{A}}\right]. \tag{A.147}$$

We start with

$$|\hat{A} - A| = O_p(\bar{s}l_n) = o_p(1), \tag{A.148}$$

by Lemma SA.2 in the Supplementary Appendix. Then by $\Gamma := \Sigma_y^{-1}$ with Assumption 8

$$A \geq Eigmin(\Gamma) = \frac{1}{Eigmax(\Sigma_y)} \geq \frac{1}{CK}. \tag{A.149}$$

Thus, clearly we obtain, since $|\hat{A}| \geq A - |\hat{A} - A|$, by multiplying and dividing by K

$$\left|\frac{A}{\hat{A}} - 1\right| = \left|\frac{A - \hat{A}}{\hat{A}}\right| \leq \frac{K|\hat{A} - A|}{K[A - |\hat{A} - A|]} = \frac{O_p(K\bar{s}l_n)}{1/C - O_p(K\bar{s}l_n)}$$

which implies with $1/C > 0$, and $K\bar{s}l_n = o(1)$ for the denominator rate

$$|\frac{A}{\hat{A}} - 1| = O_p(K\bar{s}l_n) = o_p(1). \tag{A.150}$$

Next, Lemma SA.6 in the Supplementary Appendix establishes that

$$|(\hat{A}\hat{D} - \hat{F}^2) - (AD - F^2)| = O_p(K^2 \bar{s} l_n) = o_p(1).$$

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

$$\left| \frac{\hat{A}\hat{D} - \hat{F}^2}{AD - F^2} - 1 \right| = O_p(K^2 \bar{s} l_n) = o_p(1). \tag{A.151}$$

Since

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

Combine (A.150)(A.151) in (A.147) to obtain

$$\begin{aligned} \left| \frac{\widehat{MSR}_c^2/p}{MSR_c^2/p} - 1 \right| &\leq \left| \frac{\hat{A}\hat{D} - \hat{F}^2}{AD - F^2} - 1 \right| \left| \frac{A}{\hat{A}} - 1 \right| \\ &+ \left| \frac{A}{\hat{A}} - 1 \right| + \left| \frac{\hat{A}\hat{D} - \hat{F}^2}{AD - F^2} - 1 \right| && \text{(A.152)} \\ &= O_p(K^2 \bar{s} l_n) = o_p(1), && \text{(A.153)} \end{aligned}$$

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

Proof of Theorem 7. Note that we define $\Gamma : \Sigma_y^{-1}$. We need to start with

$$\left| \frac{(\widehat{MSR}^*)^2/p}{(MSR^*)^2/p} - 1 \right| = \frac{\left| (\widehat{MSR}^*)^2/p - (MSR^*)^2/p \right|}{(MSR^*)^2/p} \tag{A.154}$$

Define the event $E_1 = \{|1_p' \hat{\Gamma} \hat{\mu}/p - 1_p' \Sigma_y^{-1} \mu/p| \leq \epsilon\}$, where $\epsilon > 0$. We condition the proof on event E_1, then at the end of the proof we show that $P(E_1) \to 1$.

Start with the condition $1_p' \Sigma_y^{-1} \mu/p \geq C > 2\epsilon > 0$;

$$\begin{aligned} \frac{1_p' \hat{\Gamma} \hat{\mu}}{p} &= \frac{1_p' \hat{\Gamma} \hat{\mu}}{p} - \frac{1_p' \Sigma_y^{-1} \mu}{p} + \frac{1_p' \Sigma_y^{-1} \mu}{p} \\ &\geq \frac{1_p' \Sigma_y^{-1} \mu}{p} - \left| \frac{1_p' \hat{\Gamma} \hat{\mu}}{p} - \frac{1_p' \Sigma_y^{-1} \mu}{p} \right| \\ &\geq \frac{1_p' \Sigma_y^{-1} \mu}{p} - \epsilon \\ &\geq C - \epsilon > 2\epsilon - \epsilon = \epsilon > 0, && \text{(A.155)} \end{aligned}$$

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

$$p^{-1} [\widehat{MSR}^2 1_{\{1_p' \hat{\Gamma} \hat{\mu} > 0\}} - MSR^2 1_{\{1_p' \Gamma \mu \geq 0\}}] = p^{-1} [\widehat{MSR}^2 - MSR^2], \tag{A.156}$$

as used in the maximum Sharpe-ratios of Theorem 5.

We consider $1_p'\Sigma_y^{-1}\mu/p \le -C < -2\epsilon < 0$. Assume that we use event E_1:

$$\begin{aligned}
\frac{1_p'\Sigma_y^{-1}\mu}{p} &= \frac{1_p'\Sigma_y^{-1}\mu}{p} - \frac{1_p'\hat{\Gamma}\hat{\mu}}{p} + \frac{1_p'\hat{\Gamma}\hat{\mu}}{p} \\
&\ge \frac{1_p'\hat{\Gamma}\hat{\mu}}{p} - |\frac{1_p'\Sigma_y^{-1}\mu}{p} - \frac{1_p'\hat{\Gamma}\hat{\mu}}{p}| \\
&\ge \frac{1_p'\hat{\Gamma}\hat{\mu}}{p} - \epsilon. \qquad \text{(A.157)}
\end{aligned}$$

Then, in (A.157), using the condition $1_p'\Sigma_y^{-1}\mu/p \le -C < -2\epsilon < 0$ (note that this also implies $1_p'\Sigma_y^{-1}\mu/p < 0$)

$$0 > -2\epsilon > -C \ge 1_p'\Sigma_y^{-1}\mu/p \ge 1_p'\hat{\Gamma}\hat{\mu}/p - \epsilon,$$

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

$$-\epsilon > -(C-\epsilon) \ge 1_p'\hat{\Gamma}\hat{\mu}/p,$$

which clearly shows that when $1_p'\Sigma_y^{-1}\mu/p < 0$, we will have $1_p'\hat{\Gamma}\hat{\mu}/p < 0$, since $-\epsilon < 0$.

So

$$p^{-1}[\widehat{MSR}_c^2 1_{\{1_p'\hat{\Gamma}\hat{\mu}<0\}} - MSR_c^2 1_{1_p'\Gamma\mu<0\}} = p^{-1}[\widehat{MSR}_c^2 - MSR_c^2], \qquad \text{(A.158)}$$

as in maximum Sharpe ratios in Theorem 6. Clearly under event E_1 with $1_p'\Gamma\mu/p \ge C > 2\epsilon > 0$, (A.154) is rewritten as

$$\left| \frac{(\widehat{MSR}^2 - MSR^2)/p}{MSR^2/p} \right| = O_p(K^2 \bar{s} l_n), \qquad \text{(A.159)}$$

where we use Theorem 5. Under event E_1, with $1_p'\Gamma\mu/p \le -C < -2\epsilon < 0$, (A.154) is rewritten as

$$\left| \frac{(\widehat{MSR}_c^2 - MSR_c^2)/p}{MSR_c^2/p} \right| = O_p(K^2 \bar{s} l_n), \qquad \text{(A.160)}$$

where we use Theorem 6.

Note that we can rewrite the event $E_1 := \{|\hat{F} - F\| \le \epsilon\}$, with $\epsilon = O(K\bar{s}l_n)$. Note that event E_1 occurs with probability approaching one by Lemma SA.3 in the Supplementary Appendix, so we have proven the desired result. **Q.E.D.**

Q.E.D.

Proof of Theorem 8. First, we start with definitions of $\hat{A} := 1_p'\hat{\Gamma}1_p/p$, $\hat{F} := 1_p'\hat{\Gamma}\hat{\mu}/p$, $A := 1_p'\Sigma_y^{-1}1_p/p$, $F := 1_p'\Sigma_y^{-1}\mu/p$.

$$\begin{aligned}
\left| \frac{\widehat{SR}_{nw}^2}{SR^2} - 1 \right| &= \left| \frac{p(1_p'\hat{\Gamma}\hat{\mu}/p)^2(1_p'\hat{\Gamma}1_p/p)^{-1}}{p(1_p'\Sigma_y^{-1}\mu/p)^2(1_p'\Sigma_y^{-1}1_p/p)^{-1}} - 1 \right| \\
&= \left| \frac{\hat{F}^2 A}{F^2 \hat{A}} - 1 \right| \\
&= \left| \frac{K(\hat{F}^2 A - F^2 \hat{A})}{K F^2 \hat{A}} \right|, \qquad \text{(A.161)}
\end{aligned}$$

where in the last step we divided and multiplied right side by K.

Now consider the numerator in (A.161):

$$\begin{aligned} |\hat{F}^2 A - F^2 \hat{A}| &= |\hat{F}^2 A - \hat{F}^2 \hat{A} + \hat{F}^2 \hat{A} - F^2 \hat{A}| \\ &\leq |\hat{F}^2(\hat{A} - A)| + |(\hat{F}^2 - F^2)\hat{A}| \\ &\leq |\hat{F}^2(\hat{A} - A)| + |\hat{F} - F||\hat{F} + F||\hat{A}|. \end{aligned} \tag{A.162}$$

Analyze the first term on the right side of (A.162):

$$\begin{aligned} \hat{F}^2 &= |\hat{F}^2 - F^2 + F^2| \\ &\leq |\hat{F}^2 - F^2| + F^2 \\ &\leq |\hat{F} - F||\hat{F} + F| + F^2. \end{aligned} \tag{A.163}$$

Then, by Lemma SA.3 in the Supplementary Appendix, via Assumption 8

$$|\hat{F} - F| = O_p(K\bar{s}l_n) = o_p(1). \tag{A.164}$$

Then,

$$\begin{aligned} |\hat{F} + F| &\leq |\hat{F}| + |F| \\ &\leq |\hat{F} - F| + 2|F| \\ &= o_p(1) + O(K^{1/2}) \\ &= O_p(K^{1/2}), \end{aligned} \tag{A.165}$$

where we use (A.164) and Lemma SA.5 in the Supplementary Appendix.
By (A.164)(A.165) and Lemma SA.5 in (A.163), we have

$$\hat{F}^2 = O_p(K). \tag{A.166}$$

Then, by Lemma SA.2 in the Supplementary Appendix and (A.166),

$$|\hat{F}^2(\hat{A} - A)| \leq \hat{F}^2|\hat{A} - A| = O_p(K\bar{s}l_n) = o_p(1). \tag{A.167}$$

Then, the second term on the right side of (A.162) is

$$|\hat{F} - F||\hat{F} + F||\hat{A}| = O_p(K\bar{s}l_n)O_p(K^{1/2})O_p(1) = O_p(K^{3/2}\bar{s}l_n) = o_p(1), \tag{A.168}$$

by (A.164)(A.165) and Lemma SA.2, Lemma SA.5 in the Supplementary Appendix, and the last equality is by Assumption 8. Use (A.167)(A.168) in (A.162) multiplied by K, with Assumption 8

$$K|\hat{F}^2 A - F^2 \hat{A}| = O_p(K^{5/2}\bar{s}l_n) = o_p(1). \tag{A.169}$$

Now consider the denominator in (A.161). Note that

$$KF^2\hat{A} = KF^2(\hat{A} - A) + KF^2 A \geq KF^2 A - K|F^2(\hat{A} - A)|.$$

So

$$KF^2 A \geq KF^2 Eigmin(\Gamma) = \frac{KF^2}{Eigmax(\Sigma_y)} \geq \frac{KF^2}{CK} \geq C > 0, \tag{A.170}$$

where the first inequality is by A definition, and then we use $\Gamma := \Sigma_y^{-1}$ in the second inequality, and for the third inequality we use Assumption 8, and the last inequality is by condition in the statement of Theorem: $F := 1_p' \Gamma \mu / p \geq C > 0$. Next

$$K|\hat{F}^2(\hat{A} - A)| = O_p(K^2 \bar{s} l_n) = o_p(1). \quad (A.171)$$

by (A.167) and Assumption 8. Combine (A.169) with (A.170)(A.171) in (A.161) to obtain the desired result. **Q.E.D.**

Proof of Theorem 9. To ease the notation in the proofs, set $AD - F^2 = z$, $A\rho_1^2 - 2F\rho_1 + D = v$. The estimates will be $\hat{z} = \hat{A}\hat{D} - \hat{F}^2$, $\hat{v} = \hat{A}\rho_1^2 - 2\hat{F}\rho_1 + \hat{D}$. Then,

$$\begin{aligned} \left| \frac{\widehat{SR}^2_{MV}}{SR^2_{MV}} - 1 \right| &= \left| \frac{\hat{z}/\hat{v}}{z/v} - 1 \right| \\ &= \left| \frac{\hat{z}}{\hat{v}} \frac{v}{z} - 1 \right| \\ &= \left| \frac{\hat{z}v - \hat{v}z}{\hat{v}z} \right|. \quad (A.172) \end{aligned}$$

First, analyze the denominator of (A.172).

$$\begin{aligned} |\hat{v}z| &= |(\hat{v} - v)z + vz|. \\ &\geq |vz| - |(\hat{v} - v)z| \\ &\geq |vz| - |\hat{v} - v||z|. \quad (A.173) \end{aligned}$$

Then, by Lemma SA.2-SA.4 in the Supplementary Appendix, triangle inequality and ρ_1 being bounded away from zero and finite, by Assumption 8,

$$|\hat{v} - v| = |(\hat{A} - A)\rho_1^2 - 2(\hat{F} - F)\rho_1 + (\hat{D} - D)| = O_p(K^2 \bar{s} l_n) = o_p(1). \quad (A.174)$$

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

$$|z| = |AD - F^2| \leq AD = O(K). \quad (A.175)$$

Thus, by (A.174)(A.175) and $z \geq C_1 > 0, v \geq C_1 > 0$ with Assumption 8: $K^3 \bar{s} l_n \to 0$ in (A.173), we have

$$|\hat{v}z| = o_p(1) + C_1^2 > 0. \quad (A.176)$$

Consider the numerator in (A.172):

$$|\hat{z}v - \hat{v}z| = |\hat{z}v - vz + vz - \hat{v}z| \leq |\hat{z} - z||v| + |z||\hat{v} - v|. \quad (A.177)$$

By Lemma SA.6 in the Supplementary Appendix, and Assumption 8

$$|\hat{z} - z| = |(\hat{A}\hat{D} - \hat{F}^2) - (AD - F^2)| = O_p(K^2 \bar{s} l_n) = o_p(1). \quad (A.178)$$

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

$$|v| = |A\rho_1 - 2F\rho_1 + D| = O(K). \quad (A.179)$$

Then, use (A.174)(A.175)(A.178)(A.179) in (A.177) by Assumption 8

$$|\hat{z}v - \hat{v}z| = O_p(K^3 \bar{s} l_n) = o_p(1). \quad (A.180)$$

Use (A.176)(A.180) in (A.172) to obtain the desired result. **Q.E.D.**

Supplementary Appendix

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

Lemma SA.1.

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

Proof of Lemma SA.1.

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

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

Recall the definition of $A := 1_p' \Gamma 1_p / p$ and $\hat{A} := 1_p' \hat{\Gamma} 1_p / p$, and $\bar{s} l_n$ is the rate of convergence in Theorem 2 in main text, and defined in Assumption 7 with the property $\bar{s} l_n \to 0$.

Lemma SA.2. *Under Assumptions 1-4, 6-7*

$$|\hat{A} - A| = O_p(\bar{s} l_n) = o_p(1).$$

Proof of Lemma SA.2.

$$\begin{aligned} |1_p'(\hat{\Gamma} - \Gamma) 1_p| / p &\leq \|(\hat{\Gamma} - \Gamma) 1_p\|_1 \|1_p\|_\infty / p \\ &\leq \max_{1 \leq j \leq p} \|\hat{\Gamma}_j - \Gamma_j\|_1 \\ &= O_p(\bar{s} l_n) = o_p(1), \end{aligned} \quad \text{(SA.2)}$$

where Hölder's inequality is used in the first inequality, Lemma SA.1 is used for the second inequality, and the last equality is obtained by using Theorem 2 and imposing Assumption 7.

Q.E.D.

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

Lemma SA.3. *Under Assumptions 1-4, 6, 7(i), 8*

$$|\hat{F} - F| = O_p(K \bar{s} l_n) = o_p(1).$$

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

$$\begin{aligned} \hat{F} - F &= [1_p'(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu)] / p & \text{(SA.3)} \\ &+ [1_p'(\hat{\Gamma} - \Gamma)\mu] / p & \text{(SA.4)} \\ &+ [1_p' \Gamma (\hat{\mu} - \mu)] / p & \text{(SA.5)} \end{aligned}$$

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

$$\begin{aligned} |1_p'(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu)|/p &\leq \|(\hat{\Gamma} - \Gamma)1_p\|_1 \|\hat{\mu} - \mu\|_\infty / p \\ &\leq [\max_{1 \leq j \leq p} \|\hat{\Gamma}_j - \Gamma_j\|_1] \|\hat{\mu} - \mu\|_\infty \\ &= O_p(\bar{s} l_n) O_p(max(K\sqrt{ln/n}, \sqrt{lnp/n})), \end{aligned} \quad \text{(SA.6)}$$

where we use Hölder's inequality in the first inequality and Lemma SA.1 with $M = \hat{\Gamma} - \Gamma$, $x = 1_p$ in the second inequality above, and the rate is from Theorem 2.

To get to the other terms, we need two extra results. Use the definition of Γ in (33)

$$\begin{aligned} \|\Gamma\|_{l_\infty} &\leq \|\Omega\|_{l_\infty} + \|\Omega L \Omega\|_{l_\infty} \\ &\leq \|\Omega\|_{l_\infty} + \|\Omega\|_{l_\infty}^2 \|L\|_{l_\infty} \\ &= O(\sqrt{\bar{s}}) + O(\bar{s}) O(r_n K^{3/2}) = O(\bar{s} r_n K^{3/2}), \end{aligned} \quad \text{(SA.7)}$$

where for the rates we use (A.100)(A.101) and Assumption 8 for the final rate. Note that $\mu = E y_t = B E f_t$, since $E u_t = 0$ by Assumption 1. So with $b_{j,k}$ representing j,k th element of $p \times k$: B matrix, and $E f_{t,k}$ representing k th element of $K \times 1$ vector $E f_t$

$$\begin{aligned} \|\mu\|_\infty &:= \max_{1 \leq j \leq p} |b_j' E f_t| = \max_{1 \leq j \leq p} |\sum_{k=1}^{K} b_{j,k} E f_{t,k}| \\ &\leq \max_{1 \leq j \leq p} \max_{1 \leq k \leq K} |b_{j,k}| [K \max_{1 \leq k \leq K} |E f_{t,k}|] \\ &= O(K), \end{aligned} \quad \text{(SA.8)}$$

where the rate is by Assumption 4, 6.

Therefore, we consider (SA.4) above.

$$|1_p'(\hat{\Gamma} - \Gamma)\mu|/p = O_p(\bar{s} l_n) O(K), \quad \text{(SA.9)}$$

where we use the same analysis that leads to (SA.6), and the rate is from Theorem 2, (SA.8).

Now consider (SA.5).

$$\begin{aligned} |1_p' \Gamma(\hat{\mu} - \mu)|/p &\leq \|\Gamma 1_p\|_1 \|\hat{\mu} - \mu\|_\infty / p \\ &\leq [\max_{1 \leq j \leq p} \|\Gamma_j\|_1] \|\hat{\mu} - \mu\|_\infty \\ &= O(\bar{s} r_n^{3/2} K) O_p(\max(K\sqrt{ln/n}, \sqrt{lnp/n})), \end{aligned} \quad \text{(SA.10)}$$

where we use Hölder's inequality in the first inequality and Lemma SA.1 with $M = \Gamma$, $x = 1_p$ in the second inequality above, and the rate is from Theorem 2, (SA.7).

Combine (SA.6)(SA.9)(SA.10) in (SA.3)-(SA.5), and note that the largest rate is coming from (SA.9) by $\bar{s} l_n$ definition in Assumption 7. **Q.E.D.**

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

Lemma SA.4. *Under Assumptions 1-4, 6, 7(i), 8*

$$|\hat{D} - D| = O_p(K^2 \bar{s} l_n) = o_p(1).$$

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

$$\hat{D} - D = [(\hat{\mu} - \mu)'(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu)]/p \quad \text{(SA.11)}$$
$$+ [(\hat{\mu} - \mu)'\Gamma(\hat{\mu} - \mu)]/p \quad \text{(SA.12)}$$
$$+ [2(\hat{\mu} - \mu)'\Gamma\mu]/p \quad \text{(SA.13)}$$
$$+ [2\mu'(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu)]/p \quad \text{(SA.14)}$$
$$+ [\mu'(\hat{\Gamma} - \Gamma)\mu]/p. \quad \text{(SA.15)}$$

Consider the first right side term above

$$\begin{aligned} |(\hat{\mu} - \mu)'(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu)|/p &\leq \|(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu)\|_1 \|\hat{\mu} - \mu\|_\infty / p \\ &\leq [\|\hat{\mu} - \mu\|_\infty]^2 [\max_j \|\hat{\Gamma}_j - \Gamma_j\|_1] \\ &= [O_p(max(K\sqrt{lnn/n}, \sqrt{lnp/n}))]^2 O_p(\bar{s}l_n) \end{aligned} \quad \text{(SA.16)}$$

where Hölder's inequality is used for the first inequality above, and the inequality Lemma SA.1, with $M = \hat{\Gamma} - \Gamma$, and $x = \hat{\mu} - \mu$ for the second inequality above, and for the rates we use Theorem 2.

We continue with (SA.12).

$$\begin{aligned} |(\hat{\mu} - \mu)'(\Gamma)(\hat{\mu} - \mu)|/p &\leq \|(\Gamma)(\hat{\mu} - \mu)\|_1 \|\hat{\mu} - \mu\|_\infty / p \\ &\leq [\|\hat{\mu} - \mu\|_\infty]^2 [\max_j \|\Gamma_j\|_1] \\ &= [O_p(max(K\sqrt{lnn/n}, \sqrt{lnp/n}))]^2 O(\bar{s}r_n K^{3/2}), \end{aligned} \quad \text{(SA.17)}$$

where Hölder's inequality is used for the first inequality above, and the inequality Lemma SA.1, with $M = \Gamma$, and $x = \hat{\mu} - \mu$ for the second inequality above, and for the rates, we use Theorem 2 and (SA.7).

Then, we consider (SA.13), using (SA.8)

$$\begin{aligned} |(\hat{\mu} - \mu)'(\Gamma)(\mu)|/p &\leq \|(\Gamma)(\hat{\mu} - \mu)\|_1 \|\mu\|_\infty / p \\ &\leq [\|\hat{\mu} - \mu\|_\infty][\max_j \|\Gamma_j\|_1] O(K) \\ &= O_p(\max(K\sqrt{lnn/n}, \sqrt{lnp/n})) O(\bar{s}r_n K^{3/2}) O(K), \end{aligned} \quad \text{(SA.18)}$$

where Hölder's inequality is used for the first inequality above, and the inequality Lemma SA.1, with $M = \Gamma$, and $x = \hat{\mu} - \mu$ for the second inequality above, and for the rates, we use Theorem 2 and (SA.8).

Then, we consider (SA.14).

$$\begin{aligned} |(\mu)'(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu)|/p &\leq \|(\hat{\Gamma} - \Gamma)(\mu)\|_1 \|\hat{\mu} - \mu\|_\infty / p \\ &\leq \|\mu\|_\infty \max_j \|\hat{\Gamma}_j - \Gamma_j\|_1 \|\hat{\mu} - \mu\|_\infty \\ &\leq [\max_j \|\hat{\Gamma}_j - \Gamma_j\|_1] \|(\hat{\mu} - \mu)\|_\infty O(K) \\ &= O_p(K\bar{s}l_n) O_p(max(K\sqrt{lnn/n}, \sqrt{lnp/n})), \end{aligned} \quad \text{(SA.19)}$$

where Hölder's inequality is used for the first inequality above, and the inequality Lemma SA.1, with $M = \hat{\Gamma} - \Gamma$, $x = \mu$ for the second inequality above, for the third inequality above, we use (SA.8), and for the rates, we use Theorem 2.

Then, we consider (SA.15):

$$\begin{aligned} |(\mu)'(\hat{\Gamma}-\Gamma)(\mu)|/p &\leq \|(\hat{\Gamma}-\Gamma)(\mu)\|_1 \|\mu\|_\infty /p \\ &\leq [\|\mu\|_\infty]^2 \max_j \|\hat{\Gamma}_j - \Gamma_j\|_1 \\ &\leq [\max_j \|\hat{\Gamma}_j - \Gamma_j\|_1] O(K^2) \\ &= O_p(K^2 \bar{s} l_n), \end{aligned} \quad \text{(SA.20)}$$

where Hölder's inequality is used for the first inequality above, and the inequality Lemma SA.1, with $M = \hat{\Gamma} - \Gamma$, $x = \mu$ for the second inequality above, for the third inequality above, we use (SA.8), and for the rate, we use Theorem 2.

Note that in (SA.11)-(SA.15) the rate in (SA.20) is the slowest due to l_n definition in (35) to obtain

$$|\hat{D} - D| = O_p(K^2 \bar{s} l_n) = o_p(1). \quad \text{(SA.21)}$$

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 SA.5. *Under Assumptions 1, 4, 6*

$$A = O(1).$$
$$|F| = O(K^{1/2}).$$
$$D = O(K).$$

Proof of Lemma SA.5.

Note that $A = 1_p' \Gamma 1_p / p \leq Eigmax(\Gamma)$. Then by p.221 of Abadir and Magnus (2005), (Exercise 8.27.b in Abadir and Magnus (2005)), $\Omega B[(cov f_t)^{-1} + B'\Omega B]^{-1} B'\Omega$ is positive semidefinite, so we can use Exercise 12.40b of Abadir and Magnus (2005), since Ω is symmetric, and by (33)

$$\begin{aligned} Eigmax(\Gamma) &:= Eigmax(\Omega - \Omega B[(cov f_t)^{-1} + B'\Omega B]^{-1} B'\Omega) \\ &\leq Eigmax(\Omega) = \frac{1}{Eigmin(\Sigma_n)} \leq \frac{1}{c}, \end{aligned} \quad \text{(SA.22)}$$

since $\Omega := \Sigma_n^{-1}$, and by Assumption 1 $Eigmin(\Sigma_n) \geq c > 0$. This last point shows that $A = O(1)$.

Now consider $D = \mu' \Gamma \mu / p$. By Theorem 5.6.2b of Horn and Johnson (2013)

$$\begin{aligned} \|\mu\|_2^2 &:= \|BEf_t\|_2^2 \leq \|B\|_{l_2}^2 \|Ef_t\|_2^2 \\ &= O(p)O(K) = O(pK), \end{aligned} \quad \text{(SA.23)}$$

by (6.3) of Fan and Lv (2008), $\|B\|_{l_2}^2 = O(p)$ under Assumption 6, and by Assumption 4, $\|Ef_t\|_2^2 = O(K)$, since $f_t : K \times 1$ vector of factors. By (SA.22)(SA.23)

$$\mu' \Gamma \mu / p \leq Eigmax(\Gamma) \|\mu\|_2^2 / p = O(pK)/p = O(K). \quad \text{(SA.24)}$$

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

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

Lemma SA.6. *Under Assumptions 1-4, 6, 7(i), 8*

$$|(\hat{A}\hat{D} - \hat{F}^2) - (AD - F^2)| = O_p(K^2 \bar{s} l_n) = o_p(1).$$

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

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

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

$$\begin{aligned} |(\hat{A}\hat{D} - \hat{F}^2) - (AD - F^2)| &\leq \{|\hat{A} - A||\hat{D} - D| + |\hat{A} - A|D \\ &+ A|\hat{D} - D| + (\hat{F} - F)^2 + 2|F||\hat{F} - F|\} \\ &= O_p(\bar{s} l_n) O_p(K^2 \bar{s} l_n) + O_p(\bar{s} l_n) O(K) \\ &+ O(1) O_p(K^2 \bar{s} l_n) + O_p(\bar{s}^2 l_n^2 K^2) + O(K^{1/2}) O_p(\bar{s} l_n K) \\ &= O_p(K^2 \bar{s} l_n) = o_p(1), \end{aligned} \tag{SA.25}$$

where we use (SA.2), Lemma SA.3, (SA.21), Lemma SA.5, and Assumption 8. **Q.E.D.**

Sufficient Conditions for Assumption 8(iii).

We propose two sufficient conditions.

$$Eigmax(Bcovf_t B') \leq \frac{CK}{2}, \quad CK \geq 2r_n \tag{SA.26}$$

See that under (SA.26)

$$Eigmax(\Sigma_y) := Eigmax(Bcovf_t B' + \Sigma_n) \leq Eigmax(BcovfB') + Eigmax(\Sigma_n) \leq \frac{CK}{2} + r_n \leq CK.$$

Sufficient condition basically tells that common factor part dominates the noise, which is sensible. We can also find more primitives on the marginals of B, and $covf_t$.

References

Abadir, K. and J. Magnus (2005). *Matrix Algebra.* Cambridge University Press.

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

Barras, L., P. Gagliardini, and O. Scaillet (2018). Skill, scale, and value creation in the mutual fund industry. *Swiss Finance Institute, DP*, 2018.66.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris (2009). Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences 106*, 12267–12272.

Callot, L., M. Caner, O. Onder, and E. Ulasan (2021). A nodewise regression approach to estimating large portfolios. *Journal of Business and Economic Statistics 39*, 520–531.

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.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica 51*, 1281–1304.

Chang, J., Y. Qiu, Q. Yao, and T. Zou (2018). Confidence regions for entries of a large precision matrix. *Journal of Econometrics 206*, 57–82.

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

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.

Ding, Y., Y. Li, and X. Zheng (2020). High-dimensional minimum variance portfolio estimation under statistical factor models. *Journal of Econometrics Forthcoming*.

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 (2011). High-dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics 39*, 3320–3356.

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.

Fan, J., Y. Liao, and X. Shi (2015). Risks of large portfolios. *Journal of Econometrics 186*, 367–387.

Fan, J. and J. Lv (2008). Sure independence screening for ultra-high dimensional feature space. *Journal of Royal Statistical Society Series B*, 849–911.

Fan, J., R. Masini, and M. Medeiros (2021). Bridging factor and sparse models. arxiv:2102.11341, arXiv.

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

Gagliardini, P., E. Ossola, and O. Scaillet (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica 84*, 985–1046.

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.

Horn, R. and C. Johnson (2013). *Matrix Analysis*. Cambridge University Press.

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

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).

Merlevede, F., M. Peligrad, and E. Rio (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields 151*, 435–474.

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.

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research 11*, 2261–2286.

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.