Sparse Portfolio

Job Market Paper

Ekaterina Seregina[*]

September 29, 2020

Abstract

The classical approach to portfolio optimization is notorious for producing undesirable extreme long and short positions due to inaccurate estimation of asset weights that fluctuate substantially over time. Besides, its asset allocations are associated with non-negligible transaction costs, high turnover and large monitoring costs. To overcome these shortcomings, we develop a novel optimization approach which produces sparse wealth allocations by setting some weights to zero using a penalty function. The proposed statistical method proceeds in two steps: first, it uses an ℓ_1-penalty on the weight vector to select stocks, second, we apply de-biasing and post-lasso to obtain the optimal asset allocation weights. The main contribution is twofold: from the theoretical perspective, this paper establishes unbiasedness and consistency of the optimal sparse allocations in a high-dimensional setting, when the number of assets exceeds the sample size. We demonstrate the importance of the de-biasing step that has been overlooked in previous studies. From the empirical perspective, the application to the constituents of the S&P500 reveals that compared to the common strategy of holding all assets, our sparse portfolio strategy leads to lower risk, lower turnover, and higher out-of-sample Sharpe ratio. We illustrate that during several economic downturns including the dot-com bubble of 2000 and the financial crisis of 2007-09, our sparse de-biased estimator was the only model that produced positive cumulative excess return (CER) and did not exceed the target level of risk. In contrast, all non-sparse models produced negative CER and violated the risk constraint. This finding suggests that our de-biased sparse estimator exhibits desirable minimax properties: it minimizes the maximum risk level of a portfolio.

Keywords: High-dimensionality, Portfolio Optimization, Factor Investing, De-biasing, Post-Lasso, Approximate Factor Model

JEL Classifications: C13, C55, C58, G11, G17

1   INTRODUCTION

The research on portfolio allocation dates back to Markowitz (1952) – a successful strategy should demonstrate good out-of-sample performance in terms of the portfolio risk and return, and the constructed portfolio should be cheap and easy to maintain and monitor. Any portfolio allocation problem requires the inverse covariance matrix, or *precision* matrix, of excess stock returns as an input. In the era of big data, a search for the optimal portfolio becomes a high-dimensional problem: the number of assets, p, is comparable to or greater than the sample size, T. This creates two major challenges to the standard optimization strategies. First, monitoring and transaction costs of a high-dimensional portfolio might be prohibitively large. Second, using a sample covariance matrix as an input to the portfolio allocation problem is either infeasible, or produces unstable solutions for weights. In this paper we propose a novel approach to a high-dimensional "investor's problem" which addresses the aforementioned issues.

All investors, ranging from beginner traders to billion-dollar corporations, encounter the same fundamental problem: which stocks to buy and how much to invest in them. One of the common strategies mentioned by Lyle and Yohn (2020) is to pick a few top-performing stocks. However, natural questions are *when do you stop* and *is there any guarantee that the best-performers of the last month or year will still demonstrate superior performance today*? These questions become especially prominent during the times of recessions, such as the one brought by COVID-19. Events like this radically alter investors' outlook on portfolio composition: e.g., cruise line and airline industry stocks were considered attractive in the past, but they are not the top choice for the investors now. Therefore, we need more theoretical guarantees related to the stock picking exercise, other than good past performance.

One popular method to create a *sparse portfolio*, that is, a portfolio with many zero entries in the weight vector, is by introducing an ℓ_1-penalty (Lasso) on the portfolio weights which shrinks some of them to zero (see Fan et al. (2019), Ao et al. (2019), Li (2015), Brodie et al. (2009) among others). This approach handles the problem of stock picking, however, it does not provide a reliable answer how much to invest in the selected stocks. Concretely, it is well-known that the Lasso-based estimator is biased (see Belloni et al. (2015); Javanmard and Montanari (2014a,b); Javanmard et al. (2018); van de Geer et al. (2014); Zhang and Zhang (2014) among others). To the best of our knowl-

edge, none of the existing papers that study sparse portfolio correct for the bias in the estimated weights. This paper proposes an estimator for portfolio weights based on the de-biased Lasso and Post-Lasso that produces an unbiased sparse portfolio. Furthermore, in contrast to the existing literature, our approach allows for several different optimal portfolio formulations depending on the investors' preferences.

At the same time, as has been mentioned above, to get optimal weights for any portfolio optimization problem one needs an estimator of the precision matrix. Instead of taking the standard approach of estimating and inverting the covariance matrix, we estimate a precision matrix directly. The proposed procedure applies a nodewise regression technique (Meinshausen and Bühlmann (2006)) to the stock returns that are driven by the common factors. We call the proposed algorithm *Factor Nodewise Regression* (*FMB*). It allows us to estimate a high-dimensional precision matrix even when the sample covariance matrix is not invertible. Furthermore, in contrast to the existing precision matrix estimators, our approach does not require any assumptions on the sparsity of the covariance or precision matrix of stock returns.

Our empirical application studies daily and monthly data for the constituents of the S&P500 and demonstrates superior performance of the de-biased estimator compared to not de-biased counterparts in terms of the return and the out-of-sample Sharpe ratio. Furthermore, we find that non-sparse high-dimensional portfolios almost always violate the risk constraint out-of-sample. In contrast, sparse portfolios are characterized by lower risk, lower turnover, and their Sharpe Ratio is comparable to risky non-sparse portfolios. Moreover, accounting for the factor structure in the portfolio allocation problem leads to improved performance in terms of the out-of -sample portfolio return and Sharpe Ratio.

Motivated by the superior performance of the factor-based portfolios compared to non-factor-based counterparts, we examine whether using factors as investment vehicles in addition to individual assets improves portfolio performance in terms of the out-of-sample Sharpe ratio. This framework is known as *factor investing* and it was shown to increase portfolio return (Ao et al. (2019)). The goal of factor investing is to decide how much weight is allocated to factors, and how much weight is allocated to individual stocks. However, to the best of our knowledge, there are no established guidelines regarding the criteria for this allocation. We propose a novel simple approach which provides clear guidelines how to determine the weight of factors and stocks in

the portfolio under factor investing. Our framework also allows to test whether using factors as investment vehicles significantly contributes to the return of the portfolio.

Our paper makes several important contributions. First, from the theoretical perspective, we contribute to the existing literature on portfolio allocation and sparse statistical modeling: to the best of our knowledge, there are no equivalent theoretical results that establish consistency of the weights of a sparse financial portfolio in a high-dimensional setting. In other words, this is the first paper that simultaneously addresses two questions relevant to investors: which stocks to invest in and how much to invest in these stocks. Second, we are not aware of any work that provides a flexible framework for constructing a sparse portfolio for the investors with various objective functions and constraints. Concretely, Ao et al. (2019) and Fan et al. (2019) focus on the portfolio formulation that has only risk constraint relaxing the requirement that the weights sum up to one. The framework of Li (2015) studies the mean-variance objective function, where the weight constraint is also relaxed, whereas Kremer et al. (2019) use the same framework imposing the constraint that weights sum up to one. In contrast, the framework developed in this paper can be applied to several different portfolio formulations, therefore, it can be used by investors with various objective functions and constraints. Third, we develop a new estimator of a high-dimensional precision matrix which uses the nodewise regression when the stock returns follow approximate factor model. In contrast to the existing literature, FMB does not impose any assumptions on the sparsity of covariance or precision matrix of stock returns. Fourth, we provide simple guidelines for factor investing and show that our approach consistently estimates the portfolio weights of individual assets and factors. To the best of our knowledge, the current literature does not have a clear statistical framework for factor investing – most approaches are ad hock and rely on the past performance of different factors documented in the empirical finance literature.

This paper is organized as follows: Section 2 introduces sparse de-biased portfolio and sparse portfolio using post-Lasso. Section 3 develops a new high-dimensional precision estimator called Factor Nodewise regression. Section 4 develops a framework for factor investing. Section 5 contains theoretical results and and Section 6 validates these results using simulations. Section 7 provides empirical application. Section 8 concludes.

For the convenience of the reader, we summarize the notation to be used throughout the paper. Let $\mathcal{S}_p$ denote the set of all $p \times p$ symmetric matrices, $\mathcal{S}_p^+$ denotes the set of all $p \times p$ positive semi-definite matrices, and $\mathcal{S}_p^{++}$ denotes the set of all $p \times p$ positive definite matrices. Given a vector $\mathbf{u} \in \mathbb{R}^d$ and parameter $a \in [1, \infty)$, let $\|\mathbf{u}\|_a$ denote l_a-norm. Given a matrix $\mathbf{U} \in \mathcal{S}_p$, let $\Lambda_{max}(\mathbf{U}) := \Lambda_1(\mathbf{U}) \geq \Lambda_2(\mathbf{U}) \geq \ldots \Lambda_{min}(\mathbf{U}) := \Lambda_p(\mathbf{U})$ be the eigenvalues of $\mathbf{U}$, and $\mathrm{eig}_K(\mathbf{U}) \in \mathbb{R}^{K \times p}$ denote the first $K \leq p$ normalized eigenvectors corresponding to $\Lambda_1(\mathbf{U}), \ldots \Lambda_K(\mathbf{U})$. Given parameters $a, b \in [1, \infty)$, let $\|\|\mathbf{U}\|\|_{a,b}$ denote the induced matrix-operator norm $\max_{\|\mathbf{y}\|_a = 1} \|\mathbf{U}\mathbf{y}\|_b$. The special cases are $\|\|\mathbf{U}\|\|_1 := \max_{1 \leq j \leq N} \sum_{i=1}^N |\mathbf{U}_{i,j}|$ for the l_1/l_1-operator norm; the operator norm (l_2-matrix norm) $\|\|\mathbf{U}\|\|_2^2 := \Lambda_{max}(\mathbf{U}\mathbf{U}')$ is equal to the maximal singular value of $\mathbf{U}$; $\|\|\mathbf{U}\|\|_\infty := \max_{1 \leq j \leq N} \sum_{i=1}^N |\mathbf{U}_{j,i}|$ for the l_∞/l_∞-operator norm. Finally, $\|\mathbf{U}\|_{\max}$ denotes the element-wise maximum $\max_{i,j} |\mathbf{U}_{i,j}|$, and $\|\|\mathbf{U}\|\|_F^2 = \sum_{i,j} u_{i,j}^2$ denotes the Frobenius matrix norm. We use $a \lesssim_P b$ to denote $a = \mathcal{O}_P(b)$, $a \vee b = \max\{a, b\}$, and $a \wedge b = \min\{a, b\}$. For an event A, we say that A wp $\to 1$ when A occurs with probability approaching 1 as T increases.

2   SPARSE PORTFOLIO

There exist several widely used portfolio weight formulations depending on the type of optimization problem solved by an investor. Suppose we observe p assets (indexed by i) over T period of time (indexed by t). Let $\mathbf{r}_t = (r_{1t}, r_{2t}, \ldots, r_{pt})' \sim \mathcal{D}(\mathbf{m}, \boldsymbol{\Sigma})$ be a $p \times 1$ vector of *excess* returns drawn from a distribution $\mathcal{D}$. In this section we do not require any assumptions on $\mathcal{D}$ but depending on the setup, we will introduce necessary restrictions in the next sections. The goal of the Markowitz theory is to choose assets weights in a portfolio *optimally*. We will study two criteria of optimality: the first is a well-known Markowitz weight-constrained optimization problem, and the second formulation relaxes constraints on portfolio weights.

The first optimization problem, which will be referred to as *Markowitz weight-constrained problem (MWC)*, searches for assets weights such that the portfolio achieves a desired expected rate of return with minimum risk, under the restriction that all weights sum up to one. The aforementioned

goal can be formulated as the following quadratic optimization problem:

$$\begin{cases} \min_{\mathbf{w}} \dfrac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \\[2mm] \text{s.t. } \mathbf{w}'\boldsymbol{\iota} = 1 \\[2mm] \mathbf{m}'\mathbf{w} \geq \mu, \end{cases} \tag{2.1}$$

where $\mathbf{w}$ is a $p \times 1$ vector of assets weights in the portfolio, $\boldsymbol{\iota}$ is a $p \times 1$ vector of ones, and μ is a desired expected rate of portfolio return. Let $\boldsymbol{\Theta} := \boldsymbol{\Sigma}^{-1}$ be the *precision matrix*. The constraint in (2.1) requires portfolio weights to sum up to one - this assumption can be easily relaxed and I will demonstrate the implications of this constraint on portfolio weights.

If $\mathbf{m}'\mathbf{w} > \mu$, then the solution to (2.1) yields the *global minimum-variance (GMV)* portfolio weights $\mathbf{w}_{GMV}$:

$$\mathbf{w}_{GMV} = (\boldsymbol{\iota}'\boldsymbol{\Theta}\boldsymbol{\iota})^{-1}\boldsymbol{\Theta}\boldsymbol{\iota}. \tag{2.2}$$

If $\mathbf{m}'\mathbf{w} = \mu$, the solution to (2.1) is

$$\mathbf{w}_{MWC} = (1 - a_1)\mathbf{w}_{GMV} + a_1\mathbf{w}_M, \tag{2.3}$$

$$\mathbf{w}_M = (\boldsymbol{\iota}'\boldsymbol{\Theta}\mathbf{m})^{-1}\boldsymbol{\Theta}\mathbf{m}, \tag{2.4}$$

$$a_1 = \frac{\mu(\mathbf{m}'\boldsymbol{\Theta}\boldsymbol{\iota})(\boldsymbol{\iota}'\boldsymbol{\Theta}\boldsymbol{\iota}) - (\mathbf{m}'\boldsymbol{\Theta}\boldsymbol{\iota})^2}{(\mathbf{m}'\boldsymbol{\Theta}\mathbf{m})(\boldsymbol{\iota}'\boldsymbol{\Theta}\boldsymbol{\iota}) - (\mathbf{m}'\boldsymbol{\Theta}\boldsymbol{\iota})^2}, \tag{2.5}$$

where $\mathbf{w}_{MWC}$ denotes the portfolio allocation with the constraint that the weights need to sum up to one and $\mathbf{w}_M$ captures all mean-related market information.

The second optimization problem, which will be referred to as *Markowitz risk-constrained (MRC)* problem, has the same objective as in (2.1), but portfolio weights are not required to sum up to one:

$$\begin{cases} \min_{\mathbf{w}} \dfrac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \\[2mm] \text{s.t. } \mathbf{m}'\mathbf{w} \geq \mu. \end{cases} \tag{2.6}$$

It can be easily shown that the solution to (2.6) is:

$$\mathbf{w}_1^* = \frac{\mu\boldsymbol{\Theta}\mathbf{m}}{\mathbf{m}'\boldsymbol{\Theta}\mathbf{m}}. \tag{2.7}$$

Alternatively, instead of searching for a portfolio with a specified desired expected rate of return and minimum risk, one can maximize expected portfolio return given a maximum risk-tolerance level:

$$\begin{cases} \max_{\mathbf{w}} \mathbf{w}'\mathbf{m} \\[2mm] \text{s.t. } \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \leq \sigma^2. \end{cases} \tag{2.8}$$

In this case, the solution to (2.8) yields:

$$\mathbf{w}_2^* = \frac{\sigma^2}{\mathbf{w'm}}\boldsymbol{\Theta}\mathbf{m} = \frac{\sigma^2}{\mu}\boldsymbol{\Theta}\mathbf{m}. \tag{2.9}$$

To get the second equality in (2.9) I used the definition of μ from (2.1) and (2.6). It follows that if $\mu = \sigma\sqrt{\theta}$, where $\theta := \mathbf{m'}\boldsymbol{\Theta}\mathbf{m}$ is the squared Sharpe ratio, then the solution to (2.6) and (2.8) admits the following expression:

$$\mathbf{w}_{MRC} = \frac{\sigma}{\sqrt{\mathbf{m'}\boldsymbol{\Theta}\mathbf{m}}}\boldsymbol{\Theta}\mathbf{m} = \frac{\sigma}{\sqrt{\theta}}\boldsymbol{\alpha}, \tag{2.10}$$

where $\boldsymbol{\alpha} := \boldsymbol{\Theta}\mathbf{m}$. Equation (2.10) tells us that once an investor specifies the desired return, μ, and maximum risk-tolerance level, σ, this pins down the Sharpe ratio of the portfolio which makes the optimization problems of minimizing risk and maximizing expected return of the portfolio in (2.6) and (2.8) identical.

This brings us to three alternative portfolio allocations commonly used in the existing literature: Global Minimum-Variance Portfolio in (2.2), weight-constrained Markowitz Mean-Variance in (2.3) and maximum-risk-constrained Markowitz Mean-Variance in (2.10). Below I summarize the aforementioned portfolio weight expressions:

GMV:
$$\mathbf{w}_{GMV} = (\boldsymbol{\iota'}\boldsymbol{\Theta}\boldsymbol{\iota})^{-1}\boldsymbol{\Theta}\boldsymbol{\iota}, \tag{2.11}$$

MWC
$$\mathbf{w}_C^* = (1-a_1)\mathbf{w}_{GMV} + a_1\mathbf{w}_{MWC}, \tag{2.12}$$

where
$$\mathbf{w}_M = (\boldsymbol{\iota'}\boldsymbol{\Theta}\mathbf{m})^{-1}\boldsymbol{\Theta}\mathbf{m},$$
$$a_1 = \frac{\mu(\mathbf{m'}\boldsymbol{\Theta}\boldsymbol{\iota})(\boldsymbol{\iota'}\boldsymbol{\Theta}\boldsymbol{\iota}) - (\mathbf{m'}\boldsymbol{\Theta}\boldsymbol{\iota})^2}{(\mathbf{m'}\boldsymbol{\Theta}\mathbf{m})(\boldsymbol{\iota'}\boldsymbol{\Theta}\boldsymbol{\iota}) - (\mathbf{m'}\boldsymbol{\Theta}\boldsymbol{\iota})^2},$$

MRC:
$$\mathbf{w}_{MRC} = \frac{\sigma}{\sqrt{\theta}}\boldsymbol{\alpha}, \tag{2.13}$$

where
$$\boldsymbol{\alpha} = \boldsymbol{\Theta}\mathbf{m},$$
$$\theta := \mathbf{m'}\boldsymbol{\Theta}\mathbf{m}.$$

So far we have considered allocation strategies that put non-zero weights to all assets in the financial portfolio. As an implication, an investor needs to buy a certain amount of each security even if there are a lot of small weights. However, oftentimes investors are interested in managing a few assets which significantly reduces monitoring and transaction costs and was shown to outperform equal weighted and index portfolios in terms of the Sharpe ratio and cumulative return (see

Fan et al. (2019), Ao et al. (2019), Li (2015), Brodie et al. (2009) among others). This strategy is based on holding a *sparse portfolio*, that is, a portfolio with many zero entries in the weight vector. In this paper we combine the ideas from the literature on de-biasing in high-dimensional models with sparse financial portfolio. We apply the Factor Nodewise regression described in Algorithm 3 within this framework to account for the factor structure in stock returns.

2.1 SPARSE DE-BIASED PORTFOLIO

Let us first introduce some notations. The sample mean and sample covariance matrix have standard formulas: $\widehat{\mathbf{m}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{r}_t$ and $\widehat{\boldsymbol{\Sigma}} = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{r}_t - \widehat{\mathbf{m}})(\mathbf{r}_t - \widehat{\mathbf{m}})'$. Our empirical application shows that risk-constrained Markowitz allocation in (2.13) outperforms GMV and MWC portfolios in (2.11)-(2.12). Therefore, we first study sparse MRC portfolios. Our goal is to construct a sparse vector of portfolio weights given by (2.13). To achieve this we use the following equivalent and unconstrained regression representation of the mean-variance optimization in (2.6) and (2.8):

$$\mathbf{w}_{MRC} = \underset{\mathbf{w}}{\arg\min}\, \mathbb{E}\left[r_c - \mathbf{w}' \mathbf{r}_t \right], \quad \text{where} \quad r_c := \frac{1 + \theta}{\theta} \mu \equiv \sigma \frac{1 + \theta}{\sqrt{\theta}}. \tag{2.14}$$

The sample counterpart of (2.14) is written as:

$$\mathbf{w}_{MRC} = \underset{\mathbf{w}}{\arg\min}\, \frac{1}{T} \sum_{t=1}^{T} (r_c - \mathbf{w}' \mathbf{r}_t)^2. \tag{2.15}$$

Ao et al. (2019) prove that the weight allocation from (2.14) is equivalent to (2.13). The sparsity is introduced through Lasso which yields the following constrained optimization problem:

$$\mathbf{w}_{\text{MRC, SPARSE}} = \underset{\mathbf{w}}{\arg\min}\, \frac{1}{T} \sum_{t=1}^{T} (r_c - \mathbf{w}' \mathbf{r}_t)^2 + 2\lambda \|\mathbf{w}\|_1. \tag{2.16}$$

Now we propose two extensions to the setup (2.16). First, the estimator $\mathbf{w}_{\text{MRC, SPARSE}}$ is infeasible since θ used for constructing r_c is unknown. Ao et al. (2019) construct an estimator of θ under normally distributed excess returns, assuming $p/T \to \rho \in (0, 1)$ and the sample size T is required to be larger than the number of assets p. Their paper uses an unbiased estimator proposed in Kan and Zhou (2007): $\hat{\theta} = ((T - p - 2)\widehat{\mathbf{m}}' \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\mathbf{m}} - p)/T$, where $\widehat{\mathbf{m}}$ and $\widehat{\boldsymbol{\Sigma}}^{-1}$ are sample mean and inverse of the sample covariance matrix respectively. One of the limitations of the model studied by Ao et al. (2019) is that it cannot handle high dimensions. In both simulations and empirical application the maximum number of stocks used by the authors is limited to 100. Another limitation of Ao et al. (2019) approach is that they do not correct the bias introduced by imposing

l_1-constraint in (2.16). However, it is well-known that the estimator in (2.16) is biased and the existing literature proposes several de-biasing techniques (see Belloni et al. (2015); Javanmard and Montanari (2014a,b); Javanmard et al. (2018); van de Geer et al. (2014); Zhang and Zhang (2014) among others).

To approach the first aforementioned limitation, we propose to use an estimator of a high-dimensional precision matrix discussed in the next section. The suggested estimator is appropriate for high-dimensional settings, it can handle cases when the sample size is less than the number of assets, and it is always non-negative by construction[1]. Consequently, the estimator of r_c is

$$\widehat{r}_c := \frac{1+\hat{\theta}}{\hat{\theta}}\mu \equiv \sigma\frac{1+\hat{\theta}}{\sqrt{\hat{\theta}}}. \tag{2.17}$$

To approach the second limitation, motivated by van de Geer et al. (2014), we propose the de-biasing technique that uses the nodewise regression estimator of the precision matrix. First, let $\mathbf{R}$ be a $T \times p$ matrix of excess returns stacked over time and $\widehat{\mathbf{r}}_c$ be a $T \times 1$ constant vector. Consider a high-dimensional linear model

$$\widehat{\mathbf{r}}_c = \mathbf{R}\mathbf{w} + \mathbf{e}, \quad \text{where} \quad \mathbf{e} \sim \mathcal{D}(\mathbf{0}, \sigma_e^2\mathbf{I}). \tag{2.18}$$

We study high-dimensional framework $p \geq T$ and in the asymptotic results we require $\log(p)/T = o(1)$. Let us rewrite (2.16):

$$\mathbf{w}_{\text{MRC, SPARSE}} = \arg\min_{\mathbf{w}\in\mathbb{R}^p} \frac{1}{T}\|\widehat{\mathbf{r}}_c - \mathbf{R}\mathbf{w}\|_2^2 + 2\lambda\|\mathbf{w}\|_1. \tag{2.19}$$

The estimator in (2.16) satisfies the following KKT conditions:

$$-\mathbf{R}'(\widehat{\mathbf{r}}_c - \mathbf{R}\widehat{\mathbf{w}})/T + \lambda\widehat{\mathbf{g}} = 0, \tag{2.20}$$

$$\|\widehat{\mathbf{g}}\|_\infty \leq 1 \quad \text{and} \quad \hat{g}_i = \text{sign}(\hat{w}_i) \quad \text{if} \quad \hat{w}_i \neq 0. \tag{2.21}$$

where $\widehat{\mathbf{g}}$ is a $p \times 1$ vector arising from the subgradient of $\|\mathbf{w}\|_1$. Let $\widehat{\boldsymbol{\Sigma}} = \mathbf{R}'\mathbf{R}/T$, then we can rewrite the KKT conditions:

$$\widehat{\boldsymbol{\Sigma}}(\widehat{\mathbf{w}} - \mathbf{w}) + \lambda\widehat{\mathbf{g}} = \mathbf{R}'\mathbf{e}/T. \tag{2.22}$$

---

[1]Our empirical results suggest that the unbiased estimator $\hat{\theta} = ((T-p-2)\widehat{\mathbf{m}}'\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{m}} - p)/T$ is oftentimes negative even after using the adjusted estimator defined in Kan and Zhou (2007) (p. 2906).

Multiply both sides of (2.22) by $\widehat{\Theta}$ which is obtained from Algorithm 3 and rearrange the terms:

$$\widehat{\mathbf{w}} - \mathbf{w} + \widehat{\Theta}\lambda\widehat{\mathbf{g}} = \widehat{\Theta}\mathbf{R}'\boldsymbol{\varepsilon}/T - \underbrace{\sqrt{T}(\widehat{\Theta}\widehat{\Sigma} - \mathbf{I}_p)(\widehat{\mathbf{w}} - \mathbf{w})}_{\Delta}/\sqrt{T}. \tag{2.23}$$

In the section with the theoretical results we show that Δ is asymptotically negligible under certain sparsity assumptions[2]. Combining (2.20) and (2.23) brings us to the de-biased estimator of portfolio weights:

$$\widehat{\mathbf{w}}_{\text{MRC, DEBIASED}} = \widehat{\mathbf{w}} + \widehat{\Theta}\lambda\widehat{\mathbf{g}} = \widehat{\mathbf{w}} + \widehat{\Theta}\mathbf{R}'(\widehat{\mathbf{r}}_c - \mathbf{R}\widehat{\mathbf{w}})/T. \tag{2.24}$$

The properties of the proposed de-biased estimator are examined in Section 5.

2.2   Sparse Portfolio Using Post-Lasso

One of the drawbacks of the de-biased portfolio weights in (2.24) is that the weight formula is tailored to a specific portfolio choice that maximizes an unconstrained Sharpe ratio (i.e. MRC in (2.13)). However, it is desirable to accommodate preferences of different types of investors who might be interested in weight allocations corresponding to GMV ((2.11)) or MWC ((2.12)) portfolios. At the same time, we are willing to stay within the framework of sparse allocations. One of the difficulties that precludes us from pursuing a similar technique as in (2.16) is the fact that once the weight constraint is added, the optimization problem in (2.16) has two solutions depending on whether $\boldsymbol{\iota}'\Theta\mathbf{m}$ is positive or negative. As shown in Maller and Turkington (2003), when $\boldsymbol{\iota}'\Theta\mathbf{m} < 0$, the minimum value cannot be achieved exactly for a specified portfolio allocation that satisfies the full investment constraint. Hence, one can design an approximate solution to approach the supremum as closely as desired.

To overcome this difficulty, we propose to use lasso regression in (2.19) for selecting a subset of stocks, and then constructing a financial portfolio using any of the weight formulations in (2.11)-(2.13). The procedure to estimate sparse portfolio using post-Lasso is described in Algorithm 1.

---

[2]Note that we cannot directly apply Theorem 2.2 of van de Geer et al. (2014) since $\mathbf{r}_c$ needs to be estimated and we first need to show consistency of the respective estimator.

Algorithm 1 Sparse Portfolio Using Post-Lasso

1: Use Lasso regression in (2.19) to select the model $\widehat{\Xi} := \text{support}(\widehat{\mathbf{w}})$

- Apply additional thresholding to remove stocks with small estimated weights:

$$\widehat{\mathbf{w}}(t) = (\widehat{w}_j \mathbb{1}\left[|\widehat{w}_j| > t\right], \ j = 1, \ldots, p),$$

where $t \geq 0$ is the thresholding level.

- The corresponding selected model is denoted as $\widehat{\Xi}(t) := \text{support}(\widehat{\mathbf{w}}(t))$. When $t = 0$, $\widehat{\Xi}(t) = \widehat{\Xi}$.

2: Choose a desired portfolio formulation in (2.11)-(2.13) and apply it to the selected subset of stocks $\widehat{\Xi}(t)$.

- When $\text{card}(\widehat{\Xi}(t)) < \widetilde{t}$, use the inverse of the sample covariance matrix as an estimator of Θ. Otherwise, apply the estimator of precision matrix described in Section 3.

---

3   Factor Nodewise Regression

In this section we first review a nodewise regression (Meinshausen and Bühlmann (2006)), a popular approach to estimate a precision matrix. After that we propose a novel estimator which accounts for the common factors in the excess returns.

Define $\mathbf{x}_t$ to be a $p \times 1$ vector at time $t = 1, \ldots, T$. Let $\mathbf{x}_t \sim \mathcal{D}(\mathbf{m}, \boldsymbol{\Sigma})$, where $\mathcal{D}$ belongs to either sub-Gaussian or elliptical families. When $\mathcal{D} = \mathcal{N}$, the precision matrix $\boldsymbol{\Sigma}^{-1} := \boldsymbol{\Theta}$ contains information about conditional dependence between the variables. For instance, if $\boldsymbol{\Theta}_{ij}$, which is the ij-th element of the precision matrix, is zero, then the variables i and j are conditionally independent, given the other variables. In the high-dimensional settings it is necessary to regularize the precision matrix, which means that some of the entries $\boldsymbol{\Theta}_{ij}$ will be zero. In other words, to achieve consistent estimation of the inverse covariance, the estimated precision matrix should be sparse.

3.1   Nodewise Regression

One of the approaches to induce sparsity in the estimation of precision matrix is to solve for $\widehat{\boldsymbol{\Theta}}$ one column at a time via linear regressions, replacing population moments by their sample counterparts. When we repeat this procedure for each variable $j = 1, \ldots, p$, we will estimate the elements of $\widehat{\boldsymbol{\Theta}}$ column by column using $\{\mathbf{x}_t\}_{t=1}^T$ via p linear regressions. Meinshausen and Bühlmann (2006) use this approach to incorporate sparsity into the estimation of the precision

matrix. They fit p separate Lasso regressions using each variable (node) as the response and the others as predictors to estimate $\widehat{\Theta}$. This method is known as the "nodewise" regression and it is reviewed below based on van de Geer et al. (2014) and Callot et al. (2019).

Let $\mathbf{x}_j$ be a $T \times 1$ vector of observations for the j-th regressor, the remaining covariates are collected in a $T \times (p-1)$ matrix $\mathbf{X}_{-j}$. For each $j = 1, \ldots, p$ we run the following Lasso regressions:

$$\widehat{\gamma}_j = \arg \min_{\gamma \in \mathbb{R}^{p-1}} \left(\|\mathbf{x}_j - \mathbf{X}_{-j}\gamma\|_2^2/T + 2\lambda_j\|\gamma\|_1 \right), \tag{3.1}$$

where $\widehat{\gamma}_j = \{\widehat{\gamma}_{j,k}; j = 1, \ldots, p, k \neq j\}$ is a $(p-1) \times 1$ vector of the estimated regression coefficients that will be used to construct the estimate of the precision matrix, $\widehat{\Theta}$. Define

$$\widehat{\mathbf{C}} = \begin{pmatrix} 1 & -\widehat{\gamma}_{1,2} & \cdots & -\widehat{\gamma}_{1,p} \\ -\widehat{\gamma}_{2,1} & 1 & \cdots & -\widehat{\gamma}_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ -\widehat{\gamma}_{p,1} & -\widehat{\gamma}_{p,2} & \cdots & 1 \end{pmatrix}. \tag{3.2}$$

For $j = 1, \ldots, p$, define

$$\hat{\tau}_j^2 = \|\mathbf{x}_j - \mathbf{X}_{-j}\widehat{\gamma}_j\|_2^2/T + \lambda_j\|\widehat{\gamma}_j\|_1 \tag{3.3}$$

and write

$$\widehat{\mathbf{T}}^2 = \text{diag}(\hat{\tau}_1^2, \ldots, \hat{\tau}_p^2). \tag{3.4}$$

The approximate inverse is defined as

$$\widehat{\Theta} = \widehat{\mathbf{T}}^{-2}\widehat{\mathbf{C}}. \tag{3.5}$$

The procedure to estimate the precision matrix using nodewise regression is summarized in Algorithm 2.

Algorithm 2 Nodewise Regression by Meinshausen and Bühlmann (2006) (MB)

1: Repeat for $j = 1, \ldots, p$:

- Estimate $\widehat{\gamma}_j$ using (3.1) for a given λ_j.

- Select λ_j using a suitable information criterion.

2: Calculate $\widehat{\mathbf{C}}$ and $\widehat{\mathbf{T}}^2$.
3: Return $\widehat{\Theta} = \widehat{\mathbf{T}}^{-2}\widehat{\mathbf{C}}$.

One of the caveats to keep in mind when using the nodewise regression method is that the estimator in (3.5) is not self-adjoint. Callot et al. (2019) show (see their Lemma A.1) that $\widehat{\Theta}$ in

(3.5) is positive definite with high probability, however, it could still occur that $\widehat{\Theta}$ is not positive definite in finite samples. To resolve this issue we use the matrix symmetrization procedure as in Fan et al. (2018) and then use eigenvalue cleaning as in Callot et al. (2017) and Hautsch et al. (2012). First, the symmetric matrix is constructed as

$$\widehat{\Theta}_{ij}^s = \widehat{\Theta}_{ij} \mathbb{1}\left[\left|\widehat{\Theta}_{ij}\right| \le \left|\widehat{\Theta}_{ji}\right|\right] + \widehat{\Theta}_{ji} \mathbb{1}\left[\left|\widehat{\Theta}_{ij}\right| > \left|\widehat{\Theta}_{ji}\right|\right], \tag{3.6}$$

where $\widehat{\Theta}_{ij}$ is the (i,j)-th element of the estimated precision matrix from (3.5). Second, we use eigenvalue cleaning to make $\widehat{\Theta}^s$ positive definite: write the spectral decomposition $\widehat{\Theta}^s = \widehat{V}'\widehat{\Lambda}\widehat{V}$, where $\widehat{V}$ is a matrix of eigenvectors and $\widehat{\Lambda}$ is a diagonal matrix with p eigenvalues $\widehat{\Lambda}_i$ on its diagonal. Let $\Lambda_m := \min\{\widehat{\Lambda}_i | \widehat{\Lambda}_i > 0\}$. We replace all $\widehat{\Lambda}_i < \Lambda_m$ with Λ_m and define the diagonal matrix with cleaned eigenvalues as $\widetilde{\Lambda}$. We use $\widetilde{\Theta} = \widehat{V}'\widetilde{\Lambda}\widehat{V}$ which is symmetric and positive definite.

3.2 Factor Nodewise Regression

The arbitrage pricing theory (APT), developed by Ross (1976), postulates that expected returns on securities should be related to their covariance with the common components or factors only. The goal of the APT is to model the tendency of asset returns to move together via factor decomposition. Let $\mathbf{r}_t = (r_{1t}, r_{2t}, \ldots, r_{pt})' \sim \mathcal{D}(\mathbf{m}, \boldsymbol{\Sigma})$ be a $p \times 1$ vector of excess returns drawn from a distribution $\mathcal{D}$, where $\mathbf{m}$ is the unconditional mean of the returns. Assume that the return generating process $(\mathbf{r}_t)$ follows a K-factor model:

$$\underbrace{\mathbf{r}_t}_{p \times 1} = \mathbf{B} \underbrace{\mathbf{f}_t}_{K \times 1} + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T \tag{3.7}$$

where $\mathbf{f}_t = (f_{1t}, \ldots, f_{Kt})'$ are the factors, $\mathbf{B}$ is a $p \times K$ matrix of factor loadings, and $\boldsymbol{\varepsilon}_t$ is the idiosyncratic component that cannot be explained by the common factors. Factors in (3.7) can be either observable, such as in Fama and French (1993, 2015), or can be estimated using statistical factor models.

In this subsection we examine how to approach the portfolio allocation problems in (2.11)-(2.13) using a factor structure in the returns. Our approach, called *Factor Nodewise Regression*, uses the estimated common factors to obtain sparse precision matrix of the idiosyncratic component. The resulting estimator is used to obtain the precision of the asset returns necessary to form portfolio weights.

As in Fan et al. (2013), we consider a spiked covariance model when the first K principal eigenvalues of Σ are growing with p, while the remaining $p - K$ eigenvalues are bounded and grow slower than p.

Rewrite equation (3.7) in matrix form:

$$\underbrace{\mathbf{R}}_{p \times T} = \underbrace{\mathbf{B}}_{p \times K} \mathbf{F} + \mathbf{E}. \tag{3.8}$$

The factors and loadings in (3.8) are estimated by solving $(\widehat{\mathbf{B}}, \widehat{\mathbf{F}}) = \operatorname{argmin}_{\mathbf{B}, \mathbf{F}} \|\mathbf{R} - \mathbf{B}\mathbf{F}\|_F^2 \quad$ s.t. $\frac{1}{T}\mathbf{F}\mathbf{F}' = \mathbf{I}_K$, $\mathbf{B}'\mathbf{B}$ is diagonal. The constraints are needed to identify the factors (Fan et al. (2018)). It was shown (Stock and Watson (2002)) that $\widehat{\mathbf{F}} = \sqrt{T}\operatorname{eig}_K(\mathbf{R}'\mathbf{R})$ and $\widehat{\mathbf{B}} = T^{-1}\mathbf{R}\widehat{\mathbf{F}}'$. Given $\widehat{\mathbf{F}}, \widehat{\mathbf{B}}$, define $\widehat{\mathbf{E}} = \mathbf{R} - \widehat{\mathbf{B}}\widehat{\mathbf{F}}$.

Since our interest is in constructing portfolio weights, our goal is to estimate a precision matrix of the excess returns. However, as pointed out by Koike (2020), when common factors are present across the excess returns, the precision matrix cannot be sparse because all pairs of the returns are partially correlated given other excess returns through the common factors. Therefore, we impose a sparsity assumption on the precision matrix of the idiosyncratic errors, Θ_ε, which is obtained using the estimated residuals after removing the co-movements induced by the factors (see Barigozzi et al. (2018); Brownlees et al. (2018); Koike (2020)).

We use the nodewise regression as a shrinkage technique to estimate the precision matrix of residuals. Once the precision Θ_f of the low-rank component is also obtained, similarly to Fan et al. (2011), we use the Sherman-Morrison-Woodbury formula to estimate the precision of excess returns:

$$\Theta = \Theta_\varepsilon - \Theta_\varepsilon \mathbf{B}[\Theta_f + \mathbf{B}'\Theta_\varepsilon \mathbf{B}]^{-1}\mathbf{B}'\Theta_\varepsilon. \tag{3.9}$$

To obtain $\widehat{\Theta}_f = \widehat{\Sigma}_f^{-1}$, we use the inverse of the sample covariance of the estimated factors $\widehat{\Sigma}_f = T^{-1}\widehat{\mathbf{F}}\widehat{\mathbf{F}}'$. To get $\widehat{\Theta}_\varepsilon$, we first apply Algorithm 2 to the estimated idiosyncratic errors, $\widehat{\varepsilon}_t$. Once we have estimated $\widehat{\Theta}_f$ and $\widehat{\Theta}_\varepsilon$, we can get $\widehat{\Theta}$ using a sample analogue of (3.9). The proposed procedure is called *Factor Nodewise Regression* and is summarized it in Algorithm 3.

Algorithm 3 Factor Nodewise Regression by Meinshausen and Bühlmann (2006) (FMB)

1: Estimate the residuals: $\widehat{\boldsymbol{\varepsilon}}_t = \mathbf{r}_t - \widehat{\mathbf{B}}\widehat{\mathbf{f}}_t$ using PCA.
   Get $\widehat{\boldsymbol{\Sigma}}_\varepsilon = \frac{1}{T}\sum_{t=1}^{T}(\widehat{\boldsymbol{\varepsilon}}_t - \bar{\boldsymbol{\varepsilon}})(\widehat{\boldsymbol{\varepsilon}}_t - \bar{\boldsymbol{\varepsilon}})'$.
2: Estimate a sparse $\boldsymbol{\Theta}_\varepsilon$ using nodewise regression: apply Algorithm 2 to $\widehat{\boldsymbol{\varepsilon}}$.
3: Estimate $\boldsymbol{\Theta}$ using the Sherman-Morrison-Woodbury formula in (3.9).

Now we can use $\widehat{\boldsymbol{\Theta}}$ obtained from (3.9) using Algorithm 3 to estimate portfolio weights in (2.11)-(2.13).

4   FACTOR INVESTING IS ALLOWED

In this section we allow an investor to hold a portfolio of assets and factors, in other words, factors are assumed to be tradable. Note that in contrast with Ao et al. (2019), the distinction between tradable and non-tradable factors is not pinned down by the fact that the excess returns are driven by the common factors. That is, factor structure of returns is allowed independently of whether factors are tradable or not. We assume that only observable factors can be tradable. Denote a $K_1 \times 1$ vector of observable factors as $\widetilde{\mathbf{f}}_t$, and $K_2 \times 1$ vector of unobservable factors as $\mathbf{f}_t^{PCA}$, where $K1 + K2 = K$. The goal of factor investing is to decide how much weight is allocated to factors $\widetilde{\mathbf{f}}_t$ and stocks $\mathbf{r}_t$. Let $r_{t,all}$ be the return of portfolio at time t:

$$r_{t,all} = \underbrace{\mathbf{w}'_{all,t}}_{1 \times (p+K_1)} \mathbf{x}_t. \tag{4.1}$$

where $\mathbf{x}_t = (\widetilde{\mathbf{f}}'_t, \mathbf{r}'_t)'$ is a $(p + K_1) \times 1$ vector of excess returns of observable factors and stocks and $\mathbf{w}_{all,t} = (\mathbf{w}'_{ft}, \mathbf{w}'_t)'$ is a vector of weights with $\mathbf{w}_{ft}$ invested in $\widetilde{\mathbf{f}}_t$ and $\mathbf{w}_t$ invested in stocks. We treat $\widetilde{\mathbf{f}}_t$ as additional K_1 investments vehicles which will contribute to the return of the total portfolio. Now consider K_2-factor model for $\mathbf{x}_t$:

$$\mathbf{x}_t = \mathbf{B}\underbrace{\mathbf{f}_t^{PCA}}_{K_2 \times 1} + \mathbf{e}_t, \quad t = 1, \dots, T \tag{4.2}$$

Rewrite equation (4.2) in matrix form:

$$\underbrace{\mathbf{X}}_{(p+K_1)\times T} = \mathbf{B}\underbrace{\mathbf{F}^{PCA}}_{K_2 \times T} + \mathbf{E}, \tag{4.3}$$

which can be estimated using the standard PCA techniques as in (3.8): $\widehat{\mathbf{F}}^{PCA} = \sqrt{T}\mathrm{eig}_{K_2}(\mathbf{X}'\mathbf{X})$ and $\widehat{\mathbf{B}} = T^{-1}\mathbf{X}\widehat{\mathbf{F}}'^{PCA}$. Given $\widehat{\mathbf{F}}^{PCA}, \widehat{\mathbf{B}}$, define $\widehat{\mathbf{E}} = \mathbf{X} - \widehat{\mathbf{B}}\widehat{\mathbf{F}}^{PCA}$.

Similarly to Algorithm 3, we use (3.9) to estimate the precision of the augmented excess returns, $\boldsymbol{\Theta}_x$. To get $\widehat{\boldsymbol{\Theta}}_{fPCA} = \widehat{\boldsymbol{\Sigma}}_{fPCA}^{-1}$, we use the inverse of the sample covariance of the estimated factors $\widehat{\boldsymbol{\Sigma}}_{fPCA} = T^{-1}\widehat{\mathbf{F}}^{PCA}\widehat{\mathbf{F}}'^{PCA}$. To get $\widehat{\boldsymbol{\Theta}}_e$, we first apply Algorithm 2 to the estimated idiosyncratic errors, $\widehat{\mathbf{e}}_t$ in (4.2). Once we have estimated $\widehat{\boldsymbol{\Theta}}_{fPCA}$ and $\widehat{\boldsymbol{\Theta}}_e$, we can get $\widehat{\boldsymbol{\Theta}}_x$ using a sample analogue of (3.9). This procedure is summarized in Algorithm 4.

Algorithm 4 Factor Investing Using FMB

1: Estimate the residuals from equation (4.2): $\widehat{\mathbf{e}}_t = \mathbf{x}_t - \widehat{\mathbf{B}}\widehat{\mathbf{f}_t^{PCA}}$ using PCA.
2: Estimate a sparse $\boldsymbol{\Theta}_e$ using nodewise regression: apply Algorithm 2 to $\widehat{e}_t$.
3: Estimate $\boldsymbol{\Theta}_x$ using the Sherman-Morrison-Woodbury formula in (3.9).

Now we can use $\widehat{\boldsymbol{\Theta}}_x$ obtained from Algorithm 4 to estimate portfolio weights $\mathbf{w}_{all,t}$ using either a de-biased technique from section 2.1 ((2.24)), or Post-Lasso (Algorithm 1). Once we obtain $\widehat{\mathbf{w}}_{all,t} = (\widehat{\mathbf{w}}'_{ft}, \widehat{\mathbf{w}}'_t)'$, we can test whether factor investing significantly contributes to the portfolio return by testing whether $\mathbf{w}_{ft} = 0$.

5 ASYMPTOTIC PROPERTIES

In this section we study asymptotic properties of the de-biased estimator of weights for sparse portfolio in (2.24) and post-lasso estimator from Algorithm 1.

Denote $S_0 := \{j; \mathbf{w}_j \neq 0\}$ to be the active set of variables, where $\mathbf{w}$ is a vector of true portfolio weights in equation (2.18). Also, let $s_0 := |S_0|$. Further, let $S_j := \{k; \gamma_{j,k} \neq 0\}$ be the active set for row γ_j for the nodewise regression in (3.1), and let $s_j := |S_j|$. Define $\bar{s} := \max_{1 \leq j \leq p} s_j$.

Consider a factor model from equation (3.7):

$$\underbrace{\mathbf{r}_t}_{p \times 1} = \mathbf{B} \underbrace{\mathbf{f}_t}_{K \times 1} + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T \tag{5.1}$$

We study the case when the factors are not known, i.e. the only observable variable in equation (5.1) is the excess returns $\mathbf{r}_t$. In this paper our main interest lies in establishing asymptotic properties of sparse de-biased portfolio weights and the out-of-sample Sharpe ratio for the high-dimensional case. We assume that the number of common factors, K, is fixed.

5.1 ASSUMPTIONS

We now list the assumptions on the model (5.1):

(A.1) (Spiked covariance model) As $p \to \infty$, $\Lambda_1 > \Lambda_2 + \ldots > \Lambda_K \gg \Lambda_{K+1} \geq \ldots \geq \Lambda_p \geq 0$, where $\Lambda_j = \mathcal{O}(p)$ for $j \leq K$, while the non-spiked eigenvalues are bounded, $\Lambda_j = o(p)$ for $j > K$.

(A.2) (Pervasive factors) There exists a positive definite $K \times K$ matrix $\breve{\mathbf{B}}$ such that $\left\|\left\| p^{-1}\mathbf{B}'\mathbf{B} - \breve{\mathbf{B}} \right\|\right\|_2 \to 0$ and $\Lambda_{\min}(\breve{\mathbf{B}})^{-1} = \mathcal{O}(1)$ as $p \to \infty$.

Similarly to Chang et al. (2018) and Callot et al. (2019), we also impose beta mixing condition.

(A.3) (Beta mixing) Let $\mathcal{F}^t_{-\infty}$ and $\mathcal{F}^{\infty}_{t+k}$ denote the σ-algebras that are generated by $\{\varepsilon_u : u \leq t\}$ and $\{\varepsilon_u : u \geq t + k\}$ respectively. Then $\{\varepsilon\}_u$ is β-mixing in the sense that $\beta_k \to 0$ as $k \to \infty$, where the mixing coefficient is defined as

$$\beta_k = \sup_t \mathbb{E} \left[\sup_{B \in \mathcal{F}^{\infty}_{t+k}} \left| \Pr\left(B | \mathcal{F}^t_{-\infty}\right) - \Pr\left(B\right) \right| \right]. \tag{5.2}$$

Some comments regarding the aforementioned assumptions are in order. Assumptions **(A.1)**-**(A.2)** are the same as in Fan et al. (2018), and assumption **(A.3)** is required to consistently estimate precision matrix for de-biasing portfolio weights. Assumption **(A.1)** divides the eigenvalues into the diverging and bounded ones. Without loss of generality, we assume that K largest eigenvalues have multiplicity of 1. The assumption of a spiked covariance model is common in the literature on approximate factor models, however, we note that the model studied in this paper can be characterized as a "very spiked model". In other words, the gap between the first K eigenvalues and the rest is increasing with p. As pointed out by Fan et al. (2018), **(A.1)** is typically satisfied by the factor model with pervasive factors, which brings us to the assumption **(A.2)**: the factors impact a non-vanishing proportion of individual time-series. Assumption **(A.3)** allows for weak dependence in the residuals of the factor model in 5.1: causal ARMA processes, certain stationary Markov chains and stationary GARCH models with finite second moments satisfy this assumption. We note that our Assumption **(A.3)** is much weaker than in Callot et al. (2019), the latter requires weak dependence of the returns series, whereas we only restrict dependence of the idiosyncratic components.

Let $\mathbf{\Sigma} = \mathbf{\Gamma}_p \mathbf{\Lambda}_p \mathbf{\Gamma}'_p$, where $\mathbf{\Sigma}$ is the covariance matrix of returns that follow factor structure described in equation (5.1). Define $\widehat{\mathbf{\Sigma}}, \widehat{\mathbf{\Lambda}}_K, \widehat{\mathbf{\Gamma}}_K$ to be the estimators of $\mathbf{\Sigma}, \mathbf{\Lambda}_p, \mathbf{\Gamma}_p$. We further let $\widehat{\mathbf{\Lambda}}_K = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_K)$ and $\widehat{\mathbf{\Gamma}}_K = (\hat{v}_1, \ldots, \hat{v}_K)$ to be constructed by the first K leading empirical eigenvalues and the corresponding eigenvectors of $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{B}}\widehat{\mathbf{B}}' = \widehat{\mathbf{\Gamma}}_K \widehat{\mathbf{\Lambda}}_K \widehat{\mathbf{\Gamma}}'_K$. Similarly to Fan et al. (2018), we require the following bounds on the componentwise maximums of the estimators:

16

(B.1) $\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_{\max} = \mathcal{O}_p(\sqrt{\log p / T})$,

(B.2) $\left\|(\widehat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Lambda}_p)\boldsymbol{\Lambda}_p^{-1}\right\|_{\max} = \mathcal{O}_p(\sqrt{\log p / T})$,

(B.3) $\left\|\widehat{\boldsymbol{\Gamma}}_K - \boldsymbol{\Gamma}_p\right\|_{\max} = \mathcal{O}_p(\sqrt{\log p / (Tp)})$.

Let $\widehat{\boldsymbol{\Sigma}}^{SG}$ be the sample covariance matrix, with $\widehat{\boldsymbol{\Lambda}}_K^{SG}$ and $\widehat{\boldsymbol{\Gamma}}_K^{SG}$ constructed with the first K leading empirical eigenvalues and eigenvectors of $\widehat{\boldsymbol{\Sigma}}^{SG}$ respectively. Also, let $\widehat{\boldsymbol{\Sigma}}^{EL1} = \widehat{\mathbf{D}}\widehat{\mathbf{R}}_1\widehat{\mathbf{D}}$, where $\widehat{\mathbf{R}}_1$ is obtained using the Kendall's tau correlation coefficients and $\widehat{\mathbf{D}}$ is a robust estimator of variances constructed using the Huber loss. Furthermore, let $\widehat{\boldsymbol{\Sigma}}^{EL2} = \widehat{\mathbf{D}}\widehat{\mathbf{R}}_2\widehat{\mathbf{D}}$, where $\widehat{\mathbf{R}}_2$ is obtained using the spatial Kendall's tau estimator. Define $\widehat{\boldsymbol{\Lambda}}_K^{EL}$ to be the matrix of the first K leading empirical eigenvalues of $\widehat{\boldsymbol{\Sigma}}^{EL1}$, and $\widehat{\boldsymbol{\Gamma}}_K^{EL}$ is the matrix of the first K leading empirical eigenvectors of $\widehat{\boldsymbol{\Sigma}}^{EL2}$. For more details regarding constructing $\widehat{\boldsymbol{\Sigma}}^{SG}$, $\widehat{\boldsymbol{\Sigma}}^{EL1}$ and $\widehat{\boldsymbol{\Sigma}}^{EL2}$ see Fan et al. (2018), Sections 3 and 4.

Theorem 1. (*Fan et al. (2018)*)
For sub-Gaussian distributions, $\widehat{\boldsymbol{\Sigma}}^{SG}$, $\widehat{\boldsymbol{\Lambda}}_K^{SG}$ and $\widehat{\boldsymbol{\Gamma}}_K^{SG}$ satisfy (B.1)-(B.3).
For elliptical distributions, $\widehat{\boldsymbol{\Sigma}}^{EL1}$, $\widehat{\boldsymbol{\Lambda}}_K^{EL}$ and $\widehat{\boldsymbol{\Gamma}}_K^{EL}$ satisfy (B.1)-(B.3).

Theorem 1 is essentially a rephrasing of the results obtained in Fan et al. (2018), Sections 3 and 4. Since there is no separate statement of these results in their paper (it is rather a summary of several theorems), we separated it as a Theorem for the convenience of the reader. As evidenced from the above Theorem, $\widehat{\boldsymbol{\Sigma}}^{EL2}$ is only used for estimating the eigenvectors. This is necessary due to the fact that, in contrast with $\widehat{\boldsymbol{\Sigma}}^{EL2}$, the theoretical properties of the eigenvectors of $\widehat{\boldsymbol{\Sigma}}^{EL}$ are mathematically involved because of the sin function.

In addition, the following structural assumption on the model is imposed:

(C.1) $\|\boldsymbol{\Sigma}\|_{\max} = \mathcal{O}(1)$ and $\|\mathbf{B}\|_{\max} = \mathcal{O}(1)$,

which is a natural structural assumption on the population quantities.

Note that in contrast to Fan et al. (2018) we do not impose sparsity on the covariance matrix of the idiosyncratic component, instead, it is more realistic to impose conditional sparsity on the precision matrix after the common factors are accounted for.

5.2 Asymptotic Properties of De-Biased Portfolio Weights

Recall that we used equation (3.9) to estimate $\boldsymbol{\Theta}$. Therefore, in order to establish consistency of the estimator in (3.9), we first show consistency of $\widehat{\boldsymbol{\Theta}}_\varepsilon$.

Theorem 2. *Suppose that Assumptions (A.1)-(A.3), (B.1)-(B.3) and (C.1) hold. Let $\omega_T := \sqrt{\log p / T} + 1/\sqrt{p}$. Then $\max_{i \le p}(1/T)\sum_{t=1}^T |\hat{\varepsilon}_{it} - \varepsilon_{it}| = \mathcal{O}_p(\omega_T^2)$ and $\max_{i,t}|\hat{\varepsilon}_{it} - \varepsilon_{it}| = \mathcal{O}_p(\omega_T) = o_p(1)$. Under the sparsity assumption $\bar{s}^2 \omega_T = o(1)$, we have*

$$\max_{1 \le j \le p}\left\|\widehat{\boldsymbol{\Theta}}_{\varepsilon,j} - \boldsymbol{\Theta}_{\varepsilon,j}\right\|_1 = \mathcal{O}_p(\bar{s}\omega_T),$$

$$\max_{1 \le j \le p}\left\|\widehat{\boldsymbol{\Theta}}_{\varepsilon,j} - \boldsymbol{\Theta}_{\varepsilon,j}\right\|_2^2 = \mathcal{O}_p(\bar{s}\omega_T^2)$$

Some comments are in order. First, the sparsity assumption $\bar{s}^2 = o(\omega_T)$ is stronger than that required for convergence of $\widehat{\boldsymbol{\Theta}}_\varepsilon$: this is necessary to ensure consistency for $\widehat{\boldsymbol{\Theta}}$ established in Theorem 3, so we impose a stronger assumption at the beginning. We also note that at the first glance, our sparsity assumption in Theorem 3 is stronger than that required by van de Geer et al. (2014) and Callot et al. (2019), however, recall that we impose sparsity on $\boldsymbol{\Theta}_\varepsilon$, not $\boldsymbol{\Theta}$ as opposed to the two aforementioned papers. Hence, this assumption can be easily satisfied once the common factors have been accounted for and the precision of the idiosyncratic components is expected to be sparse. The bounds derived in Theorem 2 help us establish the convergence properties of the precision matrix of stock returns in equation (3.9).

Theorem 3. *Under the assumptions of Theorem 2 and, in addition, assuming $\|\boldsymbol{\Theta}_{\varepsilon,j}\|_2 = \mathcal{O}(1)$, we have*

$$\max_{1 \le j \le p}\left\|\widehat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j\right\|_1 = \mathcal{O}_p(\bar{s}^2\omega_T),$$

$$\max_{1 \le j \le p}\left\|\widehat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j\right\|_2^2 = \mathcal{O}_p(\bar{s}\omega_T^2).$$

Having established the properties of the estimated precision matrix of stock returns, we are ready to state the following main result:

Theorem 4. *Under the assumptions of Theorem 3, consider the linear model 2.18 with $\mathbf{e} \sim \mathcal{D}(\mathbf{0}, \sigma_e \mathbf{I})$, where $\sigma_e^2 = \mathcal{O}(1)$. Consider a suitable choice of the regularization parameters $\lambda \asymp \omega_T$ for the lasso regression in 2.19 and $\lambda_j \asymp \omega_T$ uniformly in j for the lasso for nodewise regression in 3.1. Assume $s_0 \log(p)/\sqrt{T} = o(1)$.*

Then

$$\sqrt{T}(\widehat{\mathbf{w}}_{DEBIASED} - \mathbf{w}) = W + \Delta,$$

$$W = \widehat{\boldsymbol{\Theta}} \mathbf{R}' \mathbf{e}/\sqrt{T},$$

$$\|\Delta\|_\infty = o_p(1).$$

Some comments are in order. Our Theorem 4 is an extension of Theorem 2.4 of van de Geer et al. (2014) for non-iid case, where the latter is achieved with a help of Chang et al. (2018). Furthermore, there are several fundamental differences between Theorem 4 and Theorem 2.4 of van de Geer et al. (2014): first, we apply nodewise regression to estimate sparse precision matrix of factor-adjusted returns, which explains the difference in convergence rates. Concretely, van de Geer et al. (2014) have $\omega_T = \sqrt{\log(p)/T}$, whereas we have $\omega_T = \sqrt{\log(p)/T} + 1/\sqrt{p}$, where $1/\sqrt{p}$ arises due to the fact that factors need to be estimated. However, we note that since we deal with high-dimensional regime $p \geq T$, this additional term is asymptotically negligible, we only keep it for identification purposes. Second, in contrast with van de Geer et al. (2014), the dependent variable in the Lasso regression in (2.19) is unknown and needs to be estimated. Lemma 2 shows that $\hat{r}_c$ constructed using the precision matrix estimator from Theorem 3 is consistent and shares the same rate as the l_1-bound in Theorem 3. Third, interestingly, the sparsity assumption on the lasso regression in (2.19) is the same as in van de Geer et al. (2014): as shown in the Appendix, this condition is still sufficient to ensure that the bias term is asymptotically negligible even when the stock returns follow factor structure with unknown factors.

5.3   Asymptotic Properties of Post-Lasso Portfolio Weights

To establish the properties of the post-Lasso estimator in Algorithm 1, we focus on MRC weight formulation, since it satisfies the standard post-Lasso assumptions. For GMV and MWC formulations, the procedure described in Algorithm 1 is not "post-Lasso" in the usual sense. Concretely, the latter assumes that both steps in Algorithm 1 have the same objective function, which is violated for GMV and MWC. Consequently, we leave rigorous theoretical derivations of these two portfolio formulation for future research. For MRC, we use the post-model selection results established in Belloni and Chernozhukov (2013). Specifically, we have the following theorem:

Theorem 5. *Suppose the restricted eigenvalue condition and the restricted sparse eigenvalue condition on the empirical Gram matrix hold (see Condition RE($\bar{c}$) and Condition RSE(m) of Belloni and Chernozhukov (2013), p. 529). Let $\widehat{\mathbf{w}}$ be the post-Lasso weight estimator from Algorithm 1, we have*

$$\|\widehat{\mathbf{w}} - \mathbf{w}\|_1 \lesssim_P \begin{cases} \sigma_e\Big((s_0\omega_T) \vee (\bar{s}^2\omega_T)\Big), & \text{in general,} \\ \sigma_e s_0\Big(\sqrt{\frac{s_0}{T}} + \frac{1}{\sqrt{p}}\Big), & \text{if } s_0 \geq \bar{s}^2 \text{ and } \Xi = \hat{\Xi} \text{ wp} \to 1. \end{cases}$$

The proof of Theorem 5 easily follows from the prove of Corollary 2 of Belloni and Chernozhukov (2013) and is omitted here. Let us comment on the upper bounds for post-Lasso estimator: first, the term $(s_0\omega_T) \vee (\bar{s}^2\omega_T)$ appears since one needs to estimate the dependent variable in equation (2.18), which creates the difference between the bound in Belloni and Chernozhukov (2013) and our Theorem 5. Second, similarly to Belloni and Chernozhukov (2013), the upper bound undergoes a transition from the oracle rate enjoyed by the standard Lasso to the faster rate that improves on the latter when (1) the precision matrix of the idiosyncratic components is sparse enough and (2) the oracle model has well-separated coefficients. Noticeably, the upper bounds in Theorem 5 hold despite the fact that the first-stage Lasso regression in Algorithm 1 may fail to correctly select the oracle model Ξ as a subset, that is, $\Xi \notin \hat{\Xi}$.

6   MONTE CARLO

We study the consistency for estimating portfolio weights in (2.10) of (i) sparse portfolios that use the standard Lasso without de-biasing in (2.19), (ii) Lasso with de-biasing in (2.24), (iii) post-Lasso in Algorithm 1, and (iv) non-sparse portfolios that use FMB from Algorithm 3. Our simulation results are divided into two parts: the first part examines the performance of models (i)-(iv) under the Gaussian setting, and the second part examines the robustness of performance under the elliptical distributions (to be described later). Each part is further subdivided into two cases: low-dimensional with $p < T$ (Case 1) and high-dimensional $p > T$ (Case 2), in both cases we allow the number of stocks to increase with the sample size, i.e. $p = p_T \to \infty$ as $T \to \infty$. In the low-dimensional case we let $p = T^\delta$, $\delta = 0.85$ and $T = [2^h]$, for $h = 7, 7.5, 8, \ldots, 9.5$, in the high-dimensional case we let $p = 3 \cdot T^\delta$, $\delta = 0.85$, all else equal.

First, consider the following data generating process for stock returns:

$$\underbrace{\mathbf{r}_t}_{p \times 1} = \mathbf{m} + \mathbf{B} \underbrace{\mathbf{f}_t}_{K \times 1} + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T \tag{6.1}$$

where $\mathbf{m}_i \sim \mathcal{N}(1,1)$ independently for each $i = 1, \ldots, p$, $\boldsymbol{\varepsilon}_t$ is a $p \times 1$ random vector of idiosyncratic errors following $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$, with a Toeplitz matrix $\boldsymbol{\Sigma}_\varepsilon$ parameterized by ρ: that is, $\boldsymbol{\Sigma}_\varepsilon = (\boldsymbol{\Sigma}_\varepsilon)_{ij}$, where $(\boldsymbol{\Sigma}_\varepsilon)_{ij} = \rho^{|i-j|}$, $i, j \in 1, \ldots, p$ which leads to sparse $\boldsymbol{\Theta}_\varepsilon$, $\mathbf{f}_t$ is a $K \times 1$ vector of factors drawn from $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_f = \mathbf{I}_K/10)$, $\mathbf{B}$ is a $p \times K$ matrix of factor loadings drawn from $\mathcal{N}(\mathbf{0}, \mathbf{I}_K/100)$. We set $\rho = 0.5$ and fix the number of factors $K = 3$.

Let $\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}' + \boldsymbol{\Sigma}_\varepsilon$. To create sparse MRC portfolio weights we use the following procedure: first, we threshold the vector $\boldsymbol{\Sigma}^{-1}\mathbf{m}$ to keep the top $p/2$ entries with largest absolute values. This yields sparse vector $\boldsymbol{\alpha} = \boldsymbol{\Sigma}^{-1}\mathbf{m}$ defined in (2.13). We use $\boldsymbol{\Sigma}\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ as the values for the mean and covariance matrix parameters to generate multivariate Gaussian returns in (6.1). Note that the low rank plus sparse structure of the covariance matrix is preserved under this transformation.

Figure 1 shows the averaged (over Monte Carlo simulations) errors of the estimators of the weight $\mathbf{w}_{\text{MRC}}$ versus the sample size T in the logarithmic scale (base 2). As evidenced by Figure 1, (1) sparse estimators outperform non-sparse counterparts; (2) using de-biasing or post-Lasso improves the performance compared to the standard Lasso estimator. As expected from Theorems 4-5, the Lasso, de-biased Lasso and post-Lasso exhibit similar rates, but the two latter estimators enjoy lower estimation error. The ranking remains similar in the high-dimensional setting, however, as illustrated in Figure 1, the performance of all estimators slightly deteriorates.

Gaussian-tail assumption is too restrictive for modeling the behavior of financial returns. Hence, as a second exercise we check the robustness of our sparse portfolio allocation estimators under the elliptical distributions, which we briefly review based on Fan et al. (2018). Elliptical distribution family generalizes the multivariate normal distribution and multivariate t-distribution. Let $\mathbf{m} \in \mathbb{R}^p$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$. A p-dimensional random vector $\mathbf{r}$ has an elliptical distribution, denoted by $\mathbf{r} \sim \text{ED}_p(\mathbf{m}, \boldsymbol{\Sigma}, \zeta)$, if it has a stochastic representation

$$\mathbf{r} \stackrel{d}{=} \mathbf{m} + \zeta\mathbf{A}\mathbf{U}, \tag{6.2}$$

where $\mathbf{U}$ is a random vector uniformly distributed on the unit sphere $\mathcal{S}^{q-1}$ in $\mathbb{R}^q$, $\zeta \geq 0$ is a scalar random variable independent of $\mathbf{U}$, $\mathbf{A} \in \mathbb{R}^{p \times q}$ is a deterministic matrix satisfying $\mathbf{A}\mathbf{A}' = \boldsymbol{\Sigma}$. As pointed out in Fan et al. (2018), the representation in (6.2) is not identifiable, hence, we require $\mathbb{E}[\zeta^2] = q$, such that $\text{Cov}(\mathbf{r}) = \boldsymbol{\Sigma}$. We only consider continuous elliptical distributions with $\Pr[\zeta = 0] = 0$. The advantage of the elliptical distribution for the financial returns is its ability to model

heavy-tailed data and the tail dependence between variables.

Having reviewed the elliptical distribution, we proceed to the second part of simulation results. The data generating process is similar to Fan et al. (2018): let $(\mathbf{f}_t, \boldsymbol{\varepsilon}_t)$ from (6.1) jointly follow the multivariate t-distribution with the degrees of freedom ν. When $\nu = \infty$, this corresponds to the multivariate normal distribution, smaller values of ν are associated with thicker tails. We draw T independent samples of $(\mathbf{f}_t, \boldsymbol{\varepsilon}_t)$ from the multivariate t-distribution with zero mean and covariance matrix $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_f, \boldsymbol{\Sigma}_\varepsilon)$, where $\boldsymbol{\Sigma}_f = \mathbf{I}_K$. To construct $\boldsymbol{\Sigma}_\varepsilon$ we use a Toeplitz structure parameterized by $\rho = 0.5$, which leads to the sparse $\boldsymbol{\Theta}_\varepsilon = \boldsymbol{\Sigma}_\varepsilon^{-1}$. The rows of $\mathbf{B}$ are drawn from $\mathcal{N}(\mathbf{0}, \mathbf{I}_K/100)$. Figure 2 reports the results for $\nu = 4.2$[3]: the performance of the standard Lasso estimator significantly deteriorates, which is further amplified in the high-dimensional case where it exhibits the worst performance. Noticeably, post-Lasso still achieves the lowest estimation error, followed by de-biased estimator.

7   EMPIRICAL APPLICATION

This section is divided into three main parts. First, we examine the performance of non-sparse portfolios that use FMB from Algorithm 3 and compare their performance with equal-weighted portfolio, Index portfolio (reported as the composite S&P500 index listed as ^GSPC) and MB from Algorithm 2 which does not use the information about a factor structure in the stock returns. Second, we study the performance of sparse portfolios that are based on de-biasing and post-lasso. Third, we consider several interesting periods that include different economic downturns: we examine the merit of sparse vs non-sparse portfolios during economic downturns.

7.1   DATA

We use monthly and daily returns of the components of the S&P500 index. The data on historical S&P500 constituents and stock returns is fetched from CRSP and Compustat using SAS interface. The full sample for the monthly data has 480 observations on 355 stocks from January 1, 1980 - December 1, 2019. We use January 1, 1980 - December 1, 1994 (180 obs) as a training period and January 1, 1995 - December 1, 2019 (300 obs) as the out-of-sample test period. For the daily data the full sample size has 5040 observations on 420 stocks from January 12, 2000 - January 31,

---

[3]The results for larger degrees of freedom do not provide any additional insight, hence we do not report them here. However, they are available upon request.

2020. We use January 20, 2000 - January 24, 2002 (504 obs) as a training period and January 25, 2002 - January 31, 2020 (4536 obs) as the out-of-sample test period.

We roll the estimation window over the test sample to rebalance the portfolios monthly. At the end of each month, prior to portfolio construction, we remove stocks with less than 15 or 2 years of historical stock return data for monthly and daily returns respectively. For sparse portfolio we employ the following strategy to choose the tuning parameter λ in (2.16): we use the first two thirds of the training data (which we call the training window) to estimate weights and tune the shrinkage intensity λ in the remaining one third of the training sample to yield the highest Sharpe ratio which serves as a validation window. We estimate factors and factor loadings in the training window and validation window combined. The risk-free and rate and Fama/French factors are taken from Kenneth R. French's data library.

7.2 PERFORMANCE MEASURES

Similarly to Callot et al. (2019), we consider four metrics commonly reported in finance literature: the Sharpe ratio, the portfolio turnover, the average return and variance of a portfolio. We consider two scenarios: with and without transaction costs. Let T denote the total number of observations, the training sample consists of m observations, and the test sample is $n = T - m$. When transaction costs are not taken into account, the out-of-sample average portfolio return, variance and Sharpe ratio are

$$\hat{\mu}_{\text{test}} = \frac{1}{n} \sum_{t=m}^{T-1} \widehat{\mathbf{w}}_t' \mathbf{r}_{t+1}, \tag{7.1}$$

$$\hat{\sigma}_{\text{test}}^2 = \frac{1}{n-1} \sum_{t=m}^{T-1} (\widehat{\mathbf{w}}_t' \mathbf{r}_{t+1} - \hat{\mu}_{\text{test}})^2, \tag{7.2}$$

$$\text{SR} = \hat{\mu}_{\text{test}} / \hat{\sigma}_{\text{test}}. \tag{7.3}$$

We follow Ban et al. (2018); Callot et al. (2019); DeMiguel et al. (2009); Li (2015) to account for transaction costs (tc). In line with the aforementioned papers, we set $c = 50\text{bps}$. Define the excess portfolio at time $t + 1$ with transaction costs as

$$r_{t+1,\text{portfolio}} = \widehat{\mathbf{w}}_t' \mathbf{r}_{t+1} - c(1 + \widehat{\mathbf{w}}_t' \mathbf{r}_{t+1}) \sum_{j=1}^{p} \left| \hat{w}_{t+1,j} - \hat{w}_{t,j}^+ \right|, \tag{7.4}$$

$$\text{where} \quad \hat{w}_{t,j}^+ = \hat{w}_{t,j} \frac{1 + r_{t+1,j} + r_{t+1}^f}{1 + r_{t+1,\text{portfolio}} + r_{t+1}^f}, \tag{7.5}$$

23

where $r_{t+1,j} + r_{t+1}^f$ is sum of the excess return of the j-th asset and risk-free rate, and $r_{t+1,\text{portfolio}} + r_{t+1}^f$ is the sum of the excess return of the portfolio and risk-free rate. The out-of-sample average portfolio return, variance, Sharpe ratio and turnover are defined accordingly:

$$\hat{\mu}_{\text{test,tc}} = \frac{1}{n} \sum_{t=m}^{T-1} r_{t,\text{portfolio}}, \tag{7.6}$$

$$\hat{\sigma}_{\text{test,tc}}^2 = \frac{1}{n-1} \sum_{t=m}^{T-1} (r_{t,\text{portfolio}} - \hat{\mu}_{\text{test,tc}})^2, \tag{7.7}$$

$$\text{SR}_{\text{tc}} = \hat{\mu}_{\text{test,tc}} / \hat{\sigma}_{\text{test,tc}}, \tag{7.8}$$

$$\text{Turnover} = \frac{1}{n} \sum_{t=m}^{T-1} \sum_{j=1}^{p} \left| \hat{w}_{t+1,j} - \hat{w}_{t,j}^+ \right|. \tag{7.9}$$

7.3 RESULTS

The first set of results explores the performance of FMB from Algorithm 3 for non-sparse portfolio using monthly and daily data. We consider two scenarios, when the factors are unknown and estimated using the standard PCA (statistical factors), and when the factors are known. For the statistical factors we consider up to three PCs. For the scenario with known factors we include up to 5 Fama-French factors: FF1 includes the excess return on the market, FF3 includes FF1 plus size factor (Small Minus Big, SMB) and value factor (High Minus Low, HML), and FF5 includes FF3 plus profitability factor (Robust Minus Weak, RMW) and risk factor (Conservative Minus Agressive, CMA). In Tables 1-2, we report the monthly and daily portfolio performance for three alternative portfolio allocations in (2.11), (2.12) and (2.11). We set a return target $\mu \in \{0.7974\%, 0.0378\%\}$ for monthly and daily data respectively (both are equivalents of 10% yearly return when compounded). The target level of risk for the weight-constrained and risk-constrained Markowitz portfolio (MWC and MRC) is set at $\sigma \in \{0.05, 0.013\}$ which is the standard deviation of the monthly and daily excess returns of the S&P500 index in the first training set.

Some comments for Tables 1-2 are in order:

1. The Tables show that the MRC produces portfolio return and Sharpe ratio that are uniformly higher than those for the weight-constrained allocations MWC and GMV. This means that relaxing the constraint that portfolio weights sum up to one leads to a large increase in the out-of-sample Sharpe ratio and portfolio return. This increase, however, comes at the cost

of higher risk and higher portfolio turnover: for MRC portfolios the risk constraint is often violated.

2. Factor-based portfolios outperform non-factor-based counterparts for monthly data: factor models significantly reduce portfolio risk which increases Sharpe ratio. However, the effect of common factors deteriorates for daily data: only MRC portfolios with observable factors outperform MB. One possible explanation is a larger signal-to-noise ratio for daily stock returns, as compared to the monthly data.

3. For monthly data, FMB-based models outperform EW and index in terms of return and the out-of-sample Sharpe ratio. This is accompanied with higher risk of the factor-based models. Again, the impact of factors deteriorates for daily data: Table 2 shows that only MRC portfolios with observable factors outperform EW and Index.

4. There is no clear ranking between statistical vs observable factors: overall, the performance is comparable in terms of the out-of-sample Sharpe ratio.

As evidenced by the empirical results from Tables 1-2, the risk produced by non-sparse portfolios is relatively high compared to the risk of EW and index portfolios. Furthermore, having examined empirical performance of FMB we notice that some of the estimated portfolio weights are very close to zero. This means that an investor needs to buy a certain amount of each security even if there are a lot of small weights. However, oftentimes investors are interested in managing a few assets which significantly reduces monitoring and transaction costs and was shown to outperform equal weighted and index portfolios in terms of the Sharpe ratio and cumulative return (see Fan et al. (2019), Ao et al. (2019), Li (2015), Brodie et al. (2009) among others). This brings us to examining performance of sparse portfolios, which is reported in Table 3 for monthly data, and Table 4 for daily data. As mentioned in Section 2, shrinking portfolio weights introduces bias, here we study two ways to correct for it: the first approach applies de-biasing technique and it was described in Section 2 (see equation (2.24)). By construction, the first approach can only be applied to MRC weight formula. The second approach uses post-Lasso described in Algorithm 1: we first use Lasso-based weight estimator in (2.19) for selecting stocks with absolute value of weights above a small threshold ϵ (we use $\epsilon = 0.000001$), then we form portfolio with the selected stocks using

three alternative portfolio allocations in (2.11)-(2.13). Table 3 reports both of the aforementioned approaches for the monthly data. Some comments are in order:

1. First, similarly to the results from Table 1, the MRC produces portfolio return and Sharpe ratio that are uniformly higher than those for the weight-constrained allocations MWC and GMVP. Again, this suggests that requiring the weights to sum up to one leads to inefficiency caused by weight misallocation which deteriorates portfolio performance.

2. Second, column one demonstrates that de-biasing leads to significant performance improvement in terms of the return and out-of-sample Sharpe ratio. Note that even though the risk of de-biased portfolio is also higher, it still satisfies the risk-constraint. This result emphasizes the importance of correcting for the bias introduced by the l_1-regularization.

3. Third, comparing two bias-correction methods, de-biasing and post-lasso, we find that the latter is characterized by higher return and higher risk. However, increase in portfolio return brought by post-lasso is, overall, not sufficient to outperform de-biasing approach in terms of the out-of-sample Sharpe ratio.

4. Fourth, de-biased sparse portfolio has significantly lower risk and turnover compared to non-sparse counterparts in Table 1: we used FMB for de-biasing, hence, the direct counterparts of de-biased portfolios from Table 2 are FMB portfolios from Table 1. Sparse de-biased portfolios have lower return compared to Table 1, however, the out-of-sample Sharpe ratio is comparable, i.e. we do not see uniform superiority of either method. Therefore, incorporating sparsity allows investors to reduce portfolio risk at the cost of lower return while maintaining the Sharpe ratio comparable to holding a non-sparse portfolio. Similarly to the result from Table 1, de-biased sparse portfolio and post-lasso with statistical factors outperform EW and index.

5. Finally, the advantage of using observable Fama-French factors versus statistical factors becomes less pronounced for sparse portfolio compared to Table 1: both return and risk of the latter are slightly smaller which leads to the out-of-sample Sharpe ratio being marginally higher for observable factors.

26

Table 2 and Table 4 have similar conclusions, therefore, we only highlight one additional finding from Table 4:

1. In contrast to non-sparse portfolio in Table 2, sparse portfolios using daily data outperform EW and index in terms of risk and out-of-sample Sharpe ratio.

Table 5 compares the performance of non-sparse and sparse portfolios for monthly data for different time periods in terms of the cumulative excess return (CER) over the period of interest and risk. We note that the references to the specific crises do not intend to limit these economic downturns to only these periods. They merely provide the context for the time intervals of interest. Table 5 reveals some interesting findings:

1. First, the conclusions from Tables 1-4 are supported: MRC portfolios yield higher CER and they are characterized by higher risk, factor-based models using FMB are characterized by higher CER and much higher risk compared to sparse portfolios, EW and index.

2. Second, sparse de-biased portfolios with statistical factors are the only models that produced positive CER during the financial crisis 2007-09. Noticeably, the performance of sparse portfolios with observable factors showed small negative CER during this period. Note that all models that used MWC and GMV during that time experienced large negative CER.

3. Even when EW and index enjoy positive CER, factor-based models have superior performance in terms of CER at the cost of higher risk.

8   Conclusion and Discussion

This paper develops a novel approach to portfolio composition that addresses two prominent questions faced by investors: which stocks to hold and how much to invest in these stocks. For the first time in the literature, we provide the answers to *both* questions by developing a novel estimator for portfolio weights based on the de-biased Lasso and post-Lasso that produces an unbiased sparse portfolio. Our method handles high-dimensional cases when the the standard methods relying on the sample covariance matrix are infeasible or lead to unstable solutions. To deal with such scenarios, we develop a new technique, called Factor nodewise regression: it estimates the precision matrix for portfolio allocation problem even when the sample covariance matrix is not

invertible. In contrast to the existing approaches, our framework consistently estimates a sparse high-dimensional portfolio when the returns follow factor structure without making any assumptions on the sparsity of the covariance or precision matrix of returns. We also develop a simple framework that provides clear guidelines how to implement factor investing using the methodology developed in this paper.

Our empirical application studies daily and monthly data for the constituents of the S&P500. We find that non-sparse high-dimensional portfolios almost always violate the risk constraint out-of-sample. In contrast, sparse portfolios are characterized by lower risk, lower turnover, and their Sharpe Ratio is comparable to risky non-sparse portfolios. Furthermore, the empirical application demonstrates superior performance of de-biased portfolios compared to not de-biased counterparts in terms of the return and the out-of-sample Sharpe ratio. In addition, we find that sparse portfolios with statistical factors was the only model that produced positive cumulative excess return (CER) during several economic downturns, including the dot-com bubble of 2000 and the financial crisis of 2007-09. This finding suggests that, in addition to being consistent, our de-biased estimator of weights exhibits minimax properties. The formal theoretical justifications of the latter are left for the future research.

There are several venues for potential extensions. First, it would be interesting to examine the criteria for the "optimal" number of stocks in the sparse portfolio. One possible approach would be to compare the investors' objective functions of the sparse and non-sparse portfolios, and determine the threshold after which a sparse portfolio begins to perform better. Second, on a related note, diversification issue is linked to the research on sparse portfolio. By determining the optimal number of stocks in the portfolio, we can also address the question as to which extent increasing one's exposure to many industries improves portfolio performance, and whether there is an optimal degree of diversification. Third, our model can be extended to incorporate the information on the company's fundamentals when selecting stocks. Our conjecture is that using more stock-related information will mainly improve the selection step of both de-biased Lasso and post-Lasso.

References

Ao, M., Yingying, L., and Zheng, X. (2019). Approaching mean-variance efficiency for large portfolios. *The Review of Financial Studies*, 32(7):2890–2919.

Ban, G.-Y., El Karoui, N., and Lim, A. E. (2018). Machine learning and portfolio optimization. *Management Science*, 64(3):1136–1154.

Barigozzi, M., Brownlees, C., and Lugosi, G. (2018). Power-law partial correlation network models. *Electron. J. Statist.*, 12(2):2905–2929.

Belloni, A. and Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547.

Belloni, A., Chernozhukov, V., and Kato, K. (2015). Uniform post-selection inference for least absolute deviation regression and other z-estimation problems. *Biometrika*, 102(1):77–94.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., and Loris, I. (2009). Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences*, 106(30):12267–12272.

Brownlees, C., Nualart, E., and Sun, Y. (2018). Realized networks. *Journal of Applied Econometrics*, 33(7):986–1006.

Callot, L., Caner, M., Önder, A. O., and Ulaşan, E. (2019). A nodewise regression approach to estimating large portfolios. *Journal of Business & Economic Statistics*, 0(0):1–12.

Callot, L. A. F., Kock, A. B., and Medeiros, M. C. (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics*, 32(1):140–158.

Caner, M. and Kock, A. B. (2018). Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso. *Journal of Econometrics*, 203(1):143 – 168.

Chang, J., Qiu, Y., Yao, Q., and Zou, T. (2018). Confidence regions for entries of a large precision matrix. *Journal of Econometrics*, 206(1):57 – 82.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953.

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

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1 – 22.

Fan, J., Liao, Y., and Mincheva, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.*, 39(6):3320–3356.

Fan, J., Liao, Y., and Mincheva, M. (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., Liu, H., and Wang, W. (2018). Large covariance estimation through elliptical factor models. *Ann. Statist.*, 46(4):1383–1414.

Fan, J., Weng, H., and Zhou, Y. (2019). Optimal estimation of functionals of high-dimensional mean and covariance matrix. *arXiv preprint arXiv:1908.07460*.

Hautsch, N., Kyj, L. M., and Oomen, R. C. A. (2012). A blocking and regularization approach to high-dimensional realized covariance estimation. *Journal of Applied Econometrics*, 27(4):625–645.

Javanmard, A. and Montanari, A. (2014a). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909.

Javanmard, A. and Montanari, A. (2014b). Hypothesis testing in high-dimensional regression under the gaussian random design model: Asymptotic theory. *IEEE Transactions on Information Theory*, 60(10):6522–6554.

Javanmard, A., Montanari, A., et al. (2018). Debiasing the lasso: Optimal sample size for gaussian designs. *The Annals of Statistics*, 46(6A):2593–2622.

Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656.

Koike, Y. (2020). De-biased graphical lasso for high-frequency data. *Entropy*, 22(4):456.

Kremer, P., Lee, S., Bogdan, M., and Paterlini, S. (2019). Sparse portfolio selection via the sorted l1 - norm. *Journal of Banking & Finance*, 110:105687.

Li, J. (2015). Sparse and stable portfolio selection with parameter uncertainty. *Journal of Business & Economic Statistics*, 33(3):381–392.

Lyle, M. R. and Yohn, T. L. (2020). Fundamental analysis and mean-variance optimal portfolios. *Kelley School of Business Research Paper Forthcoming*.

Maller, R. A. and Turkington, D. A. (2003). New light on the portfolio allocation problem. *Mathematical Methods of Operations Research*, 56(3):501–511.

Markowitz, H. (1952). Portfolio selection*. *The Journal of Finance*, 7(1):77–91.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34(3):1436–1462.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341 – 360.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.*, 97(460):1167–1179.

van de Geer, S., Buhlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.*, 42(3):1166–1202.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242.

Figure 1: **Averaged errors of the estimators of w_{MRC} for Case 1 on logarithmic scale (left):** $p = T^{0.85}$, $K = 3$ **and for Case 2 on logarithmic scale (right):** $p = 3 \cdot T^{0.85}$, $K = 3$.
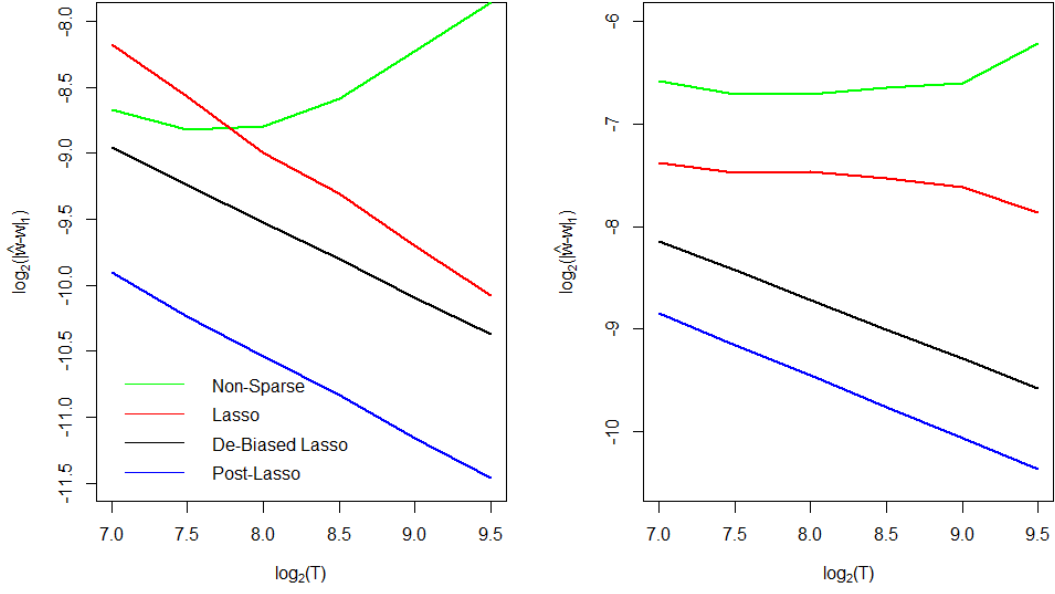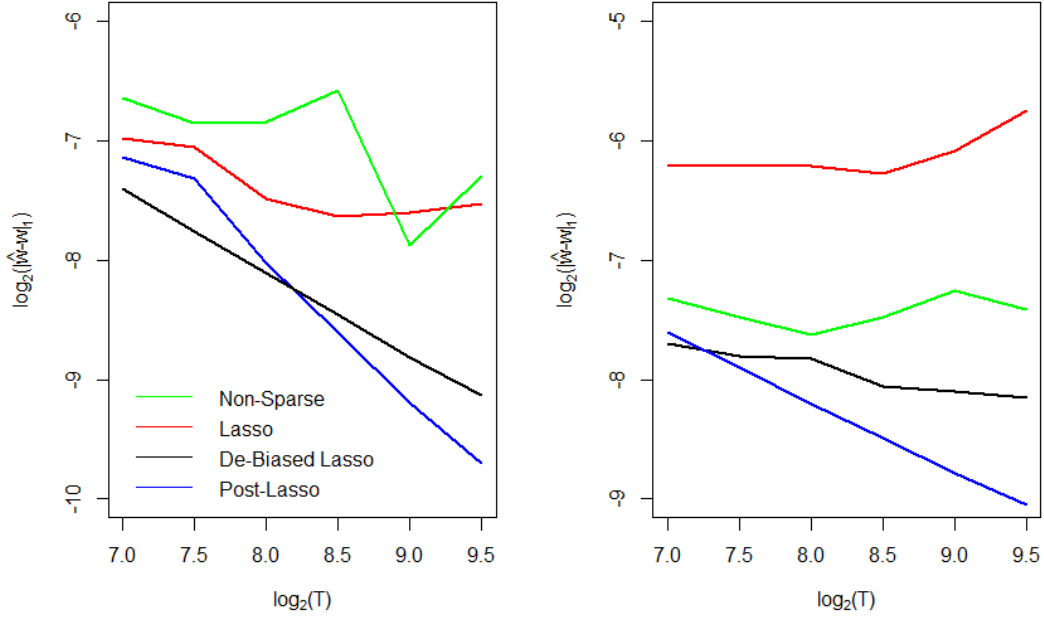

Figure 2: **Elliptical Distribution ($\nu = 4.2$): Averaged errors of the estimators of w_{MRC} for Case 1 on logarithmic scale (left):** $p = T^{0.85}$, $K = 3$ **and for Case 2 on logarithmic scale (right):** $p = 3 \cdot T^{0.85}$, $K = 3$.

	Markowitz (risk-constrained)				Markowitz (weight-constrained)				Global Minimum-Variance			
	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover
Without TC												
EW	0.0081	0.0027	0.1553	-	0.0081	0.0027	0.1553	-	0.0081	0.0027	0.1553	-
Index	0.0063	0.0021	0.1389	-	0.0063	0.0021	0.1389	-	0.0063	0.0021	0.1389	-
MB	0.0539	0.0636	0.2138	-	0.0070	0.0021	0.1539	-	0.0082	0.0020	0.1860	-
FMB (PC1)	0.0324	0.0348	0.1737	-	0.0072	0.0020	0.1616	-	0.0080	0.0019	0.1818	-
FMB (PC2)	0.0287	0.0110	0.2743	-	0.0069	0.0012	0.1968	-	0.0076	0.0012	0.2211	-
FMB (PC3)	0.0228	0.0083	0.2498	-	0.0059	0.0011	0.1817	-	0.0058	0.0011	0.1766	-
FMB (FF1)	0.0497	0.0484	0.2258	-	0.0071	0.0020	0.1582	-	0.0083	0.0019	0.1921	-
FMB (FF3)	0.0384	0.0174	0.2908	-	0.0067	0.0015	0.1754	-	0.0080	0.0013	0.2223	-
FMB (FF5)	0.0373	0.0163	0.2921	-	0.0068	0.0014	0.1788	-	0.0081	0.0013	0.2250	-
With TC												
EW	0.0080	0.0027	0.1538	0.0630	0.0080	0.0027	0.1538	0.0630	0.0080	0.0027	0.1538	0.0630
MB	0.0512	0.0637	0.2027	2.9458	0.0067	0.0021	0.1461	0.3223	0.0080	0.0020	0.1804	0.2152
FMB (PC1)	0.0287	0.0348	0.1541	3.6900	0.0064	0.0020	0.1449	0.7111	0.0073	0.0019	0.1655	0.6783
FMB (PC2)	0.0248	0.0110	0.2368	3.7190	0.0059	0.0012	0.1687	0.9872	0.0067	0.0012	0.1929	0.9686
FMB (PC3)	0.0186	0.0083	0.2044	3.8783	0.0047	0.0011	0.1457	1.1607	0.0047	0.0011	0.1412	1.1448
FMB (FF1)	0.0470	0.0485	0.2136	2.7245	0.0067	0.0020	0.1498	0.3489	0.0080	0.0019	0.1857	0.2486
FMB (FF3)	0.0356	0.0174	0.2694	2.4670	0.0062	0.0015	0.1622	0.4728	0.0076	0.0013	0.2106	0.3920
FMB (FF5)	0.0345	0.0163	0.2699	2.4853	0.0063	0.0015	0.1653	0.4847	0.0076	0.0013	0.2129	0.4057

Table 1: Monthly portfolio returns, variance, Sharpe ratio and turnover. Transaction costs are set to 50 basis points, targeted risk is set at $\sigma = 0.05$ (which is the standard deviation of the monthly excess returns on S&P 500 index from 1980 to 1995, the first training period), monthly targeted return is 0.7974% which is equivalent to 10% yearly return when compounded. In-sample: January 1, 1980 - December 31, 1995 (180 obs), Out-of-sample: January 1, 1995 - December 31, 2019 (300 obs). Green color highlights the best performance in terms of the out-of-sample Sharpe ratio, blue color - second-best performance, yellow - third best, red - the worst performance.

	Markowitz (risk-constrained)				Markowitz (weight-constrained)				Global Minimum-Variance			
	Return	Risk	SR	Turnover	Return	Risk	SR	Turnover	Return	Risk	SR	Turnover
Without TC												
EW	2.33E-04	1.90E-02	0.0123	0.0292	2.33E-04	1.90E-02	0.0123	0.0292	2.33E-04	1.90E-02	0.0123	0.0292
Index	1.86E-04	1.17E-02	0.0159		1.86E-04	1.17E-02	0.0159		1.86E-04	1.17E-02	0.0159	
MB	7.17E-04	4.80E-02	0.0149		1.20E-04	1.28E-02	0.0094		1.71E-04	1.23E-02	0.0139	
FMB (PC1)	6.98E-04	4.88E-02	0.0143		1.18E-04	1.27E-02	0.0093		1.67E-04	1.21E-02	0.0138	
FMB (PC2)	6.97E-04	4.81E-02	0.0145		1.06E-04	1.28E-02	0.0083		1.56E-04	1.22E-02	0.0128	
FMB (FF1)	1.09E-03	3.81E-02	0.0287		1.03E-04	1.70E-02	0.0061		8.13E-05	1.68E-02	0.0048	
FMB (FF3)	9.43E-04	3.62E-02	0.0260		1.02E-04	1.58E-02	0.0065		1.03E-04	1.54E-02	0.0067	
FMB (FF5)	8.31E-04	3.57E-02	0.0232		1.09E-04	1.63E-02	0.0067		1.01E-04	1.61E-02	0.0063	
With TC												
EW	2.01E-04	1.90E-02	0.0106	0.0292	2.01E-04	1.90E-02	0.0106	0.0292	2.01E-04	1.90E-02	0.0106	0.0292
MB	3.52E-04	4.81E-02	0.0073	0.3749	8.03E-05	1.28E-02	0.0063	0.0419	1.39E-04	1.23E-02	0.0113	0.0344
FMB (PC1)	3.27E-04	4.88E-02	0.0067	0.3812	7.69E-05	1.27E-02	0.0060	0.0430	1.33E-04	1.21E-02	0.0110	0.0359
FMB (PC2)	3.26E-04	4.81E-02	0.0068	0.3813	6.15E-05	1.28E-02	0.0048	0.0446	1.19E-04	1.22E-02	0.0098	0.0376
FMB (FF1)	4.97E-04	3.81E-02	0.0130	0.6001	3.58E-06	1.70E-02	0.0002	0.1502	-1.63E-05	1.68E-02	-0.0010	1.3933
FMB (FF3)	3.42E-04	3.63E-02	0.0094	0.6083	-6.43E-06	1.58E-02	-0.0004	0.1148	-2.78E-06	1.54E-02	-0.0002	0.1106
FMB (FF5)	2.21E-04	3.58E-02	0.0062	0.6188	-5.25E-06	1.63E-02	-0.0003	0.1284	-1.04E-05	1.61E-02	-0.0006	0.1232

Table 2: Daily portfolio returns, variance, Sharpe ratio and turnover. Transaction costs are set to 50 basis points, targeted risk is set at $\sigma = 0.013$ (which is the standard deviation of the daily excess returns on S&P 500 index from 2000 to 2002, the first training period), daily targeted return is 0.0378% which is equivalent to 10% yearly return when compounded. In-sample: January 20, 2000 - January 24, 2002 (504 obs), Out-of-sample: January 25, 2002 - January 31, 2020 (4536 obs).

Table 3 (rotated landscape table):

| | De-Biasing | | | | Post-Lasso | | | | | | | | | | | |
| | Markowitz (RC) | | | | Markowitz (RC) | | | | Markowitz (WC) | | | | GMV | | | |
	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover	Return	Variance	SR	Turnover
Without TC																
Lasso (PC0)	0.0007	2.30E-05	0.1406	-												
De-biased Lasso (PC0)	0.0023	0.0001	0.2266	-	0.0287	0.0148	0.2362	-	-0.0174	0.2487	-0.0350	-	-0.0187	0.2441	-0.0379	-
Lasso (PC1)	0.0008	3.46E-05	0.1350	-												
De-biased Lasso (PC1)	0.0091	0.0009	0.3117	-	0.0242	0.0141	0.2034	-	0.0100	0.0043	0.1529	-	0.0085	0.0024	0.1734	-
Lasso (PC2)	0.0006	2.74E-05	0.1122	-												
De-biased Lasso (PC2)	0.0067	0.0007	0.2542	-	0.0290	0.0101	0.2882	-	0.0075	0.0039	0.1205	-	0.0087	0.0021	0.1901	-
Lasso (PC3)	0.0007	4.16E-05	0.1051	-												
De-biased Lasso (PC3)	0.0064	0.0006	0.2603	-	0.0312	0.0125	0.2784	-	0.0076	0.0060	0.0985	-	0.0094	0.0022	0.2014	-
Lasso (FF1)	0.0007	1.49E-05	0.1902	-												
De-biased Lasso (FF1)	0.0109	0.0012	0.3213	-	0.0207	0.0142	0.1738	-	0.0031	0.0609	0.0124	-	-0.0222	0.3657	-0.0367	-
Lasso (FF3)	0.0004	1.58E-05	0.1113	-												
De-biased Lasso (FF3)	0.0072	0.0007	0.2721	-	0.0157	0.0155	0.1263	-	0.0136	0.0133	0.1182	-	-0.0226	0.3665	-0.0373	-
Lasso (FF5)	0.0002	1.78E-05	0.0577	-												
De-biased Lasso (FF5)	0.0073	0.0009	0.2467	-	0.0212	0.0127	0.1879	-	0.0093	0.0048	0.1342	-	0.0094	0.0096	0.0959	-
With TC																
Lasso (PC0)	0.0006	2.38E-05	0.1189	0.0719												
De-biased Lasso (PC0)	0.0020	0.0001	0.1953	0.7952	0.0262	0.0147	0.2155	2.1249	-0.0191	0.2490	-0.0383	1.5373	-0.0199	0.2445	-0.0402	1.0737
Lasso (PC1)	0.0007	3.46E-05	0.1141	0.0984												
De-biased Lasso (PC1)	0.0079	0.0009	0.2725	1.0562	0.0217	0.0141	0.1826	2.2720	0.0084	0.0043	0.1279	1.5709	0.0073	0.0024	0.1497	1.0919
Lasso (PC2)	0.0004	2.74E-05	0.0845	0.1136												
De-biased Lasso (PC2)	0.0055	0.0007	0.2061	1.2113	0.0268	0.0101	0.2668	2.1756	0.0059	0.0039	0.0940	1.5777	0.0076	0.0021	0.1652	1.1026
Lasso (PC3)	0.0005	4.17E-05	0.0792	0.1353												
De-biased Lasso (PC3)	0.0051	0.0006	0.2079	1.2337	0.0288	0.0125	0.2571	2.2449	0.0059	0.0060	0.0762	1.7031	0.0083	0.0022	0.1784	1.0671
Lasso (FF1)	0.0006	1.46E-05	0.1654	0.0789												
De-biased Lasso (FF1)	0.0100	0.0012	0.2949	0.8298	0.0186	0.0142	0.1559	2.1589	0.0013	0.0604	0.0052	1.7374	-0.0234	0.3668	-0.0386	1.1113
Lasso (FF3)	0.0003	1.59E-05	0.0852	0.0785												
De-biased Lasso (FF3)	0.0062	0.0007	0.2352	0.9142	0.0134	0.0155	0.1077	2.2245	0.0120	0.0132	0.1046	1.6208	-0.0236	0.3670	-0.0390	1.0482
Lasso (FF5)	0.0001	1.80E-05	0.0310	0.0861												
De-biased Lasso (FF5)	0.0062	0.0009	0.2124	0.9507	0.0184	0.0126	0.1639	2.2542	0.0076	0.0048	0.1098	1.6033	0.0083	0.0096	0.0844	1.0944

Table 3: Sparse portfolio (FMB is used for de-biasing): monthly portfolio returns, variance, Sharpe ratio and turnover. Transaction costs are set to 50 basis points, targeted risk is set at $\sigma = 0.05$ (which is the standard deviation of the monthly excess returns on S&P 500 index from 1980 to 1995, the first training period), monthly targeted return is 0.7974% which is equivalent to 10% yearly return when compounded. Factor Nodewise-regression estimator of precision matrix is used for de-biasing. Selection threshold for Post-Lasso is $\epsilon = 0.000001$. In-sample: January 1, 1980 - December 31, 1995 (180 obs), Out-of-sample: January 1, 1995 - December 31, 2019 (300 obs).

34

	Markowitz (risk-constrained)			
	Return	Risk	SR	Turnover
Without TC				
Lasso (PC1)	-9.47E-07	2.12E-04	-0.0045	
Debiased Lasso (PC1)	6.88E-05	2.15E-03	0.0320	
Lasso (PC2)	3.86E-07	2.40E-04	0.0016	
Debiased Lasso (PC2)	7.50E-05	2.42E-03	0.0310	
Lasso (FF1)	6.13E-08	5.49E-05	0.0011	
Debiased Lasso (FF1)	5.82E-05	2.28E-03	0.0256	
Lasso (FF3)	2.45E-07	1.05E-04	0.0023	
Debiased Lasso (FF3)	6.12E-05	2.31E-03	0.0265	
Lasso (FF5)	9.87E-09	6.34E-05	0.0002	
Debiased Lasso (FF5)	4.67E-05	2.23E-03	0.0209	
With TC				
Lasso (PC1)	-1.06E-06	2.12E-04	-0.0050	0.0001
Debiased Lasso (PC1)	3.92E-05	2.15E-03	0.0182	0.0293
Lasso (PC2)	2.59E-07	2.40E-04	0.0011	0.0001
Debiased Lasso (PC2)	4.37E-05	2.42E-03	0.0180	0.0311
Lasso (FF1)	-6.13E-09	5.49E-05	-0.0001	0.0001
Factor Debiased Lasso (FF1)	3.12E-05	2.28E-03	0.0137	0.0266
Lasso (FF3)	1.53E-07	1.05E-04	0.0015	0.0001
Debiased Lasso (FF3)	3.25E-05	2.31E-03	0.0141	0.0286
Lasso (FF5)	-6.47E-08	6.34E-05	-0.0010	0.0001
Debiased Lasso (FF5)	1.97E-05	2.23E-03	0.0088	0.0268

Table 4: Sparse Portfolio (FMB is used for de-biasing): daily portfolio returns, variance, Sharpe ratio and turnover. Transaction costs are set to 50 basis points, targeted risk is set at $\sigma = 0.013$ (which is the standard deviation of the daily excess returns on S&P 500 index from 2000 to 2002, the first training period), daily targeted return is 0.0378% which is equivalent to 10% yearly return when compounded. In-sample: January 20, 2000 - January 24, 2002 (504 obs), Out-of-sample: January 25, 2002 - January 31, 2020 (4536 obs)..

	Asian and Russian Financial Crisis (1997-1998)		Argentine Great Depression and Dot-com Bubble (1999-2002)		Financial Crisis (2007-2009)		Ukrainian Crisis and Chinese Stock Market Crash (2013-2015)	
	CER	Risk	CER	Risk	CER	Risk	CER	Risk
Markowitz Risk-Constrained (MRC)								
EW	0.2712	0.0547	-0.0322	0.0519	-0.4987	0.1203	0.2343	0.0344
Index	0.3222	0.0508	-0.1698	0.0539	-0.4924	0.0929	0.2155	0.0364
MB	2.1662	0.3381	-0.1140	0.2916	-3.0688	0.5101	0.9755	0.0847
FMB (PC1)	1.3285	0.0892	0.4241	0.1297	-3.0470	0.4735	1.2022	0.1658
FMB (PC2)	1.3153	0.0883	0.5016	0.1286	0.1312	0.1219	1.0534	0.1359
FMB (FF1)	2.0379	0.3029	0.0861	0.2660	-2.7247	0.4301	1.5366	0.1751
Global Minimum-Variance Portfolio (GMV)								
MB	0.2791	0.0496	-0.0470	0.0476	-0.4637	0.1015	0.1601	0.0340
FMB (PC1)	0.3960	0.0374	-0.1224	0.0510	-0.4588	0.0987	0.2656	0.0337
FMB (PC2)	0.4117	0.0364	-0.1227	0.0505	-0.3444	0.0393	0.3100	0.0321
FMB (FF1)	0.2784	0.0487	-0.0396	0.0468	-0.4570	0.0986	0.2640	0.0332
Markowitz Weight-Constrained (MWC)								
MB	0.1890	0.0496	-0.1235	0.0524	-0.5195	0.1018	0.2591	0.0339
FMB (PC1)	0.3424	0.0385	-0.1534	0.0520	-0.5147	0.0992	0.2638	0.0336
FMB (PC2)	0.3626	0.0373	-0.1550	0.0513	-0.4044	0.0440	0.3061	0.0322
FMB (FF1)	0.1906	0.0488	-0.1207	0.0517	-0.5106	0.0996	0.2613	0.0333
Sparse-Debiased MRC								
Lasso (PC1)	0.0380	0.0052	0.0186	0.0055	-0.0007	0.0185	-0.0014	0.0027
Debiased Lasso (PC1)	0.3553	0.0300	0.0674	0.0251	0.0440	0.0341	0.2066	0.0153
Lasso (PC2)	0.0378	0.0053	0.0356	0.0053	-0.0777	0.0130	0.0052	0.0038
Debiased Lasso (PC2)	0.2962	0.0261	0.0540	0.0304	0.1129	0.0408	0.1567	0.0217
Lasso (FF1)	0.0297	0.0039	0.0152	0.0039	0.0564	0.0100	0.0179	0.0035
Debiased Lasso (FF1)	0.4149	0.0277	-0.0852	0.0275	-0.0258	0.0230	0.1681	0.0240
Lasso (FF3)	0.0335	0.0041	0.0079	0.0044	-0.0328	0.0098	0.0167	0.0035
Debiased Lasso (FF3)	0.2123	0.0142	-0.0640	0.0300	-0.0406	0.0202	0.1782	0.0186

Table 5: Cumulative excess return (CER) and risk of non-sparse MRC portfolios using monthly data. Transaction costs are set to 50 basis points, targeted risk is set at $\sigma = 0.05$ (which is the standard deviation of the monthly excess returns on S&P 500 index from 1980 to 1995, the first training period), monthly targeted return is 0.7974% which is equivalent to 10% yearly return when compounded. Factor Nodewise-regression estimator of precision matrix is used for de-biasing. Selection threshold for Post-Lasso is $\epsilon = 0.000001$. In-sample: January 1, 1980 - December 31, 1995 (180 obs), Out-of-sample: January 1, 1995 - December 31, 2019 (300 obs).

Appendices

The first part of Theorem 2 was proved in Fan et al. (2018) (see their proof of Theorem 2.1) under the assumptions (A.1)-(A.3), (B.1)-(B.3) and $\log(p) = o(T)$. To prove the convergence rates for the precision matrix of the factor-adjusted returns, we follow Chang et al. (2018), Caner and Kock (2018) and Callot et al. (2019). Using the facts that $\max_{i \leq p}(1/T) \sum_{t=1}^{T} |\hat{\varepsilon}_{it} - \varepsilon_{it}| = \mathcal{O}_p(\omega_T^2)$ and $\max_{i,t} |\hat{\varepsilon}_{it} - \varepsilon_{it}| = \mathcal{O}_p(\omega_T) = o_p(1)$, we get

$$\max_{1 \leq j \leq p} \|\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j\|_1 = \mathcal{O}_p(\bar{s}\omega_T), \tag{A.1}$$

where $\hat{\boldsymbol{\gamma}}_j$ was defined in (3.1). The proof of A.1 is similar to the proof of the equation (23) of Chang et al. (2018), with $\omega_T = \sqrt{\log(p)/T}$ for their case. Similarly to Callot et al. (2019), consider the following linear model:

$$\hat{\boldsymbol{\varepsilon}}_j = \hat{\mathbf{E}}_{-j}\boldsymbol{\gamma}_j + \boldsymbol{\eta}_j, \text{ for } j = 1, \dots, p, \tag{A.2}$$

$$\mathbb{E}\left[\boldsymbol{\eta}_j'\hat{\mathbf{E}}_{-j}\right] = 0.$$

van de Geer et al. (2014) and Chang et al. (2018) showed that

$$\max_{1 \leq j \leq p} \left\|\boldsymbol{\eta}_j'\hat{\mathbf{E}}_{-j}\right\|_\infty /T = \mathcal{O}_p(\omega_T). \tag{A.3}$$

Let $\tau_j^2 := \mathbb{E}\left[\boldsymbol{\eta}_j'\boldsymbol{\eta}_j\right]$, then we have

$$\max_{1 \leq j \leq p} \left\|\boldsymbol{\eta}_j'\boldsymbol{\eta}_j/T - \tau_j^2\right\| = \mathcal{O}_p(\omega_T). \tag{A.4}$$

Note that the rate in (A.4) is the same as in Lemma 1 of Chang et al. (2018) with $\omega_T = \sqrt{\log(p)/T}$ for their case. However, the rate in (A.4) is different from the one derived in van de Geer et al. (2014) since we allow time-dependence between factor-adjusted returns.

Recall that $\hat{\tau}_j^2 = \left\|\hat{\boldsymbol{\varepsilon}}_j - \hat{\mathbf{E}}_{-j}\hat{\boldsymbol{\gamma}}_j\right\|_2^2 /T + \lambda_j \|\hat{\boldsymbol{\gamma}}_j\|_1$. Using triangle inequality, we have:

$$\max_{1 \leq j \leq p} \left|\hat{\tau}_j^2 - \tau_j^2\right| \leq \underbrace{\max_{1 \leq j \leq p} \left|\boldsymbol{\eta}_j'\boldsymbol{\eta}_j/T - \tau_j^2\right|}_{\text{I}} + \underbrace{\max_{1 \leq j \leq p} \left|\boldsymbol{\eta}_j'\hat{\mathbf{E}}_{-j}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j)/T\right|}_{\text{II}}$$

$$+ \underbrace{\max_{1 \leq j \leq p} \left|\boldsymbol{\eta}_j'\hat{\mathbf{E}}_{-j}\boldsymbol{\gamma}_j/T\right|}_{\text{III}} + \underbrace{\max_{1 \leq j \leq p} \boldsymbol{\gamma}_j'\hat{\mathbf{E}}_{-j}'\hat{\mathbf{E}}_{-j}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j)/T}_{\text{IV}}.$$

The first term was bounded in A.4, we now bound the remaining terms:

$$\text{II} \leq \max_{1 \leq j \leq p} \left\|\boldsymbol{\eta}_j'\hat{\mathbf{E}}_{-j}/T\right\|_\infty \max_{1 \leq j \leq p} \|\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j\|_1 = \mathcal{O}_p(\bar{s}\omega_T^2),$$

where we used A.1 and A.3. For III we have

$$\text{III} \leq \max_{1 \leq j \leq p} \left\|\boldsymbol{\eta}_j'\hat{\mathbf{E}}_{-j}/T\right\|_\infty \max_{1 \leq j \leq p} \|\boldsymbol{\gamma}_j\|_1 = \mathcal{O}_p(\sqrt{\bar{s}}\omega_T),$$

where we used A.3 and the fact that $\|\boldsymbol{\gamma}_j\|_1 \leq \sqrt{s_j}\|\boldsymbol{\gamma}_j\|_2 = \mathcal{O}(\sqrt{s_j})$. To bound the last term, we use KKT conditions in node-wise regression:

$$\max_{1\leq j\leq p}\left\|\widehat{\mathbf{E}}'_{-j}\widehat{\mathbf{E}}_{-j}(\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j)/T\right\|_\infty \leq \max_{1\leq j\leq p}\left\|\widehat{\mathbf{E}}'_{-j}\boldsymbol{\eta}_j/T\right\|_\infty + \max_{1\leq j\leq p}\lambda_j = \mathcal{O}_p(\omega_T),$$

where we used A.3 and $\lambda_j \asymp \omega_T$. It follows that

$$\mathrm{IV} = \mathcal{O}_p(\omega_T)\max_{1\leq j\leq p}\|\boldsymbol{\gamma}_j\|_1 = \mathcal{O}_p(\sqrt{\bar{s}}\omega_T).$$

Therefore, we now have shown that

$$\max_{1\leq j\leq p}\left|\hat{\tau}_j^2 - \tau_j^2\right| = \mathcal{O}_p(\sqrt{\bar{s}}\omega_T). \tag{A.5}$$

Using the fact that $1/\tau_j^2 = \mathcal{O}(1)$, we also have

$$1/\hat{\tau}_j^2 - 1/\tau_j^2 = \mathcal{O}_p(\sqrt{\bar{s}}\omega_T). \tag{A.6}$$

Finally, using the analysis in (B.51)-(B.53) of Caner and Kock (2018), we get

$$\max_{1\leq j\leq p}\left\|\widehat{\boldsymbol{\Theta}}_{\varepsilon,j} - \boldsymbol{\Theta}_{\varepsilon,j}\right\|_1 = \mathcal{O}_p(s_T\omega_T). \tag{A.7}$$

To prove the second rate for the precision of the factor-adjusted returns, we note that

$$\max_{1\leq j\leq p}\|\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j\|_2 = \mathcal{O}_p(\sqrt{\bar{s}}\omega_T), \tag{A.8}$$

which was obtained in Chang et al. (2018) (see their Lemma 2). We can write

$$\max_{1\leq j\leq p}\left\|\widehat{\boldsymbol{\Theta}}_{\varepsilon,j} - \boldsymbol{\Theta}_{\varepsilon,j}\right\|_2 \leq \max_{1\leq j\leq p}[\|\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j\|_2/\hat{\tau}_j^2 + \|\boldsymbol{\gamma}_j\|_2 1/\hat{\tau}_j^2 - 1/\tau_j^2] = \mathcal{O}_p(\sqrt{\bar{s}}\omega_T). \tag{A.9}$$

A.2 Proof of Theorem 3

Let $\widehat{\mathbf{J}} = \widehat{\boldsymbol{\Lambda}}^{1/2}\widehat{\boldsymbol{\Gamma}}'\widehat{\boldsymbol{\Theta}}_\varepsilon\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Lambda}}^{1/2}$ and $\widetilde{\mathbf{J}} = \widetilde{\boldsymbol{\Lambda}}^{1/2}\widetilde{\boldsymbol{\Gamma}}'\boldsymbol{\Theta}_\varepsilon\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}^{1/2}$. Also, define

$$\Delta_{\mathrm{inv}} = \widehat{\boldsymbol{\Theta}}_\varepsilon\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Lambda}}^{1/2}(\mathbf{I}_K + \widehat{\mathbf{J}})^{-1}\widehat{\boldsymbol{\Lambda}}^{1/2}\widehat{\boldsymbol{\Gamma}}'\widehat{\boldsymbol{\Theta}}_\varepsilon - \boldsymbol{\Theta}_\varepsilon\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}^{1/2}(\mathbf{I}_K + \widetilde{\mathbf{J}})^{-1}\widetilde{\boldsymbol{\Lambda}}^{1/2}\widetilde{\boldsymbol{\Gamma}}'\boldsymbol{\Theta}_\varepsilon.$$

Using Sherman-Morrison-Woodbury formulas in 3.9, we have

$$\left\|\left\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\right\|\right\|_1 \leq \left\|\left\|\widehat{\boldsymbol{\Theta}}_\varepsilon - \boldsymbol{\Theta}_\varepsilon\right\|\right\|_1 + \|\|\Delta_{\mathrm{inv}}\|\|_1. \tag{A.10}$$

As pointed out by Fan et al. (2018), $\|\|\Delta_{\mathrm{inv}}\|\|_1$ can be bounded by the following three terms:

$$\left\|\left\|(\widehat{\boldsymbol{\Theta}}_\varepsilon - \boldsymbol{\Theta}_\varepsilon)\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}^{1/2}(\mathbf{I}_K + \widetilde{\mathbf{J}})^{-1}\widetilde{\boldsymbol{\Lambda}}^{1/2}\widetilde{\boldsymbol{\Gamma}}'\boldsymbol{\Theta}_\varepsilon\right\|\right\|_1 = \mathcal{O}_p(\bar{s}\omega_T \cdot p \cdot \frac{1}{p} \cdot \sqrt{\bar{s}}),$$

$$\left\|\left\|\boldsymbol{\Theta}_\varepsilon(\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Lambda}}^{1/2} - \widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}^{1/2})(\mathbf{I}_K + \widetilde{\mathbf{J}})^{-1}\widetilde{\boldsymbol{\Lambda}}^{1/2}\widetilde{\boldsymbol{\Gamma}}'\boldsymbol{\Theta}_\varepsilon\right\|\right\|_1 = \mathcal{O}_p(\sqrt{\bar{s}} \cdot p\omega_T \cdot \frac{1}{p} \cdot \sqrt{\bar{s}}),$$

$$\left\|\left\|\boldsymbol{\Theta}_\varepsilon\widetilde{\boldsymbol{\Lambda}}^{1/2}\widetilde{\boldsymbol{\Gamma}}'((\mathbf{I}_K + \widehat{\mathbf{J}})^{-1} - (\mathbf{I}_K + \widetilde{\mathbf{J}})^{-1})\widetilde{\boldsymbol{\Gamma}}'\boldsymbol{\Theta}_\varepsilon\right\|\right\|_1 = \mathcal{O}_p(\sqrt{\bar{s}} \cdot \frac{1}{p} \cdot p\bar{s}\omega_T\sqrt{\bar{s}}).$$

To derive the above rates we used (B.1)-(B.3), Theorem 2 and the fact that $\left\|\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Gamma}}' - \mathbf{BB}'\right\|_{\mathrm{F}} = \mathcal{O}_p(p\omega_T)$. The second rate in Theorem 3 can be easily obtained using the technique described above for the l_2-norm.

A.3 Lemmas for Theorem 4

Lemma 1. *Under the assumptions of Theorem 3,*

(a) $\|\widehat{\mathbf{m}} - \mathbf{m}\|_{max} = \mathcal{O}_p(\sqrt{\log(p)/T})$, *where* $\mathbf{m}$ *is the unconditional mean of stock returns defined in Subsection 3.2, and* $\widehat{\mathbf{m}}$ *is the sample mean.*

(b) $\|\!|\boldsymbol{\Theta}|\!\|_1 = \mathcal{O}(\bar{s})$.

Proof.

(a) The proof of Part (a) is provided in Chang et al. (2018) (Lemma 1).

(b) To prove Part (b) we use Sherman-Morrison-Woodbury formula in 3.9:

$$\|\!|\boldsymbol{\Theta}|\!\|_1 \leq \|\!|\boldsymbol{\Theta}_\varepsilon|\!\|_1 + \|\!|\boldsymbol{\Theta}_\varepsilon \mathbf{B}[\mathbf{I}_K + \mathbf{B}'\boldsymbol{\Theta}_\varepsilon \mathbf{B}]^{-1}\mathbf{B}'\boldsymbol{\Theta}_\varepsilon|\!\|_1$$
$$= \mathcal{O}(\sqrt{\bar{s}}) + \mathcal{O}(\sqrt{\bar{s}} \cdot p \cdot \frac{1}{p} \cdot \sqrt{\bar{s}}) = \mathcal{O}(\bar{s}). \qquad (A.11)$$

The last equality in A.11 is obtained under the assumptions of Theorem 4. This result is important in several aspects: it shows that the sparsity of the precision matrix of stock returns is controlled by the sparsity in the precision of the idiosyncratic returns. Hence, one does not need to impose an unrealistic sparsity assumption on the precision of returns a priori when the latter follow a factor structure - sparsity of the precision once the common movements have been taken into account would suffice.

$\square$

Lemma 2. *Define* $\theta = \mathbf{m}'\boldsymbol{\Theta}\mathbf{m}/p$ *and* $\widehat{\theta} = \widehat{\mathbf{m}}'\widehat{\boldsymbol{\Theta}}\widehat{\mathbf{m}}/p$. *Under the assumptions of Theorem 3:*

(a) $\theta = \mathcal{O}(1)$.

(b) $\left|\widehat{\theta} - \theta\right| = \mathcal{O}_p(\bar{s}^2\omega_T) = o_p(1)$.

(c) $|\widehat{r}_c - r_c| = \mathcal{O}_p(\bar{s}^2\omega_T) = o_p(1)$, *where* r_c *was defined in* (2.17).

Proof.

(a) Part (a) is trivial and follows directly from $\|\!|\boldsymbol{\Theta}|\!\|_2 = \mathcal{O}(1)$.

(b) First, rewrite the expression of interest:

$$\widehat{\theta} - \theta = [(\widehat{\mathbf{m}} - \mathbf{m})'(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})(\widehat{\mathbf{m}} - \mathbf{m})]/p + [(\widehat{\mathbf{m}} - \mathbf{m})'\boldsymbol{\Theta}(\widehat{\mathbf{m}} - \mathbf{m})]/p$$
$$+ [2(\widehat{\mathbf{m}} - \mathbf{m})'\boldsymbol{\Theta}\mathbf{m}]/p + [2\mathbf{m}'(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})(\widehat{\mathbf{m}} - \mathbf{m})]/p$$
$$+ [\mathbf{m}'(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})\mathbf{m}]/p. \qquad (A.12)$$

We now bound each of the terms in A.12 using the expressions derived in Callot et al. (2019) (see their Proof of Lemma A.3), Lemma 1 and the fact that $\log(p)/T = o(1)$.

$$\left|(\widehat{\mathbf{m}} - \mathbf{m})'(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})(\widehat{\mathbf{m}} - \mathbf{m})\right|/p \leq \|\widehat{\mathbf{m}} - \mathbf{m}\|_{max}^2 \|\!|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}|\!\|_1$$
$$= \mathcal{O}_p\Big(\frac{\log(p)}{T} \cdot \bar{s}^2\omega_T\Big) \qquad (A.13)$$

39

$$\left|(\widehat{\mathbf{m}} - \mathbf{m})'\boldsymbol{\Theta}(\widehat{\mathbf{m}} - \mathbf{m})\right|/p \le \|\widehat{\mathbf{m}} - \mathbf{m}\|_{\max}^2 \||\boldsymbol{\Theta}\||_1 = \mathcal{O}_p\Big(\frac{\log(p)}{T} \cdot \bar{s}\Big). \tag{A.14}$$

$$\left|(\widehat{\mathbf{m}} - \mathbf{m})'\boldsymbol{\Theta}\mathbf{m}\right|/p \le \|\widehat{\mathbf{m}} - \mathbf{m}\|_{\max} \||\boldsymbol{\Theta}\||_1 = \mathcal{O}_p\Big(\sqrt{\frac{\log(p)}{T}} \cdot \bar{s}\Big). \tag{A.15}$$

$$\left|\mathbf{m}'(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})(\widehat{\mathbf{m}} - \mathbf{m})\right|/p \le \|\widehat{\mathbf{m}} - \mathbf{m}\|_{\max}\left\||\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\right\||_1$$
$$= \mathcal{O}_p\Big(\sqrt{\frac{\log(p)}{T}} \cdot \bar{s}^2\omega_T\Big). \tag{A.16}$$

$$\left|\mathbf{m}'(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})\mathbf{m}\right|/p \le \left\||\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\right\||_1 = \mathcal{O}_p\Big(\bar{s}^2\omega_T\Big). \tag{A.17}$$

(c) Part (c) trivially follows from Part (b).

$\square$

A.4 Proof of Theorem 4

The KKT conditions for the nodewise Lasso in (3.1) imply that

$$\hat{\tau}_j^2 = (\widehat{\boldsymbol{\varepsilon}}_j - \widehat{\mathbf{E}}_{-j}\widehat{\boldsymbol{\gamma}}_j)'\widehat{\boldsymbol{\varepsilon}}_j/T,$$

hence,

$$\widehat{\boldsymbol{\varepsilon}}_j'\widehat{\mathbf{E}}\widehat{\boldsymbol{\Theta}}_{\varepsilon,j}'/T = 1.$$

As shown in van de Geer et al. (2014), these KKT conditions also imply that

$$\left\|\widehat{\mathbf{E}}_{-j}'\widehat{\mathbf{E}}\widehat{\boldsymbol{\Theta}}_{\varepsilon,j}\right\|_\infty/T \le \lambda_j/\hat{\tau}_j^2. \tag{A.18}$$

Therefore, the estimator of precision matrix needs to satisfy the following "extended KKT" condition:

$$\left\|\widehat{\boldsymbol{\Sigma}}_\varepsilon\widehat{\boldsymbol{\Theta}}_{\varepsilon,j}' - \mathbf{e}_j\right\|_\infty \le \lambda_j/\hat{\tau}_j^2, \tag{A.19}$$

where $\mathbf{e}_j$ is the j-th unit column vector. Combining the rate in l_1 norm in Theorem 3 and (A.19), we have:

$$\left\|\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Theta}}_j' - \mathbf{e}_j\right\|_\infty \le \lambda_j/\hat{\tau}_j^2, \tag{A.20}$$

Using the definition of Δ in (2.23), it is straightforward to see that

$$\|\Delta\|_\infty/\sqrt{T} = \left\|(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}} - \mathbf{I}_p)(\widehat{\mathbf{w}} - \mathbf{w})\right\|_\infty \le \left\|\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}} - \mathbf{I}_p\right\|_\infty\|\widehat{\mathbf{w}} - \mathbf{w}\|_1. \tag{A.21}$$

Therefore, combining (A.20) and (A.21), we have

$$\|\Delta\|_\infty \le \sqrt{T}\|\widehat{\mathbf{w}} - \mathbf{w}\|_1 \max_j \lambda_j/\hat{\tau}_j^2 = \mathcal{O}_p(\sqrt{T} \cdot \Big(s_0 \vee \bar{s}^2\Big)\omega_T \cdot \omega_T) \tag{A.22}$$

$$= \mathcal{O}_p(\Big(s_0 \vee \bar{s}^2\Big)\log(p)/\sqrt{T}) = o_p(1). \tag{A.23}$$

40

Finally, we show that $\left\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Theta}\right\|_\infty = o_p(1)$. Using Theorem 3 and Lemma 1 we have $\left\|\widehat{\boldsymbol{\Theta}}_j\right\|_1 = \mathcal{O}_p(s_j)$. Also,

$$\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Theta}}' = (\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}} - \mathbf{I}_p)\widehat{\boldsymbol{\Theta}}' + \widehat{\boldsymbol{\Theta}}'. \tag{A.24}$$

And using A.20 and A.21 together with $\max_j \lambda_j s_j^2 = o_p(1)$:

$$\left\|(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}} - \mathbf{I}_p)\widehat{\boldsymbol{\Theta}}'\right\|_\infty \leq \max_j \lambda_j \left\|\widehat{\boldsymbol{\Theta}}_j\right\|_1 / \hat{\tau}_j^2 = o_p(1). \tag{A.25}$$

It follows that

$$\left\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\right\|_\infty \leq \max_j \left\|\widehat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j\right\|_2 \leq \max_j \lambda_j \sqrt{s_j} = o_p(1). \tag{A.26}$$

Combining A.24-A.26 completes the proof.