Large Portfolios under Flexible Short-Sale Constraints: Optimality, Estimation Accuracy, and Beating $1/N^*$

Han Tong[†] **Ming Yuan**[‡] **Guofu Zhou**[§] **Jason Zhou**[¶]

First version: February 2024; Current version: January 2026

*We are grateful to Andrew Detzel, Dashan Huang, Raymond Kan, Ralph Koijen, Nathan Lassance, Yingying Li, Hong Liu, Asaf Manela, Ľuboš Pástor, Dacheng Xiu, seminar participants at Shanghai Jiaotong University, Tsinghua University, Xi'an Jiaotong-Liverpool University, Washington University in St. Louis, Zhejiang University, and conference participants at 2024 Greater China Finance Conference at Xiamen University and 2025 SoFiE Financial Machine Learning Summer School at Yale University for very helpful comments.

Send correspondence to Guofu Zhou, Olin School of Business, Washington University in St. Louis, St. Louis, MO 63130; e-mail: zhou@wustl.edu; phone: 314-935-6384.

[†]Columbia University

[‡]Columbia University

[§]Washington University in St. Louis

[¶]University of Chicago

Large Portfolios under Flexible Short-sale Constraints: Optimality, Estimation Accuracy, and Beating $1/N$

Abstract

We show that the optimal portfolio under economically meaningful short-sale constraints—including no short selling—converges to the unconstrained optimal portfolio in large asset markets, provided returns admit a factor structure. Thus, short-sale constraints impose no asymptotic efficiency loss despite restricting positions. Moreover, such constraints substantially reduce estimation error: the estimated constrained optimal portfolio converges, allowing the stochastic discount factor (SDF) to be meaningfully recovered in high dimensions, whereas the unconstrained estimator may diverge or imply unrealistically large short positions. As a result, the constrained optimal portfolio generally outperforms the 1/N rule in large-N settings.

JEL classification: G11, G12

Keywords: No-short-sale, Optimal portfolio, mean-variance efficiency, diversification

1 Introduction

Mean–variance portfolio theory (Markowitz, 1952) is the foundation of various factor models in finance and remains a widely used framework in practice.[1] However, to implement the theory, both the asset expected returns and covariance matrix must be estimated, introducing the well-known estimation risk problem in portfolio selection. Since Brown (1976), Bawa, Brown, and Klein (1979) and Jorion (1986), a range of solutions have been proposed by various researchers such as MacKinlay and Pástor (2000), Ledoit and Wolf (2004) and Kan and Zhou (2007). Yuan and Zhou (2024) and the references therein provide the latest developments. However, in spite of these advances, there is a major problem. The Markowitz's mean-variance portfolio, along with almost all the portfolio theories proposed later to mitigate estimation risk, performs poorly in high dimension cases, in which the ratio of the number of assets (N) to sample size (T) is large and N is large.

Recent asset pricing studies, such as Kozak, Nagel, and Santosh (2020), Ao, Li, and Zheng (2019), Chen, Ding, Li, and Zheng (2025) and Luo, Zhou, and Zhou (2025), search for the stochastic discount factor (SDF) in high dimensions to explain asset prices, which is equivalent to estimating Markowitz's mean-variance optimal portfolio. Since the sample optimal portfolio fails to converge in high dimensions, accurately estimating the SDF becomes a central challenge. More importantly, existing studies do not impose realistic constraints on the short-sale positions, so that the resulting SDF may not be feasible in the real world.

To mitigate large estimation errors in the global minimum variance (GMV) portfolio, Jagannathan and Ma (2003) propose the novel idea of imposing the "wrong constraints" of no short-sales, which substantially improves estimation accuracy. DeMiguel, Garlappi, Nogales, and Uppal (2009) extend this approach by generalizing the no-short-sale constraint to a norm restriction. Fan, Zhang, and Yu (2012) show that, under a norm restriction, the risk of the constrained estimator converges to its population counterpart. Insights from these studies raise the question of whether imposing economically meaningful constraints can improve estimation accuracy for the SDF.

In this paper, we study the properties of optimal portfolios under both a no-short-sale

[1] Grinold and Kahn (1999), Chincarini and Kim (2023) and references therein.

restriction and a flexible form of limited short-selling. Instead of studying the GMV, we focus on the optimal portfolio itself and establish three main results.

First, we establish a surprising property of the optimal portfolio under no-short-sale constraints. In a factor model, such as Ross (1976) in which idiosyncratic risks are diversifiable, as the number of assets grows, the Sharpe ratio of the constrained portfolio converges to that of the true optimal portfolio of all the assets, which typically involves a lot of short positions.[2] In other words, even when the true optimal portfolio involves substantial short positions, a constrained investor can asymptotically achieve the same efficiency in large markets. This result overturns the conventional view that short-sale constraints necessarily impose a first-order efficiency loss. There are two important implications. (i) Long-only institutional investors can still approximate the efficient portfolio in large markets. (ii) It could support Koijen and Yogo (2019) demand-based asset pricing theory despite its no-short-sale assumption,[3] and shows that the implied pricing kernel can be made arbitrarily close to the true kernel that may include short positions. Thus, it makes sense to use institutional holdings to recover the pricing kernel.

Second, we show that imposing a flexible form of short-sale constraints resolves resolve a central econometric failure of the Markowitz portfolio. Given an arbitrary degree of short-sale level $\zeta \geq 0$, we allow the fraction of the total short positions to the total portfolio value to not exceed it. When $\zeta = 0$, the restriction reduces to the no-short-sale case, while letting $\zeta \to \infty$, it recovers the unconstrained case. When $\zeta = 0.30$, it corresponds to the popular 130/30 portfolios widely used by quantitative managers, in which 30% of short positions relative to the total portfolio are permitted.[4] Without such a constraint, existing SDF portfolios will often have too large short positions which are difficult or impossible to implement in practice.

[2] Brennan and Lo (2010) show that, under certain conditions, about one-eighth of the assets in the optimal portfolio have negative weights.

[3] Their theory stimulates a growing literature. See Gabaix and Koijen (2021), Koijen, Richmond, and Yogo (2024), Davis (2025), Haddad, Huebner, and Loualiche (2025), Haddad, He, Huebner, Kondor, and Loualiche (2025), van Binsbergen, David, and Opp (2025) and references therein.

[4] The 130/30 portfolio strategy gained significant traction among institutional investors in the early to mid-2000s. Recently, some, as indicated by an AQR report, even discussed the use of 150/50.

As long as ζ is given and fixed, we show that the Sharpe ratio of the estimated optimal portfolio will converge to that of its population counterpart under the same constraint. This is important, because it is known that the unconstrained sample optimal portfolio, the sample Markowitz, fails to converge at all when the number of assets is large relative to the sample size (N/T is close to one or above it). Our convergence result holds regardless of whether there is a factor structure or not. However, if a factor structure with exact pricing exists, then combined with our earlier result, this implies that the constrained estimated portfolio also converges to the true optimal portfolio in terms of the Sharpe ratio, or equivalently, to the pricing kernel (SDF) that prices all assets, provided that the factor model residuals are sufficiently diversifiable.

Third, we show that the flexible no-short-sale constrained portfolio generally outperforms the $1/N$ rule, which invests equally across risky assets, even in the large assets case. When N is small or fixed, the standard sample optimal portfolio will eventually dominate it. The problem is when N is large relative to T. DeMiguel, Garlappi, and Uppal (2009) provide extensive empirical evidence that existing estimated strategies rarely outperform $1/N$. Theoretically, Yuan and Zhou (2024) demonstrate that $1/N$ is also difficult to beat in high dimensions, as it is in fact optimal when returns follow a one-factor model. Surprisingly, however, the ζ constrained optimal portfolio analyzed in this paper performs theoretically no worse than the $1/N$ in the one-factor model, and outperforms in general, as long as both N and T are large.

In simulations, we confirm our first result. Under reasonable parameter assumptions, the rate of convergence to the true optimal and unconctrained portfolio is remarkably fast, even there are only 50 assets, in a 2-factor model, whether there are no-short sells or 30% short positions. The Sharpe ratio difference is less than 1%. In a 5-factor model, similar fast convergence is also observed.[5]

Empirically, our paper appears one of rare studies that examines how the optimal portfolio varies over a varying degree of short-sales. With an array of assumptions on the assets, from 100 portfolios formed on size and operating profitability or 140 anomalies, we find that

[5]The results are consistent with Yuan and Zhou (2024) who show that 20 or 30 assets perform closely to S&P500 portfolio.

indeed imposing the short-sale constraints help the converegence of the estimation optimal portfolios. If no constraints are imposed, the portfolio performs poorly in terms of Sharpe ratio, and sometime have short position over 200% of the total asset value on an individual asset, which is arguable ridiculous in practice. In our set-up, the solution is well behaved under no short-sells, and have much larger Sharpe ratios. If 30% short-sale positions are allowed, the results are generally improved. It says that a small degree of short-sales do tend to provide more flexibility in finite sample (N), but asymptotically it should not make any difference. However, it does indicate that the 130/30 policy in the real world is likely to be productive.

Our paper contributes to the large literature on portfolio selection under estimation risk. DeMiguel, Garlappi, and Uppal (2009), Duchin and Levy (2009), Kirby and Ostdiek (2012), DeMiguel, Martín-Utrera, and Nogales (2015), DeMiguel, Martín-Utrera, Nogales, and Uppal (2020), Shi, Shu, Yang, and He (2020), Kan, Wang, and Zhou (2022), Lassance, Martín-Utrera, and Simaan (2024), and Lassance, Vanderveken, and Vrins (2024), among others, propose various portfolio strategies that improve the performance of the mean-variance portfolio. In contrast to these studies, we show that the mean-variance portfolio can work well in high dimensions once we impose no-short-sale or the general ζ-constraint.

Our paper is related to Ao, Li, and Zheng (2019), who were the first to propose an estimated optimal portfolio that converges in high dimensions. However, our approach differs from theirs. Their approach is a two-step one. First, they use the LASSO to reduce dimensionality. Second, they apply the standard sample optimal portfolio to the selected assets. Hence, they require a sparsity assumption in their first step, and they do not impose restrictions in the second step. In contrast, our approach is a one-step one. We apply the optimization directly to all the assets even when $N > T$ as long as $\log(N)/T$ is small. The latter is required theoretically by almost all studies, including both theirs and ours. Moreover, our approach imposes the limited shorting constraint, so that the resulting portfolio will be guaranteed to not have extreme short-selling positions beyond what controlled by ζ.

Our paper also contributes to the extensive literature on short selling. Short selling has long been recognized as central to understanding market efficiency and anomalies (see, e.g., Stambaugh, Yu, and Yuan 2012), as well as to demand-based asset pricing (Koijen and

Yogo, 2019). We show that it can be extended to allow a limited level and this plays also a crucial role in reducing estimation errors of the optimal portfolio and the SDF. Moreover, the insights of this paper may be extended to settings where the effects of short selling can be replicated without actual short sales, although we leave a deep exploration of these applications for future research.

The rest of the paper is organized as follows. In Section 2, we discuss briefly the problems of the estimated mean-variance portfolio. In Section 3, we analyze the properties of both population and estimated optimal portfolios under no-short-sale restriction. In Sections 4 and 5, we extend the restriction and discuss the relation to the GMV. Section 5 provides simulation and empirical results. Section 6 concludes.

2 Problems of Unconstrained Portfolio

To understand the importance of the constrained estimation for the optimal portfolio, it will be useful to examine first how existing estimated portfolio rules perform in the absence of any constraints.

Consider the standard mean-variance portfolio choice problem, in which an investor chooses his/her optimal portfolio among a risk-free asset and N risky assets. Denote the gross returns of the N risky assets at time t by $\mathbf{r}_t^g$, and r_f the return on the risk-free asset. Denote by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ the expected return and covariance matrix of the excess return $\mathbf{r}_t = \mathbf{r}_t^g - r_f\mathbf{1}$. Throughout this paper, we shall write $\mathbf{1}$ and $\mathbf{0}$ as vectors of ones or zeros of conformable dimensions. A portfolio can then be represented by a weight vector $\mathbf{w} \in \mathbb{R}^N$ whose ith element indicates how much to invest in the ith risky asset. The mean and variance of the excess return of a portfolio represented by $\mathbf{w}$ is given by $\mathbf{w}^\top\boldsymbol{\mu}$ and $\mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w}$ respectively.

The investor considers to maximize a quadratic utility that balances the risk and return,

$$U(\mathbf{w}) = \boldsymbol{\mu}^\top\mathbf{w} - \frac{\gamma}{2}\mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w}, \tag{1}$$

where $\gamma > 0$ is the coefficient of risk aversion. It is well known that the optimal portfolio

weights and the associated maximum utility are

$$\mathbf{w}_* = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \qquad U_* := U(\mathbf{w}_*) = \frac{1}{2\gamma}\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \tag{2}$$

The maximum Sharpe ratio is

$$\theta_* := \theta(\mathbf{w}_*) = \max_{\mathbf{w}\in\mathbb{R}^N} \theta(\mathbf{w}) = (\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{1/2}. \tag{3}$$

However, the optimal portfolio $\mathbf{w}_*$ is not feasible because it requires the knowledge of mean and variance of asset returns which in practice must be estimated.

Given historical returns $\mathbf{r}_1, \ldots, \mathbf{r}_T$, it is natural to consider replacing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in the utility function U by their sample counterparts:

$$\widehat{U}(\mathbf{w}) = \widehat{\boldsymbol{\mu}}^\top\mathbf{w} - \frac{\gamma}{2}\mathbf{w}^\top\widehat{\boldsymbol{\Sigma}}\mathbf{w}, \tag{4}$$

where

$$\widehat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_t, \qquad \text{and} \qquad \widehat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{r}_t - \widehat{\boldsymbol{\mu}})(\mathbf{r}_t - \widehat{\boldsymbol{\mu}})^\top.$$

Similarly, the maximizer of $\widehat{U}$, referred to as the sample mean-variance portfolio hereafter, can be explicitly given as

$$\widehat{\mathbf{w}}^{\text{sample}} = \widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}/\gamma. \tag{5}$$

Usually $N \leq T-1$ is assumed, so that this sample optimal portfolio is well defined. This is known as the plug-in rule, a result of plugging the sample into the population formula.

Although intuitive, the plug-in rule $\widehat{\mathbf{w}}^{\text{sample}}$ can introduce substantial estimation risk, especially when the size of the investment universe is not negligible compared to the length of the estimation window. The problem is apparent in high dimensions. To see this, assume that $\mathbf{r}_1, \ldots, \mathbf{r}_T$ are independent copies of a random vector $\mathbf{r} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. As T increases with $\lim N/T = \eta < 1$, it can be shown (e.g., Kan and Smith 2008) that

$$U(\widehat{\mathbf{w}}^{\text{sample}}) = \left(\frac{2}{1-\eta} - \frac{1}{(1-\eta)^3}\right)U_* - \frac{\eta}{2\gamma(1-\eta)^3} + o_p(1). \tag{6}$$

This means that the sample mean-variance portfolio is inconsistent in terms of the out-of-sample utility, i.e., $U(\widehat{\mathbf{w}}^{\text{sample}})$ fails to converge to its population counterpart U_*, as long as $\eta \neq 0$. The utility $U(\widehat{\mathbf{w}}^{\text{sample}})$ deteriorates quickly and can become even negative as the ratio

$\eta = N/T$ increases or N large relative to T. This means that the investor would be better off without investing in the risky assets at all.

The Sharpe ratio has a similar problem to (6). It can be shown that (e.g., Ao, Li, and Zheng 2019; Kan, Wang, and Zheng 2024; Yuan and Zhou 2024),

$$\theta(\widehat{\mathbf{w}}^{\text{sample}}) = \left(\frac{1-\eta}{1+\eta\theta_*^{-2}}\right)^{1/2} \cdot \theta_* + o_p(1), \tag{7}$$

which says that the Sharpe ratio of the estimated optimal portfolio will also fail to converge to its population counterpart. In fact, it achieves a smaller and smaller fraction as η increases, and eventually vanishes to 0 as η approaches to 1.

Another problem with the plug-in rule is that some of the weights necessarily diverge as N/T increases. More specifically, it is easy to show that

$$\mathbb{E}(\widehat{\mathbf{w}}^{\text{sample}}) = \frac{T-1}{T-N-2} \cdot \mathbf{w}_*, \tag{8}$$

and hence, as $\eta \uparrow 1$, we have $\max_{1\le i\le N} |\widehat{w}_i^{\text{sample}}| \to \infty$. This is clearly theoretically undesirable and empirically infeasible.

One may argue that $\widehat{\mathbf{w}}^{\text{sample}}$ can be scaled to have finite mean, say by a factor of $1-\eta = (T-N)/T$. But this will not solve the problem because

$$\begin{aligned}\texttt{Var}(\widehat{w}_i^{\text{sample}}) = &\frac{(T-1)^2\left[(T-N)w_i^2 + (T-N-2)(\mathbf{\Sigma}^{-1})_{ii}\theta_*^2\right]}{(T-N-1)(T-N-2)^2(T-N-4)} \\ &+ \frac{(T-1)\left[(T-N)(N+1)-2N\right](\mathbf{\Sigma}^{-1})_{ii}}{(T-N-1)(T-N-2)^2(T-N-4)},\end{aligned} \tag{9}$$

which still diverges after the re-scaling if η approaches 1. The problem is also evident from the Sharpe ratio result. Because the Sharpe ratio is invariant to scaling, any re-scaling of the portfolio will not resolve the fundamental problem of failing to converge to its population counterpart.

When N is fixed or small, as reviewed earlier, various estimators have been proposed to mitigate the estimation error. However, they do not seem to work in high dimensions. For example, the three-fund rule of Kan and Zhou (2007) has the same no-convergence problem (See, e.g., Yuan and Zhou 2024). Therefore, it is important to provide a new approach, going beyond the unconstrained portfolio examined by most of the existing studies.

3 No Short-Sales

In this section, we study first the theoretical properties of the optimal portfolio under the no-short-sale restriction, where the parameters are assumed known and there is no estimation risk. Then, we discuss how to estimate the long-only optimal portfolio from data, analyze the estimation risk, and prove the convergence of the Sharpe ratio.

3.1 Theoretical Implications

Under the no-short-sale restriction, we seek the portfolio weights $\mathbf{w}_*^+$ that solve

$$\max_{\mathbf{w}} U(\mathbf{w}) \qquad \text{s.t.} \quad \mathbf{w} \geq \mathbf{0}, \tag{10}$$

where $\mathbf{w} \geq \mathbf{0}$ means each element of $\mathbf{w}$ is nonnegative. It is clear that the constrained problem will not yield a better solution than the unconstrained problem. Hence, we have

$$U(\mathbf{w}_*^+) \leq U_*, \qquad \theta(\mathbf{w}_*^+) \leq \theta_*,$$

where $U(\mathbf{w}_*^+)$ and $\theta(\mathbf{w}_*^+)$ are the maximum utility and Sharpe ratio under the constraint.

There are two interesting findings on the constrained problem. First, we find that the nonzero components of $\mathbf{w}_*^+$ correspond exactly to investing in a subset of assets as if there were no constraints; that is, the unconstrained portfolio restricted to this subset automatically satisfies the no-short-sale condition. Second, we find, surprisingly, under a factor model with appropriately diversifiable risks, that the cost of the restriction approaches zero in the limit as the number of assets grows without bound.

To see the first finding, denote $\mathbf{w}_A^+ = (\mathbf{w}_{*j}^+)_{j \in A}$ as the portion of $\mathbf{w}_*^+$ corresponding to a subset $A \subseteq [N] \equiv 1, \ldots, N$. Let $|A|$ denote the number of elements in A. Then $\mathbf{w}_A^+$ is an $|A|$-dimensional subvector of $\mathbf{w}_*^+ \in \mathbb{R}^N$. Similarly, denote $\boldsymbol{\mu}_A = (\mu_j) j \in A$ as the $|A|$-dimensional subvector of $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}_A = (\Sigma j_1 j_2)_{j_1, j_2 \in A}$ as the $|A| \times |A|$ principal submatrix of $\boldsymbol{\Sigma}$.

Proposition 1. *There exists a subset $A \subseteq [N]$ such that the nonzero elements of $\mathbf{w}_*^+$ are given by*

$$\mathbf{w}_A^+ = \gamma^{-1} \boldsymbol{\Sigma}_A^{-1} \boldsymbol{\mu}_A,$$

and $w_j^+ = 0$ *for* $j \notin A$.

Proof: See the appendix.

The above proposition suggests that imposing the no-short-sale constraint leads to an investment in a reduced set of assets without any constraints, and any portfolio formula, such as the unconstrained optimal weights, can be applied to this subset. It is interesting that, although $\mathbf{w}_*^+$ cannot be expressed explicitly in terms of $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$, we can still characterize its property to a certain extent. This, consistent with Dai, Jin, and Liu (2011), will shed insights for understanding why and how the estimated optimal portfolio converges, as discussed in the next subsection.

On the second finding, we note that, even if the mean-variance portfolio $\mathbf{w}_*$ does hold short positions, $\mathbf{w}^+$ can still be nearly optimal. To see this, consider the case that the returns follow a multi-factor model with exact pricing,

$$\mathbf{r}_t = Bf + \varepsilon, \tag{11}$$

where B is the factor loading, f is a K-vector of factors uncorrelated with ε, the residual. For simplicity, we assume that ε are homogeneous errors of order σ_0^2 (though we can allow for diversifiable errors), then it is clear that the asset expected returns and covariance matrix are

$$\boldsymbol{\mu} = B\lambda_f, \qquad \mathbf{\Sigma} = B\mathbf{\Sigma}_f B^\top + \sigma_0^2 \mathbf{I},$$

where λ_f denotes the factor risk premia, $\mathbf{\Sigma}_f$ is the factor covariance matrix, $\mathbf{I}$ is the $N \times N$ identity matrix, and σ_0^2 is the residual variance.

We illustrate the key economic intuition driving the result with $K = 2$. Without loss of generality, suppose that the factors are tradable, but investors both are required to invest in only the assets and are no-short-sale constrained. Ignoring scaling, without the investors' constraints, the optimal portfolio weights are $\mathbf{\Sigma}_f^{-1}\lambda_f$. Assume that

$$\mathbf{\Sigma}_f^{-1}\lambda_f = f_1 + f_2, \tag{12}$$

which is simply a sum of the two factor, an assumption will be relaxed later. Then, the investors are to find long-positions only in the assets to replicate the above portfolio of the factors as close as possible.

Let $b_i = (b_{i1}, b_{i2})$ be the factor loadings of the ith asset. Consider two index sets

$$A_1 := \{i : \ b_{i1} + b_{i2} > 0, \ b_{i2} < b_{i1}\}, \qquad A_2 := \{i : \ b_{i1} + b_{i2} > 0, \ b_{i1} < b_{i2}\}.$$

Let $N_j := |A_j|$ and assume $N_j \to \infty$ as $N \to \infty$ for $j = 1, 2$. This is feasible as long as we assume that the loadings are randomly and densely distributed around 0. Now for each j let w_j be the equal-weight portfolio on A_j (weights $1/N_j$ on members of A_j, 0 otherwise). Then

$$R_1 = \bar{b}_{11} f_1 + \bar{b}_{12} f_2 + \frac{1}{N_1} \sum_{j \in A_1} \varepsilon_j, \tag{13}$$

$$R_2 = \bar{b}_{21} f_1 + \bar{b}_{22} f_2 + \frac{1}{N_2} \sum_{j \in A_2} \varepsilon_j, \tag{14}$$

where R_1 and R_2 are the two portfolio returns, and $\bar{b}_{jk}$ are the average loadings, i.e. $\bar{b}_{jk} = \frac{\sum_{i \in A_j} b_{ik}}{N_j}$. If $(\bar{b}_{11}, \bar{b}_{12})$ and $(\bar{b}_{21}, \bar{b}_{22})$ were $(1, 0)$ and $(0, 1)$, then $R_1 + R_2$ will replicate $f_1 + f_2$ plus a noise which approaches zero as N_1 and N_2 increase to infinity. Since R_1 and R_2 each has positive weights on the assets, so does $R_1 + R_2$.

When the unit vector assumption is not true, $(1, 1)$ will be inside the cone formed by $(\bar{b}_{11}, \bar{b}_{12})$ and $(\bar{b}_{21}, \bar{b}_{22})$ simply because they are on opposite sides of $\{(x, y) : x = y\}$. Mathematically, there must be a convex combination of $(\bar{b}_{11}, \bar{b}_{12})$ and $(\bar{b}_{21}, \bar{b}_{22})$ such that it equals to $(1, 1)$, up to a scaling factor. As the convex combination coefficients are non-negative, we can then obtain a nonnegative of R_1 and R_2, or of all the assets, that will replicate the payoff of $f_1 + f_2$ up to a noise converging to zero.

The previous constructive proof assumes that $f_1 + f_2$ is the optimal portfolio. If this is not the case—for instance, if $-f_1 + f_2$ is optimal—then we simply replace A_1 with the coefficients of the opposite sign, and the proof continues to hold. More generally, for any possible sign choice between f_1 and f_2, the corresponding combinations can be replicated by suitable portfolios, up to a vanishing noise term.

Finally, if the optimal portfolio is an arbitrary (but fixed) linear combination of f_1 and f_2, then appropriate scalings of R_1 and R_2 yield the replicating portfolio. Thus, as long as the idiosyncratic noise is diversifiable and the loadings are dense, the optimal portfolio, even if it involves short positions in the assets, can be replicated up to a noise term that converges to zero as $N \to \infty$.

Since by definition, $\theta(\mathbf{w}_*^+)$ must dominate the constructive portfolio above, it must be the case that $\theta(\mathbf{w}_*^+) \to \theta_*$ under the same conditions when $K = 2$.

In general, for any K, let $b_i = (b_{i1}, \ldots, b_{iK})^\top \in \mathbb{R}^K$ be the factor loadings of the ith asset and $s = (s_1, \ldots, s_K)^\top \in \{\pm 1\}^K$ be the sign of the entries of $\mathbf{\Sigma}_f^{-1}\lambda_f$. In other words, $s_k = 1$ if the optimal mean-variance portfolio in the factor space longs the kth factor, and $s_k = -1$ if short (the factor will be ignored if it does not enter into the optimal portfolio). Then, we have the following

Theorem 1. *Assume the K-factor model. If $\exists\ \varepsilon > 0$, s.t. $\min_{1\le k\le K} |A_k| \to \infty$ as $N \to \infty$, where*

$$A_k = \{i : s_k b_{ik} > \sum_{l\neq k} |b_{il}| + \varepsilon, \quad \text{and} \quad s_l b_{il} \le 0 \quad \forall l \neq k\},$$

then

$$\theta(\mathbf{w}_*^+) \to \theta_*$$

as $N \to \infty$, i.e., the no-short-sale constrained Sharpe ratio converges to the unconstrained one.

Proof: See the appendix.

The condition $\min_{1\le k\le K} |A_k| \to \infty$ is the dense loading assumption, that is, there are infinitely many assets with loading signs on each side. In particular, the condition is satisfied when the b_i's are viewed as population samples from a normal distribution over $\mathbb{R}^K$.

Under the dense condition, Theorem 1 says that there is no loss of efficiency asymptotically by imposing the no-sale restriction. In particular, if $s_1 = \cdots = s_K = 1$, then A_k is the collection of all assets that has a positive beta for the kth factor only and it dominates all other portfolios or performs at least equally.

The no-short-sale restriction has been previously considered in the literature for the global minimum variance portfolio. Jagannathan and Ma (2003) make a novel discovery that imposing no-short-sale on the GMV portfolio, which solves

$$\min_{\mathbf{w}} \mathbf{w}^\top \mathbf{\Sigma} \mathbf{w}, \qquad \text{st.} \quad \mathbf{w}^\top \mathbf{1} = 1, \tag{15}$$

improves estimation accuracy. DeMiguel, Garlappi, Nogales, and Uppal (2009) show further improvements by generalizing the no short-sale constraints to a norm restriction that the L_1 norm of the weights is bounded. Fan, Zhang, and Yu (2012) show that the risk of the GMV under no short-sale constraints or a general norm restriction will converge to its popular counterpart,

$$\min_{\mathbf{w}} \mathbf{w}^\top \mathbf{\Sigma} \mathbf{w}, \qquad \text{st.} \quad \mathbf{w}^\top \mathbf{1} = 1, \ \mathbf{w} \geq 0. \tag{16}$$

In general, the constrained risk, solution to (16), is strictly greater than that of the unconstrained risk, solution to (15), even asymptotically. In other words, there is always a positive cost of imposing either a no-short-sale or a norm restriction on estimating the risk of the GMV. On the other hand, without any constraints, the sample GMV always diverges; it is an open question to find the risk of the GMV under a suitable condition.

In contrast, our paper focuses on the mean-variance utility maximizing portfolio, which is the portfolio that determines the SDF or maximizes the Sharpe ratio. In this case, it is interesting that the cost of "imposing the wrong assumption" of no-short-sale is virtually zero, at least in the case of a K-factor model with exact pricing, as N grows without bound.

Theorem 1 has two further implications. First, although many fund managers—particularly institutional investors—are restricted to long-only strategies, Theorem 1 shows that their portfolios can nevertheless approach the efficient portfolio when the market is well approximated by a factor model. Second, on the theoretical side, Koijen and Yogo (2019) develop a demand-based asset pricing framework that has spurred a growing body of research. For example, Davis (2025) and Haddad, Huebner, and Loualiche (2025) examine demand's implications for arbitrage and passive investments, respectively. He, Kondor, and Li (2025); van Binsbergen, David, and Opp (2025); Haddad, He, Huebner, Kondor, and Loualiche (2025) provide some of the recent advances. In their formulation, Koijen and Yogo (2019) impose a no-short-sale constraint under log utility, which serves as an approximation to mean-variance preferences. Theorem 1 lends support to this foundation: within a factor model, the no-short-sale constraint is not binding, since the resulting pricing kernel can be made arbitrarily close to the true (but unobservable) kernel, which may in principle involve short positions in the underlying assets.

We conclude this subsection with examples illustrating how quickly the constrained port-

folio converges to the true optimal portfolio. Specifically, we calibrate the parameters of a two-factor model using the market and size factors, and then extend it to a five-factor model by adding three additional factors with both the factor realizations and betas drawn from distributions similar to those in the two-factor case. The residuals are specified as homogeneous errors, consistent with our earlier setup.

Table 1 reports the results. In the two-factor model, with as few as 25 assets, the no-short-sale Sharpe ratio is 0.181214—already very close to the unconstrained Sharpe ratio of 0.181393, with an error bounded by 0.001.[6] As N increases to 50, the difference becomes even smaller, as expected. When $N = 500$, the difference shrinks further to 0.0001, and when $N = 2000$, it is only 0.00001. In the five-factor model, the errors are slightly larger than in the two-factor case, but they remain very small. At $N = 2000$, the errors are essentially the same as in the two-factor model. Overall, the numerical results confirm Theorem 1, and in fact perform much better than one might have anticipated.

Figure 1 makes these results more apparent. In the two-factor model, with as few as 100 assets, the constrained portfolio already achieves more than 99% of the true optimal Sharpe ratio. With $N = 1000$, the two ratios are virtually identical. Interestingly, in the five-factor model, the Sharpe ratio also exceeds 99% with 100 assets, though the convergence is somewhat slower than in the two-factor case. This slowdown is expected, since maximizing the Sharpe ratio becomes more challenging as the number of factors increases under the long-only constraint. This is because there will be more factors to be replicated with only positive portfolio weights. Nevertheless, the deviation from the true Sharpe ratio remains negligible. Figure 1 reinforces the evidence from Table 1 that the optimal portfolio under the no-short-sale restriction is asymptotically identical to the optimal portfolio under no constraints.

[6] The ζ-constrained portfolio, reported in the third column of the table, will be defined in the next section. It allows for limited short-sale positions, lying between the no-short-sale and unconstrained cases.

3.2 Estimation Accuracy

We now turn to estimating the optimal portfolio from data under the no-short-sale restriction, inspired by Jagannathan and Ma (2003) and the related work. Given the sample mean and covariance matrix, we seek the portfolio weights $\widehat{\mathbf{w}}^+$ that solve the sample analogue of the earlier problem,

$$\widehat{\mathbf{w}}^+ := \operatorname*{argmax}_{\mathbf{w}} \widehat{U}(\mathbf{w}) \qquad \text{subject to } \mathbf{w} \geq \mathbf{0}, \tag{17}$$

where $\widehat{U}(\mathbf{w})$ is the sample utility given by (4). While there is no analytical formula, a typical quadratic programming software can solve for $\widehat{\mathbf{w}}^+$ numerically.

The question is both whether $\widehat{\mathbf{w}}^+$ converges, motivated by the behavior of $\widehat{\mathbf{w}}^{\text{sample}}$, and if it converges, what it converges to. In contrast to an investor who estimates the parameters, an oracle who knows the true moments can achieve both the maximum utility and Sharpe ratio under the no-short-sale constraint,

$$U_*^A := \boldsymbol{\mu}_A^\top (\boldsymbol{\Sigma}_A)^{-1} \boldsymbol{\mu}_A / (2\gamma) \qquad \text{and} \qquad \theta_*^A := [\boldsymbol{\mu}_A^\top (\boldsymbol{\Sigma}_A)^{-1} \boldsymbol{\mu}_A]^{1/2},$$

where A is given by Proposition 1. Specifically, Proposition 1 ensures that

$$\mathcal{O} = \left\{ A \subseteq [N] : (\boldsymbol{\Sigma}_A)^{-1} \boldsymbol{\mu}_A \geq 0 \right\}.$$

is nonempty. Note that for any asset i, if its expected (excess) return $r_i \geq 0$, then $\{i\} \in \mathcal{O}$. Moreover, if the optimal portfolio $\mathbf{w}_*$ does not involve short positions, then $[N] \in \mathcal{O}$. Using this definition, we will show below that the investor can perform nearly as well as such the oracle asymptotically.

Theorem 2. *Suppose that $\mathbf{r}_1, \ldots, \mathbf{r}_T$ are independent $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random variables such that for some constants $0 < c_0 < C_0 < \infty$,*

(i) the maximum Sharpe ratio is bounded: $\theta_* := (\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})^{1/2} \leq C_0$;

(ii) the asset variances are bounded: $\sigma_{\max}^2 := \max_{1 \leq i \leq N} \sigma_{ii} \leq C_0$ where σ_{ii} is ith diagonal entry of $\boldsymbol{\Sigma}$;

(iii) the minimum variance portfolio has non-vanishing variance: $\sigma_g^2 := (\mathbf{1}^\top \boldsymbol{\Sigma} \mathbf{1})^{-1} \geq c_0$.

(iv) the covariance matrix $\mathbf{\Sigma}$ *is strictly positive definite: the smallest eigenvalue* $\lambda_{\min}(\mathbf{\Sigma}) \geq c_0$*;*

Then there exists a constant $c_1 > 0$ *such that*

$$U(\widehat{\mathbf{w}}^+) \geq \sup_{A \in \mathcal{O}} \left\{ U_*^A - \frac{c_1}{\gamma} \min\{|A|, \sqrt{T/\log N}\} \cdot \frac{\log N}{T} \right\},$$

and

$$\theta(\widehat{\mathbf{w}}^+) \geq \sup_{A \in \mathcal{O}} \left\{ \theta_*^A - c_1 \min\{|A|, \sqrt{T/\log N}\} \cdot \frac{\log N}{T} \right\},$$

with probability tending to one as $T \to \infty$ *and* $(\log N)/T \to 0$*.*

Proof: See the Appendix.

First, let us explain the assumptions. The normality assumption is not essential, and the results also hold when the return distribution has sub-Gaussian tails. Even if the returns follow a heavy-tailed distribution, our theoretical proof can still be applied, leading to qualitatively similar risk bounds. In the proofs, we shall indicate where specific changes can be made to accommodate more general distributions. The other assumptions are fairly standard in asset pricing. Assumption (i) is self-evident in any economy and is technically necessary to prevent the estimated Sharpe ratio from diverging to infinity. Assumption (ii) rules out explosive risks. Assumption (iii) excludes the possibility of asymptotically riskless assets that still earn positive returns. Assumption (iv) precludes asymptotically redundant assets.

To see the implications of Theorem 2, consider a special case in which $|A| = N$, then we have

Corollary 1. *Under the assumptions of Theorem 2,*

$$U(\widehat{\mathbf{w}}^+) \geq U_*^+ - \frac{c_1}{\gamma} \cdot \sqrt{\frac{\log N}{T}},$$

and

$$\theta(\widehat{\mathbf{w}}^+) \geq \theta_*^+ - c_1 \cdot \sqrt{\frac{\log N}{T}},$$

with probability tending to one as $T \to \infty$ *and* $(\log N)/T \to 0$*.*

Theorem 2 says that the estimated no-short-sale portfolio is consistent (in terms of utility or Sharpe ratio), which converges to their population counterpart at a rate no smaller than $\sqrt{\frac{\log N}{T}}$. In general, $|A|$ can be much smaller than N. Then, in this case, the convergence rate will be even greater.

If the unconstrained mean-variance portfolio $\mathbf{w}_* \geq \mathbf{0}$, or if the factor model of Theorem 1 is true, then it is clear that

$$U(\widehat{\mathbf{w}}^+) \to_p U_*, \qquad \text{and} \qquad \theta(\widehat{\mathbf{w}}^+) \to_p \theta_*,$$

that it, the estimated no-short-sale portfolio converges to the unconditional optimal portfolio that determines the SDF. Simulation examples later confirm these results.

3.3 Beating the $1/N$

The $1/N$ naive diversification strategy has a long tradition in finance. DeMiguel, Garlappi, and Uppal (2009) show that it is difficult to outperform, as the sample mean-variance portfolio is notoriously prone to underperform relative to $1/N$. Tu and Zhou (2011) find that combining the sample mean-variance portfolio with the $1/N$ rule often improves performance, but only when N is small or fixed. Lassance, Vanderveken, and Vrins (2024) and Yuan and Zhou (2024) explore further ways to outperform the $1/N$. However, in high dimensions, since the sample mean-variance portfolio fails, any combination rule will have difficulty beating the $1/N$.

Since the naive diversification rule has weights,

$$\mathbf{w}_{\text{naive}} = \frac{1}{N} \times \mathbf{1}, \tag{18}$$

it satisfies obviously the no-short-sale constraint. Hence, it is clear that

$$\theta(\mathbf{w}_{\text{naive}}) \leq \theta(\mathbf{w}^+). \tag{19}$$

That is, the $1/N$ rule will never outperform the constrained population mean-variance strategy. Now, in light of the Theorem 2, $\theta(\widehat{\mathbf{w}}^+)$ converges to the right hand side, and so it must be the case that

$$\theta(\mathbf{w}_{\text{naive}}) \leq \theta(\widehat{\mathbf{w}}^+) + \frac{c_1}{\gamma}\sqrt{\frac{\log N}{T}}, \tag{20}$$

as long as the sample size is large enough, while N can be large but $\frac{\log N}{T}$ approaches zero. This implies that the $1/N$ rule will never outperform theoretically the constrained sample mean-variance strategy $\widehat{\mathbf{w}}^{+}$ in large samples.

Assume that $\theta(\mathbf{w}_{\text{naive}})$ does not happen to be identical to the optimal portfolio weights, such as the case of more than one factors for the asset returns. In this case, bound (20) must be an inequality, and hence we have probability tending to one,

$$\theta(\widehat{\mathbf{w}}^{+}) > \theta(\mathbf{w}_{\text{naive}}) + O_p\left(\sqrt{\frac{\log N}{T}}\right), \tag{21}$$

in large samples.

In short, while DeMiguel, Garlappi, and Uppal (2009) and Yuan and Zhou (2024), among others, find it difficult to outperform the $1/N$ investment rule, our results show that imposing no-short-sale constraints generally allows one to beat it. This finding is particularly surprising, given that the constraint seems simple to implement. Of course, it is an empirical issue to determine how large N and T should be. Nevertheless, our empirical results later show that the theoretical result is relevant to reasonable practice sample sizes.

4 Limited Short-Sales

Our approach is motivated by long/short equity strategies, especially the 130/30 funds, where managers are allowed to take short positions with the total amount of shorts no greater than 30% of the overall portfolio value. The constraint can be formalized as

$$\mathbf{1}^{\top}\mathbf{w}_{-} \leq 0.3 \cdot \mathbf{1}^{\top}\mathbf{w} \tag{22}$$

where $\mathbf{w}_{-} = (\max\{-w_i, 0\})_{1\leq i\leq N}$ represents the sizes of the short positions. Since

$$\mathbf{1}^{\top}\mathbf{w}_{-} = \frac{1}{2}(\|\mathbf{w}\|_{\ell_1} - \mathbf{1}^{\top}\mathbf{w}),$$

we can express (22) in terms of the gross leverage ratio equivalently as:

$$\frac{1}{2}\frac{\|\mathbf{w}\|_{\ell_1}}{\mathbf{1}^{\top}\mathbf{w}} - 0.5 \leq 0.3. \tag{23}$$

In general, we consider portfolios with a net short position no greater than a prespecified level $\zeta \geq 0$. Then the constraint is

$$\Omega(\zeta) = \{\mathbf{w} \in \mathbb{R}^N : \frac{1}{2}\frac{\|\mathbf{w}\|_{\ell_1}}{\mathbf{1}^\top \mathbf{w}} - 0.5 \leq \zeta\}, \tag{24}$$

which is our generalized constraint. When $\zeta = 0$, it implies no-short-sale. When $\zeta = 0.2$, it is a 120/20 portfolio which some managers prefer. When ζ increases to infinity, the choice set becomes the unconstrained one.

Under the limited short-sale at any level ζ, we now solve,

$$\widehat{\mathbf{w}} := \arg\max \widehat{U}(\mathbf{w}) \qquad \text{subject to } \mathbf{w} \in \Omega(\zeta). \tag{25}$$

which yields the estimated optimal portfolio from maximizing the sample utility over all portfolios in $\Omega(\zeta)$.

Numerically, the above constrained problem can be transformed into a form solvable by typical quadratic programming packages. Let $\mathbf{w} = \mathbf{w}^+ - \mathbf{w}^-$ and $\tilde{\mathbf{w}} = [\mathbf{w}^{+,\top}, \mathbf{w}^{-,\top}]^\top$, then the ζ-constrained problem (25) can be reformulated as

$$\max_{\tilde{\mathbf{w}}} \begin{bmatrix} \widehat{\boldsymbol{\mu}}^\top & -\widehat{\boldsymbol{\mu}}^\top \end{bmatrix} \tilde{\mathbf{w}} - \tilde{\mathbf{w}}^\top \left(\widehat{\boldsymbol{\Sigma}} \otimes \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \right) \tilde{\mathbf{w}}, \tag{26}$$

subject to the constraints:

$$\mathbf{w}^+ \geq 0, \quad \mathbf{w}^- \geq 0, \quad (1+\zeta)\,\mathbf{1}^\top \mathbf{w}^- \leq \zeta \mathbf{1}^\top \mathbf{w}^+.$$

This is now a standard quadratic optimization problem.

Asymptotically, it is surprising that the rate of convergence of both $U(\widehat{\mathbf{w}})$ and $\theta(\widehat{\mathbf{w}})$ to their population counterparts stay unchanged from the no-short-sale case. More specially, we have

Theorem 3. *Under the same conditions of Theorem 2, there exists a constant $c_1 > 0$ depending only on c_0 and C_0 such that*

$$U(\widehat{\mathbf{w}}) \geq \max_{\mathbf{w} \in \Omega(\zeta)} U(\mathbf{w}) - \frac{c_1(1+2\zeta)^4}{\gamma}\sqrt{\frac{\log N}{T}}, \tag{27}$$

and

$$\theta(\widehat{\mathbf{w}}) \geq \max_{\mathbf{w}\in\Omega(\zeta)} \theta(\mathbf{w}) - c_1(1+2\zeta)^4\sqrt{\frac{\log N}{T}}, \tag{28}$$

with probability tending to one as $T \to \infty$ *and* $(\log N)/T \to 0$.

Equations (27) and (28) show that both the expected utility and the Sharpe ratio of the estimated constrained portfolio converge to their population counterparts at the same rate, $\sqrt{\log N/T}$, as before. However, in finite samples, the accuracy depends on the choice of ζ. All else equal, a larger ζ expands the set $\Omega(\zeta)$ and leads to a looser bound.

Assume that the unconstrained optimal portfolio has a finite bound on its short position. Then, when ζ is set sufficiently large yet fixed, the constrained population optimal portfolio will be then the unconstrained optimal portfolio. In this case, the bound in Theorem 3 can be refined to the following more elegant asymptotic equalities:

$$U(\widehat{\mathbf{w}}) = U_* + O_p\left(\sqrt{\frac{\log N}{T}}\right), \tag{29}$$

and

$$\theta(\widehat{\mathbf{w}}) = \theta(\mathbf{w}_*) + O_p\left(\sqrt{\frac{\log N}{T}}\right). \tag{30}$$

In other words, $\widehat{\mathbf{w}}$ is asymptotically optimal whenever $\log N = o(T)$. This contrasts with the classical convergence result, where the sample mean-variance portfolio is asymptotically optimal only when $N = o(T)$ or when N is small relative to T. Here, N can be large relative to T, provided $\log N$ is small relative to T (i.e., it does not grow exponentially with T). This is a much weaker condition on the sample size, which becomes sufficient once the limited short position constraint is imposed.

There are two technical remarks. First, unlike some large-sample theory that requires N/T to be a constant, Theorem 2 allows N and T to be dependently large. Second, it also accommodate the case where $N > T$. In this case, although $\widehat{\mathbf{\Sigma}}$ is not invertible, the solution to (26) is well defined and can be computed numerically. The asymptotic bounds in Theorem 2 still hold.

5 A Comparison

In this section, we compare our methodology with the most related ones in the literature. We highlight the major difference that we impose a flexible limited short-exposures captured by η.

5.1 SDF with dimension reduction

In large portfolios, the number of assets is substantial, and existing studies often rely on dimension-reduction techniques to render the problem tractable. Kozak, Nagel, and Santosh (2020) is among the earliest studies to estimate the stochastic discount factor (SDF). They adopt a Bayesian approach that combines dimension reduction with sample-optimal portfolio selection. Because their framework imposes no constraints on short sales, the resulting optimal portfolio—like most portfolios in the literature—can feature excessively large short positions, making it difficult to implement in real-world applications.

Ao, Li, and Zheng (2019), Chen, Ding, Li, and Zheng (2025), and Luo, Zhou, and Zhou (2025) instead employ the LASSO in a classical statistical setting. Relative to the Bayesian approach, an advantage of LASSO is that convergence to the SDF can be established asymptotically under suitable sparsity assumptions. These approaches are two-step procedures. In the first step, they estimate a regression of a constant on all asset returns,

$$c = \alpha + \beta_1 r_{1t} + \beta_2 r_{2t} + \cdots + \beta_N r_{Nt} + \varepsilon_t, \tag{31}$$

where the constant c is defined differently across studies. This first step typically reduces the number of assets from N to a much smaller subset $N^1 < T$, which is then used for portfolio selection.

In the second step, the optimal portfolio is given by

$$\widehat{\mathbf{w}}_1^{\text{sample}} = \widehat{\boldsymbol{\Sigma}}_1^{-1} \widehat{\boldsymbol{\mu}}_1 / \gamma, \tag{32}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ are the sample mean and covariance matrix of the selected N_1 assets. It is clear that there are no limited constraints on the short positions on $\widehat{\mathbf{w}}_1^{\text{sample}}$.

In contrast, our approach solves (26) directly with all the assets. The solution exists even when $N > T$. Theorem 3 says that the solution converges as long as $\log(N)/T$ approaches zero.

5.2 Global Minimum Variance Portfolio

Following Jagannathan and Ma (2003), DeMiguel, Garlappi, Nogales, and Uppal (2009) and Fan, Zhang, and Yu (2012), we consider the global minimum variance portfolio. This is not our focus here. But a comparison helps the understanding of their results and insights. For simplicity, we consider only the case of imposing the no-short-sale constraint. We show that, under our assumptions, the estimated constrained GMV converges as well.

The GMV is useful in the case when a fund manager is required to invest fully in equities or in the case the estimation of the expected returns is regarded as too noise to be used. Then, one may be interested in a portfolio of the least risk. Hence, given data, one needs only the sample covariance matrix for the investment decision, and the no-short-sale GMV portfolio weights, $\widehat{\mathbf{w}}^{+}_{\text{GMV}}$, solves the following problem,

$$\min \mathbf{w}^\top \widehat{\mathbf{\Sigma}} \mathbf{w}, \qquad \text{subject to } \mathbf{w}^\top \mathbf{1} = 1, \quad \mathbf{w} \geq 0,$$

where the $\mathbf{w} \geq 0$ is the additional restriction compared with the unconstrained GMV which still has the budget condition $\mathbf{w}^\top \mathbf{1} = 1$.

While the estimated unconstrained GMV fails to converge, the constrained portfolio $\widehat{\mathbf{w}}^{+}_{\text{GMV}}$ does, an important fact established first by Fan, Zhang, and Yu (2012). However, their conditions can be violated, say in the context of the CAPM where the leading eigenvalue of the covariance matrix diverges with increasing N. Complimentary to their study, we have

Theorem 4. *Suppose that $\mathbf{r}_1, \ldots, \mathbf{r}_T$ are independent $N(\boldsymbol{\mu}, \mathbf{\Sigma})$ random variables such that for some constants $0 < c_0 < C_0 < \infty$,*

(i) the maximum Sharpe ratio is bounded: $\theta_* := (\boldsymbol{\mu}^\top \mathbf{\Sigma}^{-1} \boldsymbol{\mu})^{1/2} \leq C_0$;

(ii) the asset variances are bounded: $\sigma^2_{\max} := \max_{1 \leq i \leq N} \sigma_{ii} \leq C_0$ where σ_{ii} is ith diagonal entry of $\mathbf{\Sigma}$;

(iii) *the minimum variance portfolio has non-vanishing variance:* $\sigma_g^2 := (\mathbf{1}^\top \mathbf{\Sigma} \mathbf{1})^{-1} \geq c_0$.

(iv) *the covariance matrix* $\mathbf{\Sigma}$ *is strictly positive definite: the smallest eigenvalue* $\lambda_{\min}(\mathbf{\Sigma}) \geq c_0$*;*

Then there exists a constants $c_1 > 0$ *such that*

$$|\mathbf{w}_{\text{GMV}}^{+,\top} \mathbf{\Sigma} \mathbf{w}_{\text{GMV}}^{+} - \widehat{\mathbf{w}}_{\text{GMV}}^{+,\top} \mathbf{\Sigma} \widehat{\mathbf{w}}_{\text{GMV}}^{+}| \leq c_1 \cdot \sqrt{\frac{\log N}{T}},$$

and

$$\theta(\mathbf{w}_{\text{GMV}}^{+}) \geq \theta(\widehat{\mathbf{w}}_{\text{GMV}}^{+}) - c_1 \cdot \sqrt{\frac{\log N}{T}},$$

with probability tending to one as $T \to \infty$ *and* $(\log N)/T \to 0$.

Note that the returns are now the raw returns instead of excess returns. This is because the riskfree asset has no role in the GMV. Other than this minor and subtle difference, the assumptions of Theorem 4 are the same as those assumption made before.

Theorem 4 says that both the risk of the estimated portfolio as well as the Sharpe ratio of $\mathbf{w}_{\text{GMV}}^{+}$, taking the riskfree rate as zero here, converge to their population counterpart at the same rate as before, $\sqrt{(\log N)/T}$. Fan, Zhang, and Yu (2012) find conditions under which $\mathbf{w}_{\text{GMV}}^{+}$ converges. Interestingly, although their Condition 1 essentially rules out factor models, their conclusion continues to hold under our more financially realistic assumptions. We further note that the Sharpe ratio also converges at the same rate.

However, it is worth emphasizing that the population Sharpe ratio $\theta(\mathbf{w}_{\text{GMV}}^{+})$ almost always differs from that of the unconstrained portfolio, and the latter often differs from θ_*, even in the limit in a factor model with large N. Thus, there is typically a cost associated with imposing the no-short-sale restriction when estimating the GMV portfolio. If, instead, one is interested in the mean-variance efficient portfolio or the SDF, the no-short-sale restriction should be imposed directly on the mean-variance portfolio, as discussed in the previous two sections, then the cost is virtually zero asymptotically in the factor model.

6 Numerical Results

In this section, we examine how the portfolio rules perform over both simulated data and real data sets.

6.1 Simulations

In our simulation, we assume that the asset excess returns follow a 5-factor model,

$$r_{it} = \alpha_i + \beta_{i1} f_{1t} + \beta_{i2} f_{2t} + \beta_{i3} f_{3t} + \beta_{i4} f_{4t} + \beta_{i5} f_{5t} + \varepsilon_{it}, \quad i = 1, \ldots, N, \tag{33}$$

where $f_{1t}, \ldots, f_{5t}$ are the factors. For simplicity, we assume that the factors are normally and independently distributed, so are the residuals. Then, if the parameters were known, it will be straightforward to simulate both the factors and the residuals, and consequently to compute the returns.

To set the parameters be some realistic values, we use the 25 well known size and book-to-market portfolios of Fama and French (1993, 2015) and their 5-factor model to calibrate the mean of asset betas and their standard deviation with monthly data from July 1926 to December 2023.[7] The betas of each asset in (33) can then be simulated from a normal distribution with these means and standard deviations. The sample means and volatilities of the Fama–French five factors can be taken as the true common parameters for the five factors in (33), and their realizations can therefore be simulated directly.

The covariance matrix of the model residuals can be determined similarly. Let $\bar{\sigma}$ and σ_σ^2 be the mean and standard deviation of the residual volatility for the 25 test assets. Then the residual volatility for each asset in (33) can be taken from

$$z \sim N(\bar{\sigma}, \sigma_\sigma^2).$$

However, if $z < \bar{\sigma}/2$, we re-draw, to make it realistic that no stocks will have too small volatility. In other words, we set

$$\sigma_{i,i} = z^2, \tag{34}$$

[7] The website, https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, provides the data.

where z is determined as above. Finally, for $1 \leq i \neq j \leq N$, we set

$$\sigma_{i,j} = \sqrt{\sigma_{i,i}\sigma_{j,j}}/(2(i-j)^2). \tag{35}$$

Hence, the covariance matrix of the residuals is now well defined. This data generating process satisfies the assumptions underlying Theorem 3.

Now we are able to consider a series of economy with $N = 25, 50, 100$ and up to 500, respectively. Given an arbitrary pair of N and T, we can simulate first a set of corresponding population parameters. Then, given the parameters, we can simulate a path of returns from (33) over T periods. Given a path of return data, we can carry out all the portfolio rules. Then we can simulate M_2, say 10,000, returns for period $T+1$ only, which will allow us to obtain the Sharpe ratios of the rules. These Sharpe ratios are clearly path dependent. In this paper, we use a total number of $M_1 = 1000$ paths, or M_1 sets of simulated data, then the resulting average Sharpe ratios should be close to true Sharpe ratios of the strategies.

Table 2 presents the results. Panel A reports the population Sharpe ratios for various values of N under four investment strategies: the unconstrained portfolio, the $1/N$ rule, the no-short-sale portfolio, and the ζ-constrained portfolio (with $\zeta = 0.3$). These are referred to as the oracle values, as they represent the outcomes investors would achieve if the true parameters were known. Note that the unconstrained values converge to the Sharpe ratio of the factors. Although the residuals here are not necessarily well diversified as the case for Table 1, the no-short-sale and ζ-constrained portfolios also appear to approach this limit. By contrast, the $1/N$ rule is not optimal in the population and can therefore be outperformed by the constrained portfolios when both N and T are large.

Panel B reports the results when $N/T = 0.25$, a case when the sample size T is large relative to N. In this case, the unconstrained portfolio achieves about 50% of the value of the $1/N$, confirming that $1/N$ is difficult to beat, especially by the traditional Markowitz portfolio. However, under no-short-sale, the constrained portfolio beats it in all the scenarios of N. The ζ-constrained portfolio performs even better, as expected if the estimation error does not change much, because it has a greater choice set.

Panel C considers the case of large N relative to T, with $N/T = 0.75$. As expected, the Markowitz portfolio performs especially poorly, achieving only about 20% of the true value.

By contrast, the other strategies perform only slightly worse than in the earlier case with $N/T = 0.25$.

Panel D examines the case of $N > T$, with $N/T = 1.25$. Here, the sample covariance matrix is not invertible, but its generalized inverse can be used in the plug-in formula. The performance is similar to the case with $N/T = 0.75$. Interestingly, when $N/T = 1.75$, the performance improves somewhat, as shown in Panel E. These results suggest that as the number of assets increases for a given sample size, performance may improve slightly because more assets are available for selection. Another likely reason is that the statistical properties of the mean-variance portfolio change when using the generalized inverse. In this case, a "double-descent phenomenon" appears, where performance can re-increase with N/T once $N/T > 1$.

Another notable finding is that, when $N > T$, the two constrained strategies can still be computed via quadratic programming despite the singularity of the sample covariance matrix. Their results are largely similar to the earlier cases, indicating that while more assets provide greater choice, the estimation errors associated with these two rules offset much of the potential gain.

6.2 Applications

We apply the constrained portfolios to several datasets. The first two consist of 100 portfolios sorted by size and operating profitability, and by size and investment, respectively. The other two are the datasets used by Kozak, Nagel, and Santosh (2020), and Luo, Zhou, and Zhou (2024) to construct the SDF. In these experiments, we incorporate proportional transaction costs, computed following both the transaction cost and portfolio literature (e.g., Balduzzi and Lynch 1999; Novy-Marx and Velikov 2016; Detzel, Novy-Marx, and Velikov 2023). Specifically, we model the transaction cost at time t as

$$\text{Transaction Cost}_t = c \times \sum_{i=1}^{N} \left| \tilde{w}_{i,t} - w^*_{i,t} \right|, \tag{36}$$

where the per-unit cost c ranges from 0 to 20 basis points, $w^*_{i,t}$ denotes portfolio weights before rebalancing, and $\tilde{w}_{i,t}$ the weights of a strategy at time t. Their performance is evaluated

with sample sizes $T = 120$ and $T = 240$, typical within estimation error literature (e.g., DeMiguel, Garlappi, and Uppal 2009; Kan, Wang, and Zhou 2022).

Table 3 reports the annualized Sharpe ratios of the same four investment strategies examined earlier, for the two sets of 100 portfolios over the out-of-sample investment period from July 1963 to December 2024. Panel A presents those sorted by size and operating profitability, while Panel B presents those sorted by size and investment. Although minor differences exist, both panels reveal the same qualitative patterns as follows.

In the absence of transaction costs, the unconstrained portfolio performs better with longer estimation windows, performing poorly when $T = 120$, but substantially better when $T = 240$. The performance of the $1/N$ portfolio is unaffected by the sample size T since its weights are independent of the data. The no-short-sale constrained portfolio also performs well, though at $T = 240$ it is slightly outperformed by both the unconstrained and the $1/N$ strategies. This is possible because the number of portfolios $N = 100$ is relatively small compared to $T = 240$ and they are not large enough; however, theory and simulations (Table 2) indicate that such out-performance disappears as both N and T grow and $(\log N)/T \to 0$. Consistent with intuition, the ζ-constrained portfolio delivers the best performance in finite samples due to an expanded choice set for the portfolio weights.

When transaction costs are introduced, the unconstrained portfolio performs poorly regardless of sample size, and the Sharpe ratios even become negative when $T = 120$. This is driven largely by its high turnover. By contrast, all other portfolios have lower turnover and perform well, and remain relatively stable across T and transaction costs, and the $1/N$ portfolio's performance being most stable.

To further explore the role of short-sale restrictions, Table 4 reports the largest short positions in each portfolio. As expected, the constrained portfolios hold significantly less extreme positions than the unconstrained portfolio. When $T = 120$, on average, the unconstrained strategy's largest short position is 263.1% and 282.2% of total portfolio value in both datasets, respectively, levels nearly impossible to implement in practice. In contrast, by definition, both the $1/N$ and no-short-sale portfolios contain no short positions, while the ζ-constrained 130/30 portfolio allows modest shorting, with an average largest negative

position of around $2-3\%$, making both of them much more practical than the unconstrained case.

We now turn to estimating the stochastic discount factor. Kozak, Nagel, and Santosh (2020) propose a novel Bayesian framework for this task, using economically motivated priors on the SDF coefficients. Their focus appears to be on the sparsity of the pricing kernel rather than on a direct comparison of estimation accuracy with the Markowitz portfolio. We estimate using a subset of 44 portfolios from their dataset, after removing the missing data, over the period from January 1964 to December 2019. Since we use up to 20 years of data for training, our out-of-sample investment period is from January 1984 to December 2019.

Table 5 reports the results. We focus on the Sharpe ratios, as higher values indicate that the implied SDFs are closer to the true, unobserved pricing kernel. The estimated unconstrained optimal portfolio performs well across various sample sizes and transaction cost assumptions. This is primarily because the number of assets ($N = 44$) is small relative to the sample size ($T \geq 120$). Interestingly, the 130/30 rule performs best in most cases. In contrast, the no-short-sale portfolios underperform, because the relatively small N limits the benefits of the constraint.

One could question why we impose short-sale constraints while the most of the underlying assets are long–short portfolios, which already short-sell themselves. The reason is that short-sale constraints are now used as a tool to improve estimation accuracy. Indeed, as the results show, imposing the 130/30 constraints does help obtain the best Sharpe ratios in most cases.

The $1/N$ rule, adjusted to hold negative weights when a port, performs poorly relative to the other strategies. This likely reflects the nature of the data used: since the anomalies capture distinct risk sources, the $1/N$ rule fails to provide sufficient diversification, and thus delivers weaker performance. This is consistent with the fact that $1/N$ is only asymptotically optimal in a one-factor setting.

Similar to the results on the two datasets of 100 portfolios, the estimated optimal unconstrained portfolio, as seen in Table 6, still includes large short positions, especially when $T = 120$. Such positions would require heavily shorting certain assets, resulting in exposures that are unrealistically large relative to the total portfolio value. In practice, this would

be difficult to implement. The 130/30 portfolio, however, limits the size of short positions, making it far more feasible to trade in general.[8]

To shed light on the above issues, we re-estimate the SDF using what is likely the largest available set of long–short portfolios. Based on the long–short portfolio returns constructed by Chen and Zimmermann (2022), we consider two cases. The first uses 109 portfolios with an out-of-sample investment period from January 1980 to December 2023. The second uses 140 portfolios with an out-of-sample period from January 1990 to December 2023.

Table 7 reports the results. Interestingly, consistent with our theory, the estimated optimal portfolio underperforms by more than 50% when $T = 120$, since N is large relative to T. Its performance improves markedly, however, as T increases to $T = 240$. In this setting, the naive $1/N$ rule performs much better than it did before. A likely explanation is that as the space of anomalies expands, their common component becomes more important and earns a positive risk premium, which the $1/N$ rule naturally captures. Note that there is no need to use the adjusted $1/N$ here, as the long–short portfolio returns in the database have already been sign-adjusted.

The most notable finding is that the average Sharpe ratios almost double compared to the earlier case with only 44 assets. This highlights the importance of considering large portfolios: as N increases, the Sharpe ratio tends to rise, bringing the estimated SDF closer to the true one.[9] Consequently, the associated pricing errors are likely to be smaller than before.

7 Conclusion

Our paper studies the role of a flexible short-sale constraints on the optimal portfolio or the stochastic discount factor. We show that under no-short-sale constraints, and in a

[8] In our applications to anomalies, the absolute size of the positions matter. The 130/30 portfolio has also reasonable maximum positive position.

[9] Over the same investment period, the greater the N, the greater the Sharpe ratio theoretically. Here the 109 portfolio case has a different investment period than the 140 case.

factor model, as the number of assets grows, its Sharpe ratio converges to that of the true optimal portfolio, even if the latter might involve short positions. This implies that long-only investors can potentially approximate efficiency in large markets and that the assumption of Koijen and Yogo (2019) demand-based asset pricing could be valid, so that the implied pricing kernel from institutional holdings could closely match the true kernel.

We further show that a limited amount of short-sell positions can fundamentally improve portfolio estimation accuracy in high-dimensional settings. We establish that, under these constraints, the estimated optimal portfolio not only converges to its population counterpart, but also closely approximates the true efficient portfolio in a factor model. These results imply that, in practice, imposing no-short-sale or limited-short-sale constraints serves as an effective form of regularization to mitigate estimation errors. Moreover, such constraints have negligible impact on the unconstrained optimality, provided that a factor model adequately explains asset returns.

We also show that the constrained optimal portfolio typically outperforms the widely studied $1/N$ rule in high-dimensional settings, except in special cases where $1/N$ is optimal in population. Our framework also generalizes naturally to limited short-selling, with direct applications to popular strategies such as 130/30 portfolios, providing both theoretical justification and practical relevance.

It may be emphasized that our flexible short-sale constraints not only resolve a central econometric failure of the Markowitz portfolio, but they provide a way to obtain realistic optimal portfolio or the stochastic discount factor in practice. Without constraints, the estimated optimal portfolio does not converge when the asset dimension is large. In contrast, imposing the constraint guarantees consistency of the estimated optimal portfolio and its Sharpe ratio. Existing studies on the optimal portfolio or the stochastic discount factor have rarely any short-sale constraints, so that the portfolio strategy may in fact be too difficult to implement in the real world. In contrast, our flexible short-sale constraints make estimated portfolio realistic and possess the consistency property to converge to the true population portfolio under the same constraints.

By highlighting the power of simple economic constraints, our work contributes to the

literature on portfolio choice under estimation risk, the convergence of optimal portfolios in high dimensions, and the role of short selling in asset pricing. Unlike approaches that rely on sparsity assumptions or external estimators, our method achieves consistency at the best available convergence rate, regardless of the underlying data-generating process. Beyond its theoretical contributions, our analysis emphasizes the broader importance of short-selling mechanisms in mitigating estimation error and suggests that similar benefits may arise in settings where short sales are restricted or indirectly replicated.

At present, numerous documented anomalies in finance (e.g., Harvey, Liu, and Zhu 2015; Chen and Zimmermann 2022) are constructed from long–short portfolios, and many shrinkage methods currently employed are unconstrained. The widely used tests of portfolio efficiency allow for unrealistic large short positions. The effects of imposing no-short-sale restrictions or more general ζ-constraints remain largely unexplored.[10] Investigating these constraints and their implications represents a promising avenue for future research.

[10] DeMiguel, Martín-Utrera, and Uppal (2024) investigate the impact of short-sale-constrained benchmarks on mutual fund performance. Their study considers a small-N case, which differs from ours.

References

Ao, Mengmeng, Yingying Li, and Xinghua Zheng, 2019, Approaching mean-variance efficiency for large portfolios, *Review of Financial Studies* 32, 2890–2919.

Balduzzi, Pierluigi, and Anthony W. Lynch, 1999, Transaction costs and predictability: some utility cost calculations, *Journal of Financial Economics* 52, 47–78.

Bawa, Vijay S., Stephen J. Brown, and Roger W. Klein, 1979, *Estimation Risk and Optimal Portfolio Choice* (North-Holland: Amsterdam).

Brennan, Thomas J., and Andrew W. Lo, 2010, Impossible frontiers, *Management Science* 56, 905–923.

Brown, Stephen J., 1976, Optimal Portfolio Choice under Uncertainty: A Bayesian Approach, Ph.D. dissertation, University of Chicago.

Chen, Andrew Y., and Tom Zimmermann, 2022, Open source cross-sectional asset pricing, *Critical Finance Review* 11, 207–264.

Chincarini, Ludwig B., and Daehwan Kim, 2023, *Quantitative Equity Portfolio Management: An Active Approach to Portfolio Construction and Management* (McGraw-Hill: New York) 2 edn.

Dai, Min, Hanqing Jin, and Hong Liu, 2011, Illiquidity, position limits, and optimal investment for mutual funds, *Journal of Economic Theory* 146, 1598–1630.

Davis, Carter, 2025, The elasticity of quantitative investment, *The Review of Financial Studies* 38, 2845–2886.

DeMiguel, Victor, Lorenzo Garlappi, Francisco J. Nogales, and Raman Uppal, 2009, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Management Science* 55, 798–812.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009, Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?, *Review of Financial Studies* 22, 1915–1953.

DeMiguel, Victor, Alberto Martín-Utrera, and Francisco J Nogales, 2015, Parameter uncertainty in multiperiod portfolio optimization with transaction costs, *Journal of Financial and Quantitative Analysis* 50, 1443–1471.

DeMiguel, Victor, Alberto Martín-Utrera, Francisco J. Nogales, and Raman Uppal, 2020, A transaction-cost perspective on the multitude of firm characteristics, *Review of Financial Studies* 33, 2180–2222.

DeMiguel, Victor, Alberto Martín-Utrera, and Raman Uppal, 2024, Rethinking mutual fund performance: From traditional alpha to achievable alpha, Working Paper.

Detzel, Andrew L., Robert Novy-Marx, and Mihail Velikov, 2023, Model comparison with transaction costs, *The Journal of Finance* 78, 1743–1775.

Duchin, Ran, and Haim Levy, 2009, Markowitz versus the talmudic portfolio diversification strategies, *Journal of Portfolio Management* 35, 71–74.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

——— , 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fan, Jianqing, Jingjin Zhang, and Ke Yu, 2012, Vast portfolio selection with gross-exposure constraints, *Journal of the American Statistical Association* 107, 592–606.

Gabaix, Xavier, and Ralph S. J. Koijen, 2021, In search of the origins of financial fluctuations: The inelastic markets hypothesis, Technical report, National Bureau of Economic Research.

Grinold, Richard C., and Ronald N. Kahn, 1999, *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk* (McGraw-Hill: New York).

Haddad, Valentin, Zhiguo He, Paul Huebner, Péter Kondor, and Erik Loualiche, 2025, Causal inference for asset pricing, Working Paper.

Haddad, Valentin, Paul Huebner, and Erik Loualiche, 2025, How competitive is the stock market? theory, evidence from portfolios, and implications for the rise of passive investing, *American Economic Review* 115, 975–1018.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2015, ... and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

He, Zhiguo, Péter Kondor, and Jessica S. Li, 2025, Demand elasticity in dynamic asset pricing, Working Paper.

Jagannathan, Ravi, and Tongshu Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* 58, 1651–1683.

Jorion, Philippe, 1986, Bayes–Stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* 21, 279–292.

Kan, Raymond, and Daniel R. Smith, 2008, The distribution of the sample minimum-variance frontier, *Management Science* 54, 1364–1380.

Kan, Raymond, Xiaolu Wang, and Xinghua Zheng, 2024, In-sample and out-of-sample sharpe ratios of multi-factor asset pricing models, *Journal of Financial Economics* 155, 103837.

Kan, Raymond, Xiaolu Wang, and Guofu Zhou, 2022, Optimal portfolio choice with estimation risk: No risk-free asset case, *Management Science* 68, 2047–2068.

Kan, Raymond, and Guofu Zhou, 2007, Optimal portfolio choice with parameter uncertainty, *Journal of Financial and Quantitative Analysis* 42, 621–656.

Kirby, Chris, and Barbara Ostdiek, 2012, It's all in the timing: Simple active portfolio strategies that outperform naive diversification, *Journal of Financial and Quantitative Analysis* 47, 437–467.

Koijen, Ralph S. J., Robert J. Richmond, and Motohiro Yogo, 2024, Which investors matter for equity valuations and expected returns?, *Review of Economic Studies* 91, 2387–2424.

Koijen, Ralph S. J., and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475–1515.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

Lassance, Nathan, Alberto Martín-Utrera, and Majeed Simaan, 2024, The risk of expected utility under parameter uncertainty, *Management Science* 70, 7644–7663.

Lassance, Nathan, Rodolphe Vanderveken, and Frédéric Vrins, 2024, On the combination of naive and mean-variance portfolio strategies, *Journal of Business & Economic Statistics* 42, 875–889.

Laurent, Beatrice, and Pascal Massart, 2000, Adaptive estimation of a quadratic functional by model selection, *The Annals of Statistics* 28, 1302–1338.

Ledoit, Olivier, and Michael Wolf, 2004, A well-conditioned estimator for large-dimensional covariance matrices, *Journal of Multivariate Analysis* 88, 365–411.

Luo, Bin, Guofu Zhou, and Ti Zhou, 2024, Which factors matter in the pricing kernel?, Working paper.

MacKinlay, Craig A., and Ľuboš Pástor, 2000, Asset pricing models: Implications for expected returns and portfolio selection, *Review of Financial Studies* 13, 883–916.

Markowitz, Harry, 1952, Portfolio selection, *Journal of Finance* 7, 77–91.

Novy-Marx, Robert, and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *Review of Financial Studies* 29, 104–147.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Shi, Fangquan, Lianjie Shu, Aijun Yang, and Fangyi He, 2020, Improving minimum-variance portfolios by alleviating overdispersion of eigenvalues, *Journal of Financial and Quantitative Analysis* 55, 2700–2731.

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288–302.

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

van Binsbergen, Jules H., Benjamin David, and Christian C. Opp, 2025, How (not) to identify demand elasticities in dynamic asset markets, Working Paper.

Yuan, Ming, and Guofu Zhou, 2024, Why naive $1/N$ diversification is not so naive, and how to beat it?, *Journal of Financial and Quantitative Analysis* 59, 3601–3632.

A Tables and Figures

Table 1: Population Sharpe ratios under different portfolio constraints

N	No-Short-Sale	ζ-Constraint (130/30)	No Constraint
Panel A: 2-Factor Model			
25	0.181214	0.181393	0.181393
50	0.181979	0.182072	0.182072
100	0.182336	0.182380	0.182380
200	0.182519	0.182538	0.182538
300	0.182577	0.182589	0.182589
400	0.182607	0.182615	0.182615
500	0.182622	0.182630	0.182630
1000	0.182657	0.182661	0.182661
1500	0.182668	0.182671	0.182671
2000	0.182674	0.182676	0.182676
Panel B: 5-Factor Model			
25	0.286779	0.287118	0.287121
50	0.287749	0.288028	0.288031
100	0.288259	0.288455	0.288455
200	0.288514	0.288654	0.288655
300	0.288628	0.288725	0.288725
400	0.288681	0.288759	0.288759
500	0.288716	0.288780	0.288780
1000	0.288785	0.288820	0.288820
1500	0.288808	0.288833	0.288833
2000	0.288820	0.288840	0.288840

This table reports the population Sharpe ratios of portfolios under no-short-sale, limited short-sell (130/30) and no constraints, in a 2-factor and 5-factor models, respectively. We calibrate the factor loadings, β_1 and β_2, of the 2-factor model using the market and size factors. For the 5-factor model, we add three factors to the 2-factor model by setting their means and standard deviations equal to the averages of the those of β_1 and β_2. We use 100 replicates in total. In each replicate, a new $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$ are generated along with different factor loadings β and different i.i.d. residual variances. The distributions (means and standard deviations) of the factor loadings and the residual covariance are fixed, calibrated from the 100 Portfolios formed via size and book-to-market.

Table 2: Sharpe Ratios of Estimated Optimal Portfolios

N	25	50	100	200	300	400	500
Panel A: Oracle Sharpe ratios							
No Constraint	0.978	0.998	1.007	1.011	1.012	1.013	1.014
Naive (1/N)	0.681	0.683	0.685	0.686	0.687	0.687	0.687
No-Short-Sales	0.870	0.825	0.876	0.879	0.888	0.901	0.939
ζ-Constraint (130/30)	0.909	0.902	0.933	0.948	0.948	0.958	0.977
Panel B: $N/T = 0.25$							
No Constraint	0.426	0.431	0.438	0.441	0.443	0.441	0.443
Naive (1/N)	0.681	0.683	0.685	0.686	0.687	0.687	0.687
No-Short-Sales	0.737	0.778	0.824	0.860	0.876	0.885	0.892
ζ-Constraint (130/30)	**0.746**	**0.797**	**0.852**	**0.895**	**0.914**	**0.926**	**0.934**
Panel C: $N/T = 0.75$							
No Constraint	0.160	0.165	0.167	0.162	0.161	0.160	0.160
Naive (1/N)	**0.681**	0.683	0.685	0.686	0.687	0.687	0.687
No-Short-Sales	0.663	**0.712**	0.765	0.815	0.837	0.854	0.862
ζ-Constraint (130/30)	0.650	0.710	**0.774**	**0.835**	**0.861**	**0.882**	**0.892**
Panel D: $N/T = 1.25$							
No Constraint	0.170	0.164	0.163	0.162	0.154	0.154	0.155
Naive (1/N)	**0.681**	**0.683**	0.685	0.686	0.687	0.687	0.687
No-Short-Sales	0.624	0.676	**0.729**	0.782	0.809	0.830	0.841
ζ-Constraint (130/30)	0.598	0.664	0.727	**0.793**	**0.826**	**0.850**	**0.863**
Panel E: $N/T = 1.75$							
No Constraint	0.214	0.254	0.275	0.291	0.288	0.288	0.287
Naive (1/N)	**0.681**	**0.683**	0.685	0.686	0.687	0.687	0.687
No-Short-Sales	0.596	0.651	**0.700**	0.758	0.787	0.811	0.823
ζ-Constraint (130/30)	0.569	0.630	0.690	**0.763**	**0.797**	**0.825**	**0.840**

Panel A of the table reports the oracle Sharpe ratios of the different strategies. The remaining panels present the Sharpe ratios of estimated optimal portfolios based on simulated data for various values of N/T, where the simulations are calibrated to the Fama–French five-factor model.

Table 3: Sharpe Ratios of 100 Portfolios.

tcost (bps)	tcost (bps) = 0			tcost (bps) = 10			tcost (bps) = 20		
T	120	180	240	120	180	240	120	180	240
Panel A: Size and Operating Profitability									
No Constraint	0.213	0.694	0.800	-0.323	0.521	0.681	-0.612	0.347	0.561
Naive (1/N)	0.697	0.697	0.697	0.696	0.696	0.696	0.694	0.694	0.694
No-Short-Sale Constraint	0.674	0.729	0.692	0.657	0.715	0.679	0.641	0.700	0.665
Leverage Constraint (130/30)	**0.863**	**0.906**	**0.862**	**0.839**	**0.886**	**0.844**	**0.814**	**0.865**	**0.827**
Panel B: Size and Investment									
No Constraint	0.305	0.552	0.827	-0.194	0.380	0.692	-0.635	0.207	0.557
Naive (1/N)	0.722	0.722	0.722	0.721	0.721	0.721	0.719	0.719	0.719
No-Short-Sale Constraint	0.673	0.785	0.716	0.654	0.768	0.702	0.635	0.751	0.687
Leverage Constraint (130/30)	**0.925**	**1.045**	**0.977**	**0.897**	**1.022**	**0.958**	**0.870**	**0.999**	**0.940**

This table reports the Sharpe ratios of 100 portfolios formed on Size and Operating Profitability in Panel A, and on Size and Investment in Panel B, over the period July 1963 to December 2024. The ratios are computed using a rolling window of length T under varying levels of the unit transaction cost c. The case tcost = 0 corresponds to results without transaction costs.

Table 4: Largest Negative Weight on 100 Portfolios

	Panel A: Size and Operating Profitability			*Panel B: Size and Investment*		
T	120	180	240	120	180	240
No Constraint	2.631	0.519	0.290	2.822	0.567	0.355
Naive (1/N)	0.000	0.000	0.000	0.000	0.000	0.000
No-Short-Sales	0.000	0.000	0.000	0.000	0.000	0.000
ζ-Constraint (130/30)	0.033	0.029	0.025	0.030	0.027	0.028

This table reports the largest negative portfolio weight across different investment strategies, using the data from Table 3. The reported values are averaged over all the out-of-sample investment periods from July 1963 to December 2024.

Table 5: Annnualized Sharpe Ratios using 44 Portfolios

tcost (bps)	tcost (bps) = 0			tcost (bps) = 10			tcost (bps) = 20		
T	120	180	240	120	180	240	120	180	240
No Constraint	1.806	1.949	**2.011**	1.598	1.830	**1.919**	1.384	1.707	1.824
Adj. Naive (1/N)	0.604	0.668	0.604	0.598	0.662	0.598	0.593	0.656	0.593
ζ-Constraint (130/30)	**1.916**	**1.967**	1.929	**1.852**	**1.927**	1.895	**1.787**	**1.886**	**1.861**
No-Short-Sales	1.361	1.358	1.344	1.324	1.333	1.325	1.288	1.307	1.305

This table reports the annualized Sharpe ratios of four investment strategies based on the 44 portfolios, a subset of data from Kozak, Nagel, and Santosh (2020). In the table here, however, the raw Naive method is replaced with an Adjusted Naive method, which assigns a weight of $-1/N$ when a portfolio's past average return is negative and $1/N$ otherwise. While the above used data from January 1964 to December 2019, the results above show the out-of-sample investment period from January 1984 to December 2019.

Table 6: Largest Negative Weight of 44 Portfolios

T	120	180	240
No Constraint	0.878	0.613	0.522
Adjusted Naive (1/N)	0.023	0.023	0.023
ζ-Constraint (130/30)	0.227	0.224	0.266
No-Short-Sales	0.000	0.000	0.000

This table reports the average maximum negative weight of different methods, based on a subset of the data from Kozak, Nagel, and Santosh (2020). The reported values are averaged over the out-of-sample investment period from January 1984 to December 2019.

Table 7: Annualized Sharpe Ratios using 109 and 140 Portfolios

tcost (bps)	tcost (bps) = 0			tcost (bps) = 10			tcost (bps) = 20		
T	120	180	240	120	180	240	120	180	240
Panel A: 109 Portfolios									
No Constraint	1.436	3.395	3.597	-0.690	3.069	3.398	-0.806	2.685	3.179
Naive (1/N)	1.858	1.858	1.858	1.848	1.848	1.848	1.838	1.838	1.838
ζ-Constraint (130/30)	3.108	**3.588**	**3.684**	2.877	3.452	**3.589**	2.563	3.302	3.486
No-Short-Sales	**3.454**	3.577	3.618	**3.373**	**3.517**	3.571	**3.288**	**3.455**	**3.523**

tcost (bps)	tcost (bps) = 0			tcost (bps) = 10			tcost (bps) = 20		
T	150	180	240	150	180	240	150	180	240
Panel B: 140 Portfolios									
No Constraint	1.156	2.126	3.004	-0.472	1.259	2.678	-0.494	-0.050	2.294
Naive (1/N)	1.770	1.770	1.770	1.761	1.761	1.761	1.751	1.751	1.751
ζ-Constraint (130/30)	2.892	3.006	3.190	2.641	2.836	3.083	2.278	2.624	2.962
No-Short-Sales	**3.165**	**3.139**	**3.326**	**3.088**	**3.076**	**3.279**	**3.007**	**3.008**	**3.230**

The table reports annualized Sharpe ratios based on long-short portfolio returns sourced from the OpenAssetPricing web repository (Chen and Zimmermann, 2022). The dataset of 109 portfolios contains data from January 1960 to December 2023, but the out-of-sample investment period is from January 1980 to December 2023. The dataset of 140 portfolios, available from January 1970 to December 2023, is evaluated over the out-of-sample period of January 1990 to December 2023.

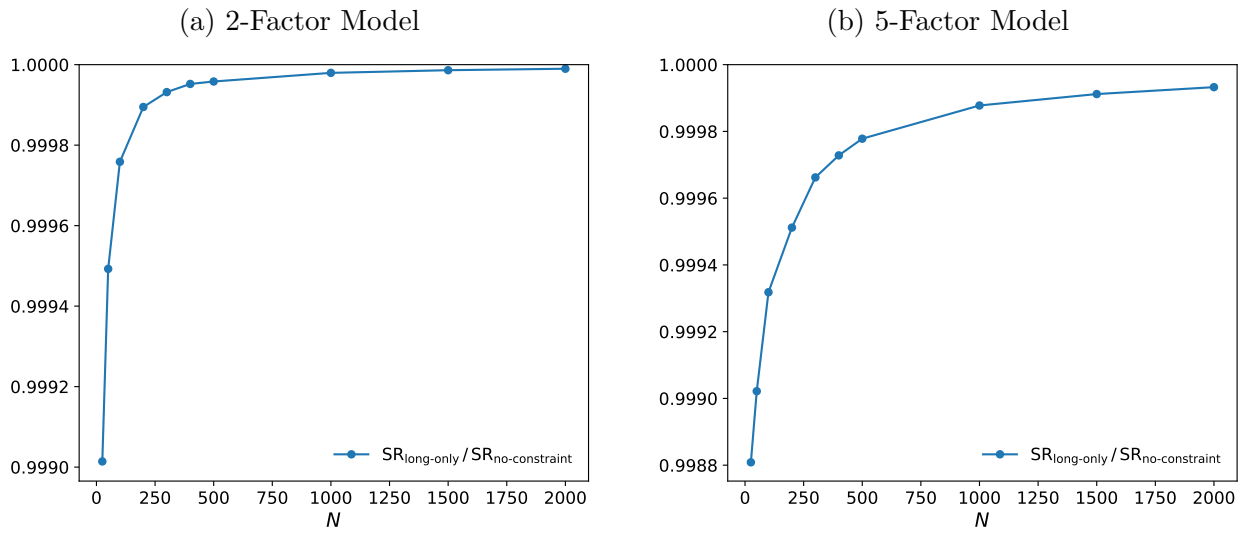


Figure 1: Ratio of the long-only Sharpe ratio to the no-constraint Sharpe ratio

This figure reports the ratio of population Sharpe ratio under the no-short-sale constraint to that under no-constraints, in a 2-factor and 5-factor models, respectively. We calibrate the factor loadings, β_1 and β_2, of the 2-factor model using the market and size factors. For the 5-factor model, we add three factors to the 2-factor model by setting their means and standard deviations equal to the averages of the those of β_1 and β_2. We use 100 replicates in total. In each replicate, a new $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$ are generated along with different factor loadings β and different i.i.d. residual variances. The distributions (means and standard deviations) of the factor loadings and the residual covariance are fixed, calibrated from the 100 Portfolios formed via size and book-to-market with data from July 1963 to December 2024.

B Proofs

In our proofs, we define $\|A\| = \sqrt{\lambda_{\max}(A^\top A)}$, where $\lambda_{\max}$ is the largest eigenvalue; $\|A\|_{\max} = \max_{1\le i,j\le N} |A_{ij}|$; and $\|A\|_{\ell_2\to\ell_\infty} = \sup_{\|\mathbf{x}\|_{\ell_2}=1} \|A\mathbf{x}\|_{\ell_\infty} = \max_{1\le i\le N} \|A_i\|_{\ell_2}$, where A_i is the i-th row of A.

B.1 Proof of Proposition 1

Proof. First, consider the no-short-sale constrained optimization:

$$\mathbf{w}_*^+ := \operatorname*{argmax}_{\mathbf{w}} \boldsymbol{\mu}^\top \mathbf{w} - \frac{\gamma}{2}\mathbf{w}^\top \boldsymbol{\Sigma}\mathbf{w} \qquad \text{s.t.} \quad \mathbf{w} \ge \mathbf{0},$$

and by the Karush-Kuhn-Tucker conditions, we have

$$\begin{aligned} \boldsymbol{\mu} - \gamma\boldsymbol{\Sigma}\mathbf{w}_*^+ + \lambda &= 0, \\ \lambda_i &\ge 0, \quad i \in [N], \\ \mathbf{w}_*^+ &\ge 0, \\ \lambda_i\, \mathbf{w}_{*i}^+ &= 0, \quad i \in [N]. \end{aligned}$$

Let A be the set of assets held in the no-short-sale constrained portfolio. Then for all $i \in A$, we have $\mathbf{w}_{*i}^+ > 0$, which implies $\lambda_i = 0$. Restricting to the subset A, we obtain

$$\boldsymbol{\mu}_A - \gamma\boldsymbol{\Sigma}_A\mathbf{w}_A^+ = 0,$$

where $\mathbf{w}_A^+$ is the subvector of $\mathbf{w}_*^+$ with length $|A|$. Similarly, defining $\mathbf{w}_{A^c}^+$ for the complement set, we have $\mathbf{w}_{A^c}^+ = 0$.

Next, consider the unconstrained optimization:

$$\mathbf{w}_A^* := \operatorname*{argmax}_{\mathbf{w}} \boldsymbol{\mu}_A^\top \mathbf{w} - \frac{\gamma}{2}\mathbf{w}^\top \boldsymbol{\Sigma}_A\mathbf{w},$$

Since $-\boldsymbol{\Sigma}_A \prec 0$, the sufficient condition for the maximum is

$$\boldsymbol{\mu}_A - \gamma\boldsymbol{\Sigma}_A\mathbf{w}_A^* = 0.$$

Note that $\mathbf{w}_A^+$ also satisfies this condition. Therefore, we conclude that $\mathbf{w}_A^+ = \mathbf{w}_A^*$. □

B.2 Proof of Theorem 1

The general proof strategy is similar to the two-factor case discussed in the main text. Without loss of generality, assume that $s_1 = \cdots = s_K = 1$ for brevity. The general case follows by the same argument after replacing each b_{jk} with $s_k b_{jk}$.

We first construct equal-weight portfolios w_j over A_j, i.e., $w_j = (w_{ji})_{i=1,\ldots,N}$ where $w_{ji} = \frac{1}{N_j}\mathbb{I}\{i \in A_j\}$ and A_j is defined in Theorem 1. For $j = 1, \ldots, K$, the returns of w_j satisfy

$$R_j = \bar{b}_{j1} f_1 + \cdots + \bar{b}_{jK} f_K + \frac{1}{N_j} \sum_{i \in A_j} \varepsilon_i,$$

where N_j is the cardinality of A_j, and $\bar{b}_{jk} = \frac{\sum_{i \in A_j} b_{ik}}{N_j}$. By definition of A_j, for all $i \in A_j$ we have $\sum_{k=1}^K s_k b_{ik} > 0$, and for all $l \neq j$, $s_l b_{il} < 0$. Since $\bar{b}_{jk} = \frac{\sum_{i \in A_j} b_{ik}}{N_j}$, it follows that $\sum_{k=1}^K s_k \bar{b}_{jk} > 0$ and for all $l \neq j$, $s_l \bar{b}_{jl} < 0$. Under the normalization $s_1 = \cdots = s_K = 1$, this is equivalent to

$$\bar{B} = \begin{pmatrix} \bar{b}_{11} & \bar{b}_{12} & \cdots & \bar{b}_{1K} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{b}_{K1} & \bar{b}_{K2} & \cdots & \bar{b}_{KK} \end{pmatrix}$$

being a diagonally dominant M-matrix, i.e., $\bar{b}_{jj} > \sum_{l \neq j} |\bar{b}_{jl}|$ and $\bar{b}_{jl} < 0$ for all $l \neq j$. Thus, $\bar{B}$ is nonsingular and inverse-positive; that is, all entries of $\bar{B}^{-1}$ are nonnegative. Also, we claim that the eigenvalue of $\bar{B}$ is lower bounded, i.e., $\bar{B} \succeq \varepsilon \mathbb{I}$.

To show this claim, we use Gershgorin's circle theorem, and we know for any eigenvalue λ of $\bar{B}$, we have

$$|\lambda - \bar{b}_{jj}| \leq \sum_{k \neq j} |\bar{b}_{jk}| = -\sum_{k \neq j} \bar{b}_{jk}.$$

Thus, we know that $\lambda \geq \min_{j \in [K]} \{\sum_{k=1}^K \bar{b}_{jk}\}$. Note that $\sum_{k=1}^K \bar{b}_{jk} = \sum_{i \in A_j} \frac{\sum_{k=1}^K b_{ik}}{N_j} \geq \varepsilon$. Then we finish the proof of the claim.

Denote

$$p = \bar{B}^{-1,\top}(\mathbf{\Sigma}_f^{-1} \lambda_f).$$

Since $p_j = e_j^\top \bar{B}^{-1,\top} \mathbf{\Sigma}_f^{-1} \lambda_f$, $R_j = e_j^\top \bar{B} f$, and f has mean λ_f and variance Σ_f, we obtain

$$\mathbb{E}\left(\sum_{j=1}^K p_j R_j \right) = \lambda_f^\top \mathbf{\Sigma}_f^{-1} \lambda_f,$$

$$\begin{aligned}\mathtt{Var}\left(\sum_{j=1}^{K} p_j R_j\right) &= \lambda_f^\top \mathbf{\Sigma}_f^{-1}\lambda_f + \sum_{k=1}^{K} p_k^2 \,\mathtt{Var}\Big(\tfrac{1}{N_k}\sum_{j\in A_k}\varepsilon_j\Big) \\ &= \lambda_f^\top \mathbf{\Sigma}_f^{-1}\lambda_f + O\big((\min\{N_1,\ldots,N_K\})^{-1}\big)\|\mathbf{\Sigma}_f^{-1}\lambda_f\|^2\|\bar{B}^{-1}\|^2.\end{aligned}$$

Since $\min\{N_1,\ldots,N_K\}\to\infty$, $\|\mathbf{\Sigma}_f^{-1}\lambda_f\|$ is fixed and $\|\bar{B}^{-1}\|\le 1/\varepsilon$, these results together imply

$$\theta\left(\sum_{j=1}^{K} p_j w_j\right) = \frac{\mathbb{E}(\sum_{j=1}^{K} p_j R_j)}{\sqrt{\mathtt{Var}(\sum_{j=1}^{K} p_j R_j)}} \;\to\; (\lambda_f^\top \mathbf{\Sigma}_f^{-1}\lambda_f)^{1/2} = \theta_*$$

as $N\to\infty$, where w_j is the equal-weight portfolio over A_j.

It remains to show that $\sum_{j=1}^{K} p_j w_j$ has no short positions. We have shown that $\bar{B}$ is inverse-positive. Moreover, all entries of $\mathbf{\Sigma}_f^{-1}\lambda_f$ are nonnegative by the assumption that $s_k = 1$ for all $k = 1,\ldots,K$. Hence, all entries of p are also nonnegative. Together with the fact that the w_j's have no short positions, we conclude that $\sum_{j=1}^{K} p_j w_j$ has no short positions.

B.3 Proof of Theorem 2

By the definition of $\widehat{\mathbf{w}}^+$, the concavity of $\widehat{U}$, and the convexity of $\{\widehat{\mathbf{w}}^+\ge 0\}$, the first-order necessary condition for a maximum implies that, for any $\mathbf{w}\ge 0$,

$$\gamma\widehat{\mathbf{w}}^{+,\top}\widehat{\mathbf{\Sigma}}(\mathbf{w}-\widehat{\mathbf{w}}^+) - \widehat{\boldsymbol{\mu}}^\top(\mathbf{w}-\widehat{\mathbf{w}}^+)\ge 0.$$

Given the definition of $\mathbf{w}_*$, this implies

$$(\mathbf{w}_*-\widehat{\mathbf{w}}^+)^\top\widehat{\mathbf{\Sigma}}(\mathbf{w}-\widehat{\mathbf{w}}^+) \le \frac{1}{\gamma}(\gamma\widehat{\mathbf{\Sigma}}\mathbf{w}_*-\widehat{\boldsymbol{\mu}})^\top(\mathbf{w}-\widehat{\mathbf{w}}^+) \le \frac{1}{\gamma}\|\gamma\widehat{\mathbf{\Sigma}}\mathbf{w}_*-\widehat{\boldsymbol{\mu}}\|_{\ell_\infty}\|\mathbf{w}-\widehat{\mathbf{w}}^+\|_{\ell_1}.$$

Let

$$\delta_{T,N} = \max\{\|\gamma\widehat{\mathbf{\Sigma}}\mathbf{w}_*-\widehat{\boldsymbol{\mu}}\|_{\ell_\infty}, \|\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\|_{\max}, \|(\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}})\mathbf{\Sigma}^{-1/2}\|_{\ell_2\to\ell_\infty}, \|\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\|_{\ell_\infty}\},$$

which yields

$$(\mathbf{w}_*-\widehat{\mathbf{w}}^+)^\top\widehat{\mathbf{\Sigma}}(\mathbf{w}-\widehat{\mathbf{w}}^+) \le \frac{\delta_{T,N}}{\gamma}\|\mathbf{w}-\widehat{\mathbf{w}}^+\|_{\ell_1}.$$

Thus,

$$(\mathbf{w}_* - \widehat{\mathbf{w}}^+)^\top \mathbf{\Sigma}(\mathbf{w} - \widehat{\mathbf{w}}^+) \le \frac{\delta_{T,N}}{\gamma}\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1} + (\mathbf{w}_* - \widehat{\mathbf{w}}^+)^\top(\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}})(\mathbf{w} - \widehat{\mathbf{w}}^+).$$

By the triangle inequality,

$$\left|(\mathbf{w}_* - \widehat{\mathbf{w}}^+)^\top(\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}})(\mathbf{w} - \widehat{\mathbf{w}}^+)\right| \le \left|(\mathbf{w} - \widehat{\mathbf{w}}^+)^\top(\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}})(\mathbf{w} - \widehat{\mathbf{w}}^+)\right| + \left|(\mathbf{w}_* - \mathbf{w})^\top(\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}})(\mathbf{w} - \widehat{\mathbf{w}}^+)\right|.$$

The first term is bounded by

$$\left|(\mathbf{w} - \widehat{\mathbf{w}}^+)^\top(\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}})(\mathbf{w} - \widehat{\mathbf{w}}^+)\right| \le \|\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}\|_{\max}\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1}^2,$$

and the second term by

$$\begin{aligned}\left|(\mathbf{w}_* - \mathbf{w})^\top(\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}})(\mathbf{w} - \widehat{\mathbf{w}}^+)\right| &\le \|(\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}})(\mathbf{w}_* - \mathbf{w})\|_{\ell_\infty}\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1} \\ &\le \|(\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}})\mathbf{\Sigma}^{-1/2}\|_{\ell_2 \to \ell_\infty}\|\mathbf{\Sigma}^{1/2}(\mathbf{w}_* - \mathbf{w})\|\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1}.\end{aligned}$$

This leads to

$$(\mathbf{w}_* - \widehat{\mathbf{w}}^+)^\top \mathbf{\Sigma}(\mathbf{w} - \widehat{\mathbf{w}}^+) \le \frac{\delta_{T,N}}{\gamma}\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1} + \delta_{T,N}\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1}^2 + \delta_{T,N}\|\mathbf{\Sigma}^{1/2}(\mathbf{w}_* - \mathbf{w})\|\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1}.$$

Observe that

$$2(\mathbf{w}_* - \widehat{\mathbf{w}}^+)^\top \mathbf{\Sigma}(\mathbf{w} - \widehat{\mathbf{w}}^+) = \|\mathbf{\Sigma}^{1/2}(\mathbf{w} - \widehat{\mathbf{w}}^+)\|^2 + \|\mathbf{\Sigma}^{1/2}(\mathbf{w}_* - \widehat{\mathbf{w}}^+)\|^2 - \|\mathbf{\Sigma}^{1/2}(\mathbf{w} - \mathbf{w}_*)\|^2.$$

Therefore, we can show that for any $\mathbf{w} \in \mathbb{R}_+^N$,

$$\begin{aligned}\|\mathbf{\Sigma}^{1/2}(\mathbf{w}_* - \widehat{\mathbf{w}}^+)\|^2 \le \|\mathbf{\Sigma}^{1/2}(\mathbf{w} - \mathbf{w}_*)\|^2 - \|\mathbf{\Sigma}^{1/2}(\mathbf{w} - \widehat{\mathbf{w}}^+)\|^2 + \frac{2\delta_{T,N}}{\gamma}\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1} \\ + 2\delta_{T,N}\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1}^2 + 2\delta_{T,N}\|\mathbf{\Sigma}^{1/2}(\mathbf{w}_* - \mathbf{w})\|\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1}. \quad (37)\end{aligned}$$

For any $\mathbf{w} \in \mathbb{R}_+^N$, denote by $\mathbf{S} = \mathbf{S}(\mathbf{w}) = \mathrm{supp}(\mathbf{w}) := \{i : w_i \neq 0\}$. Write $\widehat{\mathbf{w}}_S^+$ as an N-vector whose ith entry is $\widehat{w}_i$ if $i \in \mathbf{S}$ and 0 otherwise, and $\widehat{\mathbf{w}}_{\mathbf{S}^c}^+ = \widehat{\mathbf{w}}^+ - \widehat{\mathbf{w}}_S^+$. Then, we have

$$\begin{aligned}\|\mathbf{\Sigma}^{1/2}(\mathbf{w} - \widehat{\mathbf{w}}^+)\|^2 &\ge \|\mathbf{\Sigma}^{1/2}\widehat{\mathbf{w}}_{\mathbf{S}^c}^+\|^2 - 2(\mathbf{w} - \widehat{\mathbf{w}}_S^+)^\top \mathbf{\Sigma}\widehat{\mathbf{w}}_{\mathbf{S}^c}^+ \\ &\ge \|\mathbf{\Sigma}^{1/2}\widehat{\mathbf{w}}_{\mathbf{S}^c}^+\|^2 - 2\|\mathbf{\Sigma}(\mathbf{w} - \widehat{\mathbf{w}}_S^+)\|_{\ell_\infty}\|\widehat{\mathbf{w}}_{\mathbf{S}^c}^+\|_{\ell_1} \\ &\ge \|\mathbf{\Sigma}^{1/2}\widehat{\mathbf{w}}_{\mathbf{S}^c}^+\|^2 - 2\sigma_{\max}^2\|\mathbf{w} - \widehat{\mathbf{w}}_S^+\|_{\ell_1}\|\widehat{\mathbf{w}}_{\mathbf{S}^c}^+\|_{\ell_1}. \quad (38)\end{aligned}$$

We shall consider first the case when

$$\|\mathbf{w} - \widehat{\mathbf{w}}_S^+\|_{\ell_1} \leq \frac{\sigma_g^2}{4\sigma_{\max}^2} \|\widehat{\mathbf{w}}_{\mathbf{S}^c}^+\|_{\ell_1}. \tag{39}$$

which implies that

$$\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1} \leq \|\mathbf{w} - \widehat{\mathbf{w}}_S^+\|_{\ell_1} + \|\widehat{\mathbf{w}}_{S^c}^+\|_{\ell_1} \leq (1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}) \|\widehat{\mathbf{w}}_{S^c}^+\|_{\ell_1}. \tag{40}$$

By the definition of global minimum variance portfolio,

$$\mathbf{x}^\top \mathbf{\Sigma} \mathbf{x} \geq \sigma_g^2 \|\mathbf{x}\|_{\ell_1}^2$$

for any $\mathbf{x} \in \mathbb{R}_+^N$. Recall that $\widehat{\mathbf{w}}_{\mathbf{S}^c}^+ \in \mathbb{R}_+^N$, we have

$$2\sigma_{\max}^2 \|\mathbf{w} - \widehat{\mathbf{w}}_S^+\|_{\ell_1} \|\widehat{\mathbf{w}}_{\mathbf{S}^c}^+\|_{\ell_1} \leq \frac{\sigma_g^2}{2} \|\widehat{\mathbf{w}}_{\mathbf{S}^c}^+\|_{\ell_1}^2 \leq \frac{1}{2} \|\mathbf{\Sigma}^{1/2} \widehat{\mathbf{w}}_{\mathbf{S}^c}^+\|^2.$$

Together with (38) and (40), this implies that

$$\|\mathbf{\Sigma}^{1/2}(\mathbf{w} - \widehat{\mathbf{w}}^+)\|^2 \geq \frac{1}{2} \|\mathbf{\Sigma}^{1/2} \widehat{\mathbf{w}}_{\mathbf{S}^c}^+\|^2 \geq \frac{\sigma_g^2}{2} \|\widehat{\mathbf{w}}_{\mathbf{S}^c}^+\|_{\ell_1}^2 \geq \frac{\sigma_g^2}{2} \left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^{-2} \|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1}^2. \tag{41}$$

Let

$$A = \|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1}, \quad B = \frac{2\delta_{T,N}}{\gamma},$$

Then by Young's inequality with $\varepsilon = \frac{\sigma_g^2}{4} \left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^{-2}$, we obtain

$$A \cdot B \leq \frac{\varepsilon}{2} A^2 + \frac{1}{2\varepsilon} B^2.$$

Hence, we establish the fact that

$$\frac{2\delta_{T,N}}{\gamma} \|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1} \leq \frac{\sigma_g^2}{8} \left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^{-2} \|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1}^2 + \frac{8\delta_{T,N}^2}{\gamma^2 \sigma_g^2} \left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^2. \tag{42}$$

Similarly, we can obtain

$$\begin{aligned} 2\delta_{T,N} \|\mathbf{\Sigma}^{1/2}(\mathbf{w}_* - \mathbf{w})\| \|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1} \leq \frac{\sigma_g^2}{8} \left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^{-2} \|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1}^2 \\ + \frac{8\delta_{T,N}^2}{\sigma_g^2} \left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^2 \|\mathbf{\Sigma}^{1/2}(\mathbf{w}_* - \mathbf{w})\|^2. \end{aligned} \tag{43}$$

Then, by imputing (41), (42) and (43) into (37), we have

$$\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w}_* - \widehat{\mathbf{w}}^+)\|^2 \leq \left[1 + \frac{8\delta_{T,N}^2}{\sigma_g^2}\left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^2\right]\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w} - \mathbf{w}_*)\|^2 + \frac{8\delta_{T,N}^2}{\gamma^2\sigma_g^2}\left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^2 - \left[\frac{\sigma_g^2}{4}\left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^{-2} - 2\delta_{T,N}\right]\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1}^2.$$

Since, under the event $\mathcal{E}$,

$$\delta_{T,N} \leq \frac{\sigma_g^2}{8}\left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^{-2},$$

we have

$$\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w}_* - \widehat{\mathbf{w}}^+)\|^2 \leq \left[1 + \frac{8\delta_{T,N}^2}{\sigma_g^2}\left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^2\right]\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w} - \mathbf{w}_*)\|^2 + \frac{8\delta_{T,N}^2}{\gamma^2\sigma_g^2}\left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^2. \tag{44}$$

Note that (44) holds for any $\mathbf{w} \in \mathbb{R}_+$ satisfying (39). As such it remains valid when replacing $\mathbf{w}$ with $x\mathbf{w}$ on the right hand side for any $x > 0$. In particular, following the same argument as in the proof of Theorem 3, we, taking $x = \mathbf{w}^\top\boldsymbol{\mu}/(\gamma\mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w})$, have

$$\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w} - x\mathbf{w}_*)\|^2 = \frac{1}{\gamma^2}[\theta_*^2 - \theta^2(\mathbf{w})].$$

Moreover, the left hand side of (44) can be bounded below by $\inf_{x>0}\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w}_* - \widehat{\mathbf{w}}^+)\|^2 = [\theta_*^2 - \theta^2(\widehat{\mathbf{w}}^+)]/\gamma^2$. Together, they imply that

$$\theta^2(\widehat{\mathbf{w}}^+) \geq \sup_{\mathbf{w}\in\mathbb{R}_+^N}\theta^2(\mathbf{w}) - \left[\frac{8}{\sigma_g^2}\left(1 + \frac{\sigma_g^2}{4\sigma_{\max}^2}\right)^2(1 + \theta_*^2)\right]\cdot\delta_{T,N}^2. \tag{45}$$

It now suffices to consider the case when

$$\|\mathbf{w}_S - \widehat{\mathbf{w}}_S^+\|_{\ell_1} > \frac{\sigma_g^2}{4\sigma_{\max}^2}\|\widehat{\mathbf{w}}_{\mathbf{S}^c}^+\|_{\ell_1}.$$

Note that in this case

$$\begin{aligned}
\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1} &\leq \left(1 + \frac{4\sigma_{\max}^2}{\sigma_g^2}\right)\|\mathbf{w}_S - \widehat{\mathbf{w}}_S^+\|_{\ell_1} \\
&\leq \left(1 + \frac{4\sigma_{\max}^2}{\sigma_g^2}\right)|\mathbf{S}|^{1/2}\|\mathbf{w}_S - \widehat{\mathbf{w}}_S^+\| \\
&\leq \left(1 + \frac{4\sigma_{\max}^2}{\sigma_g^2}\right)|\mathbf{S}|^{1/2}\|\mathbf{w} - \widehat{\mathbf{w}}^+\| \\
&\leq \left(1 + \frac{4\sigma_{\max}^2}{\sigma_g^2}\right)|\mathbf{S}|^{1/2}\lambda_{\min}^{-1/2}(\boldsymbol{\Sigma})\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w} - \widehat{\mathbf{w}}^+)\|.
\end{aligned}$$

Similar to the previous case, we can derive from the last inequality that

$$\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w}_* - \widehat{\mathbf{w}}^+)\|^2 \le \left[1 + \frac{4\delta_{T,N}^2|\mathbf{S}|}{\lambda_{\min}(\boldsymbol{\Sigma})}\left(1 + \frac{4\sigma_{\max}^2}{\sigma_g^2}\right)^2\right]\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w} - \mathbf{w}_*)\|^2 + \frac{4\delta_{T,N}^2|\mathbf{S}|}{\lambda_{\min}(\boldsymbol{\Sigma})}\left(1 + \frac{4\sigma_{\max}^2}{\sigma_g^2}\right)^2 - \left[\frac{\lambda_{\min}(\boldsymbol{\Sigma})}{2|\mathbf{S}|}\left(1 + \frac{4\sigma_{\max}^2}{\sigma_g^2}\right)^{-2} - 2\delta_{T,N}\right]\|\mathbf{w} - \widehat{\mathbf{w}}^+\|_{\ell_1}^2.$$

If

$$|\mathbf{S}| \le \frac{\lambda_{\min}(\boldsymbol{\Sigma})}{4\delta_{T,N}}\left(1 + \frac{4\sigma_{\max}^2}{\sigma_g^2}\right)^{-2}, \tag{46}$$

then

$$\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w}_* - \widehat{\mathbf{w}}^+)\|^2 \le \left[1 + \frac{4\delta_{T,N}^2|\mathbf{S}|}{\lambda_{\min}(\boldsymbol{\Sigma})}\left(1 + \frac{4\sigma_{\max}^2}{\sigma_g^2}\right)^2\right]\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w} - \mathbf{w}_*)\|^2 + \frac{4\delta_{T,N}^2|\mathbf{S}|}{\lambda_{\min}(\boldsymbol{\Sigma})}\left(1 + \frac{4\sigma_{\max}^2}{\sigma_g^2}\right)^2.$$

Now by the same argument as before, we can replace $\mathbf{w}$ on the right hand side by $x\mathbf{w}$ where $x = \mathbf{w}^\top\boldsymbol{\mu}/(\gamma\mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w})$, leading to

$$\theta^2(\widehat{\mathbf{w}}^+) \ge \sup_{\mathbf{w}\in\mathbb{R}_+^N, |\mathrm{supp}(\mathbf{w})|\le k}\left\{\theta^2(\mathbf{w}) - \left[\frac{4\gamma^2}{\lambda_{\min}(\boldsymbol{\Sigma})}\left(1 + \frac{4\sigma_{\max}^2}{\sigma_g^2}\right)^2(1 + \theta_*^2)\right]\cdot k\delta_{T,N}^2\right\}. \tag{47}$$

Finally, the case when (46) does not hold follows from Theorem 3 by setting $\zeta = 0$.

B.4 Proof of Theorem 3

Using a similar method as in (37), we can show that $\forall\mathbf{w} \in \Omega(\zeta)$,

$$\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w}_* - \widehat{\mathbf{w}})\|^2 \le \|\boldsymbol{\Sigma}^{1/2}(\mathbf{w} - \mathbf{w}_*)\|^2 - \|\boldsymbol{\Sigma}^{1/2}(\mathbf{w} - \widehat{\mathbf{w}})\|^2 + \frac{2\delta_{T,N}}{\gamma}\|\mathbf{w} - \widehat{\mathbf{w}}\|_{\ell_1} + 2\delta_{T,N}\|\mathbf{w} - \widehat{\mathbf{w}}\|_{\ell_1}^2 + 2\delta_{T,N}\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w}_* - \mathbf{w})\|\|\mathbf{w} - \widehat{\mathbf{w}}\|_{\ell_1}. \tag{48}$$

Applying the Cauchy-Schwarz inequality to the last term yields

$$\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w}_* - \widehat{\mathbf{w}})\|^2 \le (1 + \delta_{T,N})\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w} - \mathbf{w}_*)\|^2 + \frac{2\delta_{T,N}}{\gamma}\|\mathbf{w} - \widehat{\mathbf{w}}\|_{\ell_1} + 3\delta_{T,N}\|\mathbf{w} - \widehat{\mathbf{w}}\|_{\ell_1}^2. \tag{49}$$

For the global minimum variance portfolio:

$$\widehat{\mathbf{w}}^\top\boldsymbol{\Sigma}\widehat{\mathbf{w}} \ge \sigma_g^2(\mathbf{1}^\top\widehat{\mathbf{w}})^2.$$

Since $\widehat{\mathbf{w}} \in \Omega(\zeta)$, (24) gives

$$\widehat{\mathbf{w}}^\top \boldsymbol{\Sigma} \widehat{\mathbf{w}} \geq (1+2\zeta)^{-2} \sigma_g^2 \|\widehat{\mathbf{w}}\|_{\ell_1}^2. \tag{50}$$

With $\mathbf{0} \in \Omega$, the definition of $\widehat{\mathbf{w}}$ implies

$$\frac{\gamma}{2} \widehat{\mathbf{w}}^\top \widehat{\boldsymbol{\Sigma}} \widehat{\mathbf{w}} - \widehat{\mathbf{w}}^\top \widehat{\boldsymbol{\mu}} \leq 0,$$

so

$$\widehat{\mathbf{w}}^\top \widehat{\boldsymbol{\Sigma}} \widehat{\mathbf{w}} \leq \frac{2}{\gamma} \widehat{\mathbf{w}}^\top \widehat{\boldsymbol{\mu}} \leq \frac{2}{\gamma} \|\widehat{\boldsymbol{\mu}}\|_{\ell_\infty} \|\widehat{\mathbf{w}}\|_{\ell_1}.$$

From assumptions (i)-(ii) in Theorem 3,

$$\|\widehat{\boldsymbol{\mu}}\|_{\ell_\infty} \leq \|\boldsymbol{\mu}\|_{\ell_\infty} + \|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_{\ell_\infty} \leq \|\boldsymbol{\mu}\| + \|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_{\ell_\infty} \leq \theta_* \sigma_{\max} + \|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_{\ell_\infty},$$

and

$$\left| \widehat{\mathbf{w}}^\top (\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \widehat{\mathbf{w}} \right| \leq \delta_{T,N} \|\widehat{\mathbf{w}}\|_{\ell_1}^2,$$

we obtain

$$\widehat{\mathbf{w}}^\top \boldsymbol{\Sigma} \widehat{\mathbf{w}} \leq \frac{2}{\gamma} (\theta_* \sigma_{\max} + \delta_{T,N}) \|\widehat{\mathbf{w}}\|_{\ell_1} + \delta_{T,N} \|\widehat{\mathbf{w}}\|_{\ell_1}^2. \tag{51}$$

Let $\mathcal{E}$ denote the event where

$$\delta_{T,N} \leq \min\{(1+2\zeta)^{-2} \sigma_g^2 / 2, \theta_* \sigma_{\max}\}.$$

Combining (50) and (51), under $\mathcal{E}$ we have

$$\|\widehat{\mathbf{w}}\|_{\ell_1} \leq \frac{2}{\gamma} \frac{\theta_* \sigma_{\max} + \delta_{T,N}}{(1+2\zeta)^{-2} \sigma_g^2 - \delta_{T,N}}. \tag{52}$$

For $\tilde{\mathbf{w}} = \operatorname{argmax}_{\mathbf{w} \in \Omega(\zeta)} U(\mathbf{w})$, we can verify

$$(1+2\zeta)^{-2} \sigma_g^2 \|\tilde{\mathbf{w}}\|_{\ell_1}^2 \leq \tilde{\mathbf{w}}^\top \boldsymbol{\Sigma} \tilde{\mathbf{w}} \leq \frac{2}{\gamma} \boldsymbol{\mu}^\top \tilde{\mathbf{w}} \leq \frac{2}{\gamma} \|\boldsymbol{\mu}\|_{\ell_\infty} \|\tilde{\mathbf{w}}\|_{\ell_1},$$

where the first inequality follows from (24) and the second from $U(\tilde{\mathbf{w}}) \geq U(\mathbf{0}) = 0$. Thus,

$$\|\tilde{\mathbf{w}}\|_{\ell_1} \leq \frac{2(1+2\zeta)^2 \|\boldsymbol{\mu}\|_{\ell_\infty}}{\gamma \sigma_g^2}.$$

With (52), this gives

$$\|\widehat{\mathbf{w}} - \tilde{\mathbf{w}}\|_{\ell_1} \leq \|\widehat{\mathbf{w}}\|_{\ell_1} + \|\tilde{\mathbf{w}}\|_{\ell_1} \leq \frac{4}{\gamma} \frac{\theta_* \sigma_{\max} + \delta_{T,N}}{(1+2\zeta)^{-2} \sigma_g^2 - \delta_{T,N}}. \tag{53}$$

Under $\mathcal{E}$, we can further bound this as

$$\|\widehat{\mathbf{w}} - \tilde{\mathbf{w}}\|_{\ell_1} \leq \frac{16(1+2\zeta)^2\theta_*\sigma_{\max}}{\gamma\sigma_g^2}. \tag{54}$$

Taking $\mathbf{w} = \tilde{\mathbf{w}}$ in (49) and using (54), we have

$$\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w}_* - \widehat{\mathbf{w}})\|^2 \leq (1+\delta_{T,N})\|\boldsymbol{\Sigma}^{1/2}(\tilde{\mathbf{w}} - \mathbf{w}_*)\|^2 + \left(\frac{32(1+2\zeta)^2\theta_*\sigma_{\max}}{\gamma^2\sigma_g^2} + \frac{768(1+2\zeta)^4\theta_*^2\sigma_{\max}^2}{\gamma^2\sigma_g^4}\right)\cdot\delta_{T,N}.$$

For any $\mathbf{w} \in \mathbb{R}^N$, it can be verified that

$$U(\mathbf{w}_*) - U(\mathbf{w}) = \frac{\gamma}{2}\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w}_* - \mathbf{w})\|^2. \tag{55}$$

This leads to

$$U(\widehat{\mathbf{w}}) \geq \sup_{\mathbf{w}\in\Omega(\zeta)} U(\mathbf{w}) - \left[\frac{16(1+2\zeta)^2\theta_*\sigma_{\max}}{\gamma\sigma_g^2} + \frac{384(1+2\zeta)^4\theta_*^2\sigma_{\max}^2}{\gamma\sigma_g^4} + \frac{\theta_*^2}{2\gamma}\right]\cdot\delta_{T,N}, \tag{56}$$

where we used (55) with $\mathbf{w} = \tilde{\mathbf{w}}$ and $U(\mathbf{w}_*) - U(\tilde{\mathbf{w}}) \leq U(\mathbf{w}_*) - U(\mathbf{0}) = \theta_*^2/2\gamma$.

For the Sharpe ratio bound, note that for any $\mathbf{w} \in \mathbb{R}^N$,

$$(\mathbf{w} - \mathbf{w}_*)^\top\boldsymbol{\Sigma}(\mathbf{w} - \mathbf{w}_*) \geq \inf_x (x\mathbf{w} - \mathbf{w}_*)^\top\boldsymbol{\Sigma}(x\mathbf{w} - \mathbf{w}_*) = \mathbf{w}_*^\top\boldsymbol{\Sigma}\mathbf{w}_* - \frac{(\mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w}_*)^2}{\mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w}},$$

with equality iff $\mathbf{w}^\top\boldsymbol{\mu} = \gamma(\mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w})$. The right-hand side equals

$$\mathbf{w}_*^\top\boldsymbol{\Sigma}\mathbf{w}_* - \frac{(\mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w}_*)^2}{\mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w}} = \frac{1}{\gamma^2}(\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}) - \frac{1}{\gamma^2}\frac{(\mathbf{w}^\top\boldsymbol{\mu})^2}{\mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w}} = \frac{1}{\gamma^2}[\theta_*^2 - \theta^2(\mathbf{w})].$$

Thus,

$$\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w}_* - \widehat{\mathbf{w}})\|^2 \geq \frac{1}{\gamma^2}[\theta_*^2 - \theta^2(\widehat{\mathbf{w}})], \tag{57}$$

and

$$\|\boldsymbol{\Sigma}^{1/2}(\mathbf{w}_* - x\tilde{\mathbf{w}})\|^2 = \frac{1}{\gamma^2}[\theta_*^2 - \theta^2(\tilde{\mathbf{w}})], \tag{58}$$

where $x = \tilde{\mathbf{w}}^\top\boldsymbol{\mu}/(\gamma\tilde{\mathbf{w}}^\top\boldsymbol{\Sigma}\tilde{\mathbf{w}})$. Moreover, we have

$$\|x\tilde{\mathbf{w}}\|_{\ell_1} = \frac{\|\theta(\tilde{\mathbf{w}})\tilde{\mathbf{w}}\|_{\ell_1}}{\gamma(\tilde{\mathbf{w}}^\top\boldsymbol{\Sigma}\tilde{\mathbf{w}})^{1/2}} \leq \frac{\theta_*}{\gamma\sigma_g}\frac{\|\tilde{\mathbf{w}}\|_{\ell_1}}{|\mathbf{1}^\top\tilde{\mathbf{w}}|} \leq \frac{(1+2\zeta)\theta_*}{\gamma\sigma_g},$$

since $\theta_* \geq \theta(\tilde{\mathbf{w}})$ and $\sigma_g \leq \frac{(\tilde{\mathbf{w}}^\top\boldsymbol{\Sigma}\tilde{\mathbf{w}})^{1/2}}{|\mathbf{1}^\top\tilde{\mathbf{w}}|}$. With (52), under $\mathcal{E}$, we obtain

$$\|\widehat{\mathbf{w}} - x\tilde{\mathbf{w}}\|_{\ell_1} \leq \frac{2}{\gamma}\frac{\theta_*\sigma_{\max} + \delta_{T,N}}{(1+2\zeta)^{-2}\sigma_g^2 - \delta_{T,N}} + \frac{(1+2\zeta)\theta_*}{\gamma\sigma_g} \leq \frac{12}{\gamma}\frac{\theta_*\sigma_{\max}(1+2\zeta)^2}{\sigma_g^2}, \tag{59}$$

since $\sigma_{\max} \geq \sigma_g$.

Taking $\mathbf{w} = x\tilde{\mathbf{w}}$ in (49) and using (59), we have

$$\|\mathbf{\Sigma}^{1/2}(\mathbf{w}_* - \widehat{\mathbf{w}})\|^2 \leq (1+\delta_{T,N})\|\mathbf{\Sigma}^{1/2}(x\tilde{\mathbf{w}} - \mathbf{w}_*)\|^2 + \left(\frac{24(1+2\zeta)^2\theta_*\sigma_{\max}}{\gamma^2\sigma_g^2} + \frac{432(1+2\zeta)^4\theta_*^2\sigma_{\max}^2}{\gamma^2\sigma_g^4}\right)\cdot\delta_{T,N}.$$

Combining (57) and (58), we conclude

$$\theta^2(\widehat{\mathbf{w}}) \geq \sup_{\mathbf{w}\in\Omega(\zeta)} \theta^2(\mathbf{w}) - \left[\frac{24(1+2\zeta)^2\theta_*\sigma_{\max}}{\sigma_g^2} + \frac{432(1+2\zeta)^4\theta_*^2\sigma_{\max}^2}{\sigma_g^4} + \theta_*^2\right]\cdot\delta_{T,N}. \tag{60}$$

The desired result then follows from

$$\delta_{T,N} = O_p\left(\sqrt{\frac{\log N}{T}}\right),$$

as shown in Lemmas 1-3.

B.5 Proof of Theorem 4

By definition,

$$\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+,\top}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+} \leq \mathbf{w}_{\mathrm{GMV}}^{+,\top}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+} \qquad \text{and} \qquad \widehat{\mathbf{w}}_{\mathrm{GMV}}^{+,\top}\mathbf{\Sigma}\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+} \geq \mathbf{w}_{\mathrm{GMV}}^{+,\top}\mathbf{\Sigma}\mathbf{w}_{\mathrm{GMV}}^{+}.$$

Therefore,

$$\begin{aligned}
0 &\leq \widehat{\mathbf{w}}_{\mathrm{GMV}}^{+,\top}\mathbf{\Sigma}\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+} - \mathbf{w}_{\mathrm{GMV}}^{+,\top}\mathbf{\Sigma}\mathbf{w}_{\mathrm{GMV}}^{+} \\
&\leq \widehat{\mathbf{w}}_{\mathrm{GMV}}^{+,\top}(\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}})\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+} - \mathbf{w}_{\mathrm{GMV}}^{+,\top}(\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}})\mathbf{w}_{\mathrm{GMV}}^{+} \\
&= (\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+} - \mathbf{w}_{\mathrm{GMV}}^{+})^\top(\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}})(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+} + \mathbf{w}_{\mathrm{GMV}}^{+}) \\
&= 2(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+} - \mathbf{w}_{\mathrm{GMV}}^{+})^\top(\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}})\mathbf{w}_{\mathrm{GMV}}^{+} \\
&\quad +(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+} - \mathbf{w}_{\mathrm{GMV}}^{+})^\top(\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}})(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+} - \mathbf{w}_{\mathrm{GMV}}^{+}).
\end{aligned}$$

Following a similar argument as before, we can show that

$$\left|(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+} - \mathbf{w}_{\mathrm{GMV}}^{+})^\top(\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}})\mathbf{w}_{\mathrm{GMV}}^{+}\right| = O_p\left(\sqrt{\frac{\log N}{T}}\right)$$

in light of Lemma 3; and

$$\left|(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+})^{\top}(\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}})(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+})\right|=O_{p}\left(\sqrt{\frac{\log N}{T}}\right),$$

in light of Lemma 2. They immediately imply that

$$\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+,\top}\boldsymbol{\Sigma}\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+,\top}\boldsymbol{\Sigma}\mathbf{w}_{\mathrm{GMV}}^{+}=O_{p}\left(\sqrt{\frac{\log N}{T}}\right). \tag{61}$$

Now consider the Sharpe ratio. Recall that

$$\boldsymbol{\mu}^{\top}(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+})=\theta(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+})\cdot\left[(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+})^{\top}\boldsymbol{\Sigma}(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+})\right]^{1/2}.$$

Thus,

$$\left[\boldsymbol{\mu}^{\top}(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+})\right]^{2}\leq\theta_{*}^{2}\cdot(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+})^{\top}\boldsymbol{\Sigma}(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+}).$$

As before, we can deduce that

$$(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+})^{\top}\boldsymbol{\Sigma}(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+})=O_{p}\left(\sqrt{\frac{\log N}{T}}\right),$$

which implies

$$\left[\boldsymbol{\mu}^{\top}(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+}-\mathbf{w}_{\mathrm{GMV}}^{+})\right]^{2}=O_{p}\left(\sqrt{\frac{\log N}{T}}\right). \tag{62}$$

Combining (61) and (62), we have

$$\theta^{2}(\mathbf{w}_{\mathrm{GMV}}^{+})-\theta^{2}(\widehat{\mathbf{w}}_{\mathrm{GMV}}^{+})=O_{p}\left(\sqrt{\frac{\log N}{T}}\right).$$

This accomplished the proof.

B.6 Auxiliary Results

Lemma 1. *Suppose that* $\mathbf{r}_1,\ldots,\mathbf{r}_T$ *are independent* $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ *random variables. If* $\sigma_{\max}\leq C$ *for some constant* C*, and* $\log N\ll T$*, then*

$$\|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}\|_{\ell_\infty}=O_{p}\left(\sqrt{\frac{\log N}{T}}\right).$$

Proof of Lemma 1. Write $\mathbf{r}_t = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}_t$, where $\mathbf{Z}_t \sim N(\mathbf{0}, \mathbf{I}_N)$. Then

$$\widehat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\bar{\mathbf{Z}}, \qquad \text{and} \qquad \widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{1/2}\mathbf{S}\boldsymbol{\Sigma}^{1/2},$$

where

$$\bar{\mathbf{Z}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t, \qquad \text{and} \qquad \mathbf{S} = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{Z}_t - \bar{\mathbf{Z}})(\mathbf{Z}_t - \bar{\mathbf{Z}})^\top$$

are the sample mean and covariance of $\mathbf{Z}_t$s. Thus

$$\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}} = \boldsymbol{\Sigma}^{1/2}(\mathbf{S} - \mathbf{I})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} - \boldsymbol{\Sigma}^{1/2}\bar{\mathbf{Z}}.$$

By triangular inequality,

$$\|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}\|_{\ell_\infty} \le \|\boldsymbol{\Sigma}^{1/2}(\mathbf{S} - \mathbf{I})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}\|_{\ell_\infty} + \|\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{Z}}\|_{\ell_\infty}.$$

Observe that

$$(\boldsymbol{\Sigma}^{1/2}(\mathbf{S} - \mathbf{I})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu})_i =_d \frac{\sigma_i\theta_*}{T-1}\sum_{t=1}^{T-1}[N_{1t}N_{2t} - \mathbb{E}(N_{1t}N_{2t})],$$

where $\sigma_i^2 = \sigma_{ii} = \mathtt{var}(r_{ti})$, i.e., the ith diagonal entry of $\boldsymbol{\Sigma}$, and (N_{1t}, N_{2t})s are independent copies of

$$\begin{bmatrix} N_1 \\ N_2 \end{bmatrix} \sim N\left(0, \begin{bmatrix} 1 & \mu_i/(\sigma_i\theta_*) \\ \mu_i/(\sigma_i\theta_*) & 1 \end{bmatrix}\right).$$

Note further that

$$N_1N_2 = \frac{1}{4}(N_1 + N_2)^2 - \frac{1}{4}(N_1 - N_2)^2.$$

It then follows that

$$\sum_{t=1}^{T-1}[N_{1t}N_{2t} - \mathbb{E}(N_{1t}N_{2t})] = \frac{1}{4}\sum_{t=1}^{T-1}\left[(N_{1t} + N_{2t})^2 - \left(2 + \frac{2\mu_i}{\sigma_i\theta_*}\right)\right] - \frac{1}{4}\sum_{t=1}^{T-1}\left[(N_{1t} - N_{2t})^2 - \left(2 - \frac{2\mu_i}{\sigma_i\theta_*}\right)\right].$$

Hence, from the χ^2 tail bounds (see, e.g., Lemma 1 of Laurent and Massart, 2000), we have

$$\frac{1}{2(T-1)}\sum_{t=1}^{T-1}\left[\left(2 + \frac{2\mu_i}{\sigma_i\theta_*}\right)^{-1}(N_{1t} + N_{2t})^2 - 1\right] \le \sqrt{\frac{x}{T-1}} + \frac{x}{T-1}$$

and

$$\frac{1}{2(T-1)}\sum_{t=1}^{T-1}\left[\left(2 - \frac{2\mu_i}{\sigma_i\theta_*}\right)^{-1}(N_{1t} - N_{2t})^2 - 1\right] \le \sqrt{\frac{x}{T-1}} + \frac{x}{T-1}$$

with probability at least $1 - 4e^{-x}$. Under this event, we obtain

$$|(\boldsymbol{\Sigma}^{1/2}(\mathbf{S} - I)\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu})_i| \le 2\sigma_i\theta_* \left(\sqrt{\frac{x}{T-1}} + \frac{x}{T-1}\right).$$

An application of union bounds over all i then yields

$$\|\boldsymbol{\Sigma}^{1/2}(\mathbf{S} - \mathbf{I})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}\|_{\ell_\infty} \le 2\sigma_{\max}\theta_* \left(\sqrt{\frac{x + \log N}{T-1}} + \frac{x + \log N}{T-1}\right)$$

with probability at least $1 - 4e^{-x}$, where $\sigma_{\max} = \max_i \sigma_i$. Together with the fact that

$$\mathbb{P}\left\{\|\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{Z}}\|_{\ell_\infty} \ge \sigma_{\max}\sqrt{\frac{2(\log N + x)}{T}}\right\} \le 2e^{-x},$$

we have

$$\|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}\|_{\ell_\infty} \le C\sigma_{\max}(1 + \theta_*)\left(\sqrt{\frac{x + \log N}{T}} + \frac{x + \log N}{T}\right)$$

with probability at least $1 - 6e^{-x}$. □

Lemma 2. *Suppose that $\mathbf{r}_1, \ldots, \mathbf{r}_T$ are independent $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random variables. If $\sigma_{\max} \le C$ for some constant C, and $\log N \ll T$, then*

$$\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\max} := \max_{1 \le i,j \le N} |\widehat{\Sigma}_{ij} - \Sigma_{ij}| = O_p\left(\sqrt{\frac{\log N}{T}}\right), \qquad \text{as } T \to \infty.$$

Proof of Lemma 2. Observe that

$$\widehat{\Sigma}_{ij} - \Sigma_{ij} =_d \frac{1}{T-1}\sum_{t=1}^{T-1}[N_{1t}N_{2t} - \mathbb{E}(N_{1t}N_{2t})],$$

where (N_{1t}, N_{2t})s are independent copies of

$$\begin{bmatrix} N_1 \\ N_2 \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ij} & \Sigma_{jj} \end{bmatrix}\right).$$

Hence, an application of the χ^2 tail bounds as before yields

$$\left|\widehat{\Sigma}_{ij} - \Sigma_{ij}\right| \le (\Sigma_{ii} + \Sigma_{jj})\left(\sqrt{\frac{x}{T-1}} + \frac{x}{T-1}\right) \le 2\sigma_{\max}^2\left(\sqrt{\frac{x}{T-1}} + \frac{x}{T-1}\right),$$

with probability at least $1 - 4e^{-x}$. Applying union bounds over all pairs (i, j), we get

$$\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\max} \le 2\sigma_{\max}^2\left(\sqrt{\frac{x + 2\log N}{T-1}} + \frac{x + 2\log N}{T-1}\right)$$

with probability at least $1 - 4e^{-x}$. □

Lemma 3. *Suppose that* $\mathbf{r}_1, \ldots, \mathbf{r}_T$ *are independent* $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *random variables. If* $\sigma_{\max} \leq C$ *for some constant* C*, and* $\log N \ll T$*, then for any fixed* $\mathbf{a}$ *such that* $\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} < \infty$*,*

$$\|(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\mathbf{a}\|_{\ell_\infty} = O_p\left(\sqrt{\frac{\log N}{T}}\right), \qquad \text{as } T \to \infty.$$

Proof of Lemma 3. Observe that

$$((\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\mathbf{a})_i =_d \frac{1}{T-1} \sum_{t=1}^{T-1} [N_{1t} N_{2t} - \mathbb{E}(N_{1t} N_{2t})],$$

where (N_{1t}, N_{2t})s are independent copies of

$$\begin{bmatrix} N_1 \\ N_2 \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \Sigma_{ii} & (\boldsymbol{\Sigma}\mathbf{a})_i \\ (\boldsymbol{\Sigma}\mathbf{a})_i & \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} \end{bmatrix}\right).$$

Following the same argument as before, we get

$$\|(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\mathbf{a}\|_{\ell_\infty} \leq (\Sigma_{ii} + \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}) \left(\sqrt{\frac{x + \log N}{T-1}} + \frac{x + \log N}{T-1}\right)$$

with probability at least $1 - 4e^{-x}$. □