# Tensor Portfolios

Tianyan Tu* Tae-Hwy Lee†

*Department of Economics, University of California, Riverside, CA, USA.

†Department of Economics, University of California, Riverside, CA, USA.

**Address correspondence to** Tae-Hwy Lee, Department of Economics, University of California, Riverside, 900 University Avenue, Riverside, CA 92521-0427, USA, or e-mail: taelee@ucr.edu, or phone: +1-951-827-1509.

**Abstract**

Motivated by the multi-dimensional nature of financial data, we propose the *tensor portfolios*, a framework exploiting the intrinsic multi-way structure of stock returns to reduce the number of free parameters required for portfolio construction. We develop three distinct methods tailored to specific structural assumptions. We systematically compare tensor and vector portfolios through Monte Carlo simulations and empirical studies. The simulation results show tensor portfolios yield significantly higher out-of-sample Sharpe ratios whenever the data exhibits a tensor structure. Empirical analysis further corroborates the effectiveness of tensor portfolios; their general outperformance over vector portfolios in read-world markets highlights the practical significance of exploiting multi-way information.

*Keywords:* Tensor Portfolio Optimization; Kronecker Separability; High-Dimensionality; Tensor Factor Model; Tensor Graphical LASSO.

*JEL Classifications:* C13, C40, C55, C58, G11, G17

## Introduction

Conventional portfolio construction typically relies on a vector-based paradigm, stacking asset returns into a single vector to determine weights through criteria such as the Sharpe ratio (Markowitz (1952), Castellano and Cerqueti (2014), and Fan et. al. (2024)). However, as the asset pool expands, these methods face a "curse of dimensionality" where the precision matrix becomes ill-conditioned, inducing significant estimation instability. To mitigate this, shrinkage and regularization approaches have been widely adopted, including LASSO (Giannone et. al. (2007), DeMiguel et. al. (2009), and Fan, J. and Yu (2012)), Graphical LASSO (GLASSO) (Goto and Xu (2015), Brownlees, Nualart, and Sun (2018), Torri, Giacometti, and Paterlini (2019), and Yuan et. al. (2020)), and factor models (Jagannathan and Ma (2003), Ledoit and Wolf (2003), Fan, Liao, and Mincheva (2011), Fan, Liao, and Mincheva (2013), Ding, Li, and Zheng (2021), Callot et. al. (2021), and Lee and Seregina (2023)). These studies confirm that such shrinkage methods effectively reduce dimensionality in high-dimensional portfolios, thereby lowering realized risk and improving Sharpe ratios.

Beyond these methodologies, we also leverage another approach of dimensionality reduction that arises directly from the data's intrinsic structure. In financial markets, forcing asset return data into a flat, one-dimensional vector often discards valuable structural information. Naturally,

an individual stock is naturally characterized along multiple distinct modes, including its sector affiliation, market capitalization, country of origin, or exposures to specific style factors. By organizing stocks along these interpretable dimensions, the conventional return vector can be reshaped into a matrix or a higher-order tensor. The key benefit of this approach is that it transforms the estimation of a massive, high-dimensional precision matrix into the estimation of multiple smaller, low-dimensional precision matrices, thereby effectively achieving dimensionality reduction.

Recognizing this inherent multi-dimensionality, we propose an alternative approach for portfolio weight estimation in high-dimensional settings: the *Tensor Portfolio*, which fully exploits the rich underlying structure of the data to achieve dimensionality reduction. First, we extend the vector-based Global Minimum Variance (GMV), Markowitz Weight-Constrained (MWC), and Markowitz Risk-Constrained (MRC) optimization frameworks to incorporate tensor structures. Second, we employ the Tensor Graphical LASSO (TLASSO) proposed by Lyu et. al. (2020) to address the ill-conditioned precision matrix problem in tensor-structured data. Finally, building upon Lee and Seregina (2023), to account for the well-established latent factor structure in stock returns, we generalize their Vector Factor Graphical LASSO (VectorFGL) framework to tensor-valued returns, yielding the Tensor Factor Graphical LASSO (TensorFGL) for portfolio construction.

Our method is designed for stock return data that are naturally organized in a multi-dimensional tensor form. Because a stock's expected performance is jointly informed by its sector affiliation and firm-specific characteristics, stocks exhibiting favorable profiles across both dimensions should receive greater allocations. For instance, a large-size stock in a high-return sector - such as Nvidia - merits a greater weight due to its dual structural advantages. Motivated by this economic intuition, we organize stock returns into a tensor indexed by time $t$, characteristic $i$, sector $j$. By assigning weights directly along these modes, the final allocation for stock $(i, j)$ is the product of the respective weights of characteristic $(\omega_i)$ and sector $(\omega_j)$: $\omega_i\omega_j$.

To our knowledge, the application of tensor-structured models to portfolio construction remains scarce. Most studies on tensor applications in economics and others social sciences focus on factor analysis in non-financial domains, such as international trade, transportation, environment, and macroeconomic fluctuations (Chen, Yang, and Zhang (2022), Chen, Han, and Yu (2024), Chen et. al. (2024), Han et. al. (2024), Han et. al. (2024), Chen, Tsay, and Chen (2020), Chen, Xiao, and Yang (2021), Yu et. al. (2022), Chen and Fan (2023)). Within finance, existing applications largely

concentrate on high-dimensional factor modeling. Specifically, studies employ Fama–French $10 \times 10$ portfolios to demonstrate how tensor models reduce free parameters and mitigate estimation errors (Wang, Liu, and Chen (2019), Chen, Tsay, and Chen (2020), Yu et. al. (2022)).

We contribute to the literature in these fields by extending the portfolio construction to the assets with tensor structure. The main advantage of constructing portfolios based on tensor structure is that it preserves the intrinsic multi-dimensional structure of the data. As argued in Chen et. al. (2024) "*Traditional methods reduce matrix observations into vectors...losing essential matrix structure and leading to less effective analyses*". Relative to vector-based portfolio construction, our approach is built on the same underlying return observations, but it leverages the additional information contained in how returns are structured across sectors and firm characteristics. By modeling the contribution of each sector and each characteristic explicitly, we no longer need to determine portfolio weights stock by stock. Instead, we assign weights along each tensor mode, i.e., sector and characteristic, which determines the weights of all constituent stocks simultaneously. This leads to a substantial reduction in the dimension of optimization problem and, in turn, improves the tractability and stability of portfolio construction in high-dimensional settings.

To demonstrate the advantages of tensor-based portfolio construction, we use Monte Carlo simulations to compare its OOS Sharpe ratios against conventional vector models under various Data Generating Processes (DGPs). The results show that whenever the data exhibit a tensor structure, tensor-based portfolios consistently achieve significantly higher OOS Sharpe ratios than their vectorized counterparts. We further complement these findings with an empirical study, we collect daily stock returns from 11 U.S. sectors over January 1, 2010 to December 31, 2019. By organizing the 30 largest stocks within 11 sectors into a $T \times 30 \times 11$ return tensor, we confirm that tensor-based portfolio exhibits a clear advantage by delivering higher OOS Sharpe ratios.

This paper is organized as follows. Section 1 introduces the frameworks for conventional vector-based portfolio construction. Section 2 introduces the frameworks for tensor portfolio construction under different data structure. Section 3 establishes theoretical results supporting the convergence of our alternating algorithm. Section 4 reports Monte Carlo simulations comparing tensor- and vector-based portfolios in terms of OOS Sharpe ratios. Section 5 presents an empirical application and evaluates the performance of tensor-based portfolios relative to conventional vector-based portfolios. Section 6 concludes.

# 1 Vector Portfolio Allocation

In the conventional vector-based portfolio paradigm, the *excess* returns (hereinafter returns) of $p$ assets at time $t$ are stacked into a single column vector, denoted as $\mathbf{r}_t \in \mathbb{R}^p$. Let $\mathbf{w} \in \mathbb{R}^p$ denote the corresponding portfolio weight vector. Over a sample of $T$ periods, the mean return of the portfolio is given by $\frac{1}{T}\sum_{t=1}^{T} \mathbf{w}'\mathbf{r}_t = \mathbf{w}'\bar{\mathbf{r}}$, where $\bar{\mathbf{r}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{r}_t$ is a vector of the sample mean return of the $p$ stocks. Accordingly, the portfolio risk, characterized by its sample standard deviation, is formulated as the square-root of $\frac{1}{T}\sum_{t=1}^{T} (\mathbf{w}'\mathbf{r}_t - \mathbf{w}'\bar{\mathbf{r}})^2 = \mathbf{w}'\mathbf{\Sigma}\mathbf{w}$, where $\mathbf{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})'$ is the sample covariance matrix of the asset returns.

We consider three common strategies for portfolio construction. The first is the Global Minimum Variance (GMV) portfolio, which aims to minimize the portfolio risk subject to the full-investment constraint $\mathbf{w}'\boldsymbol{\iota}_p = 1$, where $\boldsymbol{\iota}_p \in \mathbb{R}^p$ is a vector of ones. The GMV optimization problem is formulated as follows:

$$\min_{\mathbf{w}} \quad \mathbf{w}'\mathbf{\Sigma}\mathbf{w}, \quad \text{s.t.} \quad \mathbf{w}'\boldsymbol{\iota}_p = 1 \tag{1}$$

Denote the precision matrix as $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$, the portfolio weight obtained from the GMV problem is given by:

$$\mathbf{w} = \frac{\mathbf{\Theta}\boldsymbol{\iota}_p}{\boldsymbol{\iota}_p\mathbf{\Theta}\boldsymbol{\iota}_p'}. \tag{2}$$

The second strategy is the Markowitz weight-constrained (MWC) optimization. Similar to the GMV formulation, it aims to minimize portfolio risk subject to a full-investment constraint, with an additional requirement that the expected return must be greater than or equal to a predefined target, $\mu_{target}$:

$$\min_{\mathbf{w}} \quad \mathbf{w}'\mathbf{\Sigma}\mathbf{w}, \quad \text{s.t.} \quad \mathbf{w}'\boldsymbol{\iota}_p = 1, \quad \mathbf{w}'\bar{\mathbf{r}} \geq \mu_{target}. \tag{3}$$

The solution of MWC with binding return constraint is two-fund separation theorem (Tobin (1958)), which is a weighted average between risk minimization and return maximization:

$$\mathbf{w} = (1-a)\frac{\mathbf{\Theta}\boldsymbol{\iota}_p}{\boldsymbol{\iota}_p\mathbf{\Theta}\boldsymbol{\iota}_p'} + a\frac{\mathbf{\Theta}\bar{\mathbf{r}}}{\boldsymbol{\iota}_p\mathbf{\Theta}\bar{\mathbf{r}}}, \tag{4}$$

where $a$ denotes the weight allocated to the return-maximizing component, defined as:

$$a = \frac{\mu_{target}\left(\bar{\mathbf{r}}'\boldsymbol{\Theta}\boldsymbol{\iota}_p\right)\left(\boldsymbol{\iota}_p'\boldsymbol{\Theta}\boldsymbol{\iota}_p\right) - \left(\bar{\mathbf{r}}'\boldsymbol{\Theta}\boldsymbol{\iota}_p\right)^2}{\left(\bar{\mathbf{r}}'\boldsymbol{\Theta}\bar{\mathbf{r}}\right)\left(\boldsymbol{\iota}_p'\boldsymbol{\Theta}\boldsymbol{\iota}_p\right) - \left(\bar{\mathbf{r}}'\boldsymbol{\Theta}\boldsymbol{\iota}_p\right)^2}. \tag{5}$$

The third strategy is the Markowitz risk-constrained (MRC) optimization, in which we aim to maximize the Sharpe ratio, defined as $\mathbf{w}'\bar{\mathbf{r}}/\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}$, subject to return constraint as in MWC and the risk constraint that the risk of portfolio can never exceed a predefined target $\sigma_{target}$.

$$\max_{\mathbf{w}} \quad \frac{\mathbf{w}'\bar{\mathbf{r}}}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}}, \quad \text{s.t.} \quad \mathbf{w}'\bar{\mathbf{r}} \geq \mu_{target}, \quad \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \leq \sigma_{target}^2. \tag{6}$$

The non-convex objective function makes it mathematically challenging to solve for the portfolio weights directly, so the MRC problem is typically transformed into one of two equivalent convex optimization problems. One aims to maximize the expected return under binding risk constraint:

$$\max_{\mathbf{w}} \quad \mathbf{w}'\bar{\mathbf{r}}, \quad \text{s.t.} \quad \sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}} = \sigma_{target}, \tag{7}$$

which implies the closed-form solutions for the portfolio weights:

$$\mathbf{w} = \frac{\sigma_{target}}{\sqrt{\bar{\mathbf{r}}'\boldsymbol{\Theta}\bar{\mathbf{r}}}}\boldsymbol{\Theta}\bar{\mathbf{r}}. \tag{8}$$

The other aims to minimize the risk under binding return constraint:

$$\min_{\mathbf{w}} \quad \sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}, \quad \text{s.t.} \quad \mathbf{w}'\bar{\mathbf{r}} = \mu_{target}, \tag{9}$$

and the portfolio weights can be derived as

$$\mathbf{w} = \frac{\mu_{target}}{\bar{\mathbf{r}}'\boldsymbol{\Theta}\bar{\mathbf{r}}}\boldsymbol{\Theta}\bar{\mathbf{r}}. \tag{10}$$

When $\mu_{target}/\sigma_{target} = \sqrt{\bar{\mathbf{r}}'\boldsymbol{\Theta}\bar{\mathbf{r}}}$, (8) and (10) will yield identical result. Because the Sharpe ratio is scale invariance, even when this condition is not satisfied, they can always achieve the same Sharpe ratio.

Furthermore, the closed-form solutions in (2), (4), (8), and (10) highlight a fundamental chal-

lenge: estimating portfolio weights invariably relies on the precision matrix $\mathbf{\Theta}$. In the conventional vector-based framework, $\mathbf{\Theta}$ is a $p \times p$ matrix. Given the high-dimensional nature of real financial markets, estimating such a massive matrix frequently results in a severely ill-conditioned precision matrix.

## 2 Tensor Portfolio Allocation

We now consider the scenario where the assets discussed in Section 1 naturally exhibit a tensor structure. Specifically, the $p$-dimensional vector of return vector $\mathbf{r}_t$ can be reshaped into an $m \times n$ return matrix, $\mathbf{R}_t$, structured along distinct cross-sectional dimensions. $\mathbf{R}_t \in \mathbb{R}^{m \times n}$, for each period $t \in \{1, \ldots, T\}$, where $p = m \times n$. Accordingly, we define the weighting vectors as $\boldsymbol{\omega}_1 \in \mathbb{R}^m$ and $\boldsymbol{\omega}_2 \in \mathbb{R}^n$. These row vectors are designed to capture the weights assigned to the rows (e.g., firm size) and columns (e.g., sectors) of $\mathbf{R}_t$. Consequently, our optimization problem shifts to the estimation of $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$. Remarkably, without employing any other shrinkage methods like GLASSO or factor models, simply exploiting the data's inherent bilinear structure drastically reduces the number of free parameters to be estimated from $m \times n$ to $m+n$, which highlights the substantial dimensionality reduction offered by the tensor portfolio formulation.

In tensor portfolio, the average portfolio return across $T$ periods becomes $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\omega}_1 \mathbf{R}_t \boldsymbol{\omega}_2' = \boldsymbol{\omega}_1 \left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{R}_t\right) \boldsymbol{\omega}_2' = \boldsymbol{\omega}_1 \bar{\mathbf{R}} \boldsymbol{\omega}_2'$. Let $\mathbf{E}_t = \mathbf{R}_t - \bar{\mathbf{R}}$, then the risk of a portfolio is measured by $\nu$ such that $\nu = \sqrt{\mathbb{E}\left(\boldsymbol{\omega}_1 \mathbf{E}_t \boldsymbol{\omega}_2'\right)^2} = \sqrt{\mathbb{E}\left[\left(\boldsymbol{\omega}_1 \mathbf{E}_t \boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_1 \mathbf{E}_t \boldsymbol{\omega}_2'\right)'\right]} = \sqrt{\mathbb{E}\left[\left(\boldsymbol{\omega}_2 \mathbf{E}_t' \boldsymbol{\omega}_1'\right)\left(\boldsymbol{\omega}_1 \mathbf{E}_t \boldsymbol{\omega}_2'\right)\right]}$. The last equality holds because $\boldsymbol{\omega}_1 \mathbf{E}_t \boldsymbol{\omega}_2'$ is scalar, the following three expressions are equivalent: $\mathbb{E}\left(\boldsymbol{\omega}_1 \mathbf{E}_t \boldsymbol{\omega}_2'\right)^2$, $\mathbb{E}\left(\boldsymbol{\omega}_1 \mathbf{E}_t \boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_2 \mathbf{E}_t' \boldsymbol{\omega}_1'\right)$, and $\mathbb{E}\left(\boldsymbol{\omega}_2 \mathbf{E}_t' \boldsymbol{\omega}_1'\right)\left(\boldsymbol{\omega}_1 \mathbf{E}_t \boldsymbol{\omega}_2'\right)$. Following the definition, the Sharpe ratio is the ratio between mean return and risk, which is shown as $\boldsymbol{\omega}_1 \bar{\mathbf{R}} \boldsymbol{\omega}_2' / \nu$.

We define $\mathbf{\Sigma}_1 = \mathbb{E}\left(\mathbf{E}_t' \boldsymbol{\omega}_1' \boldsymbol{\omega}_1 \mathbf{E}_t\right)$ and $\mathbf{\Sigma}_2 = \mathbb{E}\left(\mathbf{E}_t \boldsymbol{\omega}_2' \boldsymbol{\omega}_2 \mathbf{E}_t'\right)$ as the Row-Weighted Covariance Matrices and Column-Weighted Covariance Matrices. Correspondingly, the Row- and Column-Weighted Sample Covariance Matrices are (11):

$$\hat{\mathbf{\Sigma}}_1 = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{E}_t' \boldsymbol{\omega}_1' \boldsymbol{\omega}_1 \mathbf{E}_t\right), \quad \hat{\mathbf{\Sigma}}_2 = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{E}_t \boldsymbol{\omega}_2' \boldsymbol{\omega}_2 \mathbf{E}_t'\right). \tag{11}$$

Finally, for notational convenience, we omit the hat notation for estimators throughout the remainder of the paper whenever no confusion may arise.

### 2.1 Kronecker-Separable Covariance Structure

Recent studies in tensor-structured modeling (e.g., Chen, Xiao, and Yang (2021), Han et. al. (2024), Han et. al. (2024), and Cai et. al. (2025)) rely on a *Kronecker-separable* covariance similar to Assumption 1, where the overall covariance can be decomposed into row- and column-wise components. Under this structure, the row-mode (characteristics) and column-mode (sectors) dependencies jointly determine the covariance of different stocks in a multiplicative form.

**Assumption 1.** *Covariance matrix of the vectorized return,* $\mathbf{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} vec(\mathbf{E}_t)vec(\mathbf{E}_t)'$*, is separable and in a form of Kronecker Product:* $\mathbf{\Sigma} = \mathbf{\Sigma}_c \otimes \mathbf{\Sigma}_r$*, where* $\mathbf{\Sigma}_r$ *is row-wise covariance and* $\mathbf{\Sigma}_c$ *is column-wise covariance:*

$$\mathbf{\Sigma}_r = \frac{1}{n}\mathbb{E}\big(\mathbf{E}_t\mathbf{\Sigma}_c^{-1}\mathbf{E}_t'\big) \in \mathbb{R}^{m\times m}, \quad \mathbf{\Sigma}_c = \frac{1}{m}\mathbb{E}\big(\mathbf{E}_t'\mathbf{\Sigma}_r^{-1}\mathbf{E}_t\big) \in \mathbb{R}^{n\times n}. \tag{12}$$

Assumption 1 can simplify the iterative structure. Without Assumption 1, our portfolios will build on $\mathbf{\Sigma}_1^{-1}$ and $\mathbf{\Sigma}_2^{-1}$, where the row- and column-weighted sample covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, as shown in (11), are covariances induced by projecting de-meaned stock returns onto the row- or column-weighting vectors. As a result, the estimation of $\mathbf{\Sigma}_1^{-1}$ depends on $\boldsymbol{\omega}_2$, and vice versa. In other words, at each iteration we must update both the portfolio weights and the corresponding covariance estimates, which is computationally expensive and more prone to unstable estimation. In contrast, under Assumption 1, we can use an iterative procedure to estimate the row- and column-wise covariance matrices from the sample version of (12):

$$\mathbf{\Sigma}_r = \frac{1}{nT}\sum_{t=1}^{T}\big(\mathbf{E}_t\mathbf{\Sigma}_c^{-1}\mathbf{E}_t'\big), \quad \mathbf{\Sigma}_c = \frac{1}{mT}\sum_{t=1}^{T}\big(\mathbf{E}_t'\mathbf{\Sigma}_r^{-1}\mathbf{E}_t\big), \tag{13}$$

where the initial value of each mode-specific covariance/precision is Identity matrix. This procedure is referred to the Flip-Flop approach (Dutilleul (1999) and Chen, Xiao, and Yang (2021)). Therefore, the estimation of $\mathbf{\Sigma}_r^{-1}$ and $\mathbf{\Sigma}_c^{-1}$ only involves their own alternating updates, and thus free of the impacts of updated $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$. This not only reduces the computational burden but also mitigates the uncertainty and instability in the iterative procedure.

**Remark 1** Under separability assumption, we may represent the de-meaned return matrix $\mathbf{E}_t$ as

$\boldsymbol{\Sigma}_r^{1/2}\mathbf{Z}_t\boldsymbol{\Sigma}_c^{1/2}$ with $\boldsymbol{\Sigma}_r$ and $\boldsymbol{\Sigma}_c$ shown in (12), where $\mathbf{Z}_t \in \mathbb{R}^{m\times n}$ has the second moments satisfies $\mathbb{E}(\mathbf{Z}_t\mathbf{Z}_t') = n\mathbf{I}_m$ and $\mathbb{E}(\mathbf{Z}_t'\mathbf{Z}_t) = m\mathbf{I}_n$. We do not need additional assumption on $\mathbf{Z}_t$, it can be Gaussian, heavy-tailed, etc. □

**Remark 2** The Sharpe ratio is not sensitive to the normalization in (12). Including or omitting the factors $m$ and $n$ does not affect our results on Sharpe ratio, because the normalization only introduces a common positive scale factor. This scale factor appears in both the numerator and the denominator in the normalization/constraint-enforcement steps for GMV, MWC, and MRC, and therefore cancels out. □

## 2.2 Optimization under Tensor Structures

Consider the vectorized portfolio risk under Kronecker-Separable weighting and separable covariance from Assumption 1, the risk is now derived as $(\boldsymbol{\omega}_2 \otimes \boldsymbol{\omega}_1)\,\boldsymbol{\Sigma}\,(\boldsymbol{\omega}_2 \otimes \boldsymbol{\omega}_1)' = (\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2')\,(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1')$. Consequently, the updating/solution to the GMV, MWC, and MRC optimization problems can be derived as (15), (17), (22), and (23) where the derivations of updating/solution are available in Appendix A.1.

### 2.2.1 Global Minimum Variance (GMV) Optimization

The solutions of GMV is to minimize the risk under the full investment constraint such that all the weights sum to one:

$$\min_{\boldsymbol{\omega}_1,\boldsymbol{\omega}_2} \quad \left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1'\right), \quad \text{s.t.} \quad \boldsymbol{\omega}_1\boldsymbol{\iota}_m = 1, \quad \boldsymbol{\omega}_2\boldsymbol{\iota}_n = 1, \tag{14}$$

where $\boldsymbol{\iota}_m$ and $\boldsymbol{\iota}_n$ are vectors of ones. The solution can be shown in (15):

$$\boldsymbol{\omega}_1 = \frac{\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r}{\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m}, \quad \boldsymbol{\omega}_2 = \frac{\boldsymbol{\iota}_n'\boldsymbol{\Theta}_c}{\boldsymbol{\iota}_n'\boldsymbol{\Theta}_c\boldsymbol{\iota}_n}. \tag{15}$$

**Remark 3** We do not need iterative updating to compute the portfolio weights, once we have the estimated $\boldsymbol{\Theta}_r$ and $\boldsymbol{\Theta}_c$ via Flip-Flop, we can plug them in (15), and thus get the closed solutions of GMV rather than updating. □

### 2.2.2 Markowitz Weight-Constrained (MWC) Optimization

For MWC, our target is to minimize the risk under the constraints that all the weights sum to one and the expected return is above the target $\mu_{target}$, as shown in (16):

$$\min_{\boldsymbol{\omega}_1,\boldsymbol{\omega}_2} \quad \left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1'\right), \quad \text{s.t.} \quad \boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2' \geq \mu_{target}, \quad \boldsymbol{\omega}_1\boldsymbol{\iota}_m = 1, \quad \boldsymbol{\omega}_2\boldsymbol{\iota}_n = 1. \tag{16}$$

Assume the return constraint is binding: $\boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2' = \mu_{target}$. Starting from initial values of $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$, we iteratively update them via an alternating-minimization scheme: at each step, we optimize one vector using the vector-based MWC subproblem while holding the other fixed, and repeat until convergence. This procedure yields the iteration shown in (32), which is similar to the form as the Two-Fund Separation Theorem of Tobin (1958), but according to a different mode of the tensor. As shown in (32), the updating of $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are following the weighted average of risk-minimization and return-maximization, where the weights are given by $a_r$ and $a_c$, respectively, defined in (17):

$$\boldsymbol{\omega}_1 \leftarrow (1-a_r)\frac{\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r}{\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m} + a_r\frac{(\boldsymbol{\omega}_2\bar{\mathbf{R}}')\,\boldsymbol{\Theta}_r}{(\boldsymbol{\omega}_2\bar{\mathbf{R}}')\,\boldsymbol{\Theta}_r\boldsymbol{\iota}_m}, \quad \boldsymbol{\omega}_2 \leftarrow (1-a_c)\frac{\boldsymbol{\iota}_n'\boldsymbol{\Theta}_c}{\boldsymbol{\iota}_n'\boldsymbol{\Theta}_c\boldsymbol{\iota}_n} + a_c\frac{(\boldsymbol{\omega}_1\bar{\mathbf{R}})\,\boldsymbol{\Theta}_c}{(\boldsymbol{\omega}_1\bar{\mathbf{R}})\,\boldsymbol{\Theta}_c\boldsymbol{\iota}_n}, \tag{17}$$

where the weighting $a_r$ and $a_c$ are defined as (18) such that

$$\begin{aligned} a_r &= \frac{\left[\mu_{target}\,\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m - (\boldsymbol{\omega}_2\,\bar{\mathbf{R}}')\boldsymbol{\Theta}_r\,\boldsymbol{\iota}_m\right]\left[(\boldsymbol{\omega}_2\,\bar{\mathbf{R}}')\boldsymbol{\Theta}_r\,\boldsymbol{\iota}_m\right]}{\left[(\boldsymbol{\omega}_2\,\bar{\mathbf{R}}')\boldsymbol{\Theta}_r(\bar{\mathbf{R}}\,\boldsymbol{\omega}_2')\right]\left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m\right) - \left[(\boldsymbol{\omega}_2\,\bar{\mathbf{R}}')\boldsymbol{\Theta}_r\,\boldsymbol{\iota}_m\right]^2}, \\ a_c &= \frac{\left[\mu_{target}\,\boldsymbol{\iota}_n'\boldsymbol{\Theta}_c\boldsymbol{\iota}_n - (\boldsymbol{\omega}_1\,\bar{\mathbf{R}})\boldsymbol{\Theta}_c\,\boldsymbol{\iota}_n\right]\left[(\boldsymbol{\omega}_1\,\bar{\mathbf{R}})\boldsymbol{\Theta}_c\,\boldsymbol{\iota}_n\right]}{\left[(\boldsymbol{\omega}_1\,\bar{\mathbf{R}})\boldsymbol{\Theta}_c(\bar{\mathbf{R}}'\,\boldsymbol{\omega}_1')\right]\left(\boldsymbol{\iota}_n'\boldsymbol{\Theta}_c\boldsymbol{\iota}_n\right) - \left[(\boldsymbol{\omega}_1\,\bar{\mathbf{R}})\boldsymbol{\Theta}_c\,\boldsymbol{\iota}_n\right]^2}. \end{aligned} \tag{18}$$

**Remark 4** When we ignore the return constraint, it yields the same iterations with GMV. To this end, GMV is a special case of MWC: the weight we assign to return-maximization is 0, i.e., $a_r = 0$ or $a_c = 0$. □

### 2.2.3 Markowitz Risk-Constrained (MRC) Optimization

The problem in MRC is to maximize the Sharpe ratio with constraints that the average return is higher than the desired return $\mu_{target}$ or the risk is lower than the desired risk $\sigma_{target}$. Following

the same notation in MWC optimization, we can set our optimization problem to maximize Sharpe ratio as shown in (19), which captures the average return per unit of risk:

$$\max_{\boldsymbol{\omega}_1,\boldsymbol{\omega}_2} \quad \frac{\boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2'}{\nu}, \quad \text{s.t.} \qquad \boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2' \geq \mu_{target}, \quad \left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1'\right) \leq \sigma_{target}^2. \tag{19}$$

Although this optimization problem is nonconvex, it admits an equivalent reformulation based on scale invariance of Sharpe ratio. More specifically, the portfolio weights from $\sigma_{target}$- and $\mu_{target}$-scaling may be different, but they can always lead to the same value of Sharpe ratio, so in our subsequent simulation and empirical studies, we only focus on the $\sigma_{\text{target}}$-scaling formulation. The optimization problems under two scaling methods are derived as follows:

(i) ($\sigma_{target}$-scaling) Binding the risk constraint and maximize expected return, i.e., maximize the numerator for a fixed denominator:

$$\max_{\boldsymbol{\omega}_1,\boldsymbol{\omega}_2} \quad \boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2', \quad \text{s.t.} \quad \left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1'\right) = \sigma_{target}^2. \tag{20}$$

(ii) ($\mu_{target}$-scaling) Binding return constraint and minimize portfolio risk, i.e., minimize the denominator for a fixed numerator.

$$\min_{\boldsymbol{\omega}_1,\boldsymbol{\omega}_2} \quad \left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1'\right), \quad \text{s.t.} \quad \boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2' = \mu_{target}. \tag{21}$$

When we estimate the portfolio weights in MRC using $\sigma_{target}$-scaling, the resulting updates for $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}2$ can be written as follows:

$$\boldsymbol{\omega}_1 \leftarrow \frac{\sigma_{target}\sqrt{\left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right)^{-1}}}{\sqrt{\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\left(\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right)}}\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r, \quad \boldsymbol{\omega}_2 \leftarrow \frac{\sigma_{target}\sqrt{\left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1'\right)^{-1}}}{\sqrt{\left(\boldsymbol{\omega}_1\bar{\mathbf{R}}\right)\boldsymbol{\Theta}_c\left(\bar{\mathbf{R}}'\boldsymbol{\omega}_1'\right)}}\left(\boldsymbol{\omega}_1\bar{\mathbf{R}}\right)\boldsymbol{\Theta}_c. \tag{22}$$

Alternatively, if we employ $\mu_{target}$-scaling, then the updates for $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ can be derived as:

$$\boldsymbol{\omega}_1 \leftarrow \frac{\mu_{target}}{\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\left(\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right)}\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r, \quad \boldsymbol{\omega}_2 \leftarrow \frac{\mu_{target}}{\left(\boldsymbol{\omega}_1\bar{\mathbf{R}}\right)\boldsymbol{\Theta}_c\left(\bar{\mathbf{R}}'\boldsymbol{\omega}_1'\right)}\left(\boldsymbol{\omega}_1\bar{\mathbf{R}}\right)\boldsymbol{\Theta}_c. \tag{23}$$

**Remark 5** At the portfolio weights $(\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*)$ obtained from the iterations in (22) and (23), the same Lagrangian multiplier $\lambda$ associated with the binding constraint (risk or return, depending on

which constraint is active) enters the first order conditions for both row and column modes. This forces the scaling factors in the two mode-wise updates to coincide, yielding the following equality:

$$\sqrt{\left(\boldsymbol{\omega}_2^* \bar{\mathbf{R}}'\right)\left(\boldsymbol{\omega}_2^* \boldsymbol{\Sigma}_c \boldsymbol{\omega}_2^{*\prime}\right)^{-1} \boldsymbol{\Theta}_r \left(\bar{\mathbf{R}} \boldsymbol{\omega}_2^{*\prime}\right)} = \sqrt{\left(\boldsymbol{\omega}_1^* \bar{\mathbf{R}}\right)\left(\boldsymbol{\omega}_1^* \boldsymbol{\Sigma}_r \boldsymbol{\omega}_1^{*\prime}\right)^{-1} \boldsymbol{\Theta}_c \left(\bar{\mathbf{R}}' \boldsymbol{\omega}_1^{*\prime}\right)}. \tag{24}$$

To this end, the $\sigma_{target}$- and $\mu_{target}$-scaling can yield the exactly identical weights if $\mu_{target}/\sigma_{target}$ equals to (24). Meanwhile, if this condition holds, the Sharpe ratio will also be the same as (24). □

### 2.3 Tensor Graphical LASSO

Build on Section 2.2, we apply Tensor Graphical LASSO (TLASSO) proposed by Lyu et. al. (2020), which aims to minimize (25) in each mode by keeping the precision matrix in the other mode fixed, and iterate until converge:

$$\mathcal{L}_k(\boldsymbol{\Theta}_k) = \frac{1}{p_k} \operatorname{tr}(\mathbf{S}_k \boldsymbol{\Theta}_k) - \frac{1}{p_k} \log |\boldsymbol{\Theta}_k| + \lambda_k \|\boldsymbol{\Theta}_k\|_{1,\text{off}}, \tag{25}$$

where $\mathbf{S}_r = \frac{1}{Tp_c} \sum_{t=1}^{T} \mathbf{V}_{r,t} \mathbf{V}_{r,t}'$, $\mathbf{S}_c = \frac{1}{Tp_r} \sum_{t=1}^{T} \mathbf{V}_{c,t} \mathbf{V}_{c,t}'$, $\mathbf{V}_{r,t} = \mathbf{U}_t \boldsymbol{\Theta}_c^{1/2} \in \mathbb{R}^{m \times n}$, and $\mathbf{V}_{c,t} = \mathbf{U}_t' \boldsymbol{\Theta}_r^{1/2} \in \mathbb{R}^{n \times m}$, for $p_k \in \{m, n\}$ and $k \in \{r, c\}$.

Our tuning parameter is selected following Lyu et. al. (2020) and Lee and Seregina (2023). Firstly, we choose the maximum value of grid for $\lambda_k$ which compress all the off-diagonal elements of row- and column-wise precision matrix $\boldsymbol{\Theta}_k$ to be 0. In the next step, following Condition 3.3 in Lyu et. al. (2020), we select the row- and column-wise scalings as follows

$$\vartheta_r = \sqrt{\frac{\log(m)}{T\,(mn)\,m}}, \quad \vartheta_c = \sqrt{\frac{\log(n)}{T\,(mn)\,n}}, \tag{26}$$

to set the smallest value of the grid $\lambda_{k,\min} = \vartheta_k \lambda_{k,\max}$, then we perform a grid search over the interval within $[\lambda_{k,\min}, \lambda_{k,\max}]$, and select the optimal tuning parameter according to BIC in (27):

$$\text{BIC}(\boldsymbol{\Theta}_k) = T\left[\operatorname{tr}\left(\mathbf{S}_k \boldsymbol{\Theta}_k\right) - \log\det\left(\boldsymbol{\Theta}_k\right)\right] + (\log T) \sum_{i \le j} \mathbb{1}\left[(\boldsymbol{\Theta}_k)_{ij} \neq 0\right]. \tag{27}$$

**Remark 6** Lyu et. al. (2020) show that, in TLASSO, the resulting precision matrix estimator can

be consistent even with one iteration. Accordingly, in our portfolio construction procedures that incorporate TLASSO , we perform only a single round of mode-wise iteration.

### 2.4 Tensor Portfolio with Factor Structure

The idea that stock returns exhibit a factor structure is well established in the literature (e.g., Fama and French (1993), Kozak, Nagel, and Santosh (2018), and Lee and Seregina (2023)). Building on this, we propose updating for GMV, MWC, and MRC under the factor-structured setting. Consider a case such that the fluctuations of return matrix $\mathbf{R}_t$ are driven by the latent factor tensor $\mathbf{F}_t \in \mathbb{R}^{\varphi_1 \times \varphi_2}$, where $\mathbf{A}$ and $\mathbf{B}$ are factor loadings, and $\mathbf{U}_t$ is idiosyncratic component. We assume that $\mathbf{A}$ and $\mathbf{B}$ satisfies $\mathbf{A}'\mathbf{A} = m\mathbf{I}$ and $\mathbf{B}'\mathbf{B} = n\mathbf{I}$ for identification, and $\mathbb{E}\left[\text{vec}(\mathbf{F}_t)\text{vec}(\mathbf{U}_t)'\right] = 0$. Then $\mathbf{R}_t$ can be described as $\mathbf{R}_t = \bar{\mathbf{R}} + \mathbf{A}\mathbf{F}_t\mathbf{B}' + \mathbf{U}_t$, and the risk under such structure is rewritten as (28):

$$\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{\omega}_1\mathbf{E}_t\boldsymbol{\omega}_2'\right)^2 = \frac{1}{T}\sum_{t=1}^{T}\left[\boldsymbol{\omega}_1\left(\mathbf{A}\mathbf{F}_t\mathbf{B}'\right)\boldsymbol{\omega}_2'\boldsymbol{\omega}_2\left(\mathbf{A}\mathbf{F}_t\mathbf{B}'\right)'\boldsymbol{\omega}_1' + \left(\boldsymbol{\omega}_1\mathbf{U}_t\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_2\mathbf{U}_t'\boldsymbol{\omega}_1\right)\right]. \tag{28}$$

For idiosyncratic component $\mathbf{U}_t$, we assume it exhibits a structure of Kronecker Separable Covariance similar to Assumption 1: $\boldsymbol{\Sigma}_u = \boldsymbol{\Sigma}_{u,c} \otimes \boldsymbol{\Sigma}_{u,r}$, and hence $\left(\boldsymbol{\omega}_1\mathbf{U}_t\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_2\mathbf{U}_t'\boldsymbol{\omega}_1\right) = \left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_{u,c}\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_{u,r}\boldsymbol{\omega}_1'\right)$. This structure is more consistent with those commonly adopted in related studies on tensor factor models. On the other hand, the model in Section 2.3 can be viewed as a special case of the model in Section 2.4 with zero number of factors. Given this structure, define $\tilde{\boldsymbol{\Sigma}}_{f,r} = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{F}_t\mathbf{B}'\boldsymbol{\omega}_2'\boldsymbol{\omega}_2\mathbf{B}\mathbf{F}_t'\right)$, $\tilde{\boldsymbol{\Sigma}}_{u,r} = \left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_{u,c}\boldsymbol{\omega}_2\right)\boldsymbol{\Sigma}_{u,r}$, $\tilde{\boldsymbol{\Sigma}}_{f,c} = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{F}_t'\mathbf{A}'\boldsymbol{\omega}_1'\boldsymbol{\omega}_1\mathbf{A}\mathbf{F}_t\right)$, and $\tilde{\boldsymbol{\Sigma}}_{u,c} = \left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_{u,r}\boldsymbol{\omega}_1\right)\boldsymbol{\Sigma}_{u,c}$, we can rewrite the the Row- and Column-Weighted Covariance $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ such that

$$\boldsymbol{\Sigma}_1 = \mathbf{A}\tilde{\boldsymbol{\Sigma}}_{f,r}\mathbf{A}' + \tilde{\boldsymbol{\Sigma}}_{u,r}, \quad \boldsymbol{\Sigma}_2 = \mathbf{B}\tilde{\boldsymbol{\Sigma}}_{f,c}\mathbf{B}' + \tilde{\boldsymbol{\Sigma}}_{u,c}. \tag{29}$$

Let $\tilde{\boldsymbol{\Theta}}_{f,r} = \tilde{\boldsymbol{\Sigma}}_{f,r}^{-1}$, $\tilde{\boldsymbol{\Theta}}_{f,c} = \tilde{\boldsymbol{\Sigma}}_{f,c}^{-1}$, $\tilde{\boldsymbol{\Theta}}_{u,r} = \tilde{\boldsymbol{\Sigma}}_{u,r}^{-1}$, $\tilde{\boldsymbol{\Theta}}_{u,c} = \tilde{\boldsymbol{\Sigma}}_{u,c}^{-1}$, $\boldsymbol{\Theta}_1 = \boldsymbol{\Sigma}_1^{-1}$ and $\boldsymbol{\Theta}_2 = \boldsymbol{\Sigma}_2^{-1}$, by Sherman–Morrison–Woodbury formula, we obtain the precision matrices such that:

$$\begin{aligned}
\boldsymbol{\Theta}_1 &= \tilde{\boldsymbol{\Theta}}_{u,r} - \tilde{\boldsymbol{\Theta}}_{u,r}\mathbf{A}\left(\tilde{\boldsymbol{\Theta}}_{f,r} + \mathbf{A}'\tilde{\boldsymbol{\Theta}}_{u,r}\mathbf{A}\right)^{-1}\mathbf{A}'\tilde{\boldsymbol{\Theta}}_{u,r}, \\
\boldsymbol{\Theta}_2 &= \tilde{\boldsymbol{\Theta}}_{u,c} - \tilde{\boldsymbol{\Theta}}_{u,c}\mathbf{B}\left(\tilde{\boldsymbol{\Theta}}_{f,c} + \mathbf{B}'\tilde{\boldsymbol{\Theta}}_{u,c}\mathbf{B}\right)^{-1}\mathbf{B}'\tilde{\boldsymbol{\Theta}}_{u,c}.
\end{aligned} \tag{30}$$

Since $\boldsymbol{\Theta}_1$ ($\boldsymbol{\Theta}_2$) is dependent on $\boldsymbol{\omega}_2$ ($\boldsymbol{\omega}_1$), unlike the GMV without factors in Section 2.2 and 2.3, the GMV weights are the now updated iteratively as follows:

$$\boldsymbol{\omega}_1 \leftarrow \frac{\boldsymbol{\iota}_m' \boldsymbol{\Theta}_2}{\boldsymbol{\iota}_m' \boldsymbol{\Theta}_2 \boldsymbol{\iota}_m}, \quad \boldsymbol{\omega}_2 \leftarrow \frac{\boldsymbol{\iota}_n' \boldsymbol{\Theta}_1}{\boldsymbol{\iota}_n' \boldsymbol{\Theta}_1 \boldsymbol{\iota}_n}. \tag{31}$$

Next, in MWC, the weights are updated following

$$\boldsymbol{\omega}_1 \leftarrow (1-a_r)\frac{\boldsymbol{\iota}_m' \boldsymbol{\Theta}_2}{\boldsymbol{\iota}_m' \boldsymbol{\Theta}_2 \boldsymbol{\iota}_m} + a_r \frac{\boldsymbol{\omega}_2 \bar{\mathbf{R}}' \boldsymbol{\Theta}_2}{\boldsymbol{\iota}_m' \boldsymbol{\Theta}_2 \bar{\mathbf{R}} \boldsymbol{\omega}_2'}, \quad \boldsymbol{\omega}_2 \leftarrow (1-a_c)\frac{\boldsymbol{\iota}_n' \boldsymbol{\Theta}_1}{\boldsymbol{\iota}_n' \boldsymbol{\Theta}_1 \boldsymbol{\iota}_n} + a_c \frac{\boldsymbol{\omega}_1 \bar{\mathbf{R}} \boldsymbol{\Theta}_1}{\boldsymbol{\iota}_n' \boldsymbol{\Theta}_1 \bar{\mathbf{R}}' \boldsymbol{\omega}_1'}, \tag{32}$$

and the weights we impose on return maximization component are

$$\begin{aligned} a_r &= \frac{\left(\mu_{target}\, \boldsymbol{\iota}_m' \boldsymbol{\Theta}_2 \boldsymbol{\iota}_m \;-\; \boldsymbol{\iota}_m' \boldsymbol{\Theta}_2 \bar{\mathbf{R}} \boldsymbol{\omega}_2'\right)\left(\boldsymbol{\iota}_m' \boldsymbol{\Theta}_2 \bar{\mathbf{R}} \boldsymbol{\omega}_2'\right)}{\left[(\bar{\mathbf{R}} \boldsymbol{\omega}_2')' \boldsymbol{\Theta}_2 (\bar{\mathbf{R}} \boldsymbol{\omega}_2')\right]\left(\boldsymbol{\iota}_m' \boldsymbol{\Theta}_2 \boldsymbol{\iota}_m\right) \;-\; \left(\boldsymbol{\iota}_m' \boldsymbol{\Theta}_2 \bar{\mathbf{R}} \boldsymbol{\omega}_2'\right)^2}, \\ a_c &= \frac{\left(\mu_{target}\, \boldsymbol{\iota}_n' \boldsymbol{\Theta}_1 \boldsymbol{\iota}_n \;-\; \boldsymbol{\iota}_n' \boldsymbol{\Theta}_1 \bar{\mathbf{R}}' \boldsymbol{\omega}_1'\right)\left(\boldsymbol{\iota}_n' \boldsymbol{\Theta}_1 \bar{\mathbf{R}}' \boldsymbol{\omega}_1'\right)}{\left[(\bar{\mathbf{R}}' \boldsymbol{\omega}_1')' \boldsymbol{\Theta}_1 (\bar{\mathbf{R}}' \boldsymbol{\omega}_1')\right]\left(\boldsymbol{\iota}_n' \boldsymbol{\Theta}_1 \boldsymbol{\iota}_n\right) \;-\; \left(\boldsymbol{\iota}_n' \boldsymbol{\Theta}_1 \bar{\mathbf{R}}' \boldsymbol{\omega}_1'\right)^2}. \end{aligned} \tag{33}$$

Finally, in MRC, we also have two equivalent optimization problems. For $\sigma_{target}$-scaling, in which we aim to maximize the mean return of the portfolio with binding risk constraint, the weights are updating by

$$\boldsymbol{\omega}_1 \leftarrow \frac{\sigma_{target}}{\sqrt{\left(\boldsymbol{\omega}_2 \bar{\mathbf{R}}'\right) \boldsymbol{\Theta}_2 \left(\bar{\mathbf{R}} \boldsymbol{\omega}_2'\right)}} \left(\boldsymbol{\omega}_2 \bar{\mathbf{R}}'\right) \boldsymbol{\Theta}_2, \quad \boldsymbol{\omega}_2 \leftarrow \frac{\sigma_{target}}{\sqrt{\left(\boldsymbol{\omega}_1 \bar{\mathbf{R}}\right) \boldsymbol{\Theta}_1 \left(\bar{\mathbf{R}}' \boldsymbol{\omega}_1'\right)}} \left(\boldsymbol{\omega}_1 \bar{\mathbf{R}}\right) \boldsymbol{\Theta}_1. \tag{34}$$

Alternatively, similar with (23), with binding return constraint, we can derive the updating of $\mu_{target}$-scaling as

$$\boldsymbol{\omega}_1 \leftarrow \frac{\mu_{target}}{\left(\boldsymbol{\omega}_2 \bar{\mathbf{R}}'\right) \boldsymbol{\Theta}_2 \left(\bar{\mathbf{R}} \boldsymbol{\omega}_2'\right)} \left(\boldsymbol{\omega}_2 \bar{\mathbf{R}}'\right) \boldsymbol{\Theta}_2, \quad \boldsymbol{\omega}_2 \leftarrow \frac{\mu_{target}}{\left(\boldsymbol{\omega}_1 \bar{\mathbf{R}}\right) \boldsymbol{\Theta}_1 \left(\bar{\mathbf{R}}' \boldsymbol{\omega}_1'\right)} \left(\boldsymbol{\omega}_1 \bar{\mathbf{R}}\right) \boldsymbol{\Theta}_1. \tag{35}$$

Using the property of scale invariance of Sharpe ratio, (34) and (35) can reach the same Sharpe ratio at the optimal $(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$. In addition, this iteration coincides with (34) in the portfolio weights when the ratio of the target mean return to the target risk - i.e., the target Sharpe ratio - satisfies $\mu_{target}/\sigma_{target} = \sqrt{\left(\boldsymbol{\omega}_2 \bar{\mathbf{R}}'\right) \boldsymbol{\Theta}_2 \left(\bar{\mathbf{R}} \boldsymbol{\omega}_2'\right)} = \sqrt{\left(\boldsymbol{\omega}_1 \bar{\mathbf{R}}\right) \boldsymbol{\Theta}_1 \left(\bar{\mathbf{R}}' \boldsymbol{\omega}_1'\right)}$. The derivation of the weights in this

section are available in Appendix A.2.

The estimation of (30) is available in Algorithm 1. In Step 1, there are many alternative approaches to estimating the core factor tensor $\mathbf{F}$, the corresponding loadings, and idiosyncratic components (Chen, Tsay, and Chen (2020), Chen, Yang, and Zhang (2022), Chen and Fan (2023), Han et. al. (2024), etc.). For the purpose of illustration, we adopt $\alpha$-PCA procedure of Chen and Fan (2023). Define $\mathbf{M}_r$ and $\mathbf{M}_c$ as

$$\mathbf{M}_r = \frac{1}{Tmn}\sum_{t=1}^{T}\tilde{\mathbf{E}}_t\tilde{\mathbf{E}}_t', \qquad \mathbf{M}_c = \frac{1}{Tmn}\sum_{t=1}^{T}\tilde{\mathbf{E}}_t'\tilde{\mathbf{E}}_t, \tag{36}$$

where $\tilde{\mathbf{E}}_t = \mathbf{E}_t + \tilde{\alpha}\bar{\mathbf{E}} \in \mathbb{R}^{m\times n}$, $\tilde{\alpha} = \sqrt{\alpha+1}-1$, and $\alpha \in [-1,\infty)$ is a hyperparameter that aggregates the information from the first-moment(mean) and the second-moment (contemporary covariance) of $\mathbf{E}_t$. We take $\sqrt{m}$ times the top $\varphi_r$ eigenvectors of $\mathbf{M}_r$ and $\sqrt{n}$ times the top $\varphi_c$ eigenvectors of $\mathbf{M}_c$ as the row and column factor loadings $\mathbf{A}$ and $\mathbf{B}$, then the estimated factor $\mathbf{F}_t$ is constructed as $1/(mn)\mathbf{A}'\mathbf{E}_t\mathbf{B}$, and the idiosyncratic component is derived by $\mathbf{U}_t = \mathbf{E}_t - \mathbf{A}\mathbf{F}_t\mathbf{B}'$.

**Remark 7** In our simulation and empirical study, following the derivation of risk in (28), we apply $\alpha$-PCA to the demeaned return tensor $\mathbf{E}_t = \mathbf{R}_t - \bar{\mathbf{R}}$ rather than to $\mathbf{R}_t$, this setup is consistent with the case with $\alpha = -1$ in $\alpha$-PCA. □

Our tuning parameter is selected following the same approach as Section 2.3. The differences are, firstly, BIC is about the precision matrix of idiosyncratic component (37):

$$\mathrm{BIC}(\mathbf{\Theta}_{u,k}) = T\left[\mathrm{tr}\big(\mathbf{S}_{u,k}\mathbf{\Theta}_{u,k}\big) - \log\det\big(\mathbf{\Theta}_{u,k}\big)\right] + (\log T)\sum_{i\le j}\mathbb{1}\left[(\mathbf{\Theta}_{u,k})_{ij}\neq 0\right]. \tag{37}$$

Secondly, we change the way to derive the scaling $\vartheta$ to consider the impacts from factors. We use $\vartheta_k = \varphi_k^2\sqrt{(\log p_k)/T} + \varphi_k^3/\sqrt{p_k}$, according to Theorem 2 of Lee and Seregina (2023), to set the smallest value of the grid by $\lambda_{k,\min} = \vartheta_k\lambda_{k,\max}$, where $\varphi_r$ ($\varphi_c$) is the number of factors in the row-wise (column-wise) mode.

**Algorithm 1:** Portfolio Weights under Tensor Factor Model

**Input:** Return tensor $\mathbf{R} \in \mathbb{R}^{T\times m\times n}$, where $t = 1, \ldots, T$. Tuning parameter $\lambda$. Tolerance level of iteration `tol`. The maximum number of iterations `max_iter`.

**Output:** Optimal portfolio weights $(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ for each portfolio strategy $i \in \{\text{GMV}, \text{MWC}, \text{MRC}\}$.

**Step 1** Estimate the core tensor $\mathbf{F} \in \mathbb{R}^{T\times\varphi_1\times\varphi_2}$, loading matrices $\mathbf{A} \in \mathbb{R}^{m\times\varphi_1}$ and $\mathbf{B} \in \mathbb{R}^{n\times\varphi_2}$, and the idiosyncratic component $\mathbf{U} \in \mathbb{R}^{T\times m\times n}$ from $\mathbf{R}_t = \bar{\mathbf{R}} + \mathbf{A}\mathbf{F}_t\mathbf{B}' + \mathbf{U}_t$, $t = 1, \ldots, T$.

**Step 2** Apply Tensor Graphical LASSO (TLASSO) proposed by Lyu et. al. (2020) to $\{\mathbf{U}_t\}_{t=1}^T$ to estimate the Kronecker-separable idiosyncratic covariance components $\widehat{\boldsymbol{\Theta}}_{u,r} \in \mathbb{R}^{m\times m}$, and $\widehat{\boldsymbol{\Theta}}_{u,c} \in \mathbb{R}^{n\times n}$.

**Step 3 foreach** $s \in \{GMV, MWC, MRC\}$ **do**

  Initialize $(\boldsymbol{\omega}_1^{(0)}, \boldsymbol{\omega}_2^{(0)})$;

  **for** $i = 0, 1, 2, \ldots,$ `max_iter` $- 1$ **do**

    Update $\boldsymbol{\Theta}_1^{(i+1)}$ and $\boldsymbol{\Theta}_2^{(i+1)}$ by alternating updates of $\boldsymbol{\omega}_1^{(i+1)}$ and $\boldsymbol{\omega}_2^{(i+1)}$ according to portfolio strategy $s$;

    **if** $\max\Big(\|\boldsymbol{\omega}_1^{(i+1)} - \boldsymbol{\omega}_1^{(i)}\|_\infty,\ \|\boldsymbol{\omega}_2^{(i+1)} - \boldsymbol{\omega}_2^{(i)}\|_\infty\Big) <$ `tol` **then**

      **break**;

  Set $(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) \leftarrow (\boldsymbol{\omega}_1^{(i+1)}, \boldsymbol{\omega}_2^{(i+1)})$;

**return** $(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ *for each* $s \in \{GMV, MWC, MRC\}$;

## 3 Convergence of Alternating Minimization

We utilize the convergence theory of Tseng (2001) to demonstrate why the updating rules in Section 2 are guaranteed to converge, all the proofs of Theorem and Proposition are available in Appendix B. As discussed before, regardless of GMV, MWC, or MRC, the objective function in our tensor setting is inherently a bilinear and thus non-convex optimization problem, making it difficult to obtain closed-form solutions for the mode-specific weighting vectors. We therefore employ an Alternating Minimization (AM) algorithm to estimate the portfolio weights iteratively. Specifically, at each iteration we fix the weighting vector along one mode (e.g., sectors or characteristics) and update the weighting vector along the other mode by following the conventional vector-based portfolio optimization. We then alternate between the two modes and repeat the procedure until convergence. Note that this AM procedure can also be viewed as Block Coordinate Descent (BCD), where the two blocks correspond to $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$.

In GMV and MWC, we need to minimize the risk under given constraints. In MRC, in our

simulation and empirical part, we are using $\sigma$-scaling methods, and it will generate the same results as $\mu$-scaling since they are dual problems. To this end, subject to full-investment constraints and/or return constraints, all three models are using the same objective function as (38), i.e., risk, such that

$$\{\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*\} = \arg\min_{\boldsymbol{\omega}_1, \boldsymbol{\omega}_2} f(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) = \arg\min_{\boldsymbol{\omega}_1, \boldsymbol{\omega}_2} \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{\omega}_1 \mathbf{E}_t \boldsymbol{\omega}_2)^2 . \tag{38}$$

**Theorem 1.** *The sequence of iterates* $\{\boldsymbol{\Omega}^{(i)} = (\boldsymbol{\omega}_1^{(i)}, \boldsymbol{\omega}_2^{(i)})\}_{i=0,1,\ldots}$ *generated by the AM algorithm converges to a coordinate-wise minimum of the objective function* $f(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ *defined in* (38)*, subject to the full-investment and/or target-return constraints, where* $i$ *denotes the iteration index.*

Denote the joint variables as $\mathbf{z} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) \in \mathbb{R}^m \times \mathbb{R}^n$, and let $\mathrm{dom} f$ be the effective domain of $f$, such that $\mathrm{dom} f = \{\mathbf{z} : f(\mathbf{z}) < \infty\}$. According to Tseng (2001), a coordinatewise minimum is defined as follows: *If* $\mathbf{z} \in \mathrm{dom} f$, $f(\mathbf{z} + (\mathbf{0}, \mathbf{d}_2)) \geq f(\mathbf{z})$, *and* $f(\mathbf{z} + (\mathbf{d}_1, \mathbf{0})) \geq f(\mathbf{z})$ *for all directions* $\mathbf{d}_1 \in \mathbb{R}^m$ *and* $\mathbf{d}_2 \in \mathbb{R}^n$ *such that* $\mathbf{z} + (\mathbf{d}_1, \mathbf{0}) \in \mathrm{dom} f$ *and* $\mathbf{z} + (\mathbf{0}, \mathbf{d}_2) \in \mathrm{dom} f$*, then* $\mathbf{z}$ *is a coordinatewise minimum of* $f$. Meanwhile, a stationary point is defined as follows: *If* $\mathbf{z} \in \mathrm{dom} f$ *and the directional derivative satisfies* $\boldsymbol{\nabla}_{\mathbf{d}} f(\mathbf{z}) \geq 0$ *for all feasible directions* $\mathbf{d} = (\mathbf{d}_1, \mathbf{d}_2) \in \mathbb{R}^m \times \mathbb{R}^n$*, then* $\mathbf{z}$ *is a stationary point of* $f$. Intuitively, a coordinatewise minimum ensures $f$ cannot be decreased by updating any single block alone, though joint updates might still yield improvements. A stationary point strictly requires $\boldsymbol{\nabla}_{\mathbf{d}} f(\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*) \geq 0$ along all joint feasible directions. Thus, we need Proposition 1 to upgrade the coordinatewise minimum $\mathbf{z}^* = (\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*)$ to a stationary point, thereby guaranteeing a local minimum when the Hessian is positive definite.

**Proposition 1.** *The coordinatewise minima* $(\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*)$ *in GMV, MWC and MRC are stationary points.*

**Remark 8** Since the stationary point may be a saddle, we require positive definite Hessian matrix of $f$ to ensure $(\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*)$ obtained through Alternating-Minimization in GMV, MWC and MRC are both local minima. □

Our alternating minimization methods are widely used in tensor-related research. Chen, Xiao, and Yang (2021) develop a matrix autoregression framework estimated via alternating least squares and alternating maximum likelihood, and establish the corresponding asymptotic properties. Building on the iterative methodology of Chen, Xiao, and Yang (2021), Cai et. al. (2025) extend the

factor-augmented regression model of Fan, Ke, and Wang (2020) to the tensor setting. Meanwhile, Chen, Yang, and Zhang (2022) and Han et. al. (2024) estimate tensor factor models by iteratively applying SVD to matricized (unfolded) tensors. In addition, Han et. al. (2024) and Chen, Han, and Yu (2024) develop iterative estimators for tensor factor models under the additional restriction that the factor tensor is diagonal.

Theorem 1 and Proposition 1 ensure that the alternating updates of $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ converge, however, even with a positive definite Hessian and well-posed return or risk targets, the algorithm is only guaranteed to reach a local minima rather than a global minima. To mitigate this limitation, we follow Chen, Xiao, and Yang (2021) to adopt a robust initialization strategy, thereby enhancing the stability and overall quality of the iterative solution. Firstly, if we impose separable covariance structure on the return $\mathbf{R}_t$, as in Section 2.2 and 2.3, GMV admits a closed-form solution and no longer requires iteration; hence we directly use the GMV solution as the initial value for MWC and MRC. Secondly, in the tensor portfolio with factor structure (Section 2.4), the separable covariance structure is imposed exclusively on the idiosyncratic component $\mathbf{U}_t$. Based on the iterations shown in the row and column weights in (31), (32), (34), and (35), all of them (GMV, MWC, and MRC) are highly coupled and mutually dependent, lacking any closed-form solutions. we use the equal-weight portfolio as the initial value because equal weight provides a relatively "mild" starting point, which helps prevent the iterations from being driven by extreme weights at the outset and improves numerical stability.

# 4 Simulation

In this section, we conduct Monte Carlo simulations to illustrate the benefits of using tensor-based portfolios over vector-based portfolios when the data presents a tensor structure.

## 4.1 Data Generating Processes (DGPs)

We design three categories of DGPs to compare the performance of tensor-based and vector-based portfolios based on different settings: with separable covariance but no sparsity, with both separable covariance and sparsity, and with factor structure.

Based on Remark 1, regardless of the return matrix $\mathbf{R}_t$ or the idiosyncratic component $\mathbf{U}_t$, once

a separable covariance structure holds, they can both be written in the form $\mathbf{\Sigma}_r^{1/2}\mathbf{Z}_t\mathbf{\Sigma}_c^{1/2}$. Hence, by changing the distribution of $\mathbf{Z}_t$, we can directly alter the tail behavior (the higher-order moment properties) of $\mathbf{R}_t$ or $\mathbf{U}_t$ while keeping the same separable second-moment structure governed by $\mathbf{\Sigma}_r$ and $\mathbf{\Sigma}_c$.

**Remark 9** Although our DGPs are constructed to exhibit a tensor structure, this does not imply that the comparison is unfair to vector-based models. Any dataset - regardless of whether it possesses an intrinsic tensor structure - can always be flattened (vectorized) as a vector. Hence, our simulated data are inherently compatible with a vector representation. □

**DGP 1: Tensor with Separable Covariance and No Sparse Precision Matrices on $\mathbf{R}_t$**

We generate the return tensor $\mathbf{R} \in \mathbb{R}^{T\times m\times n}$, where $m = n = 30$. Each series in $\mathbf{R}$ follows $\mathbf{R}_t = \bar{\mathbf{R}} + \mathbf{V}_t \in \mathbb{R}^{m\times n}$. We impose constraint of separable covariance on $\mathbf{R}_t$ and allow each element in $\mathbf{V}_t$ to be correlated: $\mathbf{V}_t = \mathbf{\Sigma}_{v,r}^{1/2}\mathbf{Z}_t\mathbf{\Sigma}_{v,c}^{1/2}$, where $\mathbf{\Sigma}_{v,r}$ and $\mathbf{\Sigma}_{v,c}$ capture the row- and column-wise covariance of $\mathbf{V}_t$, which is in the form of Toeplitz such that their $ij_{th}$ elements are derived as $\mathbf{\Sigma}_{ij,v,r} = \mathbf{\Sigma}_{ij,v,c} = 0.2^{|i-j|}$. the mean value of $\mathbf{R}_t$, i.e., $\bar{\mathbf{R}}$, is set as $0.01 \times \mathbf{1}_{m\times n}$, where $\mathbf{1}_{m\times n}$ indicates a matrix full of ones.

**Remark 10** The intertemporal dependence of stock return tends to be very weak. Therefore, we do not allow even weak autoregressive dynamics in our DGPs. □

**DGP 2: Tensor with Separable Covariance and Sparse Precision Matrices on $\mathbf{R}_t$**

We impose sparsity on the row- and column-wise precision matrices in DGP 1, i.e., $\mathbf{\Theta}_{v,r} = \mathbf{\Sigma}_{v,r}^{-1}$ and $\mathbf{\Theta}_{v,c} = \mathbf{\Sigma}_{v,c}^{-1}$ are sparse. Instead of generating covariance matrices, we firstly generate two identity matrices for row and column modes, then we set all the off-diagonal elements to be 0.2, and then make off-diagonal elements to be 0 randomly with probability 80% to get matrices $\boldsymbol{\mathcal{J}}_r$ and $\boldsymbol{\mathcal{J}}_c$. To ensure $\mathbf{\Theta}_{v,r}$ and $\mathbf{\Theta}_{v,c}$ are positive definite, we use $\mathbf{\Theta}_{v,r} = \boldsymbol{\mathcal{J}}_r + (|\theta_{r,min}| + 0.5)\mathbf{I}_{m\times m}$ and $\mathbf{\Theta}_{v,c} = \boldsymbol{\mathcal{J}}_c + (|\theta_{c,min}| + 0.5)\mathbf{I}_{n\times n}$, following Zhao et. al. (2012), where $\theta_{r,min}$ and $\theta_{c,min}$ are the minimum eigenvalue of $\boldsymbol{\mathcal{J}}_r$ and $\boldsymbol{\mathcal{J}}_c$, respectively. Finally, we use Cholesky decomposition on row- and column wise covariance matrices $\mathbf{\Sigma}_{v,r} = \mathbf{\Theta}_{v,r}^{-1}$ and $\mathbf{\Sigma}_{v,c} = \mathbf{\Theta}_{v,c}^{-1}$ to get $\mathbf{\Sigma}_{v,r}^{1/2}$ and $\mathbf{\Sigma}_{v,c}^{1/2}$, and generate $\mathbf{V}_t$ by $\mathbf{V}_t = \mathbf{\Sigma}_{v,r}^{1/2}\mathbf{Z}_t\mathbf{\Sigma}_{v,c}^{1/2}$. $\bar{\mathbf{R}}$ is the same as in DGP 1, which is $0.01 \times \mathbf{1}_{m\times n}$.

**DGP 3: Tensor with Factor Structure, Separable Covariance and Sparse Precision Matrices on $\mathbf{U}_t$**

Begin with factor process generated by $\mathbf{F}_t = \boldsymbol{\mathcal{E}}_t$, where $\mathbf{F}_t \in \mathbb{R}^{2\times 2}$, and we use the oracle number of factors in the simulation. The $ij_{th}$ element in $\boldsymbol{\mathcal{E}}_t$ follows $\boldsymbol{\mathcal{E}}_{ij,t} \sim \mathcal{N}(0, 0.5^2)$, then the return tensor evolves following $\mathbf{R}_t = \bar{\mathbf{R}} + \lambda_f \mathbf{A}\mathbf{F}_t\mathbf{B}' + \mathbf{U}_t$. $\lambda_f$ indicates factor strength, we set it as $(mn)^{-0.5}$, following Lam and Yao (2012) and Wang, Liu, and Chen (2019). $\mathbf{A}$ and $\mathbf{B}$ are generated from $\mathcal{N}(0,1)$, then orthonormalized by QR decomposition. The idiosyncratic component $\mathbf{U}_t$ satisfies separable covariance structure such that $\mathbf{U}_t = \boldsymbol{\Sigma}_{u,r}^{1/2}\mathbf{Z}_t\boldsymbol{\Sigma}_{u,c}^{1/2}$. The mean value of $\mathbf{R}_t$, i.e., $\bar{\mathbf{R}}$, is generated by $0.01 \times \mathbf{1}_{m\times n}$.

To generate the row- and column-wise precision matrix with sparsity, $\boldsymbol{\Theta}_{u,r}$ and $\boldsymbol{\Theta}_{u,c}$, we firstly generate two Toeplitz matrices such that their $ij_{th}$ elements are derived as $\boldsymbol{\Theta}_{ij,u,r} = \boldsymbol{\Theta}_{ij,u,c} = 0.2^{|i-j|}$, then we impose the sparsity by setting the off-diagonal elements to be 0 with the probability 1/2. Finally, we take the inverse of $\boldsymbol{\Theta}_{u,r}$ and $\boldsymbol{\Theta}_{u,c}$ to get the row- and column-wise covariance matrices $\boldsymbol{\Sigma}_{u,r}$ and $\boldsymbol{\Sigma}_{u,c}$.

**Distributions of $\mathbf{Z}_t$**

We consider two distributions for $\mathbf{Z}_t$. The first assumes that each element in $\mathbf{Z}_t$ is normally distributed. Under this setting, the $ij_{th}$ element of $\mathbf{Z}_t$, $\mathbf{Z}_{t,ij}$, is generated from $\mathcal{N}(0, 0.5^2)$. This choice is relatively idealized, but it provides a clean benchmark in which the data are light-tailed and higher-order moments are well-behaved.

Secondly, heavy-tailed $\mathbf{V}_t$ and $\mathbf{U}_t$ are very common in stock market, we therefore modify our DGPs to mimic the heavy-tailed feature commonly observed in real stock returns. We allow $\mathbf{Z}_t$ in DGPs 1 to 3 follow $t$-distribution: $\mathbf{Z}_t = \frac{0.5}{\sqrt{df/(df-2)}}\boldsymbol{\xi}_t$, where the $ij_{th}$ element in $\boldsymbol{\xi}_t$, $\boldsymbol{\xi}_{t,ij}$, is following *standard* $t$-distribution with the degree of freedom $df = 3$, by such way we can have $\mathbf{Z}_{t,ij}$ following $t$-distribution with mean at 0 and standard deviation at 0.5 with the degree of freedom of 3 and thus $\mathbf{V}_t$ and $\mathbf{U}_t$ to be heavy tailed.

## 4.2 Comparison

Estimating a dense covariance matrix using a model tailored to sparse covariance structures is not a fair comparison. Similarly, applying a model that ignores factor structure to data generated with

factor structure is also unfair. To ensure a balanced evaluation across methods, we assess each pair of models under the DGP that is most consistent with their underlying assumptions about data: *Tensor* vs. *Vector*, *TensorGL* vs. *VectorGL*, and *TensorFGL* vs. *VectorFGL*. Within each pair, the two models adopt the same settings on factor and sparsity, they differ only in whether the estimation exploits a vector or tensor structure. The description of each model is listed below:

- *Tensor*: Tensor-based model assuming a separable covariance structure for returns, without regularization.
- *Vector*: Vector-based model without regularization.
- *TensorGL*: Tensor-based model assuming a separable covariance structure for returns, with Graphical LASSO regularization.
- *VectorGL*: Vector-based model with Graphical LASSO regularization.
- *TensorFGL*: Tensor-based model imposing a separable covariance structure on the idiosyncratic component, with Factor Graphical LASSO.
- *VectorFGL*: Vector-based model with Factor Graphical LASSO.

We assess the performance of Vector- and Tensor-based model by the OOS Sharpe ratio. The return target, $\mu_{target} = \frac{1}{mn}\sum_{i,j}\bar{\mathbf{R}}^{\text{IS}}_{i,j}$, is the In-Sample (IS) mean of $\bar{\mathbf{R}}$, and the risk target, $\sigma_{target} = \text{std}\left(\frac{1}{mn}\sum_{i,j}\bar{\mathbf{R}}^{\text{IS}}_{i,j}\right)$, is the standard deviation of IS $\bar{\mathbf{R}}$.

Our simulation focus on one-step forward OOS evaluation with a fixed-window estimation scheme, where IS takes 1008, OOS takes 1008, and the total sample size $T$ is 2016, which implies we are using 4-years return data for training and another 4-years for validation (on average, the number of trading days in a year is around 252). The simulations repeat 100 times.

We conduct one-sided $t$-test to compare the differences among these models:

$$z \;=\; \frac{\bar{d}-0}{s_d/\sqrt{\tau}} \quad \text{under } H_0 : \mathbb{E}[d_i] = 0 \quad \text{vs} \quad H_1 : \mathbb{E}[d_i] > 0, \tag{39}$$

where $d_i = \text{SR}_i^{\text{tensor}} - \text{SR}_i^{\text{vector}}$, $i = 1, \ldots, \tau$, $\tau = 100$ indicates the number of simulations, and $s_d$ indicates the sample standard deviation.

The simulation results are reported in Table 1. Panel A shows that, under DGP 1, tensor portfolio delivers a significantly higher OOS Sharpe ratio than vector portfolio. A similar pattern emerges under DGP 2, where we compare *TensorGL* and *VectorGL*. As shown in Panel B, *TensorGL* continues to exhibit an overwhelming advantage over *VectorGL* in terms of the OOS Sharpe ratio. However, the performance gap is noticeably smaller than in DGP 1. A plausible explanation is that, once all competing models apply dimensional reduction, the number of parameters to be estimated becomes comparable across methods. Consequently, the intrinsic low-dimensional advantage of tensor-based approaches becomes less pronounced. In Panel C, we further compare two factor-based models under DGP 3, *TensorFGL* and *TensorFGL*. Consistent with Panel B, the tensor-based model remains significantly better, but the magnitude of the improvement, according to $z$-value, is reduced.

On the other hand, when $\mathbf{Z}_t$ is heavy-tailed distributed, we find a similar result as the DGPs with normally distributed $\mathbf{Z}_t$. Tensor-based portfolios continue to deliver significantly higher OOS Sharpe ratios than their vector-based counterparts. This pattern further confirms the robust performance gains from exploiting tensor structure in portfolio construction.

# 5 Empirical Study

Our Monte Carlo simulations show that, whenever the data exhibit a tensor structure, tensor-based portfolios tend to deliver higher OOS Sharpe ratios. In empirical application, we turn to U.S. stock return data to compare the empirical performance of tensor-based and vector-based portfolios in practice.

## 5.1 Data

We collect a sample of daily stock returns from the Center for Research in Security Prices (CRSP), covering the period of January 1, 2010 to December 31, 2019, for a total 2,516 trading days. We first assign all individual stocks to 11 sectors based on the Global Industry Classification Standard (GICS): Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financial, Information Technology, Communication Services, Utilities, and Real Estate. Within each sector, we sort stocks by market capitalization (size) on a daily basis and select the top 30

stocks, yielding a return tensor of dimension $2516 \times 30 \times 11$. Under this construction, we assume a stock's portfolio weight is jointly determined by its sector and size rank within the sector.

We do not restrict the stock pools to the stocks that exist throughout all 2,516 days, since at the time of investment an investor cannot know whether a stock will be delisted in the future. We also allow each stock's location in the tensor to vary over time, because size (and other characteristics) naturally evolve. Accordingly, our focus is not on how the weight of any fixed individual stock changes over time, but rather on how portfolio allocations vary across sectors and across characteristic of stocks.

The Sharpe ratios of individual stocks over the full sample period are presented in Figure 1. The heatmap reveals that a single sector does not consistently yield a higher Sharpe ratio than others, nor does a particular size group guarantee superior performance, instead, we observe substantial cross-sectional heterogeneity across both the sector and size dimensions. Furthermore, the Sharpe ratios exhibit a distinct block-like distribution; for example, higher Sharpe ratio are predominantly clustered among larger-size stocks within Sectors 3 to 6. This structural pattern suggests that the performance of an individual asset is not driven by a single dimension, but can be determined by the joint interaction of multiple modes (i.e., size and sector).

**Remark 11** The tensor structure employed here is a deliberate simplification, focusing on sector and firm size to illustrate the methodological advantages of our framework. However, this structure introduces specific empirical constraints. First, because stocks are dynamically re-sorted into tensor positions each period based on their characteristics (e.g., size), turnover and transaction costs calculated at the *position level* would misleadingly understate true *stock-level* trading frictions. Second, while market dependence may involve higher-order modes beyond these two dimensions, we focus on this 3-way representation to highlight the structural gains over vector-based models. Consequently, we exclude turnover-related metrics from our empirical analysis and leave the extension to higher-order tensors for future research. □

## 5.2 Description of Empirical Design

To evaluate the empirical viability of each model, we assess its out-of-sample (OOS) performance using a rolling-window estimation framework. Given the inherent noise in stock returns, relying

exclusively on in-sample (IS) fit offers limited insight into real-time investment decisions and is insufficient for validating a model's practical utility. Consequently, we implement three distinct IS/OOS sample splits: $\{504, 2012\}$, $\{1006, 1510\}$, and $\{1762, 754\}$ trading days, which correspond to IS training windows of approximately two, four, and seven years, respectively.

Unrealistic return or risk targets can force the optimization into corner solutions and may cause the algorithm to converge to a saddle point, thereby sacrificing diversification benefits. Following Lee and Seregina (2023), we set the daily return target to 0.0378% (approximately 10% annualized return). However, departing from their static risk setting, we dynamically calibrate the risk target to the S&P 500 volatility within each rolling window. This aligns with our rolling estimation design, reflecting how investors continuously adapt their risk constraints to time-varying market conditions.

### 5.3 Empirical Results

We compare the OOS performance of six portfolio constructions in terms of OOS Sharpe ratio as Section 4.2: *Tensor*, *Vector*, *TensorGL*, *VectorGL*, *TensorFGL*, and *VectorFGL*. Their OOS mean *excess* returns, realized risks, and Sharpe ratios are reported in Table 2. Overall, the results align with our expectations. By exploiting the multi-dimensional structure of the return tensor, tensor-based portfolios effectively reduce the number of parameters to be estimated, which in turn improves the stability of second-moment estimation and delivers higher OOS Sharpe ratio. This pattern is also consistent with the broader literature on high-dimensional covariance/precision estimation: when the dimension is large relative to the sample size, precision matrix estimates can be unstable, and shrinkage-type procedures can improve the OOS performance of portfolio.

As shown in Panel A, when constructing portfolios by GMV and MWC, tensor portfolios consistently outperform all vector-based benchmarks in terms of OOS Sharpe ratios across every rolling window. This indicates that tensor portfolios deliver higher realized returns per unit of risk in high-dimensional settings, even without relying on regularization or factor modeling. Consequently, these findings underscore two key advantages of the tensor portfolios: (i) by exploiting the intrinsic multi-way structure of asset returns, it significantly reduces the effective dimensionality, thereby improving the stability of covariance estimation; and (ii) unlike conventional vector-based portfolios, tensor portfolios naturally allocate capital across multiple dimensions (e.g., sector and firm size),

yielding a more favorable risk-return trade-off.

When the true DGP is unknown, the primary risk of employing a tensor structure lies in potential structural misspecification. In contrast, vector-based portfolios do not impose any additional structural restrictions; as discussed in Section 4, any multi-way data can simply be flattened into a vector. In small samples, the vectorized approach suffers from noisy covariance estimation, which frequently induces extreme, unstable portfolio weights and consequently poor OOS performance. The tensor approach drastically reduces the number of free parameters, so it acts as a powerful dimensional reduction mechanism, thereby mitigating estimation uncertainty and stabilizing the resulting weights. However, as the sample size increases, covariance estimation in the vectorized approach becomes inherently more stable. Therefore, when the sample size is large, if the true DGP deviates substantially from the assumed tensor structure, the bias introduced by structural misspecification begins to dominate, ultimately degrading the OOS performance of tensor-based portfolios.

The less pronounced advantage of the *Tensor* under MRC corroborates this point. With only a two-year IS window, i.e., IS = 504, the tensor portfolio yields a modest OOS Sharpe ratio, yet the vector-based approach can generate a negative OOS Sharpe ratio. As IS increases, however, the OOS Sharpe ratio of the *Vector* improves markedly, even marginally surpassing that of the Tensor when the IS is 1,006. Meanwhile, a similar monotonic improvement in Sharpe ratio of *Vector* can also be observed under GMV and MWC, although the magnitude is much smaller than under MRC. These suggest that over a four-year training period, the underlying data structure may deviate further from the assumed tensor structure, which induces more severe structural misspecification. At the same time, the larger sample size reduces the estimation error of *Vector*. Therefore, the improvement in *Vector*'s OOS performance is not surprising.

Interestingly, although the same precision matrices are employed in all of GMV, MWC, and MRC, *Tensor* consistently yield significantly higher OOS Sharpe ratios than their vectorized counterparts under the GMV and MWC schemes, whereas a different pattern emerges under the MRC. We also observe that OOS Sharpe ratios are generally lower under MRC than under GMV and MWC. We attribute this discrepancy to the scaling effect inherent in the constraints. Both GMV and MWC operate under the full-investment constraint, where their weight normalizations are pinned down by terms such as $\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m$ and $\boldsymbol{\iota}_n'\boldsymbol{\Theta}_c\boldsymbol{\iota}_n$, respectively. This normalization effectively

anchors the scale of the estimated precision matrices, thereby tending to partially offset common multiplicative errors or structural biases in covariance estimation. In contrast, MRC does not rely on a full-investment normalization. Its scaling is determined strictly by the risk constraint. Consequently, any estimation error stemming from potential structural misspecification can be more directly amplified or distorted through this risk-based rescaling.

In line with the discussion above, the OOS Sharpe gap reflects a trade-off between two forces: the benefit of more stable estimation induced by lower dimensionality and the potential loss from structure misspecification. Moreover, these benefits and costs can be amplified or attenuated by the scaling implicit in different weighting schemes. Once Shrinkage methods (GLASSO and TLASSO) is applied to regularize the precision matrix, this regularization stabilizes the estimated precision matrix and the resulting portfolio weights, thereby narrowing the difference attributable purely to dimensionality. Consequently, the inherent low-dimensional advantage of imposing a tensor structure is greatly attenuated, and we can expect to observe that the performance gap between tensor-based and vector-based models becomes much less pronounced under regularization.

Evidence in Panel B supports this interpretation. When sample size is relatively small, i.e., IS=504, *TensorGL* still substantially outperforms *VectorGL* across all three weighting schemes. However, as IS increases to 1006 and 1762, the larger sample size reduces estimation error in the vector-based approach, and the dimensionality advantage of imposing a tensor structure is increasingly offset by the disadvantage from potential structure misspecification. As a result, in large samples, *VectorGL* can surpass the tensor-based portfolio in terms of OOS Sharpe. Even when the tensor-based Sharpe remains higher, the difference is generally much smaller than what is observed in Panel A.

In Panel C, in which we compare two weighting schemes based on factor structure, the result implies, in terms of OOS Sharpe ratio, *TensorFGL* delivers an overwhelming advantage over *VectorFGL*. This suggests that, after accounting for the common factor structure, tensor-based models exploit the information contained in the idiosyncratic component more effectively. A possible explanation is after removing the common factor, the idiosyncratic component is closer to a separable covariance structure. In contrast, when no factor structure is modeled - i.e., the number of factors is set to zero as in Panel B - the overall data structure deviates further from separability.

There remains substantial room to further improve our *TensorFGL* approach. Although factor

structures are widely viewed as an effective form of dimension reduction, and *TensorFGL* further incorporate regularization on the idiosyncratic component to extract additional information, factor models remain unsupervised and can be highly sensitive to tuning choices (e.g., the number of factors and penalty levels).

# 6 Conclusion

We propose a new portfolio-construction framework - *Tensor Portfolio* - that more thoroughly exploits the multi-dimensional structural information embedded in return data. Based on three mainstream weighting schemes - Global Minimum Variance (GMV), Markowitz Weight-Constrained (MWC), and Markowitz Risk-Constrained (MRC) - we develop three portfolio construction methods tailored to return data with a tensor structure. Firstly, we impose a separability assumption on the covariance matrices, representing the return covariance as a Kronecker product of sector and firm-characteristic covariance matrices. Secondly, we apply Tensor Graphical LASSO proposed by Lyu et. al. (2020) to capture the sparsity in the precision matrix along each mode. Finally, building on the Factor Graphical LASSO framework of Lee and Seregina (2023), we develop Tensor Factor Graphical LASSO which combining multi-way information pooling with factor-based decomposition and regularization of the idiosyncratic component.

We use Monte Carlo simulation to illustrate whenever the data exhibit a tensor structure, tensor-based portfolios consistently achieve significantly higher OOS Sharpe ratios than their vectorized counterparts. Meanwhile, in empirical application, using U.S. individual-stock returns, we construct a return tensor by sorting stocks jointly by firm size and sector. Specifically, we use daily data from January 1, 2010 to December 31, 2019 to build a $T \times 30 \times 11$ tensor. As a result, relative to vector-based portfolios, tensor portfolios tend to achieve higher OOS Sharpe ratios, which thus confirm the advantages of exploiting the tensor structure.

## References


Brownlees, C., E. Nualart, and Y. Sun. 2018. "Realized Networks". *Journal of Applied Econometrics*. 33: 986–1006.

Cai, X., X. Kong, X. Wu, and P. Zhao. 2025. "Matrix-Factor-Augmented Regression". *Journal of Business and Economic Statistics*. 0: 1–13.

Callot, L., M. Caner, A. O. Onder, and E. Ulaşan. 2021. "A Nodewise Regression Approach to Estimating Large Portfolios". *Journal of Business and Economic Statistics*. 39: 520–531.

Castellano, R. and R. Cerqueti. 2014. "Mean–Variance Portfolio Selection in Presence of Infrequently Traded Stocks". *European Journal of Operational Research*. 234: 442–449.

Chen, B., E. Chen, S. Bolivar, and R. Chen. 2024. "Time-Varying Matrix Factor Models". URL: https://arxiv.org/abs/2404.01546.

Chen, B., Y. Han, and Q. Yu. 2024. "Estimation and Inference for CP Tensor Factor Models". URL: https://arxiv.org/abs/2406.17278.

Chen, E. and J. Fan. 2023. "Statistical Inference for High-Dimensional Matrix-Variate Factor Models". *Journal of the American Statistical Association*. 118: 1038–1055.

Chen, E., R.S. Tsay, and R. Chen. 2020. "Constrained Factor Models for High-Dimensional Matrix-Variate Time Series". *Journal of the American Statistical Association*. 115: 775–793.

Chen, R., H. Xiao, and D. Yang. 2021. "Autoregressive Models for Matrix-Valued Time Series". *Journal of Econometrics*. 222: 539–560.

Chen, R., D. Yang, and C.-H. Zhang. 2022. "Factor Models for High-Dimensional Tensor Time Series". *Journal of the American Statistical Association*. 117: 94–116.

DeMiguel, V., L. Garlappi, F.J. Nogales, and R. Uppal. 2009. "A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms". *Management Science*. 55: 798–812.

Ding, Y., Y. Li, and X. Zheng. 2021. "High-Dimensional Minimum Variance Portfolio Estimation under Statistical Factor Models". *Journal of Econometrics*. 222: 502–515.

Dutilleul, P. 1999. "The MLE Algorithm for the Matrix Normal Distribution". *Journal of Statistical Computation and Simulation*. 64: 105–123.

Fama, E.F. and K.R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds". *Journal of Financial Economics*. 33: 3–56.

Fan, J., Zhang J., and K. Yu. 2012. "Vast Portfolio Selection with Gross-Exposure Constraints". *Journal of the American Statistical Association*. 107: 592–606.

Fan, J., Y. Ke, and K. Wang. 2020. "Factor-Adjusted Regularized Model Selection (FarmSelect)". *Journal of Econometrics*. 216: 71–85.

Fan, J., Y. Liao, and M. Mincheva. 2011. "High-Dimensional Covariance Matrix Estimation in Approximate Factor Models". *The Annals of Statistics*. 39: 3320–3356.

Fan, J., Y. Liao, and M. Mincheva. 2013. "Large Covariance Estimation by Thresholding Principal Orthogonal Complements". *Journal of the Royal Statistical Society, Series B*. 75: 603–680.

Fan, Q., R. Wu, Y. Yang, and W. Zhong. 2024. "Time-Varying Minimum Variance Portfolio". *Journal of Econometrics*. 239: 105339.

Giannone, D., C. De Mol, I. Daubechies, and J. Brodie, 2007. "Sparse and Stable Markowitz Portfolios". CEPR Discussion Paper 6474.

Goto, S. and Y. Xu. 2015. "Improving Mean-Variance Optimization through Sparse Hedging Restrictions". *Journal of Financial and Quantitative Analysis*. 50: 1415–1441.

Han, Y., R. Chen, D. Yang, and C.-H. Zhang. 2024. "Tensor Factor Model Estimation by Iterative Projection". URL: https://arxiv.org/abs/2006.02611.

Han, Y., R. Chen, C.-H. Zhang, and Q. Yao. 2024. "Simultaneous Decorrelation of Matrix Time Series". *Journal of the American Statistical Association*. 119: 957–969.

Han, Y., D. Yang, C.-H. Zhang, and R. Chen. 2024. "CP Factor Model for Dynamic Tensors". *Journal of the Royal Statistical Society Series B*. 86: 1383–1413.

Jagannathan, R. and T. Ma. 2003. "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps". *The Journal of Finance*. 58: 1651–1683.

Kozak, S., S. Nagel, and S. Santosh. 2018. "Interpreting Factor Models". *The Journal of Finance*. 73: 1183–1223.

Lam, C. and Q. Yao. 2012. "Factor Modeling for High-Dimensional Time Series: Inference for the Number of Factors". *The Annals of Statistics*. 40: 694–726.

Ledoit, O. and M. Wolf. 2003. "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection". *Journal of Empirical Finance*. 10: 603–621.

Lee, T.-H. and E. Seregina. 2023. "Optimal Portfolio Using Factor Graphical Lasso". *Journal of Financial Econometrics*. 22: 670–695.

Lyu, X., W.W. Sun, Z. Wang, H. Liu, J. Yang, and G. Cheng. 2020. "Tensor Graphical Model: Non-Convex Optimization and Statistical Inference". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 42: 2024–2037.

Markowitz, H. 1952. "Portfolio Selection". *The Journal of Finance*. 7: 77–91.

Tobin, J. 1958. "Liquidity Preference as Behavior Towards Risk". *The Review of Economic Studies*. 25: 65–86.

Torri, G., R. Giacometti, and S. Paterlini. 2019. "Sparse Precision Matrices for Minimum Variance Portfolios". *Computational Management Science*. 16: 375–400.

Tseng, P. 2001. "Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization". *Journal of Optimization Theory and Applications*. 109: 475–494.

Wang, D., X. Liu, and R. Chen. 2019. "Factor Models for Matrix-Valued High-Dimensional Time Series". *Journal of Econometrics*. 208: 231–248.

Yu, L., Y. He, X. Kong, and X. Zhang. 2022. "Projected Estimation for Large-Dimensional Matrix Factor Models". *Journal of Econometrics*. 229: 201–217.

Yuan, X., W. Yu, Z. Yin, and G. Wang. 2020. "Improved Large Dynamic Covariance Matrix Estimation with Graphical Lasso and Its Application in Portfolio Selection". *IEEE Access*. 8: 189179–189188.

Zhao, T., H. Liu, K. Roeder, J. Lafferty, and L. Wasserman. 2012. "The HUGE Package for High-Dimensional Undirected Graph Estimation in R". *Journal of Machine Learning Research*. 13: 1059–1062.

Table 1: Simulation Results

| | Normal $\mathbf{Z}_t$ | | | Heavy-Tailed $\mathbf{Z}_t$ | | |
|---|---|---|---|---|---|---|
| Method | GMV | MWC | MRC | GMV | MWC | MRC |
| *Panel A: DGP 1* | | | | | | |
| Tensor | 6.493 | 6.494 | 5.326 | 6.381 | 6.378 | 5.306 |
| Vector | 2.198 | 2.205 | 0.828 | 2.156 | 2.170 | 0.837 |
| $z$ | $78.706^{***}$ | $79.028^{***}$ | $61.600^{***}$ | $68.728^{***}$ | $68.999^{***}$ | $60.652^{***}$ |
| *Panel B: DGP 2* | | | | | | |
| TensorGL | 26.404 | 26.352 | 26.012 | 26.089 | 26.006 | 25.700 |
| VectorGL | 25.492 | 25.492 | 14.079 | 24.617 | 24.616 | 13.137 |
| $z$ | $36.925^{***}$ | $33.401^{***}$ | $95.502^{***}$ | $32.107^{***}$ | $21.777^{***}$ | $84.115^{***}$ |
| *Panel C: DGP 3* | | | | | | |
| TensorFGL | 11.639 | 11.639 | 10.472 | 11.552 | 11.553 | 10.429 |
| VectorFGL | 11.536 | 11.536 | 4.814 | 11.203 | 11.203 | 4.936 |
| $z$ | $24.248^{***}$ | $23.928^{***}$ | $86.487^{***}$ | $28.352^{***}$ | $28.390^{***}$ | $83.511^{***}$ |

(1) Each entry reports the average OOS Sharpe ratio across 100 Monte Carlo simulations.
(2) All numbers are rounded to three decimals for presentation; thus, some entries may appear identical after rounding even though the underlying averages differ slightly. In Panel B, under normal $\mathbf{Z}_t$, $\text{VectorGL}_{GMV} = 25.4920$ vs. $\text{VectorGL}_{MWC} = 25.4924$. In Panel C, under normal $\mathbf{Z}_t$, $\text{TensorFGL}_{GMV} = 11.6387$ vs. $\text{TensorFGL}_{MWC} = 11.6390$, and $\text{VectorFGL}_{GMV} = 11.5361$ vs. $\text{VectorFGL}_{MWC} = 11.5362$; under Heavy-Tailed $\mathbf{Z}_t$, and $\text{VectorFGL}_{GMV} = 11.20345$ vs. $\text{VectorFGL}_{MWC} = 11.20338$.
(3) In the $z$-rows, ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. (4) The Sharpe ratio is annualized using 252 trading days, i.e., the realized Sharpe ratio times $\sqrt{252}$.

Table 2: OOS *Excess* Return, Risk, and Annualized Sharpe, 1/1/2010 - 12/31/2019

| | | | GMV | | | MWC | | | MRC | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| IS | OOS | Method | Return | Risk | SR | Return | Risk | SR | Return | Risk | SR |
| *Panel A: Tensor, Vector* | | | | | | | | | | | |
| 504 | 2012 | Tensor | **6.050e-04** | **6.997e-03** | **1.371** | **6.040e-04** | **7.067e-03** | **1.358** | **3.500e-05** | **4.743e-03** | **0.119** |
| | | Vector | 4.881e-04 | 9.503e-03 | 0.815 | 5.018e-04 | 9.479e-03 | 0.840 | -5.753e-05 | 1.289e-02 | -0.071 |
| 1006 | 1510 | Tensor | **5.490e-04** | **7.132e-03** | **1.222** | **5.220e-04** | **7.295e-03** | **1.137** | 2.670e-04 | **5.444e-03** | 0.778 |
| | | Vector | 4.480e-04 | 7.575e-03 | 0.940 | 4.400e-04 | 7.571e-03 | 0.923 | **3.710e-04** | 6.408e-03 | **0.919** |
| 1762 | 754 | Tensor | **5.940e-04** | **6.624e-03** | **1.422** | **5.710e-04** | **6.758e-03** | **1.341** | **5.820e-04** | 6.484e-03 | **1.426** |
| | | Vector | 4.090e-04 | 6.764e-03 | 0.960 | 4.000e-04 | 6.767e-03 | 0.937 | 3.510e-04 | **5.254e-03** | 1.061 |
| *Panel B: TensorGL, VectorGL* | | | | | | | | | | | |
| 504 | 2012 | TensorGL | **7.140e-04** | 8.017e-03 | **1.413** | **6.940e-04** | 8.113e-03 | **1.357** | 1.100e-05 | **1.140e-04** | **1.490** |
| | | VectorGL | 5.250e-04 | **6.298e-03** | 1.324 | 5.160e-04 | **6.285e-03** | 1.304 | **7.230e-04** | 8.804e-03 | 1.304 |
| 1006 | 1510 | TensorGL | **5.650e-04** | 8.090e-03 | 1.109 | **5.370e-04** | 8.348e-03 | 1.020 | 9.000e-06 | **1.220e-04** | **1.114** |
| | | VectorGL | 5.420e-04 | **6.600e-03** | **1.303** | 5.300e-04 | **6.593e-03** | **1.276** | **3.520e-04** | 5.626e-03 | 0.993 |
| 1762 | 754 | TensorGL | **5.940e-04** | 7.580e-03 | **1.244** | **5.660e-04** | 7.695e-03 | **1.167** | 1.000e-05 | **1.160e-04** | 1.328 |
| | | VectorGL | 4.740e-04 | **6.297e-03** | 1.196 | 4.550e-04 | **6.328e-03** | 1.143 | **3.870e-04** | 4.429e-03 | **1.389** |
| *Panel C: TensorFGL, VectorFGL* | | | | | | | | | | | |
| 504 | 2012 | TensorFGL | **7.190e-04** | **8.052e-03** | **1.417** | **6.980e-04** | **8.182e-03** | **1.354** | 1.000e-05 | **1.120e-04** | **1.484** |
| | | VectorFGL | 5.160e-04 | 8.447e-03 | 0.970 | 5.230e-04 | 8.427e-03 | 0.986 | **2.730e-04** | 6.976e-03 | 0.621 |
| 1006 | 1510 | TensorFGL | **5.640e-04** | **8.125e-03** | **1.101** | **5.300e-04** | **8.483e-03** | **0.991** | 9.000e-06 | **1.210e-04** | **1.113** |
| | | VectorFGL | 2.560e-04 | 9.638e-03 | 0.422 | 2.660e-04 | 9.631e-03 | 0.438 | **2.090e-04** | 7.148e-03 | 0.465 |
| 1762 | 754 | TensorFGL | **5.880e-04** | **7.630e-03** | **1.223** | **5.720e-04** | **7.830e-03** | **1.160** | 1.000e-05 | **1.170e-04** | **1.306** |
| | | VectorFGL | 2.230e-04 | 9.903e-03 | 0.358 | 2.110e-04 | 9.905e-03 | 0.338 | **1.840e-04** | 5.392e-03 | 0.541 |

(1) Bold numbers indicate the highest mean excess return, the lowest risk, and the highest Sharpe ratio (SR) in OOS.
(2) The Sharpe ratio is annualized using 252 trading days, i.e., the realized Sharpe ratio times $\sqrt{252}$.

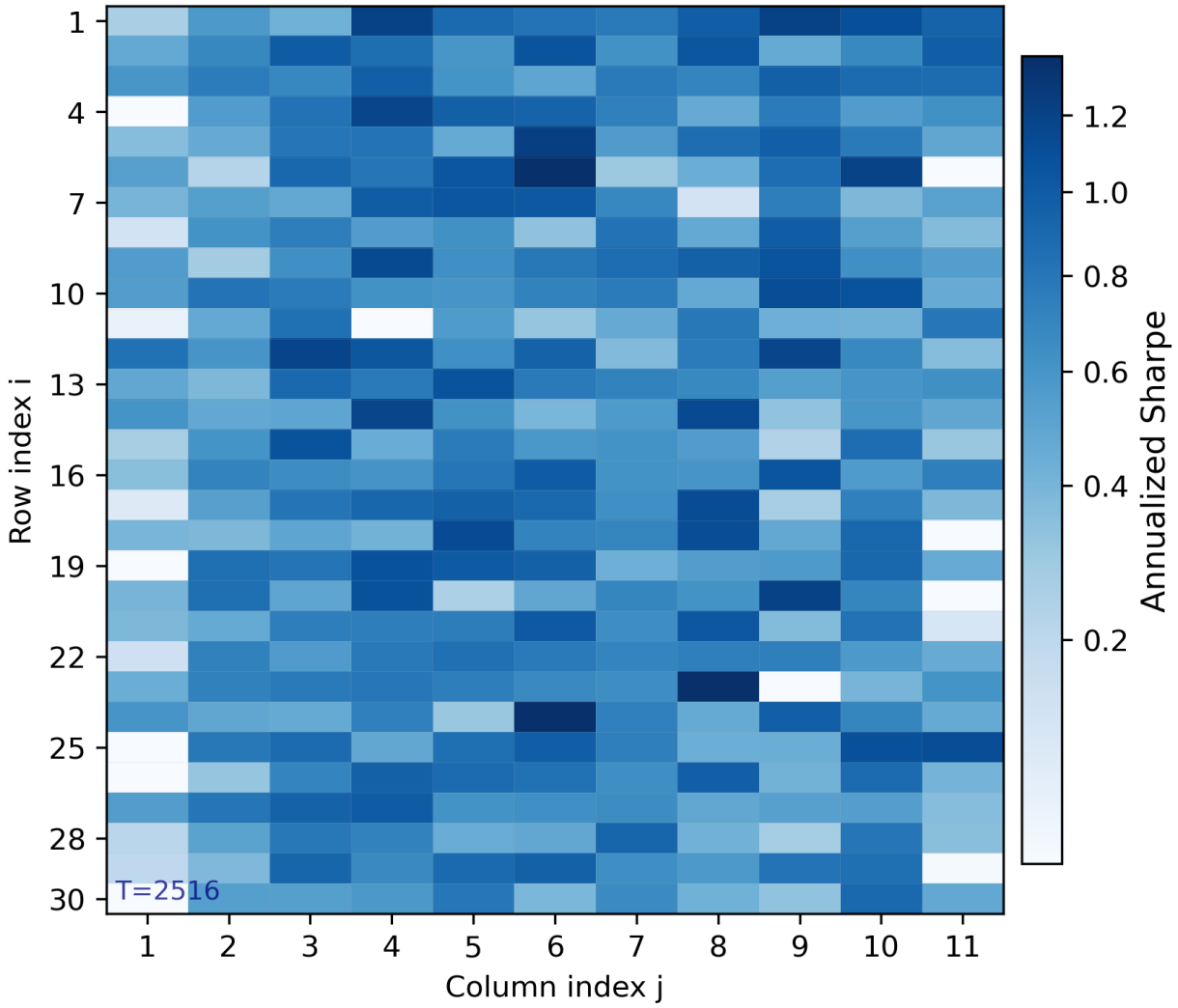


Figure 1: Individual Stock Sharpe Ratios, 1/1/2010–12/31/2019

# Appendix

Appendix A presents the derivations of tensor portfolio weights in Section 2. Appendix B shows the proofs about Theorem, Lemma, and Propositions in Section 3.

## A Derivation of Tensor Portfolio Weights

In this section, we provide the derivation for the updating of weighting vector in one mode when keeping the other one fixed. Note that we only show how to update $\boldsymbol{\omega}_1$ when fixing $\boldsymbol{\omega}_2$, since the $\boldsymbol{\omega}_2$ can be solved in the same way.

### A.1 Tensor Portfolio with Separable Covariance Matrices on De-meaned Return

We show the process to derive the portfolio weights on GMV, MWC, and MRC with separability assumption, as in Section 2.2 and 2.3.

**GMV** Under separable covariance structure shown in Assumption 1, we can express the risk as $(\boldsymbol{\omega}_2 \otimes \boldsymbol{\omega}_1)\,\boldsymbol{\Sigma}\,(\boldsymbol{\omega}_2 \otimes \boldsymbol{\omega}_1)' = (\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2')\,(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1')$. Hence, we can construct the following Lagrangian:

$$\mathcal{L} = \frac{1}{2}\left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1'\right) - \zeta_1\left(\boldsymbol{\omega}_1\boldsymbol{\iota}_m - 1\right) - \zeta_2\left(\boldsymbol{\omega}_2\boldsymbol{\iota}_n - 1\right). \tag{40}$$

Denote $\delta = \boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'$, we then have the first order condition w.r.t. $\boldsymbol{\omega}_1$ as $\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right) = \zeta_1\boldsymbol{\iota}_m'$, and thus $\zeta_1 = \delta/\left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m\right)$. Finally, replacing $\zeta_1$ in the first order condition, we are able to show the closed-form solution of $\boldsymbol{\omega}_1^*$ as (41):

$$\boldsymbol{\omega}_1^* = \frac{\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r}{\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m}. \tag{41}$$

**MWC** The Lagrangian with full investment and return constraints is constructed as follows:

$$\mathcal{L} = \frac{1}{2}\left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1'\right) - \lambda\left(\boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2' - \mu_{target}\right) - \zeta_1\left(\boldsymbol{\omega}_1\boldsymbol{\iota}_m - 1\right) - \zeta_2\left(\boldsymbol{\omega}_2\boldsymbol{\iota}_n - 1\right), \tag{42}$$

We derive the first order condition w.r.t. $\boldsymbol{\omega}_1$ as $\delta\boldsymbol{\omega}_1 = \lambda\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_r + \zeta_1\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r$. Using the full investment

and return constraints, we obtain a linear system as (43), and the solved $\lambda$ and $\zeta_1$ are shown in (44):

$$\begin{bmatrix} \left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\left(\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right) & \boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\left(\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right) \\ \left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\boldsymbol{\iota}_m & \boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m \end{bmatrix} \begin{bmatrix} \lambda \\ \zeta_1 \end{bmatrix} = \begin{bmatrix} \delta\mu_{target} \\ \delta \end{bmatrix}, \tag{43}$$

$$\begin{aligned} \lambda &= \frac{\left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m\right)\delta\mu_{target} - \left[\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\left(\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right)\right]\delta}{\left[\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\left(\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right)\right]\left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m\right) - \left[\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\boldsymbol{\iota}_m\right]^2}, \\ \zeta_1 &= \frac{\left[\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\left(\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right)\right]\delta - \left[\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\boldsymbol{\iota}_m\right]\delta\mu_{target}}{\left[\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\left(\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right)\right]\left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m\right) - \left[\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\boldsymbol{\iota}_m\right]^2}. \end{aligned} \tag{44}$$

Replacing $\lambda$ and $\zeta_1$ in the first order condition, the updating of $\boldsymbol{\omega}_1$ can be derived as

$$\boldsymbol{\omega}_1 \leftarrow (1-a_r)\,\frac{\boldsymbol{\iota}_m'\,\boldsymbol{\Theta}_r}{\boldsymbol{\iota}_m'\,\boldsymbol{\Theta}_r\boldsymbol{\iota}_m} \;+\; a_r\,\frac{\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r}{\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\,\boldsymbol{\iota}_m}, \tag{45}$$

where the weight $a_r$ is defined as (46):

$$a_r = \frac{\left(\mu_{target}\,\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m - \boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\,\bar{\mathbf{R}}\,\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\,\bar{\mathbf{R}}\,\boldsymbol{\omega}_2'\right)}{\left[\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r(\bar{\mathbf{R}}\,\boldsymbol{\omega}_2')\right]\left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\boldsymbol{\iota}_m\right) - \left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_r\,\bar{\mathbf{R}}\,\boldsymbol{\omega}_2'\right)^2}\quad . \tag{46}$$

**MRC** Begin with maximizing mean return with binding risk constraint, we have the following Lagrangian:

$$\mathcal{L} = \boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2' - \frac{\lambda}{2}\left[\left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1'\right) - \sigma_{target}^2\right]. \tag{47}$$

The first order condition w.r.t. $\boldsymbol{\omega}_1$ implies that $\boldsymbol{\omega}_1 = \frac{1}{\delta\lambda}\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_r\right)$. We then can rewrite risk constraint such that

$$\left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1'\right)\left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right) = \left(\frac{1}{\delta\lambda}\right)^2\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_r\right)\boldsymbol{\Sigma}_r\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_r\right)'\delta = \sigma_{target}^2, \tag{48}$$

and solve the inverse of $\lambda$ as

$$\frac{1}{\lambda} = \sigma_{target}\sqrt{\frac{\delta}{\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_r\bar{\mathbf{R}}\boldsymbol{\omega}_2'}}. \tag{49}$$

Replacing $\lambda$ in first order condition, we can show the updating for $\boldsymbol{\omega}_1$ by (50).

$$\boldsymbol{\omega}_1 \leftarrow \frac{\sigma_{target}\sqrt{\left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right)^{-1}}}{\sqrt{\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\left(\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right)}}\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r. \tag{50}$$

When we minimize risk under binding return constraint, the Lagrangian becomes

$$\mathcal{L} = \frac{1}{2}\left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_c\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\omega}_1\boldsymbol{\Sigma}_r\boldsymbol{\omega}_1'\right) - \lambda\left(\boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2' - \mu_{target}\right). \tag{51}$$

Our first order condition w.r.t. $\boldsymbol{\omega}_1$ is derived as $\delta\boldsymbol{\omega}_1 = \lambda\boldsymbol{\omega}_2\bar{\mathbf{R}}\boldsymbol{\Theta}_r$. Imposing the return constraint, we have $\lambda = \delta\mu_{target}/\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_r\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right)$. Finally, replacing $\lambda$ in the first order condition, we solve the updating for $\boldsymbol{\omega}_1$ as

$$\boldsymbol{\omega}_1 \leftarrow \frac{\mu_{target}}{\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r\left(\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right)}\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_r. \tag{52}$$

## A.2 Tensor Portfolio with Factor Structure and Separable Covariance on Idiosyncratic Components

We derive the portfolio weights for GMV, MWC, and MRC in Section 2.4. Since the row- and column-weighted covariance matrices under the factor structure coincide with those in the no-factor case, it suffices to derive the covariance expressions without imposing a factor structure. We show their equivalent as follows. Imposing the factor structure on $\mathbf{E}_t$, then the row-weighted sample covariance matrices can be shown as

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_2 &= \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{E}_t\boldsymbol{\omega}_2'\boldsymbol{\omega}_2\mathbf{E}_t'\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbf{A}\left(\mathbf{F}_t\mathbf{B}'\boldsymbol{\omega}_2'\boldsymbol{\omega}_2\mathbf{B}\mathbf{F}_t'\right)\mathbf{A}' + \frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_t\boldsymbol{\omega}_2'\boldsymbol{\omega}_2\mathbf{U}_t' \\
&= \mathbf{A}\tilde{\boldsymbol{\Sigma}}_{f,r}\mathbf{A}' + \frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_t\boldsymbol{\omega}_2'\boldsymbol{\omega}_2\mathbf{U}_t' \\
&= \mathbf{A}\tilde{\boldsymbol{\Sigma}}_{f,r}\mathbf{A}' + \tilde{\boldsymbol{\Sigma}}_{u,r}.
\end{aligned} \tag{53}$$

The last equality holds by using the property that $\mathbf{U}_t\boldsymbol{\omega}_2'\boldsymbol{\omega}_2\mathbf{U}_t' = \mathbf{I}_m\mathbf{U}_t\boldsymbol{\omega}_2'\boldsymbol{\omega}_2\mathbf{U}_t'\mathbf{I}_m'$, where $\mathbf{I}_m$ is an Identity matrix with rank $m$, then we have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_t\boldsymbol{\omega}_2'\boldsymbol{\omega}_2\mathbf{U}_t' &= \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{\omega}_2\otimes\mathbf{I}_m\right)\mathrm{vec}(\mathbf{U}_t)\mathrm{vec}(\mathbf{U}_t)'\left(\boldsymbol{\omega}_2\otimes\mathbf{I}_m\right)' \\ &= \left(\boldsymbol{\omega}_2\otimes\mathbf{I}_m\right)\left(\boldsymbol{\Sigma}_{u,c}\otimes\boldsymbol{\Sigma}_{u,r}\right)\left(\boldsymbol{\omega}_2\otimes\mathbf{I}_m\right)' \\ &= \left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_{u,c}\boldsymbol{\omega}_2'\right)\left(\mathbf{I}_m\boldsymbol{\Sigma}_{u,r}\mathbf{I}_m'\right) \\ &= \left(\boldsymbol{\omega}_2\boldsymbol{\Sigma}_{u,c}\boldsymbol{\omega}_2'\right)\boldsymbol{\Sigma}_{u,r} \\ &= \tilde{\boldsymbol{\Sigma}}_{u,r}.\end{aligned} \tag{54}$$

**GMV** Start from the optimization problem of minimizing portfolio risk subject to the full-investment constraints:

$$\min_{\boldsymbol{\omega}_1,\boldsymbol{\omega}_2}\quad \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{\omega}_1\mathbf{E}_t\boldsymbol{\omega}_2'\right)^2,\quad \text{s.t.}\quad \boldsymbol{\omega}_1\boldsymbol{\iota}_m = 1,\quad \boldsymbol{\omega}_2\boldsymbol{\iota}_n = 1. \tag{55}$$

Begin with fixing $\boldsymbol{\omega}_1$ and updating $\boldsymbol{\omega}_2$, derive the Lagrange as

$$\mathcal{L} = \frac{1}{2T}\sum_{t=1}^{T}\left(\boldsymbol{\omega}_1\mathbf{E}_t\boldsymbol{\omega}_2'\right)^2 - \zeta_1\left(\boldsymbol{\omega}_1\boldsymbol{\iota}_m - 1\right) - \zeta_2\left(\boldsymbol{\omega}_2\boldsymbol{\iota}_n - 1\right), \tag{56}$$

and the first order condition w.r.t. $\boldsymbol{\omega}_1$ implies that $\boldsymbol{\omega}_1\boldsymbol{\Sigma}_2 = \zeta_1\boldsymbol{\iota}_m'$. By imposing the full investment constraint, we can solve $\zeta_1 = \left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m\right)^{-1}$. Replacing $\zeta_1$ in the first order condition, we can solve the updating of $\boldsymbol{\omega}_1$ when we fix $\boldsymbol{\omega}_2$ as (57):

$$\boldsymbol{\omega}_1 \leftarrow \frac{\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2}{\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m}. \tag{57}$$

**MWC** Different with GMV, we impose a return constraint in addition to the full-investment constraints, as shown below.

$$\min_{\boldsymbol{\omega}_1,\boldsymbol{\omega}_2}\quad \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{\omega}_1\mathbf{E}_t\boldsymbol{\omega}_2'\right)^2,\quad \text{s.t.}\quad \boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2' \geq \mu_{target},\quad \boldsymbol{\omega}_1\boldsymbol{\iota}_m = 1,\quad \boldsymbol{\omega}_2\boldsymbol{\iota}_n = 1. \tag{58}$$

When the return constraint is binding, fixing $\boldsymbol{\omega}_2$ and updating $\boldsymbol{\omega}_1$, the optimization problem yields

MWC solutions by solving the Lagrangian below

$$\mathcal{L} = \frac{1}{2T}\sum_{t=1}^{T}\left(\boldsymbol{\omega}_1\mathbf{E}_t\boldsymbol{\omega}_2'\right)^2 - \lambda\left(\boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2' - \mu_{target}\right) - \zeta_1(\boldsymbol{\omega}_1\boldsymbol{\iota}_m - 1) - \zeta_2(\boldsymbol{\omega}_2\boldsymbol{\iota}_n - 1). \tag{59}$$

We obtain the first order condition w.r.t. $\boldsymbol{\omega}_1$ such that $\boldsymbol{\omega}_1 = \zeta_1\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2 + \lambda\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2$. Imposing the full-investment and return constraints, we have a linear system shown as (60), in which the solutions of $\lambda$, $\zeta_1$ can be derived as (61)

$$\begin{bmatrix} \boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m & \boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m \\ \boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\bar{\mathbf{R}}\boldsymbol{\omega}_2' & \boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2\bar{\mathbf{R}}\boldsymbol{\omega}_2' \end{bmatrix}\begin{bmatrix} \zeta_1 \\ \lambda \end{bmatrix} = \begin{bmatrix} 1 \\ \mu_{target} \end{bmatrix}, \tag{60}$$

$$\begin{aligned} \zeta_1 &= \frac{\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right) - \mu_{target}\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m\right)}{\left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m\right)\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right) - \left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m\right)^2}, \\ \lambda &= \frac{\mu_{target}\left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m\right) - \left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m\right)}{\left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m\right)\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right) - \left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m\right)^2}. \end{aligned} \tag{61}$$

Replacing $\zeta_1$ and $\lambda$ in the first order condition, we can solve the updating for $\boldsymbol{\omega}_1$ as (62), and weight to balance risk-minimization and return-maximization $a_r$ is shown in (63):

$$\boldsymbol{\omega}_1 \leftarrow (1 - a_r)\frac{\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2}{\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m} + a_r\frac{\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2}{\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\,\bar{\mathbf{R}}\,\boldsymbol{\omega}_2'}, \tag{62}$$

$$a_r \;=\; \frac{\left(\mu_{target}\,\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m \;-\; \boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\,\bar{\mathbf{R}}\,\boldsymbol{\omega}_2'\right)\left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\,\bar{\mathbf{R}}\,\boldsymbol{\omega}_2'\right)}{\left[(\bar{\mathbf{R}}\,\boldsymbol{\omega}_2')'\boldsymbol{\Theta}_2(\bar{\mathbf{R}}\,\boldsymbol{\omega}_2')\right]\left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\boldsymbol{\iota}_m\right) \;-\; \left(\boldsymbol{\iota}_m'\boldsymbol{\Theta}_2\,\bar{\mathbf{R}}\,\boldsymbol{\omega}_2'\right)^2}. \tag{63}$$

**MRC** In MRC, our target is to maximize the Share ratio subject to return and risk constraints.

$$\max_{\boldsymbol{\omega}_1,\boldsymbol{\omega}_2} \quad \frac{\boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2'}{\nu}, \quad \text{s.t.} \quad \boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2' \geq \mu_{target}, \quad \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{\omega}_1\mathbf{E}_t\boldsymbol{\omega}_2'\right)^2 \leq \sigma_{target}^2. \tag{64}$$

Consider the case where we maximize our mean return and make risk constraint binding, then the updating for $\boldsymbol{\omega}_1$ can be solved from the Lagrangian shown in (65):

$$\mathcal{L} = \boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2' - \frac{\lambda}{2T}\sum_{t=1}^{T}\left[\left(\boldsymbol{\omega}_1\mathbf{E}_t\boldsymbol{\omega}_2'\right)^2 - \sigma_{target}^2\right]. \tag{65}$$

The first order condition w.r.t. $\boldsymbol{\omega}_1$ implies that $\boldsymbol{\omega}_1 = \frac{1}{\lambda}\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2$. By imposing the binding risk constraint, we have $\sigma^2_{target} = \boldsymbol{\omega}_1\boldsymbol{\Sigma}_2\boldsymbol{\omega}_1' = \frac{1}{\lambda^2}\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2\bar{\mathbf{R}}\boldsymbol{\omega}_2'$, which implies that

$$\lambda = \frac{\sqrt{\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2\bar{\mathbf{R}}\boldsymbol{\omega}_2'}}{\sigma_{target}}. \tag{66}$$

Replacing $\lambda$ in the first order condition, we can solve the updating for $\boldsymbol{\omega}_1$ in (67):

$$\boldsymbol{\omega}_1 \leftarrow \frac{\sigma_{target}}{\sqrt{\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_2\left(\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right)}}\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_2. \tag{67}$$

When the return constraint is binding, our Lagrangian becomes (68):

$$\mathcal{L} = \frac{1}{2T}\sum_{t=1}^{T}\left(\boldsymbol{\omega}_1\mathbf{E}_t\boldsymbol{\omega}_2'\right)^2 - \lambda\left(\boldsymbol{\omega}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2' - \mu_{target}\right). \tag{68}$$

The first order condition w.r.t. $\boldsymbol{\omega}_1$ implies that $\boldsymbol{\omega}_1\boldsymbol{\Sigma}_2 = \lambda\boldsymbol{\omega}_2\bar{\mathbf{R}}'$. We can impose the return constraint by the same way with MWC and GMV and solve $\lambda = \mu_{target}/\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\boldsymbol{\Theta}_2\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right)$. Replacing $\lambda$ in the first order condition, we derive the updating for $\boldsymbol{\omega}_1$ as (69):

$$\boldsymbol{\omega}_1 \leftarrow \frac{\mu_{target}}{\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_2\left(\bar{\mathbf{R}}\boldsymbol{\omega}_2'\right)}\left(\boldsymbol{\omega}_2\bar{\mathbf{R}}'\right)\boldsymbol{\Theta}_2. \tag{69}$$

# B Proof of Theorem, Lemma, and Proposition

In Appendix B, we show the proofs of Theorem, Lemma, and Proposition in Section 3.

**Lemma 1.** *The level set* $\boldsymbol{\Omega} = \{(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) : f(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) \leq f(\boldsymbol{\omega}_1^0, \boldsymbol{\omega}_2^0)\}$ *is compact, where* $\boldsymbol{\omega}_1^0$ *and* $\boldsymbol{\omega}_2^0$ *are the initial values of* $\boldsymbol{\omega}_1$ *and* $\boldsymbol{\omega}_2$.

*Proof.* A compact set is closed and bounded. The level set $\boldsymbol{\Omega}$ is bounded because in the real financial market, the weight imposed in an asset should be in a reasonable range. Hence, there exists a constant $M < \infty$ such that, at $r_{th}$ iteration, where $r \in \{1, \cdots, T\}$, $\sup_{r>0}\|\boldsymbol{\omega}_k^{(r)}\| \leq M$, for $k \in \{1, 2\}$, and boundedness follows.

In the next. we need to show $\boldsymbol{\Omega}$ is closed in both GMV, MWC, and MRC. Begin with GMV

and MWC. Suppose $\mathbf{\Omega}$ is not a closed set, then there exists a $\epsilon > 0$ such that

$$f(\boldsymbol{\omega}_1 + (0, \cdots, \epsilon, \cdots, 0), \boldsymbol{\omega}_2) > f(\boldsymbol{\omega}_1^0, \boldsymbol{\omega}_2^0), \tag{70}$$

where $(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ satisfies full investment constraints. However, since $[\boldsymbol{\omega}_1 + (0, \cdots, \epsilon, \cdots, 0)]\boldsymbol{\iota}_m = 1 + \epsilon > 1$, it implies that $\boldsymbol{\omega}_1 + (0, \cdots, \epsilon, \cdots, 0)$ violates the full investment constraint, and we can conclude that the level sets associated with GMV and MWC are closed.

In MRC, there is no full-investment constraint and with the return constraint only. We define $g(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ as a continuous and bilinear mapping in $(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ such that $g(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) := \boldsymbol{\omega}_1 \bar{\mathbf{R}} \boldsymbol{\omega}_2' = \mu_{\text{target}}$. The feasible set, $\mathcal{C}$, induced by return constraint can be expressed as the pre-image of $\mu_{target}$. Since $\mu_{target} \in \mathbb{R}^+$ and $g(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ is continuous, $\mathcal{C}$ is a closed set.

$$\mathcal{C} = \{(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) : g(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) = \mu_{\text{target}}\} = g^{-1}(\{\mu_{\text{target}}\}). \tag{71}$$

When fixing one block and updating the other, the feasible sets can be expressed as (72), which are also closed.

$$\mathcal{C}_1(\boldsymbol{\omega}_2) = \{\boldsymbol{\omega}_1 : \boldsymbol{\omega}_1 \left(\bar{\mathbf{R}} \boldsymbol{\omega}_2'\right) = \mu_{\text{target}}\}, \quad \mathcal{C}_2(\boldsymbol{\omega}_1) = \{\boldsymbol{\omega}_2 : \left(\boldsymbol{\omega}_1 \bar{\mathbf{R}}\right) \boldsymbol{\omega}_2' = \mu_{\text{target}}\}. \tag{72}$$

□

**Proof of Theorem 1**

*Proof.* We rewrite the Theorem 4.1(c) in Tseng (2001) according to our case with two blocks $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ as following: Every cluster point $\mathbf{z} = (\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*)$ of $\{(\boldsymbol{\omega}_1^{(i)}, \boldsymbol{\omega}_2^{(i)})\}$ is a coordinatewise minimum point of $f$ if the following four prerequisites are satisfied:

1. The level set $\mathbf{\Omega} = \{(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) : f(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) \leq f(\boldsymbol{\omega}_1^{(i)}, \boldsymbol{\omega}_2^{(i)})\}$ is compact;
2. $f$ is continuous on $\mathbf{\Omega}$;
3. $f(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ has at most one minimum when update $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$;
4. The updating follows the cyclic rule.

We prove the first prerequisite through Lemma 1. The second and third prerequisites can be directly implied by the properties of $f(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{\omega}_1 \mathbf{E}_t \boldsymbol{\omega}_2)^2$ such that continuous and Gâteaux differentiable on its domain, and also strictly convex in each block while fixing the other.

For the last prerequisite, according to Tseng (2001), cyclic rule is defined as: *We update block $i$ at iteration $i, i+N, i+2N, \cdots$, for $i = 1, \cdots, N$, where $N$ is the total number of blocks and $i$ denotes the index of the block being updated. In our model, $N = 2$ and $i$ takes 2 values: $i = 1$ corresponds to $\boldsymbol{\omega}_1$, and $i = 2$ corresponds to $\boldsymbol{\omega}_2$.*

The cyclic rule emphasizes that every block in $f(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ must be updated at least once within a finite number of iterations and following a fixed cyclic order. At iteration 1, we fix $\boldsymbol{\omega}_2$ and update $\boldsymbol{\omega}_1$; that is, we set $i = 1$ at iteration 1. At iteration 2, we set $i = 2$ and update $\boldsymbol{\omega}_2$. At iteration 3, we return to $i = 1$ and update $\boldsymbol{\omega}_1$ again. Continuing in this way, we choose $i = 1$ at odd iterations $1,\ 1+2,\ 1+2\times 2,\ 1+3\times 2,\ \cdots$, and we choose $i = 2$ at even iterations $2,\ 2+2,\ 2+2\times 2,\ 2+3\times 2,\ \cdots$. Therefore, our algorithm automatically satisfies the cyclic rule.

The convergence of iterations can be reached based on the Theorem 4.1(c) in Tseng (2001), by which we can conclude that the iterations in all of GMV, MWC, and MRC will eventually converge to a coordinatewise point. □

**Proof of Proposition 1**

*Proof.* We will only show that any coordinatewise $(\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*)$ in MWC is stationary point, then we can use the same way to show $(\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*)$ in GMV and MRC are stationary point, but without full investment constraints. Consider a problem in MWC such that we minimize risk $f(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{\omega}_1 \mathbf{E}_t \boldsymbol{\omega}_2')^2$ subject to return constraint $\boldsymbol{\omega}_1 \bar{\mathbf{R}} \boldsymbol{\omega}_2' \geq \mu$, and two full investment constraints $\boldsymbol{\omega}_1 \boldsymbol{\iota}_m = 1$ and $\boldsymbol{\omega}_2 \boldsymbol{\iota}_n = 1$. For the purpose of simplicity, we denote $f^* = f(\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*)$ to indicates the level of risk at $(\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*)$. Consider the following Lagrangian:

$$\mathcal{L} = f(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) - \lambda\left(\boldsymbol{\omega}_1 \bar{\mathbf{R}} \boldsymbol{\omega}_2' - \mu\right) - \zeta_1\left(\boldsymbol{\omega}_1 \boldsymbol{\iota}_m - 1\right) - \zeta_2\left(\boldsymbol{\omega}_2 \boldsymbol{\iota}_n - 1\right). \tag{73}$$

The F.O.C. implies that

$$\begin{aligned} \boldsymbol{\nabla}_{\omega_1} f^* &= \lambda \bar{\mathbf{R}} \boldsymbol{\omega}_2^{*\prime} + \zeta_1 \boldsymbol{\iota}_m, \\ \boldsymbol{\nabla}_{\omega_2} f^* &= \lambda \bar{\mathbf{R}}' \boldsymbol{\omega}_1^{*\prime} + \zeta_2 \boldsymbol{\iota}_n. \end{aligned} \tag{74}$$

Suppose there are small feasible perturbations $\epsilon\boldsymbol{\Delta}_1 \in \mathbb{R}^m$ and $\epsilon\boldsymbol{\Delta}_2 \in \mathbb{R}^n$ imposed on $\boldsymbol{\omega}_1^*$ and $\boldsymbol{\omega}_2^*$, where $\epsilon > 0$. Under full investment constraints, $(\boldsymbol{\omega}_1^* + \epsilon\boldsymbol{\Delta}_1)\,\boldsymbol{\iota}_m = 1$, and $(\boldsymbol{\omega}_2^* + \epsilon\boldsymbol{\Delta}_2)\,\boldsymbol{\iota}_n = 1$, which implies that, along any feasible direction, $\boldsymbol{\Delta}_1\boldsymbol{\iota}_m = \boldsymbol{\Delta}_2\boldsymbol{\iota}_n = 0$. Then, under return constraint,

$$(\boldsymbol{\omega}_1^* + \epsilon\boldsymbol{\Delta}_1)\,\bar{\mathbf{R}}\,(\boldsymbol{\omega}_2^* + \epsilon\boldsymbol{\Delta}_2)' = \boldsymbol{\omega}_1^*\bar{\mathbf{R}}\boldsymbol{\omega}_2^{*\prime} + \epsilon\left(\boldsymbol{\Delta}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2^{*\prime} + \boldsymbol{\omega}_1^*\bar{\mathbf{R}}\boldsymbol{\Delta}_2'\right) \geq \mu. \tag{75}$$

With binding return constraint at $(\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*)$, we have $\boldsymbol{\omega}_1^*\bar{\mathbf{R}}\boldsymbol{\omega}_2^{*\prime} = \mu$, then we must have $\boldsymbol{\Delta}_1^*\bar{\mathbf{R}}\boldsymbol{\omega}_2^{*\prime} + \boldsymbol{\omega}_1^*\bar{\mathbf{R}}\boldsymbol{\Delta}_2' \geq 0$. Around $(\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*)$, and using the F.O.C. with Kuhn-Tucker condition that $\lambda \geq 0$, the first-order derivative of the risk function along the feasible joint direction $\mathbf{d} = (\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2)$ can be shown to be non-negative as (76):

$$\begin{aligned}
\boldsymbol{\nabla}_{\mathbf{d}} f\,(\boldsymbol{\omega}_1^*, \boldsymbol{\omega}_2^*) &= \boldsymbol{\Delta}_1\boldsymbol{\nabla}_{\boldsymbol{\omega}_1} f^* + \boldsymbol{\Delta}_2\boldsymbol{\nabla}_{\boldsymbol{\omega}_2} f^* \\
&= \lambda\left(\boldsymbol{\Delta}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2^{*\prime} + \boldsymbol{\Delta}_2\bar{\mathbf{R}}'\boldsymbol{\omega}_1^{*\prime}\right) + \zeta_1(\boldsymbol{\Delta}_1\boldsymbol{\iota}_m) + \zeta_2(\boldsymbol{\Delta}_2\boldsymbol{\iota}_n) \\
&= \lambda\left(\boldsymbol{\Delta}_1\bar{\mathbf{R}}\boldsymbol{\omega}_2^{*\prime} + \boldsymbol{\omega}_1^*\bar{\mathbf{R}}\boldsymbol{\Delta}_2'\right) \geq 0,
\end{aligned} \tag{76}$$

which implies that, for any feasible perturbation jointly imposed on $\boldsymbol{\omega}_1^*$ and $\boldsymbol{\omega}_2^*$, the first order derivative is non-negative, which is consistent with the definition of stationary point in Tseng (2001). □