Which Factors Matter in the Pricing Kernel?

Bin Luo [1] Guofu Zhou [2] Ti Zhou [3]

First version: April 2023
This version: August 2024

[†]We thank Wenbin Cao, Justin Case, Guanhao (Gavin) Feng, Jiawen Gu, Jinji Hao, Xin He, Fuwei Jiang, Junqing Kang, Stavros Panageas, Xinyu Wu, Qi Xu, Jingda Yan, Yifan Ye, Chu Zhang, Zhengyu Zhang, Hao Zhou, and the participants of Renmin University of China, Anhui University of Finance and Economics, the 20th International Symposium on Financial System Engineering and Risk Management, the 6th Conference on Big Data, Artificial Intelligence and Fintech, the 6th China Econometricians Forum,the 7th Annual China FinTech Conference, 2024 China Fintech Research Conference, International Association for Applied Econometrics 2024 Annual Conference, 2024 China International Risk Forum, 2024 Wellington Finance Summit, the 5th Greater China Area Finance Conference for their helpful comments and suggestions.

[1]Southern University of Science and Technology, luob2021@mail.sustech.edu.cn.
[2]Washington University in St. Louis. zhou@wustl.edu.
[3]Corresponding author, Southern University of Science and Technology, zhout@sustech.edu.cn.

Which Factors Matter in the Pricing Kernel?

ABSTRACT

We propose a general approach to selecting factors that provide independent information for mean-variance efficiency and show that the estimated portfolio converges to the true efficient portfolio at a rate of $\sqrt{\log(N)/T}$. Applying this to 174 characteristic-based factors, we find that about 40 matter, with announcement return and earnings predictability being the most important, and momentum, investment, and intangibles playing major roles too. The resulting portfolio weights are sparse and well-behaved, with an out-of-sample Sharpe ratio of 3.76 (excluding transaction costs), significantly outperforming alternative methods. Our approach also provides a novel test of well-known factor models, rejecting all of them.

***JEL classification*: G11; G12; C10; C58**

***Keywords*: Mean-variance efficiency, Firm characteristics, Regularized regressions, Sharpe ratio, Asset Pricing**

1 Introduction

The finance and accounting literature has accumulated a large number of factors that affect stock returns (see, e.g., Harvey, Liu, and Zhu (2016), Hou, Xue, and Zhang (2020), and Chen and Zimmermann (2022) for a summary), which Cochrane (2011) describes as a "factor zoo". These factors are usually constructed based on characteristic-sorted portfolios, such as the prominent size and value factors in the Fama and French (1993) three-factor model.[1] Existing studies have focused on how firm characteristics predict cross-sectional stock returns, and little attention has been paid to their role in achieving mean-variance efficiency. This issue is important because, in theory, only those factors that contribute to mean-variance efficiency, i.e., have non-zero coefficients in an efficient portfolio, should enter the pricing kernel that prices all assets (Hansen and Jagannathan, 1991).

In this paper, we develop a general approach to estimate the mean-variance efficient portfolio weights, especially when the number of assets is so large that it is close to or even larger than the number of observations. In this high-dimensional case, determining the mean-variance frontier is a challenging task because it requires knowledge of both the mean and the inverse of the covariance matrix of asset returns. In particular, it is known that the sample covariance matrix is either nearly singular or noninvertible. Therefore, certain sparsity assumptions are required to solve the problem.

Our first contribution is to show that any mean-variance optimal portfolio problem is equivalent to running a specific OLS regression. We derive this result from the equivalence between an expected quadratic utility maximization problem, whose solution is a mean-variance efficient portfolio, and the mean squared loss objective of the OLS regression. In this regression, the reciprocal of the investor's risk aversion is regressed on the excess returns of the factors with no intercept. Theoretically, the regression coefficients are equal to the weights of the efficient portfolio that maximizes both Sharpe ratio and expected utility.

Our second contribution is to propose the use of the LASSO regularization (Tibshirani, 1996) for the regression-representation of the mean-variance portfolio optimization problem, so as to overcome the dimensionality hurdle. By imposing sparsity on the regression coeffi-

[1] More recent examples include the profitability and investment factors in the Fama-French five-factor model (Fama and French, 2015) and in the q-factor model of Hou, Xue, and Zhang (2015), among others. Other important factor models in the recent literature include Stambaugh and Yuan (2017), Fama and French (2018), Daniel, Hirshleifer, and Sun (2020), and Hou, Mo, Xue, and Zhang (2021).

cients, the LASSO-regularized regression can obtain the efficient portfolio weights even when the number of assets N is much *larger* than the sample size T. More importantly, we show theoretically that under mild regularity assumptions similar to Ao, Li, and Zheng (2019), the expected return and volatility of the portfolio estimated using the LASSO-regularized regression converge in probability to those of the theoretically efficient portfolio. The same is true for the Sharpe ratio.

Our estimated portfolio, the Regularized Efficient Portfolio, hereafter simply REP, has three interesting and important features. First, it achieves sparsity and thereby identifies whether a factor is useful or not, and so the non-zero coefficients indicate those factors that are valuable in constructing the efficient portfolio. Second, REP is an *end-to-end integrated* approach in the sense that it obtains the efficient portfolio weights directly from the observed returns, without estimating the expected returns and the covariance matrix as in traditional portfolio optimization. Third, the asymptotic convergence rate of REP is desirable, which is of the order of $\sqrt{\log(N)/T}$. This allows practical applications for determining the sample size required for good accuracy. In short, based on our LASSO-regularized regression framework, we are able to eliminate redundant factors and identify those that provide independent and useful information in asset pricing.

Our method provides, in addition, a generic approach for comparing and testing factor models in a high-dimensional setting. We consider the well-known asset pricing models, such as Fama and French (1993, 2015, 2018), Carhart (1997), Hou, Xue, and Zhang (2015), Hou et al. (2021), Stambaugh and Yuan (2017), and Daniel, Hirshleifer, and Sun (2020), and examine their performance in terms of out-of-sample (OOS) Sharpe ratio. First, we compute the OOS Sharpe ratio of the factors from a given asset pricing model. Then, by adding these factors to a set of competing factors, we recompute another OOS Sharpe ratio. If the given asset pricing model is correctly specified, the model should price all the competing factors, and hence there should be no statistically significant difference between the two OOS Sharpe ratios. This is exactly the idea of Barillas and Shanken (2017), who emphasize the irrelevance of test assets in the context of comparing factor models. Chib, Zhao, and Zhou (2024) implement this idea in a Bayesian framework by computing all possible combinations in the case where the total number of factors, both original and competing, is small, e.g. no more than 15. While there is no good way to deal with large numbers of factors in these

studies, our paper provides a simple approach.

Our third contribution is empirical. Using a large set of firm characteristics from Chen and Zimmermann (2022) spanning from January 1970 to December 2022, we construct 174 characteristic-based factors and perform an out-of-sample (OOS) portfolio analysis based on a rolling window of 120 months. The OOS period begins in January 1980, and we rebalance the portfolio every month. The assets include the 174 factors and the market portfolio. We have a number of interesting findings.

First, we find that the portfolio estimated by REP has a high (hereafter annualized) OOS Sharpe ratio of 3.76, dominating the other benchmark portfolio rules, such as the MAXSER rule in Ao, Li, and Zheng (2019) whose Sharpe ratio is 3.24. The Sharpe ratio of the REP portfolio far exceeds the Sharpe ratios of the market portfolio and a strategy that invests in both the market and the equally-weighted portfolio of the 174 factors, 0.53 and 0.66, respectively. Second, by dropping the characteristic factors one by one and examining the reductions in Sharpe ratio of the REP portfolio based on the remaining factors, we find that AnnouncementReturn and EarningsPredictability are the two most important factors. Third, from a group perspective, the factors constructed from characteristics belonging to the groups of intangibles, momentum, and investment contribute the most to mean-variance efficiency. Fourth, we conduct asset pricing tests and find that the REP portfolio prices the 174 characteristic factors well in terms of large cross-sectional OLS R^2, large generalized least square (GLS) R^2, and small and insignificant time series regression intercepts, i.e., α. Therefore, the estimated REP portfolio plays the other side of the dual role of an efficient portfolio, i.e., a pricing kernel. Fifth, we demonstrate how our approach can be used to test factor models in a high-dimensional setting. We show that the nine famous factor models in the literature do not provide incrementally pricing information to the 174 characteristic factors. Finally, we document that the superior performance of the REP portfolio is robust to the choice of the risk-aversion levels and the size of the estimation window, among others. Moreover, at a proportional transaction cost of 60 bps, the OOS Sharpe ratio of the REP rule is 1.81, which remains economically significant.

Our study is related to several strands of literature. The first strand, pioneered by Britten-Jones (1999), develops the regression-based methodology to estimate mean-variance efficient portfolio weights as regression coefficients. Britten-Jones (1999) shows that regress-

ing the scalar "1" on the excess returns of the assets and normalizing the OLS coefficients recovers the tangency portfolio weight. However, his estimation relies on OLS, which is not feasible when $N > T$, and the normalization procedure further exacerbates the estimation risk (Ao, Li, and Zheng, 2019), which is also confirmed in our empirical analysis. Brodie, Daubechies, De Mol, Giannone, and Loris (2009) and Ao, Li, and Zheng (2019) propose regression-based approaches like ours, but they require prior estimate of unknown parameters, such as the mean of asset returns or the maximum Sharpe ratio, which is difficult for a large portfolio, especially when $N > T$. In contrast, our approach requires no prior estimation of any unknown parameters, works whether $N > T$ or not, and has a theoretically known and fast convergence rate.

The second strand of literature is on mean-variance portfolio choice under estimation risk. It is well known that estimated mean-variance portfolio rules based on sample moments (the "plug-in" rule) often take extreme positions in assets and incur substantial estimation risk, leading to poor OOS performance (Jobson and Korkie, 1980; Green and Hollifield, 1992; Michaud, 1989). Existing studies propose a variety of approaches to improve the OOS performance of estimated mean-variance rules, including shrinkage-based estimation of the mean and covariance matrix (e.g., Frost and Savarino, 1986; Jorion, 1986; Ledoit and Wolf, 2004, 2017, 2022), constraints on portfolio weights (Jagannathan and Ma, 2003; DeMiguel, Garlappi, and Uppal, 2009; Frost and Savarino, 1988; DeMiguel, Martin-Utrera, Nogales, and Uppal, 2020), and the approaches that explicitly account for estimation risk in portfolio optimization, including Bayesian portfolio choice based on predictive distributions[2] and the portfolio combination strategies (Kan and Zhou, 2007; Tu and Zhou, 2011). However, this stream of approaches is often limited to the case of small N relative to T and usually performs poorly or even becomes infeasible in our high-dimensional case.[3]

The third strand of literature concerns the relation between firm characteristics and the cross-section of individual stock returns. The Fama-MacBeth cross-sectional regression (Fama and MacBeth, 1973) is the classic empirical tool to study this relation (Haugen and Baker, 1996; Lewellen, 2015; Green, Hand, and Zhang, 2017). Recently, there has been a

[2]See the early work of Zellner and Chetty (1965) and Brown (1979), and the recent review by Avramov and Zhou (2010).

[3]The linear and nonlinear covariance shrinkage methods of Ledoit and Wolf (2004), Ledoit and Wolf (2017), Ledoit and Wolf (2022) are applicable in the high-dimensional setting and we consider them as the benchmark methods for comparison.

growing number of studies using machine learning methods to address this question, which can jointly deal with a large number of characteristics and allow for complex relations, such as Rapach, Strauss, and Zhou (2013), Light, Maslov, and Rytchkov (2017), Chinco, Clark-Joseph, and Ye (2019), Freyberger, Neuhierl, and Weber (2020), Gu, Kelly, and Xiu (2020), and Leippold, Wang, and Zhou (2022), among others. While some of the studies also examine the importance of characteristics in forming long-short portfolio strategies, their focus is on the return predictive power of the characteristics. Since the long-short portfolios are not necessarily optimal even in the limit (and are unlikely to be so since they ignore other deciles), these studies will not have direct implications for the pricing kernel. In contrast, what we are doing here is directly estimating the efficient portfolio or pricing kernel.

The fourth strand of literature deals with dimension reduction. Kozak, Nagel, and Santosh (2018) apply the principal components (PCA) method to shrink the dimension of stock return factors, and He, Huang, Li, and Zhou (2023) use a reduced-rank approach. Kozak, Nagel, and Santosh (2020) and Chib, Zhao, and Zhou (2024) conduct further similar Bayesian analyses. Kelly, Pruitt, and Su (2019) propose instrumented PCA, which allows for dynamic factor loadings that vary with firm characteristics, to extract latent factors from individual stock returns. The standard PCA extracts factors that best capture the time series variation of asset returns, but does not use information from the cross section of expected returns. The risk-premium PCA of Lettau and Pelger (2020) combines the PCA objective with a regularization term that penalizes the pricing errors of the factors. A limitation of these approaches is that they often reduce to a set of latent factors that are linear combinations of all factors, rather than individual factors. In contrast, our approach does exactly the latter, which answers whether or not a given factor is in the pricing kernel.

The paper is organized as follows. Section 2 describes the theory of REP, highlighting its ability to approach mean-variance efficiency in a high-dimensional setting, and provides a geometric interpetation for REP. Section 3 describes the data used in the empirical analysis and also discuss other benchmark portfolio rules for comparison. Section 4 presents the empirical results based on U.S stock return data and results of simulation analysis. Section 5 concludes the paper.

2 Methodology

This section describes the theory and methodology of the REP rule. First, we introduce an expected quadratic utility maximization problem that yields a theoretically efficient portfolio. Second, we show that this efficient portfolio can alternatively be obtained by an equivalent OLS regression. Third, we show that the portfolio estimated by the OLS regression with LASSO regularization converges to the theoretically efficient portfolio under mild conditions.

2.1 Regression Representation of Efficient Portfolio weights

Expected Quadratic Utility Maximization

Throughout the paper, we use bold mathematical symbols to denote matrices or vectors, while non-bold symbols are used to denote scalars. Suppose there are N assets whose random excess returns over the risk-free rate are denoted by $\boldsymbol{r}$, hereafter simply referred to as "returns". Denote the population mean and covariance matrix of $\boldsymbol{r}$ as $\boldsymbol{\mu} \equiv \mathbb{E}(\boldsymbol{r}) = (\mu_1, \cdots, \mu_N)'$ and $\boldsymbol{\Sigma} \equiv \mathbb{E}[(\boldsymbol{r}-\boldsymbol{\mu})(\boldsymbol{r}-\boldsymbol{\mu})'] = \{\sigma_{jk}\}_{j,k=1}^{N}$, respectively.

Consider the following standard portfolio optimization problem, where an investor chooses the portfolio weight $\boldsymbol{\omega}$ to maximize the mean-variance utility,

$$\boldsymbol{\omega}^0 = \arg\max_{\boldsymbol{\omega}} \mathbb{E}\left(\boldsymbol{\omega}'\boldsymbol{r}\right) - \frac{\gamma}{2}\mathrm{Var}\left(\boldsymbol{\omega}'\boldsymbol{r}\right) = \arg\max_{\boldsymbol{\omega}} \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}, \tag{1}$$

where γ is her risk-aversion level. The optimal portfolio weight is well known,

$$\boldsymbol{\omega}^0 = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \tag{2}$$

which is proportional to the weight of the ex-ante tangency portfolio of the risky assets. The portfolio $\boldsymbol{\omega}^0$ thus falls on the mean-variance efficient frontier and its Sharpe ratio achieves the theoretical maximum, $\sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}$.

In contrast to the standard problem, here we consider an equivalent problem that maximizes expected quadratic utility,

$$\boldsymbol{\omega}^\star = \arg\max_{\boldsymbol{\omega}} \mathbb{E}\left[\boldsymbol{\omega}'\boldsymbol{r} - \frac{\gamma}{2}\left(\boldsymbol{\omega}'\boldsymbol{r}\right)^2\right] = \arg\max_{\boldsymbol{\omega}} \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}'\left(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}'\right)\boldsymbol{\omega}. \tag{3}$$

The corresponding optimal portfolio weight is easily shown to be

$$\boldsymbol{\omega}^\star = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\gamma\,(1+\zeta)}, \quad \zeta \equiv \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \tag{4}$$

which is simply a scaling of the earlier optimal portfolio weight. Thus, $\boldsymbol{\omega}^\star$ is also on the efficient frontier. Denote the expected return and variance associated with $\boldsymbol{\omega}^\star$ as

$$r^\star \equiv (\boldsymbol{\omega}^\star)'\boldsymbol{\mu} = \frac{\zeta}{\gamma(1+\zeta)}, \tag{5}$$

$$(\sigma^\star)^2 \equiv (\boldsymbol{\omega}^\star)'\boldsymbol{\Sigma}\boldsymbol{\omega}^\star = \frac{\zeta}{\gamma^2\,(1+\zeta)^2}. \tag{6}$$

The Sharpe ratio of $\boldsymbol{\omega}^\star$ attains the theoretical maximum as $r^\star/\sigma^\star = \sqrt{\zeta} = \sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}$.

An Equivalent Regression Representation

Interestingly, the expected quadratic utility optimization problem (3) has an equivalent unconstrained regression representation in terms of the optimal portfolio $\boldsymbol{\omega}^\star$. That is, we have

Proposition 1. *The mean-variance efficient portfolio weight $\boldsymbol{\omega}^\star$ can be recovered by solving the least squares problem,*

$$\boldsymbol{\omega}^\star = \arg\min_{\boldsymbol{\omega}} \mathbb{E}\left(\frac{1}{\gamma} - \boldsymbol{\omega}'\boldsymbol{r}\right)^2. \tag{7}$$

Therefore, $\boldsymbol{\omega}^\star$ is the population regression coefficient of the following OLS regression,

$$\frac{1}{\gamma} = \boldsymbol{\omega}'\boldsymbol{r} + \epsilon. \tag{8}$$

In addition, by varying γ from zero to infinity, all portfolios in the efficient frontier can be recovered by the OLS regression, and their weights change proportionally to each other.[4]

Proof. All proofs are in the appendix. □

The equivalent regression representation is notable in several aspects. First, the dependent variable is a constant instead of a random variable and depends only on the investor's

[4]A polar case is when $\gamma = \infty$, the investor invests only in the risk-free asset, and the regression coefficient is also zero, consistent with the optimal weight.

risk aversion. While the mean-variance utility optimization problem (1) also admits a regression representation with a constant response, the response depends on the squared maximum Sharpe ratio ζ, which is unknown in practice and difficult to estimate in a high-dimensional setting where $T < N$.[5] Second, the regression has no intercept and no constraints. Third, the independent variables are the asset returns and the regression coefficients are the optimal asset weights $\boldsymbol{\omega}^\star$. In particular, the regression coefficients change in proportion to the inverse of the risk-aversion level γ. In other words, $\boldsymbol{\omega}^\star/2$ solves regression (7) when γ is doubled. However, this does not affect the Sharpe ratio, which is invariant to scaling.

2.2 Geometric Interpretation of the Equivalent Regression

In this subsection, we aim at providing an economic interpretation of the equivalence between OLS regression (7) and the expected quadratic utility maximization problem (3), inspired by Gibbons, Ross, and Shanken (1989)'s interpretation for their famous portfolio efficiency test. Figure 1 illustrates the geometry. The red solid line and the cyan hyperbola represent the mean-variance efficient frontier when there is a risk-free asset and when there are only risky assets, respectively. Denote by Θ the set of feasible portfolios, i.e., the region on and below the straight efficient frontier (red solid line).

[**Insert Figure 1 here**]

We first focus on Figure 1(a). On the y-axis is a point G whose distance from the origin is $1/\gamma$, corresponding to the dependent variable. For an arbitrary portfolio $P \in \Theta$, its expected return μ_P and volatility σ_P are equal to the lengths of the line segments AP and PB, respectively. The Pythagorean Theorem implies the following relation between the squared distance between the points G and P and the least squares objective (7),

$$|GP|^2 = |AG|^2 + |AP|^2 = (|OG| - |OA|)^2 + |AP|^2 = \left(\frac{1}{\gamma} - \mu_P\right)^2 + \sigma_P^2 = \mathbb{E}\left(\frac{1}{\gamma} - r_P\right)^2.$$

It can also be shown that

$$|GP|^2 = \frac{1}{\gamma^2} - \frac{2}{\gamma}\left(\mu_P - (\mu_P^2 + \sigma_P^2)\right) = \frac{1}{\gamma^2} - \frac{2}{\gamma}\mathbb{E}\left(r_P - \frac{\gamma}{2}r_P^2\right).$$

[5] The corresponding dependent variable in the equivalent regression representation is $(1+\zeta)/\gamma$.

Thus, mathematically, maximizing expected quadratic utility is equivalent to minimizing $|GP|^2$. Consequently, the expected quadratic utility maximization problem (3) boils down to finding $P^\star \in \Theta$ such that the mean squared difference between $r_{P^\star}$ and $1/\gamma$ is the minimum

$$P^\star = \underset{P\in\Theta}{\arg\max}\ \mathbb{E}\left(r_P - \frac{\gamma}{2}r_P^2\right) = \underset{P\in\Theta}{\arg\min}\ |GP|^2 = \underset{P\in\Theta}{\arg\min}\ \mathbb{E}\left(\frac{1}{\gamma} - r_P\right)^2. \tag{9}$$

Figure 1(b) characterizes the position of $P^\star$. Specifically, the line originating from point G and perpendicular to the straight efficient frontier identifies $P^\star$, which must fall at the intersection of the perpendicular line and the straight efficient frontier (the blue perpendicular point). This is because the line segment $GP^\star$ (the blue dashed line) is the shortest of all feasible line segments GP. Note also that the feasible set Θ is the linear space spanned by the asset excess returns. Therefore, the optimal portfolio $P^\star$ with the shortest length $|GP^\star|$ must be the projection of G onto Θ, which is exactly where $P^\star$ lies.

Figure 1(c) illustrates how the optimal portfolio $P^\star$ varies with γ. Consider three investors with risk aversion of 1, 5, and 10. The three green dashed lines, starting from the points $(0, 1/\gamma), \gamma = 1, 5, 10$ on the y-axis, and perpendicular to the efficient frontier characterize the optimal portfolios for the three investors. The three optimal portfolios fall on the three green perpendicular points on the efficient frontier. Note that all optimal portfolios have the same Sharpe ratio as the ex ante tangency portfolio, but have different leverage ratios, with $\gamma = 1$ ($\gamma = 10$) being the highest (lowest). Figure 1(c) also reveals that since point $P^\star$ must fall below G, the expected return and volatility of $P^\star$ are implicitly controlled by γ.

2.3 Relation with Other Regression-based Methods

The literature has proposed a number of equivalent regression representations of mean-variance portfolio optimization problems. Britten-Jones (1999) shows first that regressing the scalar "1" on the asset excess returns and normalizing the resulting OLS coefficients (i.e., making them sum to 1) will reocover the tangency portfolio weight. This corresponds to the special case in our setting where $\gamma = 1$. However, Britten-Jones (1999) does not study the expected quadratic utility maximization problem (3). More improtantly, his estimation procedure relies on OLS, which is not feasible when $T < N$, and the normalization of the regression coefficients further exacerbates the estimation risk (Ao, Li, and Zheng, 2019),

which is also confirmed in our empirical analysis.

Both Brodie et al. (2009) and Ao, Li, and Zheng (2019) study the mean-variance portfolio problem in the classical form

$$\boldsymbol{\omega}^\star = \arg\min_{\boldsymbol{\omega}} \boldsymbol{\omega}'\boldsymbol{\mu} \quad \text{subject to} \quad \boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \leq \sigma^2. \tag{10}$$

Brodie et al. (2009) propose a *constrained* regression representation of the dual form of problem (10), where the target expected return r is the dependent variable. In the implementation, they replace the population mean $\boldsymbol{\mu}$ in the constraint with its sample analog $\bar{\boldsymbol{r}}$. Ao, Li, and Zheng (2019) derive an *unconstrained* regression representation of problem (10) where the dependent variable is $\sigma\frac{1+\zeta}{\sqrt{\zeta}}$ and ζ is the squared maximum Sharpe ratio (Equation (4)), and they replace ζ with its sample analog when running the regression. The constant dependent variables in both studies involve unknown parameters and require further estimation, which introduces additional estimation risk. In particular, $\bar{\boldsymbol{r}}$ used in Brodie et al. (2009) is not a consistent estimate of $\boldsymbol{\mu}$ in a high-dimensional setting (Ao, Li, and Zheng, 2019), and the estimate of ζ used in Ao, Li, and Zheng (2019) is not feasible when $T < N$.

Our representation (1) is fundamentally different from existing approaches because our dependent variable $1/\gamma$ is known and no estimation is required. This greatly simplifies the estimation, allows our approach to handle the high-dimensional case, and leads to a much faster convergence rate to the theoretically efficient portfolio, as shown next in Section 2.4.

2.4 LASSO-regularized Regression

In reality, the theoretically efficient portfolio $\boldsymbol{\omega}^\star$ is not feasible. To estimate $\boldsymbol{\omega}^\star$, one may run regression (8) using realized asset returns. Suppose we observe a sample of asset returns over T time periods. Denote the collection of asset returns as a $T \times N$ matrix $\boldsymbol{R} = (\boldsymbol{R}_1, \cdots, \boldsymbol{R}_T)'$, where $\boldsymbol{R}_t = (R_{t1}, \cdots, R_{tN})'$ is an $N \times 1$ vector of asset returns at time t. Denote by $\mathbf{1}$ a conformable vector of ones. The OLS regression (8) based on the realized asset return data produces the estimated efficient portfolio weight

$$\widehat{\boldsymbol{\omega}}^{\text{OLS}} = \arg\min_{\boldsymbol{\omega}} \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\gamma} - \boldsymbol{\omega}'\boldsymbol{R}_t\right)^2. \tag{11}$$

Essentially, $\widehat{\boldsymbol{\omega}}^{\text{OLS}}$ is the "plug-in" estimate of $\boldsymbol{\omega}^\star$, with the true parameters replaced by their sample counterparts. Although simple, plug-in portfolio rules are generally biased, and the bias tends to increase with the ratio of the number of assets to the sample size, N/T (Kan and Zhou, 2007; Bai, Liu, and Wong, 2009; EL Karoui, 2010; Ao, Li, and Zheng, 2019). For this reason, the plug-in portfolio fails to approach mean-variance efficiency in a high-dimensional setting where N grows proportionally to T. Numerically, OLS does not provide a unique solution when $N > T$.

The common method to overcome this problem is to impose penalty terms on the regression coefficients to achieve sparsity. We employ the LASSO regularization (Tibshirani, 1996) to obtain a sparse estimate of $\boldsymbol{\omega}^\star$ by imposing an upper bound on the ℓ_1-norm of the regression coefficients, defined as $\|\boldsymbol{\omega}\|_1 \equiv \sum_{j=1}^{N} |\omega_j|$. We call our approach *Regularized Efficient Portfolio* (REP) and establish its asymptotic properties in Theorem 1.

Theorem 1. *Given the following five assumptions,*

A1*: The asset returns are independent and follow the normal distribution* $\boldsymbol{r} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*;*

A2*: There exists* $M < \infty$ *such that* $\max(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \max_{1\leq j\leq N} |\mu_j|) \leq M$*;*

A3*: There exists* $L < \infty$ *such that* $\max_{1\leq i\leq N} |\sigma_{ii}| \leq L$*;*

A4*: The theoretical optimal portfolio satisfies* $\|\boldsymbol{\omega}^\star\|_1 \leq \lambda$ *for some constant* λ*;*

A5*: Both the number of assets* N *and the sample size* T *go to infinity and satisfy the relation* $\sqrt{\log(N)/T} \to 0$ *as* $T \to \infty$*,*

then the following LASSO-regularized regression

$$\min_{\boldsymbol{\omega}} \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\gamma} - \boldsymbol{\omega}'\boldsymbol{R}_t\right)^2 \quad \textit{subject to} \quad \|\boldsymbol{\omega}\|_1 \leq \tilde{\lambda}, \quad \textit{where} \quad \tilde{\lambda} \geq \lambda, \tag{12}$$

yields the REP rule $\widetilde{\boldsymbol{\omega}}$ *whose expected return and volatility converge in probability to those of the theoretically mean-variance efficient portfolio* $\boldsymbol{\omega}^\star$*, i.e.,*

$$|\boldsymbol{\mu}'\widetilde{\boldsymbol{\omega}} - \boldsymbol{\mu}'\boldsymbol{\omega}^\star| \xrightarrow{P} 0, \tag{13}$$

and

$$\left|\sqrt{\widetilde{\boldsymbol{\omega}}'\boldsymbol{\Sigma}\widetilde{\boldsymbol{\omega}}} - \sqrt{(\boldsymbol{\omega}^\star)'\boldsymbol{\Sigma}\boldsymbol{\omega}^\star}\right| \xrightarrow{P} 0. \tag{14}$$

Moreover, by continuous mapping, the Sharpe ratio and the expected quadratic utility of $\widetilde{\boldsymbol{\omega}}$

converge in probability to those of $\boldsymbol{\omega}^\star$ *as well.*[6]

Theorem 1 states that the REP rule is asymptotically mean-variance efficient. A detailed proof of Theorem 1 is given in the Appendix.

A few remarks about Theorem 1 are in order. The first four assumptions of Theorem 1 are the same as those used in Theorem 1 of Ao, Li, and Zheng (2019, p. 2900). Assumption A1 assumes normality of asset returns, which is the standard working assumption underlying mean-variance analysis (Cochrane, 2009). Assumption A2 requires boundedness of the maximum Sharpe ratio and expected asset returns. Assumption A4 requires the bounded ℓ_1-norm of $\boldsymbol{\omega}^\star$ and hence controls for the sparsity of the optimal portfolio weight. Together, assumptions A2 and A4 rule out the extreme case where $\boldsymbol{\omega}^\star$ takes non-trivial positions in most assets, such that the sum of weights on individual assets grows unboundedly as N becomes large, and consequently the estimation error does not vanish asymptotically.[7] This is not to say that most weights are zero. For example, value-weighted and equally-weighted portfolios are allowed. Assumption A5 is new and differs from the corresponding one in Ao, Li, and Zheng (2019), which characterizes the rates at which T and N go to infinity. Note that we consider the asymptotics where both N and T go to infinity and $\log(N)/T$ converges to zero.

Theorem 1 shares some similarities with Theorem 1 of Ao, Li, and Zheng (2019). Both the Sharpe ratios of REP and the MAXSER rule of Ao, Li, and Zheng (2019)'s asymptotically converge to the theoretical maximum Sharpe ratio $\sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}$. Moreover, while MAXSER solves the mean-variance problem (10), which explicitly controls for risk, our REP rule chooses the optimal leverage along the efficient frontier for investors with different levels of risk aversion, thus controlling for risk implicitly. We emphasize, however, that the formulation of our equivalent regression representation (7) and the implications of Theorem 1 are fundamentally different from Ao, Li, and Zheng (2019) in two ways.

First, REP can solve the high-dimensional case where $T < N$, whereas MAXSER cannot because the constant dependent variable in its regression depends on the squared maximum Sharpe ratio $\zeta = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$. Accurately estimating ζ is challenging since it involves the inverse

[6]The theorem holds whenever $\tilde{\lambda} \geq \lambda$.

[7]Estimated portfolio rules contain estimation errors. For instance, the estimation error of the sample mean accumulates at the order of $O(\sqrt{N/T})$. Intuitively, if the estimated rule takes non-negligible positions in most assets, the estimated errors will accumulate and may not vanish in a high-dimensional setting where N grows proportionally to T.

of the covariance matrix. Ao, Li, and Zheng (2019) employ Kan and Zhou (2007)'s adjusted estimator of ζ, which is only viable when $T > N$. In addition, to further reduce the dimension of the assets, they utilize a subpool technique to select a smaller pool of assets and then apply the MAXSER rule.[8] While this procedure reduces estimation risk by focusing on a smaller cross section, it may lose valuable information. In contrast, the dependent variable in REP, $1/\gamma$, is a known constant. Thus, REP can be implemented directly by running LASSO-regularized regression (12) using all assets, making it an *end-to-end* approach.

Second, compared to MAXSER, REP significantly relaxes the convergence condition. Assumption A5 in Theorem 1 implies that the Sharpe ratio of $\widetilde{\boldsymbol{\omega}}$ converges to $\sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}$ as T and N go to infinity and $\sqrt{\log(N)/T} \to 0$. In contrast, the convergence condition in Ao, Li, and Zheng (2019) is that $N/T \to \rho \in (0, 1)$, which implies that N grows proportionally to T. The assumption of REP is much weaker, since it allows for the case where N is a polynomial function of T, and thus much larger than T. The REP rule achieves a faster convergence rate because its dependent variable does not require estimation, while Ao, Li, and Zheng (2019) does.[9] Also, their condition is stronger than ours. To see this,

$$N \approx \rho T \quad \Rightarrow \quad \sqrt{\log(N)/T} \approx \sqrt{\log(\rho T)/T} \to 0 \quad \text{as} \quad T \to \infty. \tag{15}$$

2.4.1 No Risk-free Asset Case

Thus far, we focus on the setting where there is a risk-free asset. In Appendix C.1, we demonstrate that our regression-based approach can be extended to the case where there are only risky assets. In this setting, a mean-variance efficient portfolio problem can still be represented by OLS regression with a dependent variable different from that of regression (7). We propose a similar LASSO-regularized regression to consistently estimate efficient portfolio weights and show that the convergence rate remains of the order of $\sqrt{log(N)/T}$.

[8] Specifically, the subpool technique is as follows. First, the size of the subpool, $N^- < N$, is prespecified, which is 50 in Ao, Li, and Zheng (2019). Then they randomly draw 50 assets from the original pool of assets and estimate their ζ. This process is repeated 1,000 times. Finally, the subpool with ζ close to the 95% quantile of all ζ is selected, and the MAXSER rule is applied to the selected subpool of assets.

[9] A consistent estimate of the dependent variable in their equivalent regression, $\sigma\frac{1+\zeta}{\sqrt{\zeta}}$, requires the condition $N/T \to \rho \in (0, 1)$.

2.5 Tuning the Hyperparameter λ

To implement REP, we must determine the hyperparameter λ of the LASSO-regularized regression, which controls the sparsity of the estimated efficient portfolio. A higher (lower) value of λ results in a less (more) sparse portfolio. As λ increases from 0 to some threshold, we get the so-called solution path, i.e., a series of solutions, each of which is an estimated portfolio. Following Ao, Li, and Zheng (2019), we employ an efficient algorithm called "Least Angle Regression (LARS)" (Efron, Hastie, Johnstone, and Tibshirani, 2004) to obtain the solution path.[10] The first and last on the solution path are the zero vector (the most sparse portfolio) and the OLS solution with the minimum ℓ_1-norm, denoted as $\boldsymbol{\omega}^{LS,\ell_1}$, respectively.[11] In particular, when $T > N$, the OLS solution is unique and $\boldsymbol{\omega}^{LS,\ell_1}$ is exactly $\boldsymbol{\omega}^{OLS}$. Define the ℓ_1 ratio of a solution $\boldsymbol{\omega}$ on the solution path as

$$\xi \equiv \frac{||\boldsymbol{\omega}||_1}{||\boldsymbol{\omega}^{LS,\ell_1}||_1}. \tag{16}$$

By construction, there is a one-to-one correspondence between ξ and λ along the solution path, and ξ increases monotonically from 0 to 1. Therefore, we regard ξ as our hyperparameter.

[**Insert Figure 2 here**]

Panel A of Figure 2 displays the values of the mean squared loss objective (12) of all estimated portfolios on the solution path against the values of ξ over $[0, 1]$. The sample contains 120 i.i.d. observations of excess returns on 100 stocks simulated from the multivariate normal distribution with mean and covariance matrix calibrated from real data (see Section 4.8). As shown in Panel A, the least squares objective is a U-shaped function of ξ. Intuitively, when ξ is close to zero, the estimated portfolio is overly sparse, excluding stocks with attractive risk-return profiles. As ξ moves away from zero, the relaxed sparsity constraint improves mean-variance efficiency and leads to a reduction in mean squared loss. However, mean squared loss eventually increases because the estimated portfolio is no longer sparse, and estimation risk begins to dominate diversification gains. Panel B of Figure 2 shows that the corresponding Sharpe ratios of the estimated portfolios on the solution path exhibit an

[10] Specifically, we use the modified LARS algorithm to obtain the solution path of the LASSO estimates.
[11] Recall that when $T < N$, the OLS solution may not be unique. See Lemma 7 in Tibshirani (2013).

inverse U-shaped pattern. More importantly, as indicated by the vertical dashed line, the portfolio with the minimum mean square loss almost coincides with the portfolio with the maximum Sharpe ratio.

The lesson from Figure 2 is that the optimal degree of sparsity should best balance the mean-variance efficiency provided by diversification gains against estimation risk. Therefore, we choose the optimal ξ such that the associated estimated portfolio has the minimum mean squared loss objective.[12] Note that for illustrative purposes, the values of mean squared loss and Sharpe ratio in Figure 2 are calculated based on the population mean and covariance matrix of the 100 stocks. In reality, the true parameters are unknown, so we perform a 10-fold cross-validation (CV) to select ξ that has the minimum realized mean squared loss. We repeat the CV procedure 10 times and take the average of ξ as the final input for REP.

In the follow sections, we apply our REP approach to characteristic-based U.S. equity factors to construct optimal mean-variance efficient portfolios. We examine the OOS performance of the REP portfolio and examine which characteristics provide independent information in achieving mean-variance efficiency.

3 Data and Benchmark Strategies

In this section, we describe the data used in the empirical analysis and illustrate benchmark strategies for comparison.

3.1 Data

The data used in the empirical analysis consist of stock returns and firm characteristics. We obtain monthly returns on common stocks (share codes 10 and 11) listed on the NYSE, AMEX, and NASDAQ from the CRSP database. The sample period spans from January 1970 to December 2022. Stock returns are adjusted for delisting. We remove stocks whose market capitalization falls below the 20th percentile of the cross section. This procedure yields 2,502,843 stock-month observations for 23,216 unique stocks. For characteristics data, we collect 212 commonly used firm characteristics from the database of Chen and Zimmer-

[12]This is equivalent to choosing ξ for the portfolio that maximizes expected quadratic utility, as shown in section 2.1.

mann (2022)[13]. After filtering out characteristics with too many missing observations, we obtain 174 firm characteristics. Table A.1 provides detailed information on the characteristics, where they are classified into six groups according to Hou, Xue, and Zhang (2020): Momentum, Value vs. Growth, Investment, Profitability, Intangibles, and Trading Friction.

Characteristics-based factors

We construct the characteristic-based factors using the 174 firm characteristics. In particular, the factors are characteristic-managed portfolios, following the pioneering work of Brandt, Santa-Clara, and Valkanov (2009). This approach has been widely used in the literature (Hjalmarsson and Manchev, 2012; Kelly, Pruitt, and Su, 2019; DeMiguel et al., 2020).

Suppose there are N_{t+1} stocks in month-$t+1$. Denote their excess returns by $\widetilde{\boldsymbol{R}}_{t+1} = (\tilde{R}_{t+1,1}, \cdots, \tilde{R}_{t+1,N_{t+1}})'$ and their lagged characteristics in month-t by a $N_{t+1} \times K$ matrix $\boldsymbol{Z}_t = (\boldsymbol{Z}_{t,1}, \cdots, \boldsymbol{Z}_{t,K})$, where K is 174 and column j of $\boldsymbol{Z}_t$ is a $N_{t+1} \times 1$ vector, containing the j-th characteristic of the stocks in the cross section, $\boldsymbol{Z}_{t,j} := \left(Z_{t,j}^{(1)}, \cdots, Z_{t,j}^{(N_{t+1})}\right)', j = 1, \cdots, K$. Each column of $\boldsymbol{Z}_t$ is winsorized at the top and bottom 1% level in the cross section to reduce the effects of outliers, and is standardized to have a (cross-sectional) mean of zero and a unit standard deviation.[14] The excess returns on the 174 characteristic-managed portfolios are given by

$$\boldsymbol{R}_{c,t+1} = \frac{\boldsymbol{Z}_t' \widetilde{\boldsymbol{R}}_{t+1}}{N_{t+1}}. \tag{17}$$

The j-th column of $\boldsymbol{R}_{c,t+1}$ is the excess return on the managed portfolio constructed based on the j-th characteristics, where the weights on the stocks are equal to their values in characteristic j. Since each characteristic is cross-sectionally standardized, the characteristic-managed (CM) portfolios are self-financed long-short strategies that take long (short) positions in firms whose characteristics are above (below) the cross-sectional mean. The CM portfolios are analogous to the sorted long-short portfolios but use more assets in the cross section. For example, Figure 3(a) and Figure 3(b) visually illustrate the construction difference between the long-short portfolio sorted by book-to-market ratio and the B/M-managed

[13]The website of the database is `https://www.openassetpricing.com`.

[14]Before performing these transformations, we fill in missing observations in $\boldsymbol{Z}_t$ as follows. If characteristic j is missing for stock i, we replace it with the mean of the characteristics of the industry to which stock j belongs. If the industry mean is also missing, we use the cross-sectional mean of all stocks. Note that all of these procedures are performed using cross-sectional information only.

portfolio, respectively.

[**Insert Figure 3 here**]

In addition to the CM portfolios, we also consider the value-weighted market whose return in month-$t+1$ is denoted by $R_{m,t+1}$. Denote by θ_m and $\boldsymbol{\theta}_c$ the weights on the market portfolio and the 174 CM portfolios, respectively. The excess return on the whole portfolio in month-$t+1$, $R_{p,t+1}$, is

$$R_{p,t+1} = \theta_m R_{m,t+1} + \boldsymbol{\theta}_c' \boldsymbol{R}_{c,t+1} = \theta_m \underbrace{\left(\boldsymbol{w}_{m,t}' \widetilde{\boldsymbol{R}}_{t+1}\right)}_{R_{m,t+1}} + \boldsymbol{\theta}_c' \underbrace{\left(\frac{1}{N_{t+1}} \boldsymbol{Z}_t' \widetilde{\boldsymbol{R}}_{t+1}\right)}_{\boldsymbol{R}_{c,t+1}}. \tag{18}$$

Alternatively, equation (18) implies that the return on the whole portfolio can also be derived using the weights and returns of individual stock returns[15]

$$R_{p,t+1} = \boldsymbol{w}_{p,t}' \widetilde{\boldsymbol{R}}_{t+1}, \tag{19}$$

where

$$\boldsymbol{w}_{p,t} = \theta_m \boldsymbol{w}_{m,t} + \frac{1}{N} \boldsymbol{Z}_t \boldsymbol{\theta}_c. \tag{20}$$

This alternative representation of the whole portfolio weights can be used to calculate the trading properties, such as the turnover and leverage.

3.2 Other Benchmark Portfolio Rules

Besides the REP portfolio, we also consider a few portfolio rules used as benchmarks in both empirical analysis. Table 1 lists the benchmark rules along with their respective abbreviations. Column 1 of Table 1 gives the descriptions of the benchmark portfolios and divides them into two categories. The first category of benchmark portfolios come from the literature directly. They are MAXSER (Ao, Li, and Zheng, 2019), and the OLS regression-based

[15] The net position of $\boldsymbol{w}_{p,t}$ is equal to θ_m. To see this,

$$\mathbf{1}'\boldsymbol{w}_{p,t} = \theta_m(\underbrace{\mathbf{1}'\boldsymbol{w}_{m,t}}_{=1}) + \frac{1}{N}(\underbrace{\mathbf{1}'\boldsymbol{Z}_t}_{=0})\boldsymbol{\theta}_c = \theta_m,$$

portfolio by Britten-Jones (1999) (BJ). We also construct the equally-weighted portfolios DeMiguel, Garlappi, and Uppal (2009).

[**Insert Table 1 here**]

The second category of benchmark portfolios are the MV portfolios computed as follows.

$$\boldsymbol{\omega}_{MV} = \frac{\sigma}{\sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}.$$

The estimation of the covariance matrix $\boldsymbol{\Sigma}$ applies several shrinkage methods to form different benchmark portfolios in the second category, including the sample estimate (or a pure plug-in portfolio, MV-Sample), the linear shrinkage method (MV-LS, Ledoit and Wolf (2004)), nonlinear quadratic-inverse shrinkage (MV-NLS, Ledoit and Wolf (2022))[16] and nonlinear quadratic-inverse shrinkage towards the Fama-French three-factor model (MV-NLS-FF3)[17]. The expected returns $\boldsymbol{\mu}$ for these MV benchmark portfolios are simply the sample mean of the simulated stock returns $\boldsymbol{r}_t,\ t = 1, \cdots, T$. The related literature and codes of these covariance shrinkage methods can be found on Michael Wolf's personal website.

Our REP portfolio is constructed based on a given risk-aversion level of $\gamma = 10$, while the remaining benchmark portfolios are implemented based on a given risk level of $\sigma = 0.04$ except for the BJ portfolio whose construction does not necessitate the pre-specification of either the risk-aversion level γ or the risk level σ. Some of the benchmark portfolios discuss above are not viable in high-dimensional cases when $T < N$, including MAXSER, BJ and MV-Sample. To implement these benchmark portfolios in high-dimensional cases, we apply the subpool technique, as described in (Ao, Li, and Zheng, 2019), which firstly selects a smaller pool of assets and then applies these approaches to the subpool. The size of the subpool is fixed at 50. The shrinkage-based approach, including MV-LS, MV-QIS and MV-QIS-FF3, are capable of handling high-dimensional cases, and thus we apply these approaches to the full cross section.

[16]The nonlinear shrinkage method of Ledoit and Wolf (2022) is similar to, but faster than, the one of Ledoit and Wolf (2017).

[17]The nonlinear shrinkage method towards a specific factor model is implemented primarily by applying the nonlinear shrinkage method, QIS, to the preconditioned data adjusted for this factor structure. See Section 4.2 in Ledoit and Wolf (2017) for the entire details.

4 Empirical Results

In this section, we perform empirical analyses to examine the performance of the REP rule and to answer the question of what factors contribute independent information to mean-variance efficiency.

4.1 Empirical Design

We focus on an out-of-sample. Suppose the sample period is from 1 to T. We estimate the REP weight $\widehat{\boldsymbol{\theta}}_t$ in month t based on the LASSO-regularized regression using the data over the past q months ($t = q, \cdots, T-1$)

$$\widehat{\boldsymbol{\theta}}_t = \arg\min_{\boldsymbol{\theta}} \frac{1}{q} \sum_{s=t-q+1}^{t} \left(\frac{1}{\gamma} - \boldsymbol{\theta}' \boldsymbol{R}_t \right)^2 \quad \text{subject to} \quad \|\boldsymbol{\theta}\|_1 \leq \lambda. \tag{21}$$

In particular, the full sample is from January 1970 to December 2022, $T = 636$ months in total. The length of the rolling window is $q = 120$ months, resulting in an out-of-sample period from January 1980 to December 2022, a total of 516 months. The REP portfolio and the other benchmark portfolios listed in Table 1 are constructed using the market portfolio plus the 174 CM portfolios, and a subpool technique of 50 assets is used to the construct the benchmark portfolios if the methods are infeasible in high dimension.

4.2 Out-of-sample Performance

Table 2 summarizes the out-of-sample performance of the REP portfolio and the other nine benchmark portfolios. The performance measures include quadratic utility gains with respect to the market portfolio, the Sharpe ratios, the out-of-sample average returns, standard deviations, skewness, kurtosis and maximum drawdowns.

[Insert Table 2 here]

From Table 2, we can see that the out-of-sample Sharpe ratio of the REP portfolio is the highest at 3.76, suggesting that the REP portfolio is the most out-of-sample mean-variance efficient portfolio. The significance of the performance of the REP portfolio is justified by the Sharpe ratio tests (Jobson and Korkie, 1981; Memmel, 2003) reported in Column 3.

The Sharpe ratio of the REP portfolio is significantly higher than those of most of the other benchmark portfolios at 1%, and higher than those of the MV-NLS and MV-NLS-FF3 portfolios at 5%. The MAXSER portfolio is shown be to be able to control risk in the literature, and its out-of-sample standard deviation (21.35%) is close to that of the REP portfolio (20.97%), suggesting that the choice of the risk-aversion level $\gamma = 10$ for the REP portfolio is similar to setting a (monthly) given risk level of 4%. The maximum drawdown of the REP portfolio is the lowest of all the portfolios. The mean-variance portfolios with different covariance matrix estimates (MV portfolios) are constructed based on a given risk level of 4%, but their standard deviations are explicitly larger than those of the REP and the MAXSER portfolios. The MV-Sample portfolio suffers from severe estimation errors in high-dimensional investment problem. The underperformance of the EW-CM portfolio suggests that naively using the information in the CM portfolios does not improve the portfolio performance very much. The BJ portfolio performs the worst due to its inability to handle the high dimensionality of the OLS regression. In short, the REP portfolio has a superior out-of-sample performance compared to the other benchmark portfolios, in terms of Sharpe ratio and risk control.

In addition, Table A.3 in the appendix reports the risk-adjusted performance of the REP portfolio with respect to eight well-known factor models in the literature. They are the CAPM of Sharpe (1964) (CAPM), Fama-French three- (FF-3), five- (FF-5), and six- (FF-6) factor models (Fama and French, 1993, 2015, 2018), the 4 factor model of Carhart (1997) (Carhart-4), the q factor model of Hou et al. (2021) (HXZ-q5), the mispricing factor model of Stambaugh and Yuan (2017) (SY-4), the behavioral factor model of Daniel, Hirshleifer, and Sun (2020) (DHS-3). In addition to the eight factor models, we also collect the union factors from them to create a new aggregated factor model called "ALL"[18]. The results show that the REP portfolio has positive α's with respect to these factor models that are significant at 1% level.

[**Insert Table 3 here**]

Table 3 reports the trading properties of the proposed portfolios. To compute the trading properties for all of the portfolios listed in Table 1, which determine the weights in the market

[18]The "ALL" factors are: MKT, SMB, HML, MOM, RMW, CMA, IA, ROE, EG, MGMT, PERF, PEAD and FIN. The MKT, SMB and MOM factors are from Fama-French models.

and the 174 CM portfolios, denote the estimated weight of a portfolio rule in month-t by $\widehat{\boldsymbol{\theta}}_t \equiv \left(\widehat{\theta}_{m,t}, \widehat{\boldsymbol{\theta}}'_{c,t}\right)'$. Then the weights in the CM portfolios (and the market portfolio) can be transformed into the weights in the individual stocks according Equation (20):

$$\widehat{\boldsymbol{w}}_t = \widehat{\theta}_{m,t}\boldsymbol{w}_{m,t} + \frac{1}{N_t}\boldsymbol{Z}_t\widehat{\boldsymbol{\theta}}_{c,t}, \quad t = q, \cdots, T-1. \tag{22}$$

We can then calculate the trading-related information based on Eq.(22) such as leverage, turnover and so on.

Table 3 shows that the REP portfolio has an average absolute position in individual stocks of 0.17%, which is close to that of MAXSER and explicitly lower than those of the MV portfolios. The average maximum and minimum positions of the REP portfolio are 1.74% and -1.60%, and the average long (short) and average fractions of long (short) positions are 4.18 (61.82%) and -3.75 (38.18%), respectively. In short, the positions and leverage of the REP portfolio are not that extreme.

4.3 Importance of Individual Characteristics

Recall that the goal of this paper is to answer which characteristics are important in constructing mean-variance efficient portfolios in high dimension. To do this, we drop one of the CM portfolios each time and re-estimate the REP portfolio based on the remaining portfolios. We then compute the out-of-sample Sharpe ratio of the re-estimated REP portfolio, and compare it to that of the REP portfolio without the drop. In particular, we calculate the ratios of the reductions to the original Sharpe ratio without dropping, and focus on the significance of the reductions using the Sharpe ratio tests (Jobson and Korkie, 1981; Memmel, 2003). The larger and the more significant the Sharpe ratio reduction is, the more important the associated characteristic is. Originally, the out-of-sample Sharpe ratio of the REP portfolio without the drop is 3.76.

[**Insert Table 4 here**]

Among the 175 candidate portfolios (the market plus the 174 CM portfolios), Table 4 shows that 42 portfolios are at least 10% significant, suggesting that most of the characteristics are redundant. The two most important characteristics are earnings predictability

(EarningsPredictability) and announcement return (AnnouncementReturn) with 1% significant reductions of 7.67% and 7.61%, respectively, far greater than the reductions of other characteristics by a large margin. The second tier important characteristics are those whose reductions are within the range of 1.00-2.00, such as enterprise multiple (EntMult, 1.98%), changes in noncurrent operating liability (ChNCOL, 1.94%), (industry-adjusted) change in capital investment (ChInvIA, 1.83%), return seasonality of years 6 to 10 (MomSeason06YrPlus, 1.72%), revenue surprise (RevenueSurprise, 1.57%), earnings forecast to price (sfe, 1.38%), the persistence of earnings (EarningsPersistence, 1.27%), failure probability (FailureProbability, 1.25%), idiosyncratic risks with respect to the q factor model (IdioVolQF, 1.20%), quarterly return on assets (roaq, 1.15%), cash-based operating profitability (CBOperProf, 1.12%). The remaining important characteristics have smaller but significant Sharpe ratio reductions. Of the 42 significant CM portfolios, 5 of them belong to Momentum, 7 belong to Value vs. Growth, 4 belong to Investment, 6 belong to Profitability, 14 belong to Intangible and 6 belong to Trading Friction. Column 4 reports the fraction of nonzero weights for each CM portfolio over the out-of-sample period.[19] The two most important characteristics, AnnoucementReturn and EarningsPredictability, have high selection rates of 100.00% and 71.32%, respectively. Within each group, the fractions of non-zero weights decrease, though not strictly, as the Sharpe ratio reductions decrease.

[**Insert Figure 4, 5 and 6 here**]

The signals of the characteristics can be persistent or decaying over the out-of-sample period. A characteristic can be persistent if it is a risk premium, and decaying if it is a pricing error. Figure 4, 5 and 6 plot the out-of-sample evolution of the REP weights in the CM portfolio whose Sharpe ratio reductions are 1%, 5% and 10% significant, respectively (reported in Table 4). The different colors of the plot lines indicate the groups to which the corresponding characteristics belong. The black dashed horizontal lines in the subplots are the levels of the in-sample REP weights, and the black crosses (if any) represent the dates of the corresponding literature in which the characteristics or anomalies are published. The shaded areas represent the NBER recessions.

[19]The out-of-sample period is from January 1980 to December 2022, a total of 516 months. However, the denominator for calculating the fraction of non-zero weights for each CM portfolio is not necessarily 516. The denominators are equal to the total number of non-missing returns of the corresponding CM portfolios over the out-of-sample period.

The signals of some characteristics are persistent over the out-of-sample period, such as AnnouncementReturn and RevenueSurprise. The signals of some characteristics fade just after the publication of the literature such as sfe, The trends of some characteristics increase steadily such as ChNCOL, EarningsPredictability, RevenueSurprise, suggesting that these characteristics are likely to be risk premia. Moreover, the signals of ChNCOL, ChInvAI, EarningsPredictability, ReturnSkewCAPM, RevenueSurprise, Leverage, RetNOA, and AnalystValue become prominent after their corresponding literature proposes them. Most of the significant characteristics are only significantly useful for a small portion of the entire out-of-sample period.

4.4 Importance of Characteristics Groups

To quantitatively study the mean-variance importance of the characteristics in different groups, we construct the REP portfolio by removing the CM portfolios within the same group each time.

[**Insert Table 5 here**]

Table 5 documents the contributions of the characteristics in different groups to the out-of-sample mean-variance efficiency. Intangible, Momentum and Investment are the three most important groups, with out-of-sample Sharpe ratio reductions of 0.96, 0.66, 0.25, and reduction percentages of 25.52%, 17.53% and 6.77% respectively. The Sharpe ratio tests (Jobson and Korkie, 1981; Memmel, 2003) show that the Sharpe ratio reductions of Momentum, Investment and Intangible are significant at 1%. The Value vs. Growth, Profitability and Trading Friction groups are similarly informative, although their Sharpe ratio reductions are not significant at 1%. The market portfolio is the least useful. Column 5 reports the out-of-sample average of the non-zero ratios for the seven groups. The non-zero ratio for a particular group in a month is equal to the number of the non-missing REP weights in that group divided by the number of the CM portfolios in that group. The market portfolio is almost always selected in the REP portfolio. Nearly half of the momentum characteristics are useful and about a third of the value vs. growth, intangible and trading friction characteristics contribute to the out-of-sample mean-variance efficiency. About a quarter of the investment and profitability characteristics are selected by the LASSO-regularized regressions.

4.5 The Asset Pricing Performance of REP

Recall that the mean-variance efficient portfolio (or tangency portfolio) plays a dual role, having the highest Sharpe ratio and being equivalent to the pricing kernel, so it is necessary to examine whether the REP portfolio can price the CM portfolios cross-sectionally. To address this concern, we examine the performance of the REP portfolio in pricing the 174 CM portfolios cross-sectionally. In particular, we first obtain the balanced sample in which all of the 174 CM portfolios have non-missing returns in each month observation. This leads us to a smaller sample of 521 months from the original full sample of 636 months from January 1970 to December 2022. We then find the intersection period that aligns with the out-of-sample period of the REP portfolio returns and the balanced sample, and call it the "matched sample". We treat the out-of-sample returns of the REP portfolio as a single factor, and regress the time-series returns of the 174 CM portfolios on the REP factor to price them cross-sectionally. We report the average OLS adjusted R^2 and the average GLS R^2 of the asset pricing regressions, the latter being a stricter and more appropriate measure for testing whether the factors are on the efficient frontier. We also report the average absolute α and the average absolute t-statistics.

We also examine the same performance measures of nine factor models proposed in the previous part for comparison. Some may further argue that, perhaps, the asset pricing outperformance of the REP factor over these factor models can be due to the availability of the 174 characteristics. To address this concern, we also include the EW-CM portfolio as a single factor to cross-sectionally price the 174 CM portfolios in the matched sample. If the EW-CM factor performs well in asset pricing, this suggests that the outperformance of the REP portfolio is largely due to the incremental information of the 174 CM portfolios. On the other hand, if the asset pricing performance of the EW-CM factor is poor, then the asset pricing superiority of the REP factor is largely due to its subtle design and methodology.

Table 6 reports the results of the cross-sectional asset pricing performance. We focus first on Column 2-5, which document the results for the first set of test assets, i.e., the 174 CM portfolios. The REP prices the 174 CM portfolios the best, with excellent and the highest adjusted OLS R^2 of 88.44% and GLS R^2 of 34.94%. In addition, the average absolute α of the REP factor is the smallest and the t-statistics is the only one less than 1.00. These quantities suggest that the REP factor is a successful pricing kernel in pricing the 174 CM

portfolios, and the highest GLS R^2 implies that the REP factor is more close to the true efficient frontier compared to other factor models. The the ALL factor ranks second with an adjusted OLS R^2 of 81.69% and a GLS R^2 of 15.73%, although its GLS R^2 is less than half of that of the REP factor. The HXZ-q5 performs the best among the factor models in the literature, having the highest adjusted OLS R^2, the smallest average absolute α of 0.07, and a t-statistic of 1.96. The asset pricing performance of the EW-CM factor is similar to that of the CAPM model, implying that the outperformance of the REP factor is mainly due to its methodology and not just to the additional information provided by the CM portfolios.

[**Insert Table 6 here**]

A reasonable concern assets that the good performance of the REP factor in pricing the 174 CM portfolios is natural because the constructed REP factor comes from these CM portfolios. Furthermore, it can be further doubted whether the outstanding ability of the REP factor to price assets can be extended to portfolios and assets beyond the 174 CM portfolios. To address this concern, we propose another set of test assets that include not only the 174 CM portfolios, but also the REP factor itself and the ALL factors. In other words, all of the factor models listed in Column 1 now face a broader set of test assets, and the REP factor has to price those portfolios that are beyond its constituents. Column 6-9 of Table 6 report the asset pricing performance measures for such set of test assets. We can see that, as before, the REP factor has the highest adjusted OLS R^2 and GLS R^2, and the lowest average absolute α and t-statistic, overwhelmingly dominating all other factor models in all measures. It is interesting to note that from the first set of test assets (Column 2-5) to the second one (Column 6-9), all factor models except the REP factor visibly decrease in the adjusted R^2, suggesting that the REP factor is a key component of the anamoly source. Those factor models that can price the REP factor well have good asset pricing performance, which is reflected in the increase of the adjusted OLS R^2 of the REP factor in Column 6.

[**Insert Figure 7 here**]

The subplots of Figure 7 show the realized returns of the test assets against their expected returns predicted by the REP and EW-CM portfolios and the factor models. The black straight line is sloped at 45 degree. We can see that all the CM portfolios are located

more tightly around the 45-degree sloped line in the REP subplot, suggesting that the REP eliminates the alphas in pricing the 174 CM portfolios. The EW-CM portfolio and the other factor models in the literature cannot eliminate the alphas efficiently. Figure 8 shows the same plots for the second set of test assets and the conclusions are similar.

[**Insert Figure 8 here**]

In summary, the REP portfolio can simultaneously price the 174 characteristic factor and also factors from popular asset pricing models in the literature. This result suggests that the REP portfolio plays the other side of the dual role of a mean-variance efficient portfolio, i.e., a pricing kernel.

4.6 Comparing and Testing Asset Pricing Models

Our REP approach also provides a generic tool for comparing and evaluating factor models, which is important because of the proliferation of factors in the literature (Harvey, Liu, and Zhu, 2016; Cochrane, 2011). Consistent with our previous analysis, we consider the eight well-known factor models. The idea is as follows. First, based on the REP approach, we obtain the OOS Sharpe ratio of the factors from a given factor model, say the FF-3 model. Next, by combing the three factors of the FF-3 model with the 174 characteristic-based factors, we obtain another OOS Sharpe ratio based on the REP approach. If the FF-3 model is true, it should price all the 174 characteristic-based factors, and hence there should be no statistically significant difference between the two OOS Sharpe ratios. By testing whether the two OOS Sharpe ratio are equal, we can comparing factors models in a high-dimensional setting. This is exactly the rationale of Barillas and Shanken (2017), who emphasize the irrelevance of test assets in the context of comparing factor models.

Table 7 reports the results of comparing the factor models. As shown in Column 2, the out-of-sample periods of different factor models can be different because of the data availability. Thus, we create a balanced sample by intersecting the out-of-sample periods of the factor models with the sample of the 174 characteristic-based factor.[20] Column 4 reports the OOS Sharpe ratios of the factors in a given factor model estimated by REP. Column 5 reports the OOS Sharpe ratios of the factors plus the 174 characteristic-based

[20]The balanced sample of the characteristic-based factors alone is 521 months where the returns of the 174 factors are *all* non-missing.

factors estimated by REP. Comparing columns 4 and 5, we observe that by adding the 174 factors, the OOS Sharpe ratio of all factor models are substantially improved. Column 6 reports the p-value of the Sharpe ratio tests (Jobson and Korkie, 1981; Memmel, 2003). The null hypothesis that the Sharpe ratio in column 4 is greater than that in column 5 and the hypothesis is strongly rejected at least at 1% level. Therefore, we conclude that all the well-known factors in the literature are rejected in terms of pricing the 174 factors. We find similar conclusion when we consider the union of factors from all factor models.

4.7 Robustness Analysis

To show the reliability and stability of our REP approach, we perform extensive robustness analysis to check whether the remarkable out-of-sample performance is robust to some variations in implementation. We aggregate the results in the robustness analysis in Table 8. The results in Panel A do not take transaction costs into account, while the results in Panel B do. In each panel, Row 1-6 report the Sharpe ratios, average returns, standard deviations, maximum drawdowns, skewness and kurtosis. The last row in Panel B reports the average turnover over the out-of-sample period.

To calculate the portfolio returns with transaction costs, we use apply the transformation from the weights in the CM portfolios to the weights in the individual stocks $\widehat{\boldsymbol{w}}_t, t = q, \cdots, T-1$, as in Equation (22). The total transaction costs from month-t to $t+1$ (generated at the beginning of month-$t+1$) are then

$$tc_{t+1} = \left\| c \left(\widehat{\boldsymbol{w}}_{t+1} - \frac{\widehat{\boldsymbol{w}}_t \circ \left(\mathbf{1} + r_{f,t}\mathbf{1} + \widetilde{\boldsymbol{R}}_{t+1} \right)}{1 + r_{f,t} + \widehat{\boldsymbol{w}}_t' \widetilde{\boldsymbol{R}}_{t+1}} \right) \right\|_1, \tag{23}$$

where the scaler $c = 0.006$. That is, the transaction cost is constant at 60bps. The excess return regarding transaction costs in month-$t+1$ is

$$\widehat{R}^{tc}_{p,t+1} = \left[(1 + r_{f,t} + \hat{R}_{p,t+1}) (1 - tc_{t+1}) \right] - (1 + r_{f,t}). \tag{24}$$

This gives us a time series of excess returns regarding transaction costs over the out-of-sample period, from which we can compute the performance measures when the transaction costs are taken into account. Figure 9 plots the log accumulative wealth evolution for the

REP, the MAXSER and the market portfolios under transaction costs. The REP portfolio has the highest ending values and rises steadily during the recessions. The ending values of the MAXSER and the market portfolios are half or even less than that of the REP portfolio, and the their accumulative wealth can decrease during the recessions.

[**Insert Table 8 here**]

Column 2 of Panel A in Table 8 places the out-of-sample performance of the REP portfolio in the baseline case for ease of comparison. Column 2 of Panel B shows that the REP portfolio with a transaction cost of 60bps has a high out-of-sample Sharpe ratio of 1.81, although its annualized average turnover is as high as 64.93.

Different Risk-aversion Levels

Columns 2-4 of Table 8 report the robustness results for different levels of risk aversion $\gamma = 5$ and 15. Recall that the coefficients of the LASSO-regularized regression change proportionally to the dependent variables. Without transaction costs, The Sharpe ratios, skewness and kurtosis are scale invariant so these measures are the same as in the baseline case. The average returns and standard deviations change proportionally to the reciprocal of the risk-aversion levels, except for the maximum drawdown because its calculation is associated with gross returns.

With transaction costs, the Sharpe ratios for the two cases of risk-aversion levels are similar and around 1.80, suggesting that the out-of-sample mean-variance efficiency achieved by the REP portfolio is robust to the choice of risk-aversion levels. The average returns, standard deviations and maximum drawdowns change, though not exactly, in proportion to the reciprocal of the risk-aversion levels.

Lengths of Rolling Windows

Columns 5-7 of Table 8 report the robustness results for the REP constructed based on different estimation windows. Overall, the Sharpe ratios are similar for rolling windows of $q = 80$, 120 and recursive windows without transaction costs, although they are lower than that of the baseline case. With transaction costs, the performance of the cases using rolling windows are around 1.5-1.6, lower than that of the baseline case but better than that of the recursive case.

Rebalancing Frequency

Columns 8-10 of Table 8 report the robustness results for the buy-and-hold REP portfolios that are rebalanced quarterly, semi-annually and annually. With or without transaction costs, the out-of-sample performance of the REP portfolios declines as the rebalancing is becomes frequent. This is reasonable because the slow rebalancing leads to poor use of information.

4.8 Simulation Analysis

In this part, we report the simulation results to study the ability of REP to achieve mean-variance efficiency and to examine the asymptotic properties noted in Theorem 1.

The design of our simulation study closely follows Ao, Li, and Zheng (2019) who simulate the excess returns of $N = 100$ stocks based on the Fama-French 3 (FF-3) factor model with non-zero alpha.

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{F}_t + \boldsymbol{\epsilon}_t. \tag{25}$$

Calibrated quantities include $\boldsymbol{\alpha}$ and $\boldsymbol{B}$ of the 100 assets with respect to the FF3 factors, the mean $\boldsymbol{\mu_F}$ and the covariance matrix $\boldsymbol{\Sigma_F}$ of the FF-3 factors, and the covariance matrix $\boldsymbol{\Sigma_\epsilon}$ of the error terms. See Ao, Li, and Zheng (2019) for the calibration details.

After the calibrations, the returns on the 100 stocks are simulated via the data generating process (25). For a sample size of T, we simulate $T+1$ observations and estimate the portfolio weights based on the first T observations. Then, the constructed portfolios are invested in the $T+1$-th observation as an out-of-sample investment. We run this simulation $S = 10,000$ times, resulting in a time series of $S = 10,000$ "out-of-sample" returns for each portfolio. We focus on the out-of-sample Sharpe ratios and investigate the significance of the Sharpe ratio difference between the REP portfolio and other benchmark portfolios based on the test of Jobson and Korkie (1981) and Memmel (2003). We study simulation cases with sample sizes of $T = 120$ and 240 as done in Ao, Li, and Zheng (2019). In addition, we present a case of $T = 80$ because we want to demonstrate the advantages of REP when the dimension exceeds the sample size.

[**Insert Table 9 here**]

Table 9 shows the simulation results. The theoretical maximum annualized Sharpe ratio is 1.5217. The results show that the REP portfolio outperforms all of the proposed benchmark portfolios in terms of Sharpe ratios for various sample sizes. The MAXSER portfolio is a strong competitor to the REP portfolio, but it has inferior Sharpe ratios in all of the cases of different sample sizes. The Sharpe ratio tests strongly suggest that the Sharpe ratios of the REP portfolios cannot be significantly beaten by other benchmark portfolios at 1%, except for the MV-NLS-FF3 in $T = 80$, which is still significant at 5% though.

The BJ portfolio drastically underperforms all other portfolios (much worse than MV-PlugIn) and has a negative simulated out-of-sample Sharpe ratio in $T = 120$. The underperformance of the BJ portfolio is due to the poor OLS in high-dimensional cases and the aggregated estimation errors when normalizing the weights. In an untabulated result, we find that the performance of the BJ portfolio without weight normalization improves significantly and is very close to the performance of MV-Sample, suggesting that the weight normalization worsens the performance of the BJ portfolio.

As for the covariance shrinkage methods, the pure plug-in mean-variance portfolio (MV-PlugIn) performs the worst. The nonlinear shrinkage portfolio MV-NLS performs better than the linear shrinkage portfolios (MV-LS). It is reasonable that the MV-NLS-FF3 portfolio performs better than the MV-NLS portfolio because the MV-NLS-FF3 uses the additional information from the Fama-French three-factor model. In short, Table 9 shows that the REP achieves a good mean-variance efficient performance even when the dimension is high.

5 Conclusion

In this paper, we establish the equivalence between mean-variance efficient portfolio problems and the corresponding OLS regressions. We then propose a novel end-to-end regularized regression approach, called REP, which can estimate the efficient portfolio weights even when the number of assets N greatly exceeds the sample size T. We show theoretically that the expected return and volatility of the REP portfolio converge to those of the theoretically efficient portfolio at the rate of $\sqrt{\log(N)/T}$, significantly faster than the MAXSER rule of Ao, Li, and Zheng (2019). Our approach thus provides a powerful tool for selecting factors that provide independent information to achieve mean-variance efficiency.

Applying our approach to 174 characteristic-based U.S. equity factors over the OOS period from January 1980 to December 2022, we find that the REP portfolio produces a high annualized OOS Sharpe ratio of 3.76, dominating all other benchmark portfolio rules in our study. Among the 174 factors, AnnouncementReturn and EarningsPredictability are the two most important characteristic factors in terms of their contributions to the portfolio's Sharpe ratio. From a group perspective, the factors constructed from characteristics in the groups of Intangibles, Momentum, and Investment are the most important in improving mean-variance efficiency.

We also examine the pricing performance of the REP portfolio. Empirical results show that the REP portfolio prices the 174 factors well, with a cross-sectional OLS R^2 close to one and insignificant time series regression αs. This evidence suggests that the estimated REP portfolio can serve as a pricing kernel, thus playing the other side of the dual role of an efficient portfolio, i.e., asset pricing. Moreover, our REP approach provides a generic way to test factor models in a high-dimensional setting, and we show that the nine famous factor models in the literature do not add any information to the 174 characteristic portfolios.

We conduct extensive robustness analysis and find that the superior OOS performance of the REP portfolio is robust to the choice of risk aversion levels, estimation windows, and transaction costs. In addition, simulation analysis shows that the REP portfolio outperforms all other competing rules with higher OOS Sharpe ratios under various settings, including the case where $T < N$.

While we focus on the mean-variance efficient portfolio or pricing kernel for the stock market, our approach is general and can be applied to corporate bonds, currencies, and various other asset classes whenever it is of interest to find the efficient portfolio in a high-dimensional space. These will be the future research directions.

References

Ao, M., Y. Li, and X. Zheng. 2019. Approaching mean-variance efficiency for large portfolios. *Review of Financial Studies* 32:2890–919.

Avramov, D., and G. Zhou. 2010. Bayesian portfolio analysis. *Annu. Rev. Financ. Econ.* 2:25–47.

Bai, Z., H. Liu, and W.-K. Wong. 2009. Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory. *Mathematical Finance* 19:639–67.

Barillas, F., and J. Shanken. 2017. Which alpha? *The Review of Financial Studies* 30:1316–38.

———. 2018. Comparing asset pricing models. *Journal of Finance* 73:715–54.

Brandt, M. W., P. Santa-Clara, and R. Valkanov. 2009. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies* 22:3411–47.

Britten-Jones, M. 1999. The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance* 54:655–71.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris. 2009. Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences* 106:12267–72.

Brown, S. J. 1979. *Optimal portfolio choice under uncertainty: A Bayesian approach.* Ph.D. Thesis, University of Chicago.

Carhart, M. M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52:57–82.

Chatterjee, S. 2013. Assumptionless consistency of the lasso. *arXiv preprint arXiv:1303.5817* .

Chen, A. Y., and T. Zimmermann. 2022. Open source cross-sectional asset pricing. *Critical Finance Review* 11:207–64.

Chib, S., L. Zhao, and G. Zhou. 2024. Winners from winners: A tale of risk factors. *Management Science* 70:396–414.

Chinco, A., A. D. Clark-Joseph, and M. Ye. 2019. Sparse signals in the cross-section of returns. *Journal of Finance* 74:449–92.

Cochrane, J. 2009. *Asset pricing: Revised edition.* Princeton University Press.

Cochrane, J. H. 2011. Presidential address: Discount rates. *Journal of Finance* 66:1047–108.

Daniel, K., D. Hirshleifer, and L. Sun. 2020. Short-and long-horizon behavioral factors. *Review of Financial Studies* 33:1673–736.

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.

DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial studies* 22:1915–53.

DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal. 2020. A transaction-cost perspective on the multitude of firm characteristics. *Review of Financial Studies* 33:2180–222.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani. 2004. Least angle regression. *Annals of Statistics* 32:407–51.

EL Karoui, N. 2010. High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation. *Annals of Statistics* 38:3487–566.

Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47:427–65.

———. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33:3–56.

———. 1997. Industry costs of equity. *Journal of Financial Economics* 43:153–93.

———. 2008. Dissecting anomalies. *Journal of Finance* 63:1653–78.

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

———. 2018. Choosing factors. *Journal of Financial Economics* 128:234–52.

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–36.

Freyberger, J., A. Neuhierl, and M. Weber. 2020. Dissecting characteristics nonparametrically. *Review of Financial Studies* 33:2326–77.

Frost, P. A., and J. E. Savarino. 1986. An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis* 21:293–305.

———. 1988. For better performance: Constrain portfolio weights. *Journal of Portfolio Management* 15:29–.

Gibbons, M. R., S. A. Ross, and J. Shanken. 1989. A test of the efficiency of a given portfolio. *Econometrica* 57:1121–52.

Green, J., J. R. Hand, and X. F. Zhang. 2017. The characteristics that provide independent information about average us monthly stock returns. *Review of Financial Studies* 30:4389–436.

Green, R. C., and B. Hollifield. 1992. When will mean-variance efficient portfolios be well diversified? *Journal of Finance* 47:1785–809.

Gu, S., B. Kelly, and D. Xiu. 2020. Empirical asset pricing via machine learning. *Review of Financial Studies* 33:2223–73.

Hansen, L. P., and R. Jagannathan. 1991. Implications of security market data for models of dynamic economies. *Journal of Political Economy* 99:225–62.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. ... and the cross-section of expected returns. *Review of Financial Studies* 29:5–68.

Haugen, R. A., and N. L. Baker. 1996. Commonality in the determinants of expected stock returns. *Journal of Financial Economics* 41:401–39.

He, A., D. Huang, J. Li, and G. Zhou. 2023. Shrinking factor dimension: A reduced-rank approach. *Management Science* 69:5501–22.

Hjalmarsson, E., and P. Manchev. 2012. Characteristic-based mean-variance portfolio choice. *Journal of Banking & Finance* 36:1392–401.

Hou, K., H. Mo, C. Xue, and L. Zhang. 2021. An augmented q-factor model with expected growth. *Review of Finance* 25:1–41.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *Review of Financial Studies* 28:650–705.

———. 2020. Replicating anomalies. *Review of Financial Studies* 33:2019–133.

Jagannathan, R., and T. Ma. 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance* 58:1651–83.

Jobson, J., and B. M. Korkie. 1981. Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance* 36:889–908.

Jobson, J. D., and B. Korkie. 1980. Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association* 75:544–54.

Jorion, P. 1986. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* 21:279–92.

Kan, R., and G. Zhou. 2007. Optimal portfolio choice with parameter uncertainty. *Journal of*

Financial and Quantitative Analysis 42:621–56.

Kelly, B. T., S. Pruitt, and Y. Su. 2019. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics* 134:501–24.

Kozak, S., S. Nagel, and S. Santosh. 2018. Interpreting factor models. *Journal of Finance* 73:1183–223.

———. 2020. Shrinking the cross-section. *Journal of Financial Economics* 135:271–92.

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

———. 2017. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Review of Financial Studies* 30:4349–88.

———. 2022. Quadratic shrinkage for large covariance matrices. *Bernoulli* 28:1519–47.

Leippold, M., Q. Wang, and W. Zhou. 2022. Machine learning in the chinese stock market. *Journal of Financial Economics* 145:64–82.

Lettau, M., and M. Pelger. 2020. Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies* 33:2274–325.

Lewellen, J. 2015. The cross-section of expected stock returns. *Critical Finance Review* 4:1–44.

Light, N., D. Maslov, and O. Rytchkov. 2017. Aggregation of information about the cross section of stock returns: A latent variable approach. *Review of Financial Studies* 30:1339–81.

Lintner, J. 1965. Security prices, risk, and maximal gains from diversification. *Journal of Finance* 20:587–615.

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

Memmel, C. 2003. Performance hypothesis testing with the Sharpe ratio. *Finance Letters* 1.

Michaud, R. O. 1989. The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal* 45:31–42.

Rapach, D. E., J. K. Strauss, and G. Zhou. 2013. International stock return predictability: What is the role of the united states? *Journal of Finance* 68:1633–62.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19:425–42.

Stambaugh, R. F., and Y. Yuan. 2017. Mispricing factors. *Review of Financial Studies* 30:1270–315.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58:267–88.

Tibshirani, R. J. 2013. The lasso problem and uniqueness. *Electronic Journal of Statistics* 7:1456–90.

Tu, J., and G. Zhou. 2011. Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics* 99:204–15.

Zellner, A., and V. K. Chetty. 1965. Prediction and decision problems in regression models from the bayesian point of view. *Journal of the American Statistical Association* 60:608–16.

Table 1: **List of Portfolio Rules**

Table 1 lists the portfolios studied in the empirical study, along with their respective abbreviations. Column 1 gives the descriptions of the studied portfolios and divides them into two categories. The first category encompasses portfolios that come from the literature directly. The second category encompasses the plug-in mean-variance (MV) portfolios whose covariance matrices are estimated via different methods. Column 2 lists the corresponding abbreviations for these portfolios. The REP, the MV-LS, the MV-NLS and the MV-NLS-FF3 portfolios are high-dimensional ones, i.e., feasible when the dimension N exceeds the sample size T. The MAXSER, BJ and MV-SampleCov portfolios are infeasible when $T < N$ and hence a subpool technique of 50 assets is applied to the construction of these portfolios. The construction of the Market, 1/N, EW-CM portfolios has nothing to do with the relatiion between T and N.

(1)	(2)
Portfolio rules	Abbreviation
Regularized Efficient Portfolio	REP
MAXSER of Ao, Li and Zheng (2019)	MAXSER
Market portfolio	Market
Equally-weighted individual stocks	1/N
Equally-weighted CM portfolios combined with the market	EW-CM
OLS regression approach of Britten-Jones (1999)	BJ
MV portfolios with different covariance matrix (cov) estimates	
MV with sample cov	MV-Sample
MV with linear shrinkage cov of Ledoit and Wolf (2004)	MV-LS
MV with nonlinear shrinkage cov of Ledoit and Wolf (2022)	MV-NLS
MV with nonlinear shrinkage cov towards FF3 model of Ledoit and Wolf (2022)	MV-NLS-FF3

Table 2: **Summary of Out-of-sample Performance**

Table 2 summaries the out-of-sample performance for the REP portfolio and the other benchmark portfolios. The out-of-sample period is from January 1980 to December 2022. Column 1 lists the abbrevations of the portfolios studied. Column 2 reports the annualized Sharpe ratios over the out-of-sample period, and each entry of Column 3 reports the p-value of the Sharpe ratio test (Jobson and Korkie, 1981; Memmel, 2003) whose null hypothesis is that the Sharpe ratio of the REP portfolio is no greater than that of the corresponding portfolio listed in Column 1. Column 4-8 document the annualized average returns, annualized standard deviations, maximum drawdowns, skewness and kurtosis of the portfolios over the out-of-sample period.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Portfolios	SR (Ann.)	$p(SR_{REP} \leq SR)$	Mean (Ann.%)	Std (Ann.%)	MaxDD (%)	Skewness	Kurtosis
REP	3.76		78.80	20.97	31.81	−0.14	4.36
MAXSER	3.24	0.00	69.27	21.35	33.40	0.08	4.17
Market	0.53	0.00	8.27	15.65	49.99	−0.66	5.00
1/N	0.40	0.00	7.84	19.68	58.18	−0.58	5.61
EW-CM	0.66	0.00	9.85	15.01	48.44	−0.67	5.15
BJ	0.19	0.00	2.94	15.11	51.59	3.03	63.03
Mean-variance portfolios with different covariance matrix estimates							
MV-Sample	2.34	0.00	104.38	44.69	98.12	−0.05	16.24
MV-LS	3.43	0.00	96.39	28.11	61.75	0.00	6.37
MV-NLS	3.55	0.03	119.11	33.57	56.86	0.03	4.97
MV-NLS-FF3	3.51	0.02	113.93	32.45	45.91	0.10	5.08

Table 3: **Trading Properties of the Portfolio Rules**

Table 3 reports the trading properties of the portfolio and the other benchmark portfolios over the out-of-sample period from January 1980 to December 2022. Column 1 lists the portfolio rules. Columns 2-4 report the average absolute, average maximum, and average minimum positions (in percent) in individual stocks. Column 5-8 report the average leverage and average fractions of long and short positions.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Portfolios	$\Sigma\|w_i\|/N_t$ (%)	$max\, w_i$ (%)	$min\, w_i$ (%)	$\Sigma w_i I(w_i \geq 0)$	$\Sigma w_i I(w_i < 0)$	$\Sigma I(w_i \geq 0)/N_t$ (%)	$\Sigma I(w_i < 0)/N_t$ (%)
REP	0.17	1.74	−1.60	4.18	−3.75	61.82	38.18
MAXSER	0.15	1.38	−1.44	3.64	−3.47	62.57	37.43
Market	0.02	3.15	0.00	1.00	0.00	100.00	0.00
1/N	0.03	0.03	0.03	1.00	0.00	100.00	0.00
EW-CM	0.02	3.15	−0.02	1.04	−0.04	80.32	19.68
BJ	0.05	0.46	−0.48	1.24	−1.24	56.86	43.14
Mean-variance portfolios with different covariance matrix estimates							
MV-Sample	0.34	2.91	−3.24	8.42	−8.20	60.76	39.24
MV-LS	0.17	2.02	−1.55	4.59	−4.04	61.77	38.23
MV-NLS	0.24	2.40	−2.17	6.30	−5.69	60.58	39.42
MV-NLS-FF3	0.28	2.67	−2.60	7.12	−6.44	61.73	38.27

Table 4: **Importance of Characteristics over the Out-of-sample Period**

Table 4 reports reductions of the out-of-sample Sharpe ratios (OOS SR) of the REP portfolios if one of the CM portfolios is removed. Originally, the REP portfolio is constructed based on the market and the 174 CM portfolios, and the original OOS SR is 3.76. Each time a CM portfolios (or a market portfolio) is removed, the REP portfolio based on the remaining portfolios is re-constructed. Column 1 lists the significantly important characteristics during the dropping tests. Column 3 reports the ratios of the SR reductions to the original Sharpe ratio (in percentage). ***, ** and * in Column 3 represent 1%, 5% and 10% significance of the SR tests between the original REP portfolios and the dropped portfolios. Only those characteristics that are at least 10% significant are shown in the table for readability. Column 2 indicates the characteristics that show up for the first time for each group, and the characteristics in Column 1 are ordered first by groups and then by descending SR reductions. Column 5 reports the fractions of the non-missing REP weights of these CM portfolios over the out-of-sample period.

(1)	(2)	(3)	(4)
CM Portfolios	Group	SR Drop (%)	Fraction (%)
AnnouncementReturn	Momentum	7.61***	100.00
RevenueSurprise		1.57**	72.09
IndMom		1.27**	45.16
High52		0.87*	39.34
ResidualMomentum		0.81**	33.72
EntMult	Value vs. Growth	1.98***	42.05
sfe		1.38**	62.02
AnalystValue		0.62*	38.37
PayoutYield		0.46*	29.46
sgr		0.38*	4.26
IntrinsicValue		0.29*	8.53
Leverage		0.18**	11.24
ChNCOL	Investment	1.94***	34.11
ChInvIA		1.83***	59.11
AbnormalAccrualsPercent		0.47**	12.98
NetEquityFinance		0.31*	20.16
FailureProbability	Profitability	1.25**	38.76
roaq		1.15***	32.95
CBOperProf		1.12**	51.36
RetNOA		0.82**	24.81
Tax		0.65***	12.21
OperProf		0.58*	29.46
EarningsPredictability	Intangible	7.67***	71.32
MomSeason06YrPlus		1.72**	77.71
EarningsPersistence		1.27*	27.13
ForecastDispersion		0.97**	34.50
GrAdExp		0.80**	41.47
MomSeason11YrPlus		0.79*	55.43
PriceDelayTstat		0.51**	19.92
PriceDelayRsq		0.48*	32.81
realestate		0.47**	18.60
OrgCap		0.47*	43.99
HerfBE		0.45***	8.72
KZ		0.42**	18.99
AccrualQualityJune		0.21*	8.91
HerfAsset		0.16**	2.52
IdioVolQF	Trading Friction	1.20*	64.24
ReturnSkewCAPM		0.63***	14.73
ReturnSkew3F		0.59**	18.41
betaRR		0.42*	36.82
VolSD		0.29*	9.50
IdioVolCAPM		0.14**	0.97

Table 5: **Contribution to Mean-variance Efficiency: Importance of Groups**

Table 5 documents the reductions of the REP portfolio's out-of-sample Sharpe ratio when the CM portfolios in a single group are all dropped. The original out-of-sample Sharpe ratio of the REP portfolio without dropping any CM portfolios is 3.76. The REP portfolio is re-estimated each time the CM portfolios in the same group are dropped, and the resultant REP portfolios are invested in the same out-of-sample period from January 1980 to December 2022. Column 1 lists the characteristics groups and the market portfolio itself is treated as an independent group. Column 2 reports the Sharpe ratio reductions and Column 3 reports the percentage of the Sharpe ratio reductions with respect to the original Sharpe ratio. Column 4 documents the p-values of the Sharpe ratio tests of Jobson and Korkie (1981) and Memmel (2003), whose null hypotheses are that the original Sharpe ratio (3.76) is lower than those of the REP portfolios after the removal of the CM portfolios. Column 5 reports the out-of-sample average of the non-zero ratios for the seven groups. The non-zero ratio for a specific group in a month is equal to the number of the non-missing REP weights in this group, divided by the number of the CM portfolios with non-missing month-end returns in that group.

Original SR: 3.76				
(1)	(2)	(3)	(4)	(5)
	SR drops	Drop (%)	p-JK	Fraction (%)
Market	0.02	0.46	0.42	98.06
Momentum	0.66	17.53	<0.01	47.13
Value v.s. Growth	0.11	2.86	0.14	34.69
Investment	0.25	6.77	<0.01	22.58
Profitability	0.12	3.26	0.06	26.27
Intangible	0.96	25.52	<0.01	32.19
Trading Friction	0.13	3.38	0.17	32.80

Table 6: **Asset Pricing Performance of REP**

Table 6 documents the cross-sectional asset pricing performance of the REP portfolio, the EW-CM portfolio and the other nine factor models over the balanced sample. The REP and the EW-CM portfolio is estimated based on the balanced sample and the resultant portfolios are then invested in the balanced sample. The time series of the REP returns is treated as a single factor to price the test assets, so is the EW-CM portfolio. The other nine factor models are the CAPM of Sharpe (1964) (CAPM), Fama-French three- (FF-3), five- (FF-5), and six- (FF-6) factor models (Fama and French, 1993, 2015, 2018), the four-factor model of Carhart (1997) (Carhart-4), the q-factor model of Hou et al. (2021) (HXZ-q5), the mispricing factor model of Stambaugh and Yuan (2017) (SY-4), the behavioral factor model of Daniel, Hirshleifer, and Sun (2020) (DHS-3), and the union of all models (ALL). Two sets of test assets are considered. The first set of test assets,reported in Column 2-5, is the 174 CM portfolios. The second set, reported in Column 6-9, is the 174 characteristic-based portfolio plus the REP portfolio and the factors in the ALL factor model. Measures of asset pricing performance for each set of test assets include adjusted OLS-R^2, GLS-R^2, the average absolute of the α's and the average absolute of the t-statistics.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	CM portfolios				CM portfolios+Factors			
Factors	OLS R^2_{adj} (%)	GLS R^2 (%)	$A(\lvert\alpha_i\rvert)$	$A(\lvert t_{\alpha_i}\rvert)$	OLS R^2_{adj} (%)	GLS R^2 (%)	$A(\lvert\alpha_i\rvert)$	$A(\lvert t_{\alpha_i}\rvert)$
REP	88.45	34.94	0.06	0.89	97.04	21.55	0.07	1.30
EW-CM	31.65	0.00	0.17	3.70	1.14	0.45	0.25	3.91
CAPM	33.78	0.01	0.17	3.71	1.85	0.29	0.24	3.91
FF-3	33.22	0.05	0.15	3.80	3.68	0.63	0.21	3.99
Carhart-4	52.44	1.06	0.12	3.14	17.06	1.14	0.17	3.33
FF-5	57.11	2.08	0.11	3.04	32.58	1.78	0.16	3.21
FF-6	60.81	2.62	0.09	2.75	32.28	2.04	0.14	3.03
HXZ-q5	77.83	6.23	0.07	1.89	55.32	4.73	0.11	1.99
SY-4	66.45	3.90	0.87	2.01	34.77	3.41	0.12	2.15
DHS-3	62.62	8.17	0.09	2.05	43.68	3.30	0.13	2.22
ALL	81.69	15.73	0.08	2.11	73.46	6.25	0.10	2.52

Table 7: **Comparing and Testing Asset Pricing Models**

Table 7 summarizes the difference in the out-of-sample Sharpe ratios between the REP portfolios constructed either by the factors from a given asset pricing model in the literature, or by the 174 characteristic-based factors plus the former. Column 1 lists the nine factor models. They are the CAPM of Sharpe (1964) (CAPM), Fama-French three (FF-3), five- (FF-5), and six- (FF-6) factor models (Fama and French, 1993, 2015, 2018), the four-factor model of Carhart (1997) (Carhart-4), the q-factor model of Hou et al. (2021) (HXZ-q5), the mispricing factor model of Stambaugh and Yuan (2017) (SY-4), the behavioral factor model of Daniel, Hirshleifer, and Sun (2020) (DHS-3), and the union of all models (ALL). Column 2 report the out-of-sample periods of these factor models where the initial estimation window is 120 months. The out-of-sample periods can be different across models because of the data availability. Column 3 reports the length of the out-of-sample periods for each factor model. Column 4 reports the out-of-sample Sharpe ratios of the REP rule based on the factors from a given factor model. Column 5 reports the out-of-sample Sharpe ratios of the REP rule based on the 174 characteristic-based factors plus the factors from a given factor model. Column 6 reports the p-values of the Sharpe ratio tests of Jobson and Korkie (1981) with the corrections of Memmel (2003), whose null hypotheses are that the Sharpe ratios in Column 4 are greater than those in Column 5.

(1)	(2)	(3)	(4)	(5)	(6)
Factor models	Time periods of factor models	Length of factor time series	Ann.SR factors	Ann.SR factors + CM	p-JK
CAPM	1980.01-2022.12	516	0.34	3.76	0.00
FF-3	1980.01-2022.12	516	0.51	3.75	0.00
FF-4	1980.01-2022.12	516	0.76	3.76	0.00
FF-5	1980.01-2022.12	516	1.06	3.67	0.00
FF-6	1980.01-2022.12	516	1.10	3.64	0.00
HXZ-q5	1980.01-2022.12	516	1.70	3.68	0.00
SY-4	1980.01-2016.12	444	1.55	4.25	0.00
DHS-3	1982.07-2022.12	486	1.46	3.69	0.00
ALL	1982.07-2016.12	414	1.97	4.26	0.00

Table 8: **Robustness Analysis**

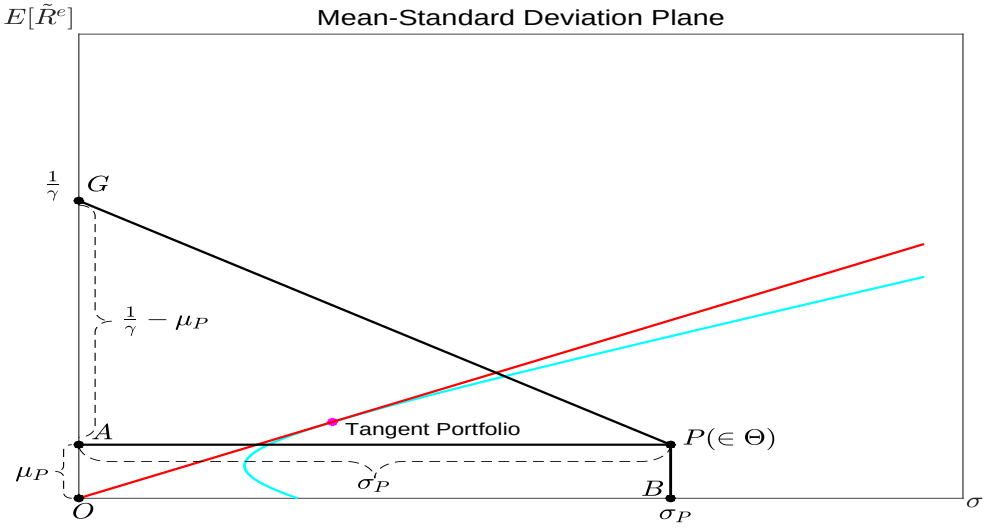
Table 8 aggregates the out-of-sample performance of the REP portfolios constructed based on different risk-aversion levels, different estimation windows and different rebalanced frequency. Results in Penal A do not consider transaction costs while results in Panel B do, and the transaction cost is set to 60bps for each stock. In each panel, Column 1 first lists the evaluation measures and they are: Sharpe ratios, average returns, standard deviations, maximum drawdowns, skewness, kurtosis, annualized average turnover (Panel B only). Column 2 reports the performance of the REP portfolio in the baseline case. Column 3 and 4 report the robustness results of different risk-aversion levels. Column 5-7 report the robustness results of different estimation windows. The out-of-sample periods of the $T = 80$ and the recursive cases start from January 1980, the same to the baseline case, while that of the $T = 240$ case starts from January 1990. Column 8-10 report the performance of the REP portfolios rebalanced quarterly, semi-annually and annually.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Without Transaction Costs									
	Baseline	Risk-aversion		Estimation Window			Rebalancing Frequency		
	$\gamma = 10$ $T = 120$ Monthly	$\gamma = 5$	$\gamma = 15$	$T = 80$	$T = 240$	Recursive	Qtly	Semi-Ann.	Ann.
SR	3.76	3.76	3.76	3.33	3.23	3.39	2.80	2.19	1.45
Mean (%)	78.80	157.60	52.53	76.24	71.70	78.76	57.06	42.94	29.52
Std (%)	20.97	41.94	13.98	22.86	22.19	23.27	20.35	19.58	20.33
MaxDD (%)	31.81	56.18	22.04	40.51	48.32	21.51	36.78	55.92	75.47
Skewness	−0.14	−0.14	−0.14	−0.05	0.23	2.57	−0.01	−0.18	−0.51
Kurtness	4.36	4.36	4.36	6.67	10.72	29.46	4.61	4.73	8.80
Panel B: With Transaction Costs									
	Baseline	Risk-aversion		Estimation Window			Rebalancing Frequency		
	$\gamma = 10$ $T = 120$ Monthly	$\gamma = 5$	$\gamma = 15$	$T = 80$	$T = 240$	Recursive	Qtly	Semi-Ann.	Ann.
SR	1.81	1.75	1.83	1.52	1.64	1.13	1.71	1.52	1.10
Mean (%)	37.25	70.28	25.23	34.06	35.74	31.40	38.76	32.47	23.52
Std (%)	20.54	40.20	13.78	22.43	21.75	27.77	22.65	21.37	21.48
MaxDD (%)	66.65	91.77	50.51	65.19	55.92	97.45	55.89	66.13	80.54
Skewness	−0.21	−0.25	−0.21	−0.17	0.10	−4.15	0.03	−0.18	−0.44
Kurtness	4.52	4.59	4.51	6.63	11.21	43.41	3.84	4.18	7.40
Turnover	64.93	128.96	43.53	66.08	56.64	73.54	29.48	17.07	9.86

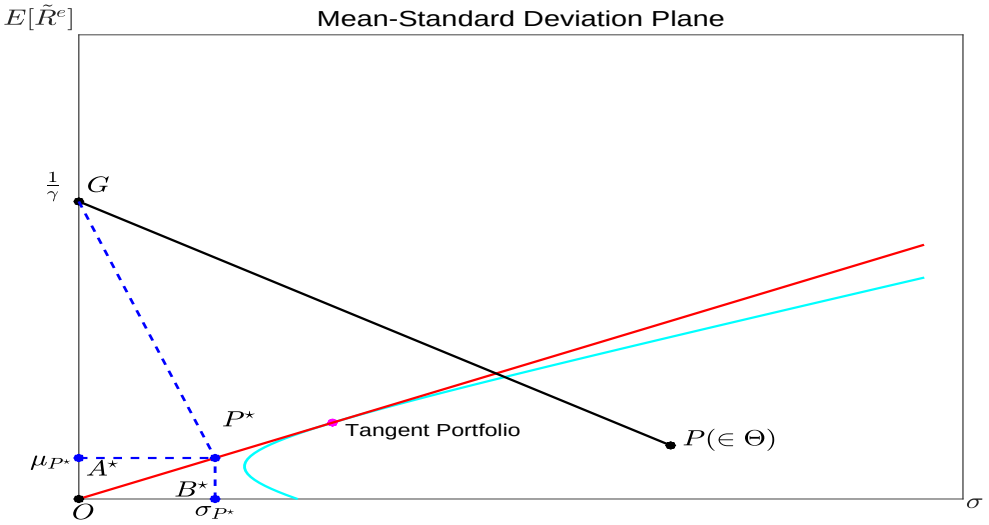
Table 9: Simulation Results

Table 9 reports the results of the simulation for $S = 10,000$ times. The study simulates excess returns for $N = 100$ assets based on Fama-French 3 factor models, as described in Ao, Li, and Zheng (2019). For a sample size T, a total of $T + 1$ observations are simulated for the 100 assets in each simulation. The first T observations are used to construct the REP and the benchmark portfolios. Subsequently, the constructed portfolios are invested in the $T + 1$-th observation, resulting in a time series of $L = 10,000$ "out-of-sample" returns for each portfolio. The sample sizes considered include $T = 80, 120$ and 240. For each case of sample size, the reported outcomes are the annualized OOS Sharpe ratios and the p-values of the Sharpe ratio test by Jobson and Korkie (1981) with the correction of Memmel (2003), whose null hypotheses are that the Sharpe ratio of the REP portfolio is lower than those of the benchmark portfolios. The REP portfolio IS constructed based on a risk aversion $\gamma = 10$, while the MAXSER, MV-PlugIn, MV-LS, MV-NLS, and MV-NLS-FF3 portfolios are constructed based on a given risk level $\sigma = 0.04$. The scales of γ and σ have no impact of the Sharpe ratios. The construction of the BJ portfolio does not necessitate the pre-specification of either γ or σ. The theoretical annualized maximium Sharpe ratio is 1.5217.

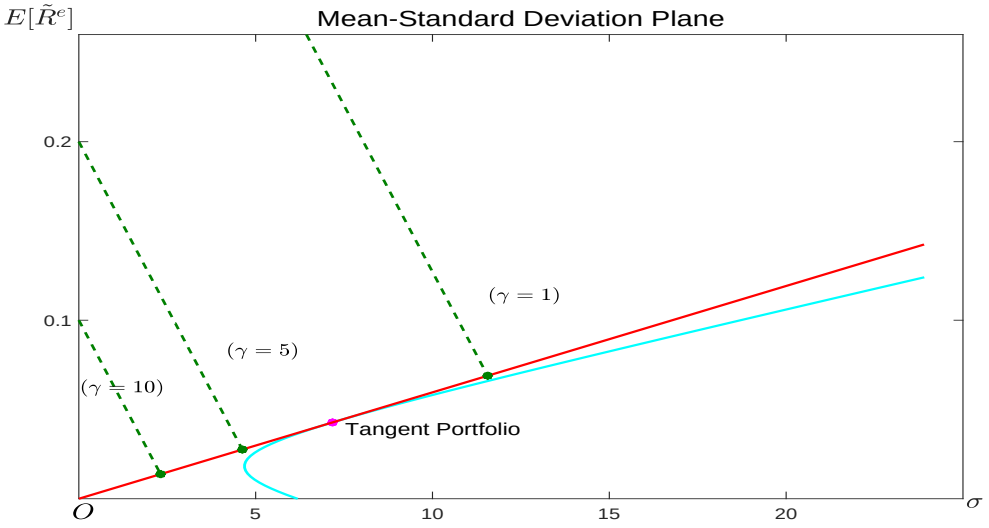
The theoretical maximum Sharpe ratio: 1.5217						
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Portfolio	$T = 80$		$T = 120$		$T = 240$	
($\gamma = 10$ or $\sigma = 0.04$)	Sharpe ratio	p-value	Sharpe ratio	p-value	Sharpe ratio	p-value
REP	0.6475	-	0.9258	-	1.1384	-
MAXSER	0.5658	0.0010	0.7128	< 0.0001	1.0426	< 0.0001
1/N	0.4601	< 0.0001	0.4593	< 0.0001	0.4628	< 0.0001
BJ	0.0484	< 0.0001	-0.0318	< 0.0001	0.0264	< 0.0001
Mean-variance portfolios with different covariance matrix estimates						
MV-Sample	0.3664	< 0.0001	0.3113	< 0.0001	0.7114	< 0.0001
MV-LS	0.4770	< 0.0001	0.5461	< 0.0001	0.7655	< 0.0001
MV-NLS	0.5617	< 0.0001	0.6912	< 0.0001	0.8563	< 0.0001
MV-NLS-FF3	0.5995	0.0333	0.7108	< 0.0001	0.8866	< 0.0001



(a)



(b)



(c)

Figure 1: **Equivalence between the Expected Quadratic Utility Maximization and the OLS Regression**

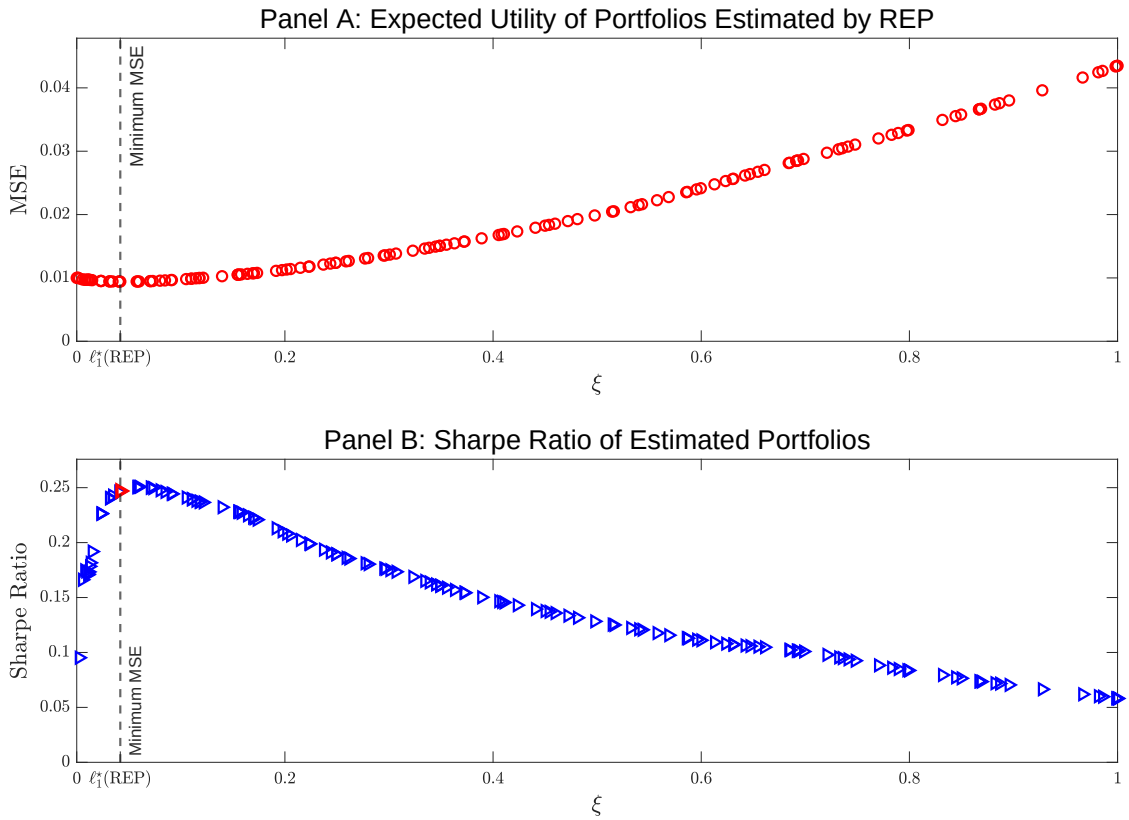
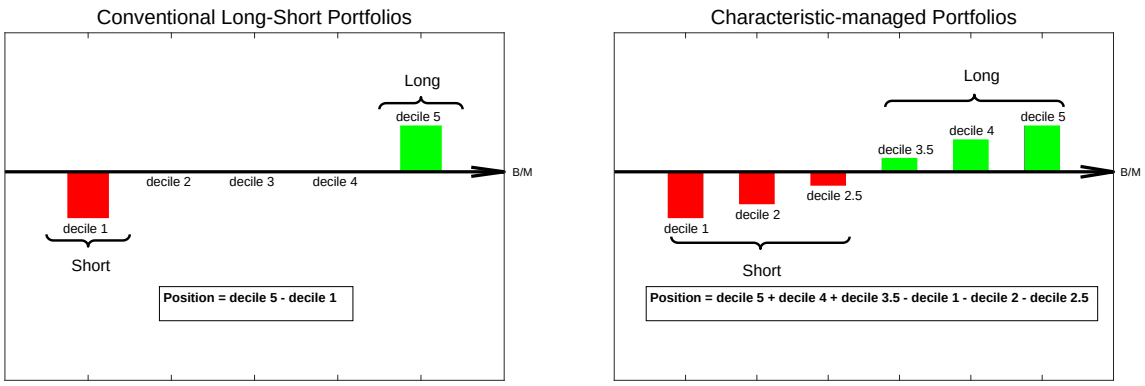


Figure 2: **Expected Quadratic Utility and Sharpe Ratios of Estimated Portfolios on LASSO Solution Path**

Figure 2 plots the expected quadratic utility and Sharpe ratio of all estimated portfolios on the LASSO solution path estimated by the LARS algorithm. The sample contains 120 i.i.d. observations of excess returns on 100 stocks simulated from the multivariate normal distribution with mean and covariance matrix calibrated from real data. The x-axis is the ratio of the ℓ_1 norm of an estimated solution to that of the OLS solution with the minimum ℓ_1 norm. The vertical line in both panels indicates the optimal value of the hyperparameter ξ, which is the ℓ_1 ratio of the estimated portfolio with the minimum mean squared error. The expected quadratic utility and Sharpe ratios are computed based on the mean and covariance matrix of 100 stocks.



(a) B/M Long-short Portfolio (b) B/M-managed Portfolio

Figure 3: **Illustrations of Difference between Long-short and CM Portfolios**

Figure 3 illustrates visibly the difference between the CM portfolios and the conventional long-short portfolios. Figure 3 (a) refers to the long-short portfolio sorted by the book-to-market ratios that longs and shorts stocks in both extreme deciles. Figure 3 (b) illustrates that holding a B/M-managed portfolio is equivalent to longing those stocks whose book-to-market ratios exceed the *mean* of the cross section and shorting those below the *mean*. The positions in these cross-sectional stocks are weighted-averaged by the absolute values of the corresponding (normalized) book-to-market ratios.

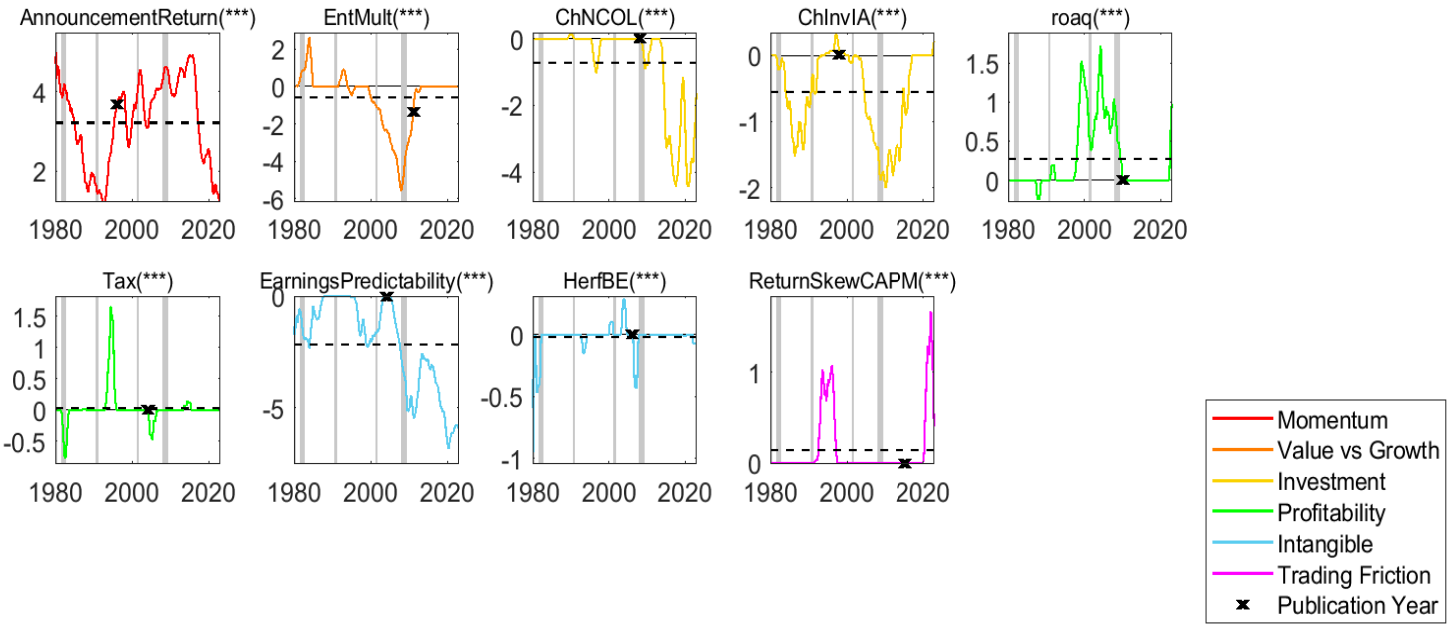


Figure 4: Out-of-sample Evolution of the REP Weights, ***

Figure 4 plots the out-of-sample evolution of the REP weights in the CM portfolios whose corresponding out-of-sample Sharpe ratio reductions are 1% significant (the significance reported in Table 4). The different colors of the plot lines indicate the groups to which the corresponding characteristics belong. The black dashed horizontal lines in the subplots are the levels of the average values of the corresponding REP weights over the out-of-sample period. The black crosses (if any) represent the publication years of the corresponding literature. The shaded areas represent the NBER recessions.

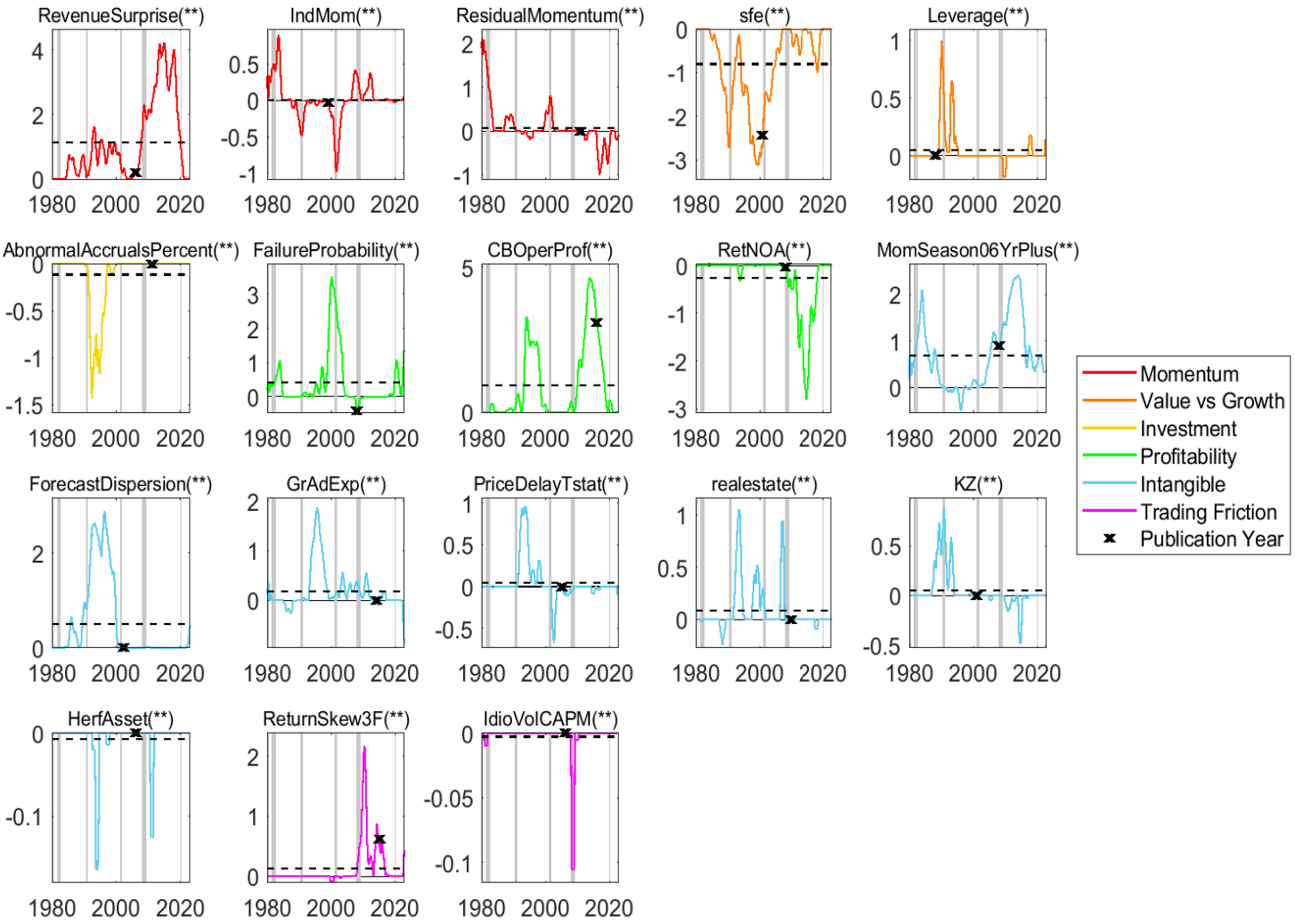


Figure 5: Out-of-sample Evolution of the REP Weights, **

Figure 5 plots the out-of-sample evolution of the REP weights in the CM portfolios whose corresponding out-of-sample Sharpe ratio reductions are 5% significant (the significance reported in Table 4). The different colors of the plot lines indicate the groups to which the corresponding characteristics belong. The black dashed horizontal lines in the subplots are the levels of the average values of the corresponding REP weights over the out-of-sample period. The black crosses (if any) represent the publication years of the corresponding literature. The shaded areas represent the NBER recessions.

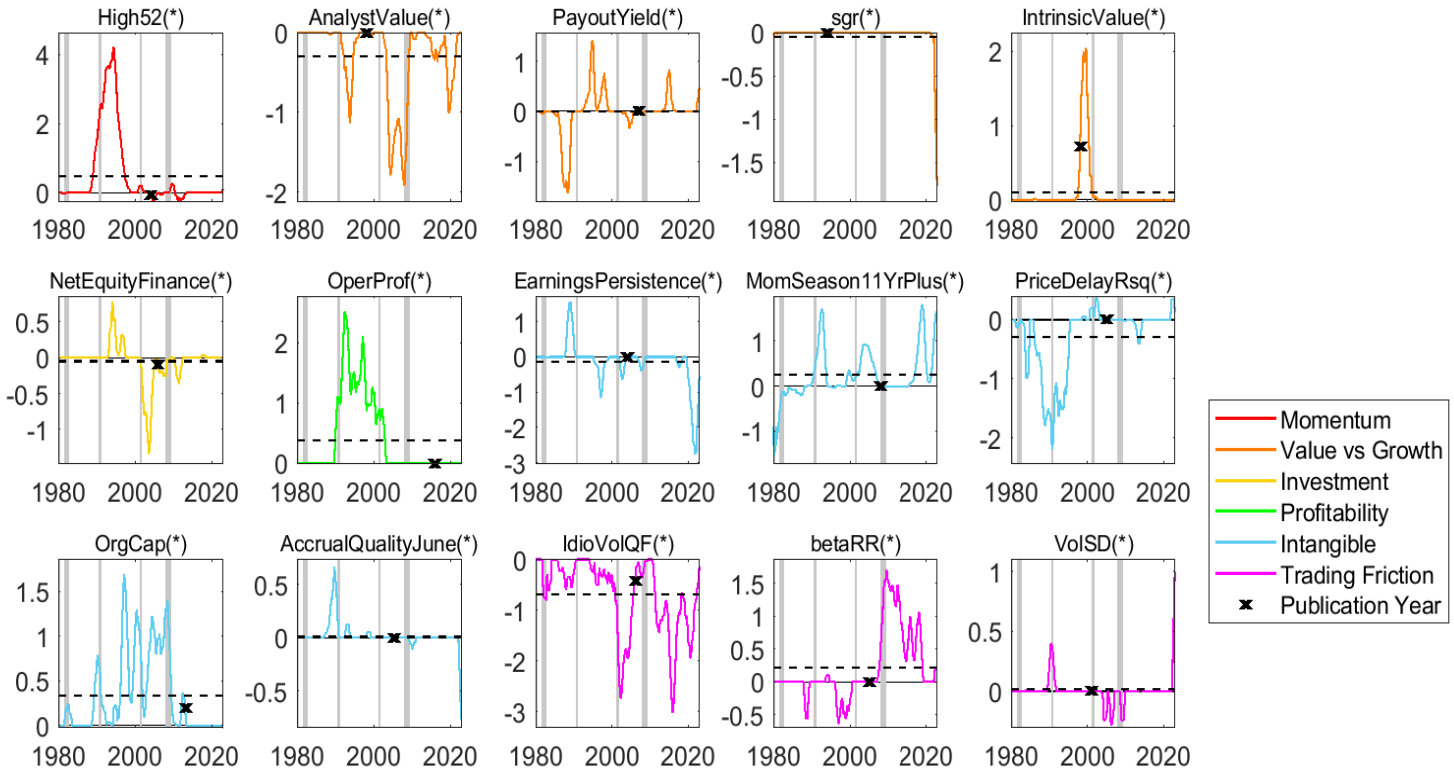


Figure 6: Out-of-sample Evolution of the REP Weights, *

Figure 6 plots the out-of-sample evolution of the REP weights in the CM portfolios whose corresponding out-of-sample Sharpe ratio reductions are 10% significant (the significance reported in Table 4). The different colors of the plot lines indicate the groups to which the corresponding characteristics belong. The black dashed horizontal lines in the subplots are the levels of the average values of the corresponding REP weights over the out-of-sample period. The black crosses (if any) represent the publication years of the corresponding literature. The shaded areas represent the NBER recessions.

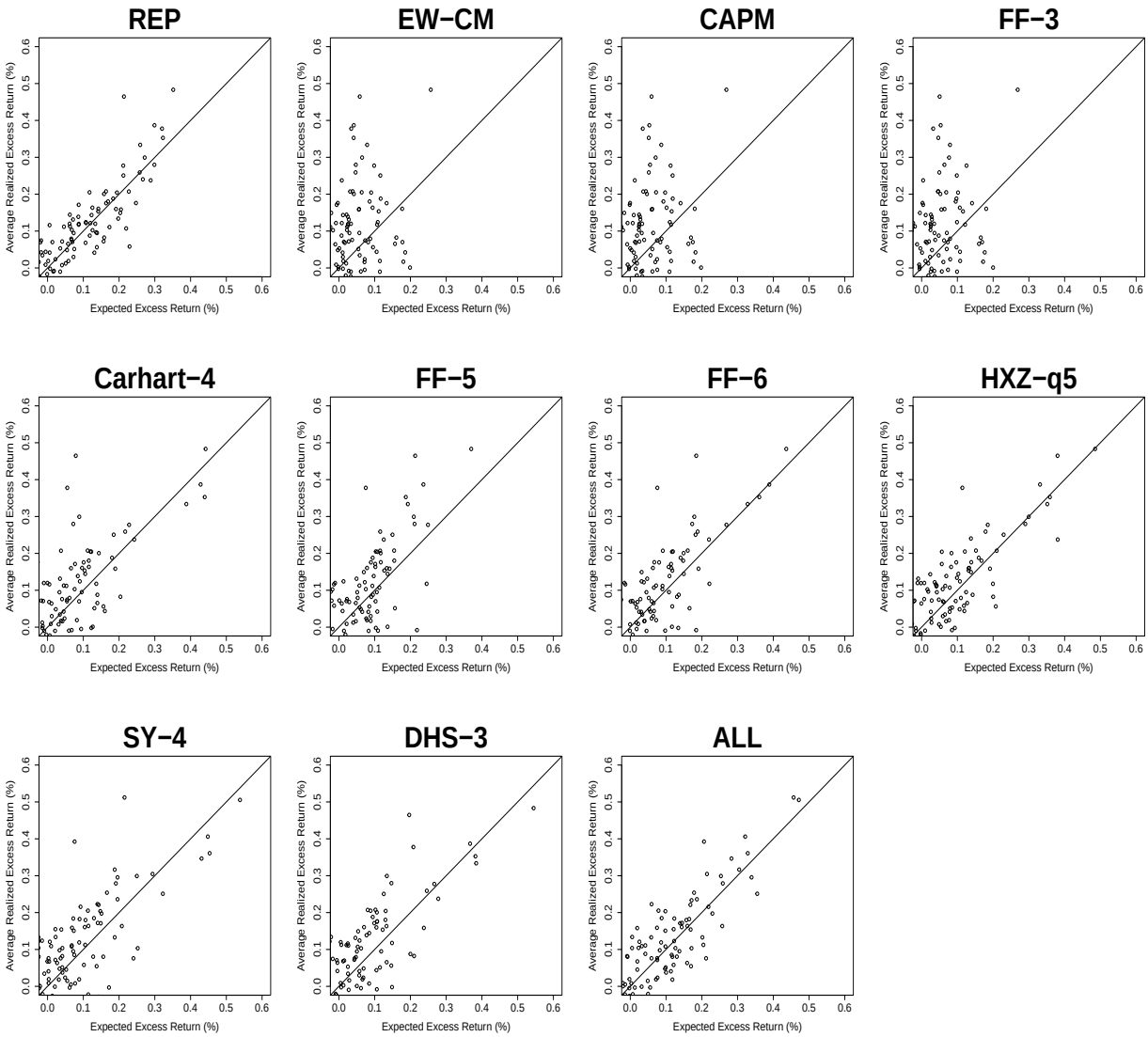


Figure 7: **Pricing the 174 CM Portfolios Cross-sectionally**

Figure 7 plots the performance of the REP, the EW-CM portfolios and the nine factor models in cross-sectionally pricing the 174 CM portfolios in the balanced sample. The vertical axis of each sub-figure is the average realized returns of the 174 CM portfolios, while the horizontal axis is their expected returns predicted by the the REP, the EW-CM portfolios and the nine factor models. The nine factor models are: the CAPM model of Sharpe (1964) (CAPM), Fama-French 3 (FF-3), 5 (FF-5) and 6 (FF-6) factor models (Fama and French, 1992, 2015, 2018), the 4 factor model of Carhart (1997) (Carhart-4), the q factor model of Hou et al. (2021) (HXZ-q5), the mispricing factor model of Stambaugh and Yuan (2017) (SY-4), the behavioral factor model of Daniel, Hirshleifer, and Sun (2020) (DHS-3), and the union factor model (ALL). The full sample is from January 1980 to December 2022 and, the balanced sample is obtained by removing those month observations from the full sample if any one of the CM portfolios has missing returns. The REP and the EW-CM portfolios are estimated once using the balanced sample and then the estimated weights are invested in the balanced sample to obtain the time series of returns for the two portfolios. The time series of the returns are treated as two separate factors to price the test assets.

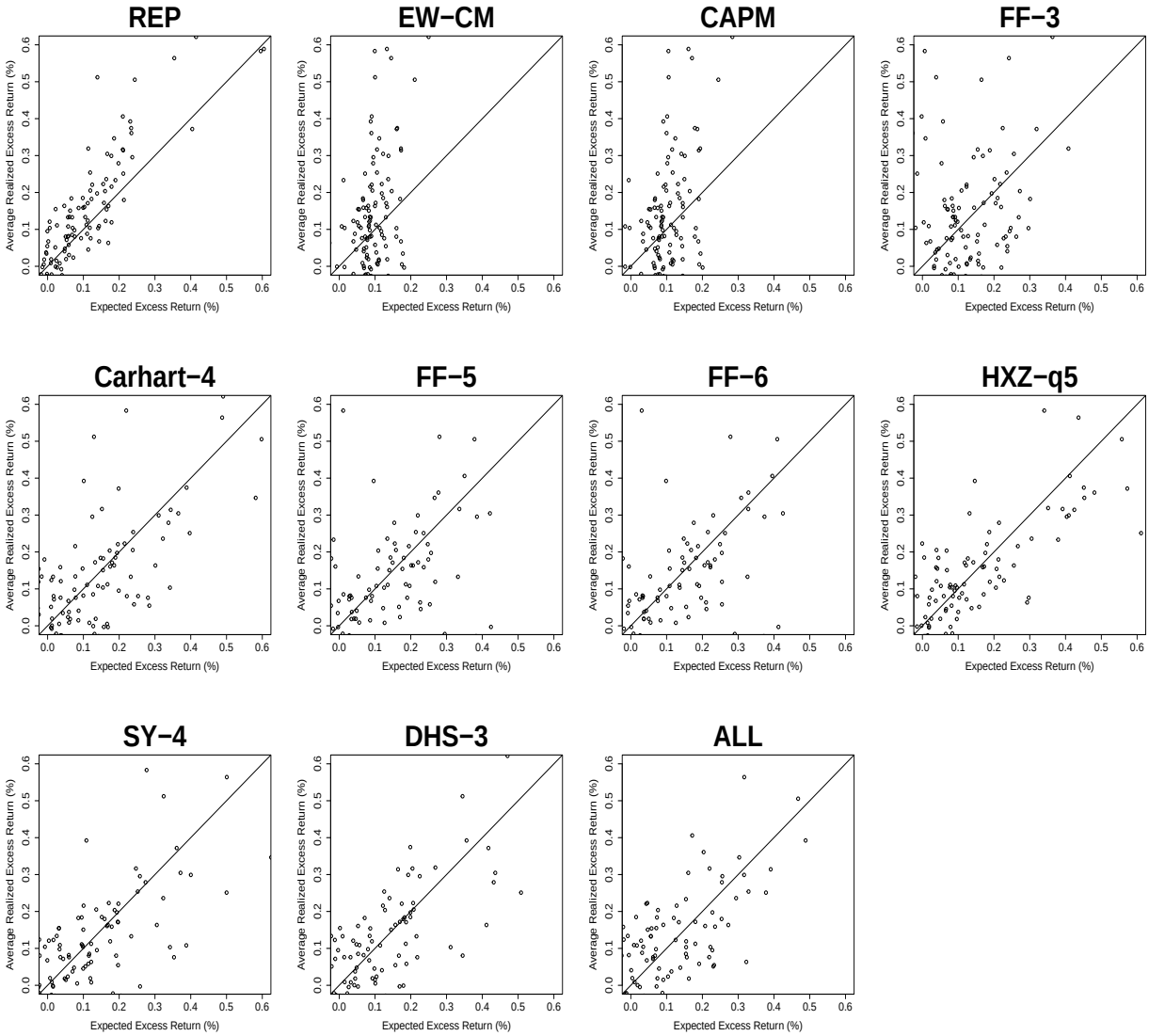


Figure 8: **Pricing the ALL Factors and the 174 CM Portfolios Cross-sectionally**

Figure 8 plots the performance of the REP, the EW-CM portfolios and the nine factor models in cross-sectionally pricing the 174 CM portfolios plus the union factors in the balanced sample. The vertical axis of each sub-figure is the average realized returns of the 174 CM portfolios, while the horizontal axis is their expected returns predicted by the the REP, the EW-CM portfolios and the nine factor models. The nine factor models are: the CAPM model of Sharpe (1964) (CAPM), Fama-French 3 (FF-3), 5 (FF-5) and 6 (FF-6) factor models (Fama and French, 1992, 2015, 2018), the 4 factor model of Carhart (1997) (Carhart-4), the q factor model of Hou et al. (2021) (HXZ-q5), the mispricing factor model of Stambaugh and Yuan (2017) (SY-4), the behavioral factor model of Daniel, Hirshleifer, and Sun (2020) (DHS-3), and the union factor model (ALL). The full sample is from January 1980 to December 2022 and, the balanced sample is obtained by removing those month observations from the full sample if any one of the CM portfolios has missing returns. The REP and the EW-CM portfolios are estimated once using the balanced sample and then the estimated weights are invested in the balanced sample to obtain the time series of returns for the two portfolios. The time series of the returns are treated as two separate factors to price the test assets.

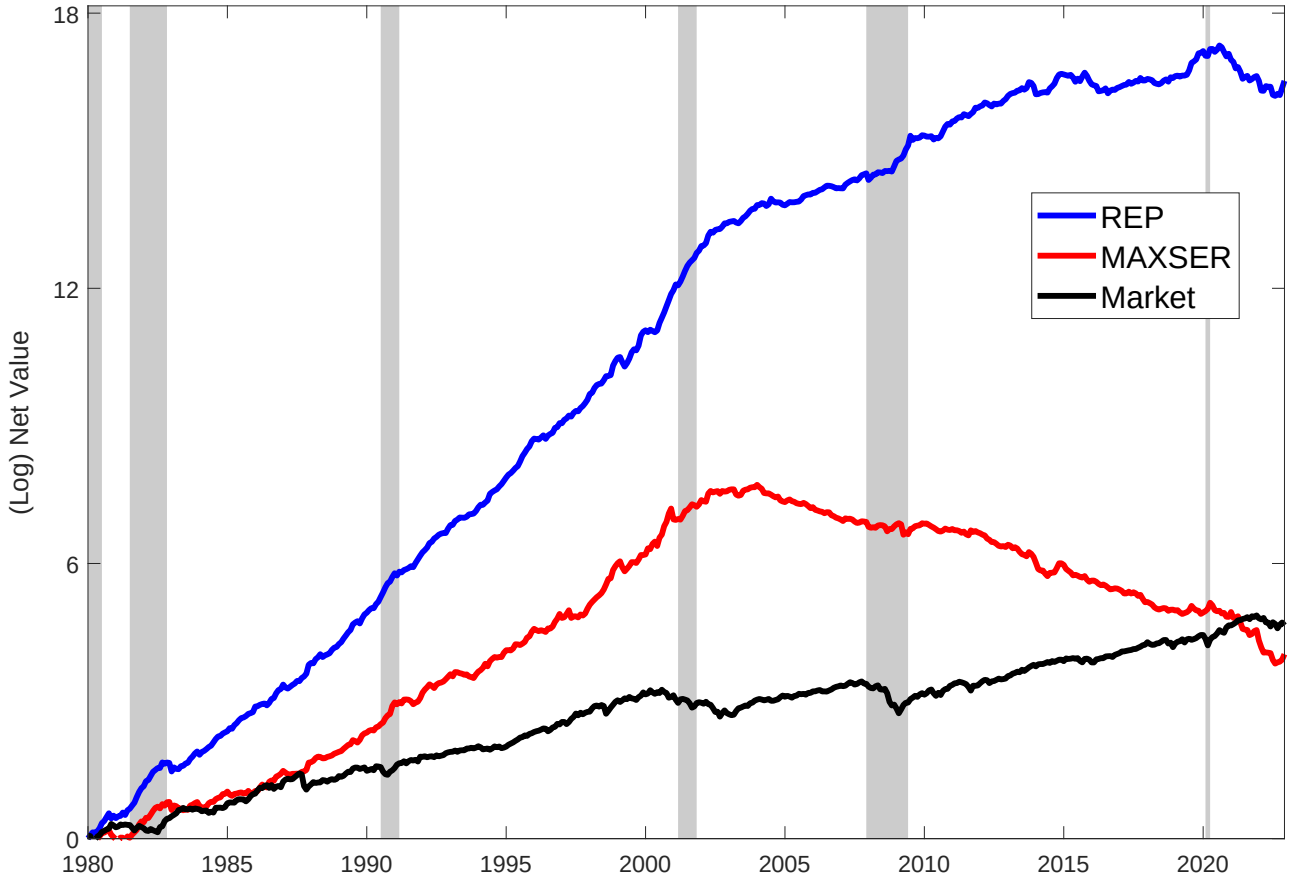


Figure 9: **Log Accumulative Wealth Evolution (with Transaction Costs)**

Figure 9 plots the log accumulative wealth evolution with transaction costs for the REP, the MAXSER and the market portfolio over the out-of-sample period from January 1980 to December 2022. The transaction cost is set to 60bps for each stock. Gross returns are computed using the risk-free rates obtained from Fama and French (1992). The shaded areas represent the NBER recessions.

Appendices

A.1 Supplementary Material

Table A.1: Basic Information of Characteristics

(1)	(2)	(3)	(4)
Acronym	Brief Introduction	Literature	Group
EarningsSurprise	Earnings Surprise	Foster, Olsen, and Shevlin (1984)	1
AnnouncementReturn	Earnings announcement return	Chan, Jegadeesh, and Lakonishok (1996)	1
REV6	Earnings forecast revisions	Chan, Jegadeesh, and Lakonishok (1996)	1
Mom6m	Momentum (6 month)	Jegadeesh and Titman (1993)	1
Mom12m	Momentum (12 month)	Jegadeesh and Titman (1993)	1
IndMom	Industry Momentum	Grinblatt and Moskowitz (1999)	1
RevenueSurprise	Revenue Surprise	Jegadeesh and Livnat (2006)	1
ChTax	Change in Taxes	Thomas and Zhang (2011)	1
AnalystRevision	EPS forecast revision	Hawkins, Chamberlin, Daniel (1984)	1
NumEarnIncrease	Earnings streak length	Loh and Warachka (2012)	1
High52	52 week high	George and Hwang (2004)	1
ResidualMomentum6m	6 month residual momentum	Blitz, Huij and Martens (2011)	1
ResidualMomentum	Momentum based on FF3 residuals	Blitz, Huij and Martens (2011)	1
retConglomerate	Conglomerate return	Cohen and Lou (2012)	1
BEME	Book-to-market ratio	Rosenberg, Reid, and Lanstein (1985)	2
AM	Total assets to market	Fama and French (1992)	2
LRreversal	Long-run reversal	De Bondt and Thaler (1985)	2
EP	Earnings-to-Price Ratio	Basu (1977)	2
sfe	Earnings Forecast to price	Elgers, Lo and Pfeiffer (2001)	2
CF	Cash flow to market	Lakonishok, Shleifer, Vishny (1994)	2
DivYieldAnn	Last year's dividends over price	Naranjo, Nimalendran, Ryngaert (1998)	2
PayoutYield	Payout Yield	Boudoukh et al. (2007)	2
NetPayoutYield	Net Payout Yield	Boudoukh et al. (2007)	2
MeanRankRevGrowth	Revenue Growth Rank	Lakonishok, Shleifer, Vishny (1994)	2
sgr	Annual sales growth	Lakonishok, Shleifer, Vishny (1994)	2
EntMult	Enterprise Multiple	Loughran and Wellman (2011)	2
SP	Sales-to-price	Barbee, Mukherji and Raines (1996)	2
cfp	Operating Cash flows to price	Desai, Rajgopal, Venkatachalam (2004)	2
IntanBM	Intangible return using BM	Daniel and Titman (2006)	2
IntrinsicValue	Intrinsic or historical value	Frankel and Lee (1998)	2
AnalystValue	Analyst Value	Frankel and Lee (1998)	2
EBM	Enterprise component of BM	Penman, Richardson and Tuna (2007)	2
NetDebtPrice	Net debt to price	Penman, Richardson and Tuna (2007)	2
EquityDuration	Equity Duration	Dechow, Sloan and Soliman (2004)	2
Leverage	Market leverage	Bhandari (1988)	2
Investment	Investment to revenue	Titman, Wei and Xie (2004)	2
AssetGrowth	Asset growth	Cooper, Gulen and Schill (2008)	3

Continued on next page

Table A.1 – continued from previous page

(1)	(2)	(3)	(4)
Acronym	Brief Introduction	Literature	Group
InvestPPEInv	change in ppe and inv/assets	Lyandres, Sun and Zhang (2008)	3
NOA	Net Operating Assets	Hirshleifer et al. (2004)	3
dNoa	change in net operating assets	Hirshleifer, Hou, Teoh, Zhang (2004)	3
GrLTNOA	Growth in long term operating assets	Fairfield, Whisenant and Yohn (2003)	3
grcapx1y	Investment growth (1 year)	Anderson and Garcia-Feijoo (2006)	3
grcapx	Change in capex (two years)	Anderson and Garcia-Feijoo (2006)	3
grcapx3y	Change in capex (three years)	Anderson and Garcia-Feijoo (2006)	3
ShareIss1Y	Share issuance (1 year)	Pontiff and Woodgate (2008)	3
ChInvIA	Change in capital inv (ind adj)	Abarbanell and Bushee (1998)	3
CompEquIss	Composite equity issuance	Daniel and Titman (2006)	3
CompositeDebtIssuance	Composite debt issuance	Lyandres, Sun and Zhang (2008)	3
InvGrowth	Inventory Growth	Belo and Lin (2012)	3
ChInv	Inventory Growth	Thomas and Zhang (2002)	3
Accruals	Accruals	Sloan (1996)	3
TotalAccruals	Total accruals	Richardson et al. (2005)	3
DelCOA	Change in current operating assets	Richardson et al. (2005)	3
DelCOL	Change in current operating liabilities	Richardson et al. (2005)	3
ChNNCOA	Change in Net Noncurrent Op Assets	Soliman (2008)	3
ChNCOA	Change in Noncurrent Operating Assets	Soliman (2008)	3
ChNCOL	Change in Noncurrent Operating Liab	Soliman (2008)	3
DelNetFin	Change in net financial assets	Richardson et al. (2005)	3
DelSTI	Change in short-term investment	Richardson et al. (2005)	3
DelLTI	Change in long-term investment	Richardson et al. (2005)	3
DelFINL	Change in financial liabilities	Richardson et al. (2005)	3
DelEqu	Change in equity to assets	Richardson et al. (2005)	3
AbnormalAccruals	Abnormal Accruals	Xie (2001)	3
PctAcc	Percent Operating Accruals	Hafzalla, Lundholm, VanWinkle (2011)	3
AbnormalAccrualsPercent	Percent Abnormal Accruals	Hafzalla, Lundholm, VanWinkle (2011)	3
XFIN	Net external financing	Bradshaw, Richardson, Sloan (2006)	3
NetEquityFinance	Net equity financing	Bradshaw, Richardson, Sloan (2006)	3
NetDebtFinance	Net debt financing	Bradshaw, Richardson, Sloan (2006)	3
RoE	net income / book equity	Haugen and Baker (1996)	4
ChangeRoE	Change in Return on equity	Balakrishnan, Bartov and Faurel (2010)	4
roaq	Return on assets (qtrly)	Balakrishnan, Bartov and Faurel (2010)	4
ChangeRoA	Change in Return on assets	Balakrishnan, Bartov and Faurel (2010)	4
RetNOA	Return on Net Operating Assets	Soliman (2008)	4
PM	Profit Margin	Soliman (2008)	4
AssetTurnover	Asset Turnover	Soliman (2008)	4
CapTurnover	Capital turnover	Haugen and Baker (1996)	4
GP	gross profits / total assets	Novy-Marx (2013)	4
GPlag	gross profits / total assets	Novy-Marx (2013)	4
OperProf	operating profits / book equity	Ball et al. (2016)	4
OperProfLag	operating profits / book equity	Fama and French (2006)	4
OperProfRD	Operating profitability R&D adjusted	Ball et al. (2016)	4

Continued on next page

Table A.1 – continued from previous page

(1)	(2)	(3)	(4)
Acronym	Brief Introduction	Literature	Group
OperProfRDLagAT	Oper prof R&D adj lagged assets	Ball et al. (2016)	4
CBOperProf	Cash-based operating profitability	Ball et al. (2016)	4
CBOperProfLagAT	Cash-based oper prof lagged assets	Ball et al. (2016)	4
PS	Piotroski F-score	Piotroski (2000)	4
FailureProbability	Failure probability	Campbell, Hilscher and Szilagyi (2008)	4
FailureProbabilityJune	Failure probability	Campbell, Hilscher and Szilagyi (2008)	4
ZScore	Altman Z-Score	Dichev (1998)	4
Tax	Taxable income to income	Lev and Nissim (2004)	4
BookLeverage	Book leverage (annual)	Fama and French (1992)	4
OrgCapNoAdj	Org cap w/o industry adjustment	Eisfeldt and Papanikolaou (2013)	5
OrgCap	Organizational capital	Eisfeldt and Papanikolaou (2013)	5
AdExp	Advertising Expense	Chan, Lakonishok and Sougiannis (2001)	5
GrAdExp	Growth in advertising expenses	Lou (2014)	5
RD	R&D over market cap	Chan, Lakonishok and Sougiannis (2001)	5
rd_sale	R&D to sales	Chan, Lakonishok and Sougiannis (2001)	5
OPLeverage	Operating leverage	Novy-Marx (2010)	5
hire	Employment growth	Bazdresch, Belo and Lin (2014)	5
BrandInvest	Brand capital investment	Belo, Lin and Vitorino (2014)	5
AOP	Analyst Optimism	Frankel and Lee (1998)	5
Herf	Industry concentration (sales)	Hou and Robinson (2006)	5
HerfAsset	Industry concentration (assets)	Hou and Robinson (2006)	5
HerfBE	Industry concentration (equity)	Hou and Robinson (2006)	5
FirmAge	Firm age based on CRSP	Barry and Brown (1984)	5
PriceDelayRsq	Price delay r square	Hou and Moskowitz (2005)	5
PriceDelaySlope	Price delay coeff	Hou and Moskowitz (2005)	5
PriceDelayTstat	Price delay SE adjusted	Hou and Moskowitz (2005)	5
GrSaleToGrInv	Sales growth over inventory growth	Abarbanell and Bushee (1998)	5
GrSaleToGrReceivables	Change in sales vs change in receiv	Abarbanell and Bushee (1998)	5
GrGMToGrSales	Gross margin growth to sales growth	Abarbanell and Bushee (1998)	5
GrSaleToGrOverhead	Sales growth over overhead growth	Abarbanell and Bushee (1998)	5
ETR	Effective Tax Rate	Abarbanell and Bushee (1998)	5
LaborforceEfficiency	Laborforce efficiency	Abarbanell and Bushee (1998)	5
nanalyst	Number of analysts	Elgers, Lo and Pfeiffer (2001)	5
tang	Tangibility	Hahn and Lee (2009)	5
realestate	Real estate holdings	Tuzel (2010)	5
KZ	Kaplan Zingales index	Lamont, Polk and Saa-Requejo (2001)	5
WW	Whited-Wu index	Whited and Wu (2006)	5
Cash	Cash to assets	Palazzo (2012)	5
AccrualQualityJune	Accrual Quality in June	Francis, LaFond, Olsson, Schipper (2005)	5
AccrualQuality	Accrual Quality	Francis, LaFond, Olsson, Schipper (2005)	5
OrderBacklog	Order backlog	Rajgopal, Shevlin, Venkatachalam (2003)	5
EarningsPersistence	Earnings persistence	Francis, LaFond, Olsson, Schipper (2004)	5
EarningsPredictability	Earnings Predictability	Francis, LaFond, Olsson, Schipper (2004)	5
EarningsSmoothness	Earnings Smoothness	Francis, LaFond, Olsson, Schipper (2004)	5

Continued on next page

Table A.1 – continued from previous page

(1)	(2)	(3)	(4)
Acronym	Brief Introduction	Literature	Group
EarningsValueRelevance	Value relevance of earnings	Francis, LaFond, Olsson, Schipper (2004)	5
EarningsTimeliness	Earnings timeliness	Francis, LaFond, Olsson, Schipper (2004)	5
EarningsConservatism	Earnings conservatism	Francis, LaFond, Olsson, Schipper (2004)	5
AssetLiquidityBook	Asset liquidity over book assets	Ortiz-Molina and Phillips (2014)	5
AssetLiquidityMarket	Asset liquidity over market	Ortiz-Molina and Phillips (2014)	5
ForecastDispersion	EPS Forecast Dispersion	Diether, Malloy and Scherbina (2002)	5
MomSeasonShort	Return seasonality last year	Heston and Sadka (2008)	5
Mom12mOffSeason	Momentum without the seasonal part	Heston and Sadka (2008)	5
MomSeason	Return seasonality years 2 to 5	Heston and Sadka (2008)	5
MomOffSeason	Off season long-term reversal	Heston and Sadka (2008)	5
MomSeason06YrPlus	Return seasonality years 6 to 10	Heston and Sadka (2008)	5
MomOffSeason06YrPlus	Off season reversal years 6 to 10	Heston and Sadka (2008)	5
MomSeason11YrPlus	Return seasonality years 11 to 15	Heston and Sadka (2008)	5
MomOffSeason11YrPlus	Off season reversal years 11 to 15	Heston and Sadka (2008)	5
MomSeason16YrPlus	Return seasonality years 16 to 20	Heston and Sadka (2008)	5
MomOffSeason16YrPlus	Off season reversal years 16 to 20	Heston and Sadka (2008)	5
BrandCapital	Brand capital to assets	Belo, Lin and Vitorino (2014)	5
size	Market capitalization	Banz (1981)	6
IdioVol3F	Idiosyncratic risk (3 factor)	Ang et al. (2006)	6
IdioVolAHT	Idiosyncratic risk (AHT)	Ali, Hwang, and Trombley (2003)	6
IdioVolCAPM	Idiosyncratic risk (CAPM)	Ang et al. (2006)	6
IdioVolQF	Idiosyncratic risk (q factor)	Ang et al. (2006)	6
Beta	CAPM beta	Fama and MacBeth (1973)	6
BetaFP	Frazzini-Pedersen Beta	Frazzini and Pedersen (2014)	6
BetaDimson	Dimson Beta	Dimson (1979)	6
std_turn	Share turnover volatility	Chordia, Subra, Anshuman (2001)	6
DolVol	Past trading volume	Brennan, Chordia, Subra (1998)	6
VolSD	Volume Variance	Chordia, Subra, Anshuman (2001)	6
Illiquidity	Amihud's illiquidity	Amihud (2002)	6
zerotradeAlt1	Days with zero trades	Liu (2006)	6
zerotrade	Days with zero trades	Liu (2006)	6
zerotradeAlt12	Days with zero trades	Liu (2006)	6
MaxRet	Maximum return over month	Bali, Cakici, and Whitelaw (2010)	6
ReturnSkew	Return skewness	Bali, Engle and Murray (2015)	6
ReturnSkewCAPM	Idiosyncratic skewness (CAPM)	Bali, Engle and Murray (2015)	6
ReturnSkew3F	Idiosyncratic skewness (3F model)	Bali, Engle and Murray (2015)	6
ReturnSkewQF	Idiosyncratic skewness (Q model)	Bali, Engle and Murray (2015)	6
Coskewness	Coskewness	Ang, Chen and Xing (2006)	6
BetaTailRisk	Tail risk beta	Kelly and Jiang (2014)	6
betaRR	Return-market return illiquidity beta	Acharya and Pedersen (2005)	6
betaCC	Illiquidity-illiquidity beta	Acharya and Pedersen (2005)	6
betaRC	Return-market illiquidity beta	Acharya and Pedersen (2005)	6
betaCR	Illiquidity-market return beta	Acharya and Pedersen (2005)	6
betaNet	Net liquidity beta	Acharya and Pedersen (2005)	6

Continued on next page

Table A.1 – continued from previous page

(1)	(2)	(3)	(4)
Acronym	Brief Introduction	Literature	Group
SRreversal	Short-run reversal	Jegadeesh (1990)	6
DownsideBeta	Downside beta	Ang, Chen and Xing (2006)	6
BidAskSpread	Bid-ask spread	Amihud and Mendelson (1986)	6
BetaBDLeverage	Broker-Dealer Leverage Beta	Adrian, Etula and Muir (2014)	6
BetaLiquidityPS	Pastor-Stambaugh liquidity beta	Pastor and Stambaugh (2003)	6

Table A.2: Summary Statistics of CM Portfolios

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Acronym	Nobs	Mean (%)	Median (%)	Std (%)	SR	Skew	Kurt
EarningsSurprise	636	0.1673	0.1828	0.5200	0.3218	-2.2860	22.8014
AnnouncementReturn	617	0.3468	0.3804	0.6037	0.5745	-0.3639	8.0312
REV6	556	0.2058	0.2632	0.9356	0.2200	-1.6673	12.1195
Mom6m	636	0.3236	0.4041	1.8498	0.1750	-1.7341	21.2184
Mom12m	636	0.4007	0.4978	1.8292	0.2190	-1.2107	14.5211
IndMom	636	0.2226	0.2321	1.4088	0.1580	0.1303	12.5790
RevenueSurprise	636	0.1132	0.1086	0.5214	0.2171	-1.3764	12.9209
ChTax	636	0.1852	0.1747	0.6124	0.3024	-0.3173	7.7158
AnalystRevision	562	0.1193	0.1453	0.3832	0.3114	-2.9066	33.4557
NumEarnIncrease	636	0.2089	0.2056	0.6299	0.3317	-0.3283	9.1353
High52	636	0.3805	0.6105	2.4959	0.1525	-1.5389	13.7840
ResidualMomentum6m	636	0.1575	0.1831	0.9122	0.1727	-0.4539	9.4252
ResidualMomentum	636	0.2589	0.2591	0.9641	0.2685	-0.2516	8.4375
retConglomerate	563	0.4418	0.3093	1.6338	0.2704	1.2157	15.5096
BEME	636	0.2205	0.1897	1.2813	0.1721	0.3347	5.7079
AM	636	0.0991	0.0514	1.4063	0.0705	-0.1622	7.1553
LRreversal	636	0.0908	0.0250	1.1336	0.0801	-1.1236	7.5320
EP	636	0.1470	0.1045	1.1995	0.1225	-0.0590	6.9495
sfe	561	0.0787	0.0827	1.7042	0.0462	-0.5899	10.4693
CF	636	0.2238	0.2012	1.5738	0.1422	-0.6440	9.4445
DivYieldAnn	636	0.1724	0.1003	1.5400	0.1119	-0.1053	5.6182
PayoutYield	636	0.1061	0.0613	0.8395	0.1263	0.6927	7.0689
NetPayoutYield	636	0.2022	0.1849	1.0304	0.1962	-0.2075	6.2330
MeanRankRevGrowth	636	0.0005	0.0210	0.6369	0.0008	-0.3469	3.9674
sgr	636	0.1924	-0.1919	0.7439	0.2586	0.1992	4.8788
EntMult	636	0.1264	-0.1042	0.8845	0.1429	0.6304	8.9556
SP	636	0.1688	0.0594	1.3560	0.1245	1.5980	15.8149
cfp	636	0.2075	0.1909	1.3489	0.1538	0.4493	10.2245

Continued on next page

Table A.2 – continued from previous page

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Acronym	Nobs	Mean (%)	Median (%)	Std (%)	SR	Skew	Kurt
IntanBM	636	0.0473	0.0116	1.2988	0.0364	-0.4197	11.2643
IntrinsicValue	558	0.0612	0.0367	1.4315	0.0428	-0.6012	9.8892
AnalystValue	558	0.0745	-0.0023	1.3757	0.0542	0.0220	9.8113
EBM	636	0.0673	0.0599	0.4494	0.1498	0.0995	7.1692
NetDebtPrice	636	0.0041	-0.0526	1.2572	0.0033	-0.3663	10.9365
EquityDuration	636	0.2394	-0.2016	1.3571	0.1764	0.6426	7.5522
Leverage	636	0.0817	0.0426	1.3290	0.0615	-0.2305	6.6150
Investment	636	0.0536	-0.0473	0.4817	0.1113	-0.3632	6.8205
AssetGrowth	636	0.2736	-0.2398	0.7698	0.3554	-0.1951	5.7012
InvestPPEInv	636	0.2087	-0.2013	0.7850	0.2658	0.2015	4.7836
NOA	636	0.1658	-0.1552	0.8799	0.1884	-0.1001	4.8959
dNoa	636	0.2472	-0.2184	0.7006	0.3529	0.1218	5.8258
GrLTNOA	636	0.0297	0.0226	0.6198	0.0480	0.6932	6.7938
grcapx1y	636	0.0358	-0.0623	0.4390	0.0815	1.1977	10.9240
grcapx	636	0.1211	-0.1218	0.4906	0.2469	-0.0235	5.8050
grcapx3y	636	0.1137	-0.0863	0.4629	0.2456	-0.4484	4.8847
ShareIss1Y	636	0.1175	-0.0674	0.6022	0.1952	-0.6191	11.1331
ChInvIA	636	0.0583	-0.0545	0.5680	0.1026	-0.5037	8.3687
CompEquIss	636	0.1343	-0.1036	0.8203	0.1637	-1.8633	16.6856
CompositeDebtIssuance	636	0.0489	-0.0661	0.4347	0.1124	0.2356	4.8838
InvGrowth	636	0.1239	-0.1362	0.6826	0.1816	-0.0915	7.1985
ChInv	636	0.1381	-0.1351	0.5716	0.2416	0.0232	4.9978
Accruals	636	0.0955	-0.1071	0.6472	0.1476	-0.4310	6.7672
TotalAccruals	636	0.0566	0.0012	0.6726	0.0842	-1.6754	13.5617
DelCOA	636	0.1435	-0.0947	0.7216	0.1989	0.0279	3.8542
DelCOL	636	0.0998	-0.0987	0.6995	0.1427	0.0315	4.4749
ChNNCOA	636	0.0962	-0.1007	0.4376	0.2198	0.9648	10.2302
ChNCOA	636	0.2198	-0.1826	0.5873	0.3742	-0.7471	7.2866
ChNCOL	636	0.1405	-0.1260	0.5008	0.2804	-0.0297	5.1217

Continued on next page

Table A.2 – continued from previous page

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Acronym	Nobs	Mean (%)	Median (%)	Std (%)	SR	Skew	Kurt
DelNetFin	636	0.1094	0.1198	0.4742	0.2306	-0.4541	6.3796
DelSTI	636	0.0188	0.0184	0.4596	0.0410	-0.2505	6.8781
DelLTI	636	0.0313	-0.0271	0.3781	0.0827	-0.8687	7.8481
DelFINL	636	0.1655	-0.1936	0.4624	0.3578	0.5601	6.4670
DelEqu	636	0.1729	-0.1146	0.9079	0.1904	-0.5283	5.9803
AbnormalAccruals	612	0.0622	-0.0459	0.4376	0.1421	-1.7248	16.2063
PctAcc	636	0.0698	-0.0809	0.4894	0.1427	-0.0799	5.5018
AbnormalAccrualsPercent	612	0.0693	-0.0804	0.3002	0.2307	0.7602	7.3767
XFIN	612	0.3403	-0.2536	1.2787	0.2661	-0.1064	7.5363
NetEquityFinance	612	0.1933	-0.1699	1.1124	0.1738	0.5491	8.1976
NetDebtFinance	612	0.1258	-0.1331	0.5146	0.2445	-0.9626	13.9606
RoE	636	0.1272	0.2321	1.0372	0.1226	-1.1600	9.0751
ChangeRoE	636	0.1330	0.1374	0.5568	0.2389	-1.2779	17.0842
roaq	636	0.2463	0.3345	1.5268	0.1613	-1.3328	11.4078
ChangeRoA	636	0.1302	0.1665	0.6700	0.1943	-1.1797	10.6646
RetNOA	636	0.0164	0.0343	0.6124	0.0267	-1.1082	9.7232
PM	636	0.1683	0.2397	1.1388	0.1478	-1.2056	8.2731
AssetTurnover	636	0.0818	0.0645	0.6303	0.1298	0.5702	6.0720
CapTurnover	636	0.0398	0.0307	0.8270	0.0481	0.1179	3.3348
GP	636	0.1475	0.1243	0.8702	0.1695	-0.1439	3.5284
GPlag	636	0.0612	0.0760	0.9499	0.0644	0.1242	3.6044
OperProf	636	0.1386	0.1332	0.9183	0.1509	0.1377	8.9239
OperProfLag	636	0.0689	0.0554	0.8342	0.0826	-0.5386	9.6757
OperProfRD	636	0.2459	0.2419	1.0697	0.2299	-0.1095	7.5258
OperProfRDLagAT	636	0.1473	0.1530	0.9252	0.1592	0.0522	4.1599
CBOperProf	636	0.2837	0.3098	0.9450	0.3002	-0.1764	10.2895
CBOperProfLagAT	636	0.2140	0.2514	0.8384	0.2552	-0.2466	6.0956
PS	612	0.0186	0.0142	0.8486	0.0219	-0.7890	9.0344
FailureProbability	636	0.2097	-0.2669	2.2308	0.0940	1.2212	11.6634

Continued on next page

Table A.2 – continued from previous page

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Acronym	Nobs	Mean (%)	Median (%)	Std (%)	SR	Skew	Kurt
FailureProbabilityJune	636	0.1527	-0.1911	2.0222	0.0755	0.8076	10.6013
ZScore	636	0.0901	-0.1118	0.9189	0.0981	1.8488	25.4184
Tax	636	0.1227	0.1173	0.7333	0.1674	-0.8734	11.3312
BookLeverage	636	0.0188	-0.0212	1.0919	0.0173	-0.3296	7.2180
OrgCapNoAdj	636	0.0871	0.0702	1.0777	0.0809	0.9279	10.7253
OrgCap	636	0.0877	0.0501	0.7066	0.1242	1.5271	17.3207
AdExp	636	0.0315	-0.0123	0.9377	0.0336	1.8240	20.4992
GrAdExp	636	0.0678	-0.0321	0.6281	0.1079	-0.1794	6.9507
RD	636	0.2029	0.0590	1.4376	0.1412	1.8167	12.1470
rd_sale	636	0.0140	-0.1319	1.2250	0.0115	2.1264	20.4253
OPLeverage	636	0.1005	0.0633	0.8095	0.1241	0.3078	4.4038
hire	636	0.1769	-0.1428	0.7341	0.2410	-0.2413	5.0216
BrandInvest	636	0.0475	0.0584	0.9537	0.0498	-0.8014	11.9836
AOP	558	0.0159	-0.0058	0.6615	0.0240	0.1947	5.6249
Herf	636	0.0011	0.0101	0.7490	0.0015	-0.4019	14.5418
HerfAsset	636	0.0153	0.0089	0.8013	0.0191	-0.6045	14.2305
HerfBE	636	0.0170	-0.0315	0.6469	0.0263	-0.7273	13.4585
FirmAge	636	0.1161	0.1426	1.1842	0.0981	-0.1277	7.8896
PriceDelayRsq	632	0.0564	0.0431	1.1113	0.0507	-0.2167	4.6643
PriceDelaySlope	632	0.0209	-0.0039	0.4891	0.0427	-0.0537	7.6987
PriceDelayTstat	632	0.0142	0.0437	0.6939	0.0205	0.0700	8.0239
GrSaleToGrInv	636	0.0600	0.0536	0.4336	0.1384	0.4369	7.1073
GrSaleToGrReceivables	636	0.0211	0.0441	0.3951	0.0534	0.6381	9.4088
GrGMToGrSales	636	0.0786	0.0988	0.4629	0.1698	-0.4618	4.6878
GrSaleToGrOverhead	636	0.0643	-0.0332	0.5641	0.1140	0.3810	11.5191
ETR	636	0.0206	-0.0504	0.4165	0.0494	0.8581	9.3890
LaborforceEfficiency	636	0.0717	-0.0906	0.4862	0.1474	0.4043	5.3370
nanalyst	563	0.0328	0.0409	0.9866	0.0332	-0.1386	6.0629
tang	636	0.0925	0.0670	0.9889	0.0935	1.0560	12.4638

Continued on next page

Table A.2 – continued from previous page

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Acronym	Nobs	Mean (%)	Median (%)	Std (%)	SR	Skew	Kurt
realestate	636	0.0258	0.0654	0.5031	0.0512	-0.3628	4.9295
KZ	636	0.0112	-0.0338	0.6144	0.0182	0.3069	4.2766
WW	636	0.1397	-0.1699	1.8056	0.0774	0.5437	8.5919
Cash	615	0.0659	0.0760	1.6534	0.0399	1.3489	17.3613
AccrualQualityJune	636	0.0910	-0.1015	1.2018	0.0757	0.4166	6.7419
AccrualQuality	636	0.0910	-0.0968	1.1732	0.0776	0.3411	6.4989
OrderBacklog	624	0.0065	0.0019	0.6576	0.0098	-0.1603	3.7585
EarningsPersistence	636	0.0114	0.0091	0.6795	0.0167	-0.1797	6.0410
EarningsPredictability	636	0.0940	-0.1212	0.4832	0.1945	0.2019	6.5410
EarningsSmoothness	636	0.0141	0.0150	1.3911	0.0101	0.6678	9.6889
EarningsValueRelevance	636	0.0535	0.0511	0.4766	0.1123	-0.2511	4.7012
EarningsTimeliness	636	0.0073	0.0072	0.3739	0.0195	0.1214	5.6245
EarningsConservatism	636	0.0131	-0.0223	0.2646	0.0494	0.2348	18.7961
AssetLiquidityBook	636	0.0087	-0.0318	1.3602	0.0064	1.1787	14.7746
AssetLiquidityMarket	636	0.2641	0.1673	1.1573	0.2282	0.6458	8.4601
ForecastDispersion	563	0.1203	-0.1502	0.8856	0.1359	0.9582	8.5817
MomSeasonShort	636	0.1669	0.1669	1.0775	0.1549	-0.1634	12.0264
Mom12mOffSeason	636	0.3521	0.4459	1.9895	0.1770	-1.7200	23.1641
MomSeason	636	0.1125	0.1377	0.7652	0.1470	0.4914	11.8495
MomOffSeason	636	0.2350	-0.1310	1.1556	0.2034	-0.7462	8.6860
MomSeason06YrPlus	636	0.1463	0.1181	0.6237	0.2346	0.2170	10.3696
MomOffSeason06YrPlus	636	0.1268	-0.1261	0.7079	0.1791	0.0062	4.9978
MomSeason11YrPlus	636	0.1176	0.1053	0.5445	0.2159	0.3508	7.2033
MomOffSeason11YrPlus	636	0.0446	-0.0457	0.6141	0.0727	-0.8977	7.4414
MomSeason16YrPlus	636	0.1250	0.1031	0.6049	0.2066	0.0998	7.5632
MomOffSeason16YrPlus	636	0.0379	-0.0176	0.6942	0.0546	-0.1188	4.2793
BrandCapital	636	0.0019	-0.0116	0.6331	0.0030	0.1432	4.1756
size	636	0.0436	0.1833	1.4085	0.0309	-1.0182	7.7007
IdioVol3F	636	0.4510	-0.5714	2.2441	0.2010	1.0666	10.2396

Continued on next page

Table A.2 – continued from previous page

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Acronym	Nobs	Mean (%)	Median (%)	Std (%)	SR	Skew	Kurt
IdioVolAHT	636	0.3701	-0.5015	2.5830	0.1433	0.7856	8.5247
IdioVolCAPM	636	0.4603	-0.5758	2.2854	0.2014	1.0579	10.3777
IdioVolQF	601	0.4427	-0.5716	2.1624	0.2048	1.0819	11.2216
Beta	636	0.0790	-0.1038	2.4782	0.0319	0.8751	9.1777
BetaFP	636	0.1184	-0.2366	2.2850	0.0518	0.6818	7.0452
BetaDimson	636	0.0833	-0.0975	1.4682	0.0567	0.3532	10.8005
std_turn	636	0.2053	-0.2207	1.7572	0.1168	0.8755	10.5308
DolVol	636	0.0063	0.0035	1.2581	0.0050	0.1213	5.2087
VolSD	636	0.1117	-0.1688	1.0365	0.1078	1.1501	10.4722
Illiquidity	636	0.0231	-0.0419	0.9036	0.0255	0.3486	5.4366
zerotradeAlt1	636	0.0338	0.0143	0.9609	0.0352	0.0762	6.4127
zerotrade	636	0.0521	0.0301	1.0002	0.0521	0.0550	5.8889
zerotradeAlt12	636	0.0694	0.0908	0.9505	0.0730	-0.1653	4.8003
MaxRet	636	0.4730	-0.6000	1.9415	0.2436	0.9958	11.4690
ReturnSkew	636	0.0818	-0.0673	0.5818	0.1406	0.8292	16.8898
ReturnSkewCAPM	636	0.0559	-0.0267	0.5606	0.0997	0.8942	21.5101
ReturnSkew3F	636	0.0223	-0.0102	0.5116	0.0437	0.5060	23.0851
ReturnSkewQF	601	0.0296	-0.0101	0.4819	0.0615	0.3646	22.5234
Coskewness	636	0.1729	-0.1410	0.8608	0.2009	-0.2807	16.7624
BetaTailRisk	636	0.0942	0.0142	1.3352	0.0705	0.9258	7.9243
betaRR	636	0.0451	-0.0621	2.0198	0.0223	0.7112	8.4836
betaCC	636	0.0064	-0.0056	0.7823	0.0082	0.1715	5.4704
betaRC	636	0.0404	0.0738	1.4302	0.0283	-1.2713	11.3796
betaCR	636	0.0114	-0.0058	0.5574	0.0204	0.2902	6.1852
betaNet	636	0.0023	-0.0051	0.7369	0.0031	-0.0526	5.2315
SRreversal	636	0.2960	-0.2156	1.6864	0.1755	-1.0808	18.6680
DownsideBeta	636	0.0899	-0.0881	2.2795	0.0395	0.2933	8.1328
BidAskSpread	636	0.2779	-0.3731	2.0080	0.1384	1.3647	13.4456
BetaBDLeverage	596	0.0666	0.0385	0.8397	0.0793	0.9916	13.0266

Continued on next page

Table A.2 – continued from previous page

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Acronym	Nobs	Mean (%)	Median (%)	Std (%)	SR	Skew	Kurt
BetaLiquidityPS	636	0.0304	0.0484	0.6836	0.0445	-0.4831	8.1642

Table A.3: **Risk-adjusted Performance of the REP Portfolio**

Table A.3 documents the out-of-sample risk-adjusted performance of the REP portfolio with respect to nine factor models. The nine factor models are the CAPM model of Sharpe (1964) (CAPM), Fama-French 3 (FF-3), 5 (FF-5) and 6 (FF-6) factor models (Fama and French, 1992, 2015, 2018), the 4 factor model of Carhart (1997) (Carhart-4), the q factor model of Hou et al. (2021) (HXZ-q5), the mispricing factor model of Stambaugh and Yuan (2017) (SY-4), the behavioral factor model of Daniel, Hirshleifer, and Sun (2020) (DHS-3), and the union factor model (ALL). Column 1 of the table lists the reported items including the α's, factor betas and the R^2. Column 2-10 reports the risk-adjusted performance of these nine factor models. The quantities encapsulated in parenthesis are the t-statistics of the α's or factor betas. The out-of-sample period of these factor models are all from January 1980 to December 2022, except for the SY-4 and the DHS-3 factor models. Due to the limited availability on the authors' websites, the out-of-sample period of the SY-4 model is from January 1980 to December 2016, and that of the DHS-3 model is from July 1982 to December 2022. We evaluate the risk-adjusted performance of the REP portfolio with respect to these two factor models based on the intersected out-of-sample periods. The union factor model (ALL) is consist of the union of the eight factor models in the literature and hence its out-of-sample period is the global intersection, i.e., from July 1982 to December 2016.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	CAPM	FF-3	Carhart-4	FF-5	FF-6	HXZ-q5	SY-4	DHS-3	ALL
α (%)	6.50%	6.51%	6.36%	6.18%	6.10%	5.68%	6.41%	5.73%	5.77%
	(21.81)	(21.99)	(20.89)	(20.51)	(19.66)	(17.58)	(18.96)	(17.25)	(16.53)
MKT	0.09	0.11	0.15	0.20	0.22	0.26	0.28	0.26	0.31
	(1.40)	(1.73)	(2.27)	(2.91)	(3.07)	(3.59)	(3.60)	(3.48)	(3.98)
SMB		-0.16	-0.16	-0.04	-0.05	0.09	0.12		0.13
		(-1.63)	(-1.60)	(-0.35)	(-0.45)	(0.95)	(1.05)		(1.24)
HML		-0.03	0.04	-0.27	-0.18				-0.08
		(-0.23)	(0.37)	(-2.02)	(-1.37)				(-0.51)
MOM			0.18		0.14				-0.05
			(2.33)		(1.76)				(-0.43)
RMW				0.46	0.43				0.56
				(3.08)	(2.84)				(2.19)
CMA				0.48	0.41				-0.08
				(2.39)	(2.03)				(-0.21)
IA						0.15			0.41
						(0.94)			(1.07)
ROE						0.02			-0.30
						(0.12)			(-1.51)
EG						0.89			0.61
						(4.48)			(2.96)
MGMT							0.45		0.21
							(3.36)		(0.95)
PERF							0.27		0.03
							(2.95)		(0.18)
PEAD								0.86	0.81
								(5.32)	(4.63)
FIN								0.30	-0.11
								(3.05)	(-0.58)
R^2	0.47%	1.02%	2.67%	4.66%	5.59%	8.37%	6.57%	9.64%	14.95%

B.1 Proof of Proposition 1

Proof. Let

$$\mathbb{L}(\boldsymbol{\omega}) = \mathbb{E}\left(\epsilon^2\right) = \mathbb{E}\left(\frac{1}{\gamma} - \boldsymbol{\omega}'\boldsymbol{r}\right)^2 = \frac{1}{\gamma^2} - \frac{2}{\gamma}\boldsymbol{\omega}'\boldsymbol{\mu} + \boldsymbol{\omega}'\left(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}'\right)\boldsymbol{\omega}. \tag{B.1}$$

The first order condition with respective to $\boldsymbol{\omega}$ is

$$\frac{\partial \mathbb{L}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = -\frac{2}{\gamma}\boldsymbol{\mu} + 2\left(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}'\right)\boldsymbol{\omega} = \mathbf{0}, \tag{B.2}$$

where $\mathbf{0}$ is a conformable vector of zeros. By the Sherman-Morrison formula, the optimal portfolio obtained by the OLS regression is

$$\boldsymbol{\omega}_{OLS}^{\star} = \frac{1}{\gamma}\left(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}'\right)^{-1}\boldsymbol{\mu} = \frac{1}{\gamma}\left(\boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}\boldsymbol{\prime}\boldsymbol{\Sigma}^{-1}}{1 + \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}\right)\boldsymbol{\mu} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\gamma\left(1 + \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\gamma\left(1 + \zeta\right)}, \tag{B.3}$$

which is exactly equal to $\omega^{\star}$, the optimal portfolio for the expected quadratic utility maximization problem (4). □

B.2 Proof of Theorem 1

B.2.1 Notation

Mathematical notations: The relation $\stackrel{\mathcal{D}}{=}$ means identically distributed. Define the $\boldsymbol{\Sigma}$-norm of a vector $\boldsymbol{x} \in \mathbb{R}^N$ by $\|\boldsymbol{x}\|_{\boldsymbol{\Sigma}}^2 = \boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}$.

Mathematical symbols: The population maximum Sharpe ratio is denoted as $\zeta \equiv \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$. The theoretical mean-variance efficient portfolio is

$$\boldsymbol{\omega}^{\star} \equiv \frac{1}{\gamma}\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{1 + \zeta}, \tag{B.4}$$

together with its theoretical expected return and variance

$$r^{\star} \equiv \left(\boldsymbol{\omega}^{\star}\right)'\boldsymbol{\mu} = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\gamma(1+\zeta)} = \frac{\zeta}{\gamma(1+\zeta)}, \tag{B.5}$$

$$\sigma^2 \equiv \left(\boldsymbol{\omega}^{\star}\right)'\boldsymbol{\Sigma}\boldsymbol{\omega}^{\star} = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\gamma^2(1+\zeta)^2} = \frac{\zeta}{\gamma^2(1+\zeta)^2}. \tag{B.6}$$

For a sample of realized excess returns

$$\boldsymbol{R} \equiv \begin{pmatrix} R_{11} & \cdots & R_{1N} \\ \vdots & \ddots & \vdots \\ R_{T1} & \cdots & R_{TN} \end{pmatrix},$$

the time series of returns of the theoretical mean-variance efficient portfolio $\boldsymbol{\omega}^\star$ and the REP $\widetilde{\boldsymbol{\omega}}$ are denoted by

$$\boldsymbol{R}^\star \equiv \boldsymbol{R}\boldsymbol{\omega}^\star, \qquad \widetilde{\boldsymbol{R}} \equiv \boldsymbol{R}\widetilde{\boldsymbol{\omega}}, \tag{B.7}$$

respectively.

B.2.2 Lemmas

We list four lemmas that will be used in the proof of Theorem 1.

Lemma 1. (Lemma 3 in Chatterjee (2014)). *Suppose that $\zeta_i \sim N(0, \sigma_i^2)$ for $i = 1, \cdots, m$, which need not be independent. Then*

$$\mathbb{E}\left(\max_{1\le i\le m} |\zeta_i|\right) \le \max_{1\le i\le m} \sigma_i \cdot \sqrt{2\log(2m)}. \tag{B.8}$$

Lemma 2. (Lemma 2 in Ao, Li, and Zheng (2019)). *Suppose that $\xi_i \sim \chi^2(T)$ for $i = 1, \cdots, m$, which need not be independent. If $\sqrt{\log(2m)/2T} \le 1/4$, then*

$$\mathbb{E}\left(\max_{1\le i\le m} |\xi_i - T|\right) \le 2\sqrt{2T\log(2m)}. \tag{B.9}$$

Lemma 3. (Lemma 3 in Ao, Li, and Zheng (2019)). *Suppose that $X_j \sim N(0,1)$ for $j = 1, \cdots, p$, and $Y_k \sim N(0,1)$ for $k = 1, \cdots, q$. The two sequences $(X_j)_{j=1}^p$ and $(Y_k)_{k=1}^q$ are independent, but X_j's do not need to be independent, neither do Y_k's. Let $\left(X_j^t\right)_{t=1}^T$ and $\left(Y_k^t\right)_{t=1}^T$ be i.i.d copies of $(X_j)_{j=1}^p$ and $(Y_k)_{k=1}^q$, respectively. Further assume that $\log(2pq)/T \le 1/2$. Then*

$$\mathbb{E}\left(\max_{j,k}\left|\sum_{t=1}^T X_j^t Y_k^t\right|\right) \le 2\sqrt{T\log(2pq)}. \tag{B.10}$$

Lemma 4. (Lemma 4 in Ao, Li, and Zheng (2019)). *Suppose that $\boldsymbol{X}_t = (X_{t,1}, \cdots, X_{t,N})', t = 1, \cdots, T$, are i.i.d random vectors from multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)'$ and $\boldsymbol{\Sigma} = (\sigma_{jk})_{1\le j,k\le N}$. For $j, k = 1, \cdots, N$, define $Y_{j,k} \equiv \mathbb{E}(X_{1,j}X_{1,k}) - \frac{1}{T}\sum_{t=1}^T X_{t,j}X_{t,k}$. If*

$\max_{1\le j\le N}|\mu_j| \le M$ *and* $\max_{1\le j\le N}|\sigma_{jj}| \le L$, *then*

$$\mathbb{E}\left(\max_{1\le j,k\le N}|Y_{j,k}|\right) \le 2L\sqrt{\frac{2\log(2N)}{T}} + \left(2L + 2M\sqrt{2L}\right)\sqrt{\frac{\log(2N^2)}{T}}. \tag{B.11}$$

B.2.3 Proof of Theorem 1

Proof. The proof can be divided into several parts.

(1). We first want to show that

$$\frac{1}{T}\mathbb{E}\left(\left\|\boldsymbol{R}^\star - \widetilde{\boldsymbol{R}}\right\|_2^2\right) = O(\log(N)/T). \tag{B.12}$$

Given $\boldsymbol{R}$, define

$$\boldsymbol{S}_N = \{\boldsymbol{R}\boldsymbol{\omega} : |\omega_1| + \cdots + |\omega_N| \le \lambda\}. \tag{B.13}$$

It is straightforward to see that $\boldsymbol{S}_N$ is a compact convex subset of $\mathbb{R}^T$. Note that $\widetilde{\boldsymbol{R}} \equiv \boldsymbol{R}\widetilde{\boldsymbol{\omega}}$ is the projection of $(1/\gamma)\mathbf{1}$ onto $\boldsymbol{S}_N$, where $\mathbf{1}$ is $T \times 1$ vector of one. By Assumption A4, $\boldsymbol{R}^\star \equiv \boldsymbol{R}\boldsymbol{\omega}^\star$ is also in $\boldsymbol{S}_N$. Therefore, the angle between $(\boldsymbol{R}^\star - \widetilde{\boldsymbol{R}})$ and $[(1/\gamma)\mathbf{1} - \widetilde{\boldsymbol{R}}]$ is obtuse, that is,

$$\left(\boldsymbol{R}^\star - \widetilde{\boldsymbol{R}}\right)'\left(\frac{1}{\gamma}\mathbf{1} - \widetilde{\boldsymbol{R}}\right) \le 0. \tag{B.14}$$

Hence, by (B.14) and the definition of $r^\star$, we have

$$\begin{aligned}
\left\|\boldsymbol{R}^\star - \widetilde{\boldsymbol{R}}\right\|_2^2 &\le \left(\boldsymbol{R}^\star - \widetilde{\boldsymbol{R}}\right)'\left(\boldsymbol{R}^\star - \frac{1}{\gamma}\mathbf{1}\right) \\
&= \left(\boldsymbol{R}^\star - \widetilde{\boldsymbol{R}}\right)'(\boldsymbol{R}^\star - r^\star\mathbf{1}) + \left(\boldsymbol{R}^\star - \widetilde{\boldsymbol{R}}\right)'\left(r^\star\mathbf{1} - \frac{1}{\gamma}\mathbf{1}\right) \\
&= \left(\boldsymbol{R}^\star - \widetilde{\boldsymbol{R}}\right)'(\boldsymbol{R}^\star - \mathbf{1}\boldsymbol{\mu}'\boldsymbol{\omega}^\star) - \frac{1}{\gamma(1+\zeta)}\left(\boldsymbol{R}^\star - \widetilde{\boldsymbol{R}}\right)'\mathbf{1}. \\
&= \sum_{j=1}^N\left(\omega_j^\star - \widetilde{\omega}_j\right)\sum_{t=1}^T\sum_{k=1}^N\omega_k^\star\left(R_{tk} - \mu_k\right)R_{tj} - \frac{1}{\gamma(1+\zeta)}\sum_{j=1}^N\left(\omega_j^\star - \widetilde{\omega}_j\right)\sum_{t=1}^T R_{tj} \\
&= \sum_{j=1}^N\left(\omega_j^\star - \widetilde{\omega}_j\right)\sum_{t=1}^T\sum_{k=1}^N\omega_k^\star\left[\left(R_{tk} - \mu_k\right)\left(R_{tj} - \mu_j\right) - \sigma_{jk}\right] \\
&\quad + \sum_{j=1}^N\left(\omega_j^\star - \widetilde{\omega}_j\right)\sum_{t=1}^T\sum_{k=1}^N\omega_k^\star\left(R_{tk} - \mu_k\right)\mu_j \\
&\quad + \sum_{j=1}^N\left(\omega_j^\star - \widetilde{\omega}_j\right)\sum_{t=1}^T\sum_{k=1}^N\omega_k^\star\sigma_{jk} - \frac{1}{\gamma(1+\zeta)}\sum_{j=1}^N\left(\omega_j^\star - \widetilde{\omega}_j\right)\sum_{t=1}^T R_{tj} \\
&= \sum_{j=1}^N\left(\omega_j^\star - \widetilde{\omega}_j\right)D_{j,1} + \sum_{j=1}^N\left(\omega_j^\star - \widetilde{\omega}_j\right)D_{j,2} + D_3,
\end{aligned} \tag{B.15}$$

where

$$D_{j,1} \equiv \sum_{t=1}^{T}\sum_{k=1}^{N} \omega_k^\star \left[\left(R_{tk} - \mu_k \right) \left(R_{tj} - \mu_j \right) - \sigma_{jk} \right], \tag{B.16}$$

$$D_{j,2} \equiv \sum_{t=1}^{T}\sum_{k=1}^{N} \omega_k^\star \left(R_{tk} - \mu_k \right) \mu_j, \tag{B.17}$$

$$D_3 \equiv \sum_{j=1}^{N} \left(\omega_j^\star - \widetilde{\omega}_j \right) \sum_{t=1}^{T}\sum_{k=1}^{N} \omega_k^\star \sigma_{jk} - \frac{1}{\gamma(1+\zeta)} \sum_{j=1}^{N} \left(\omega_j^\star - \widetilde{\omega}_j \right) \sum_{t=1}^{T} R_{tj}. \tag{B.18}$$

Now we examine these three terms, respectively.

(*i*) For $D_{j,1}$, first define

$$\rho_j \equiv \mathrm{Corr}\left(\sum_{k=1}^{N} \omega_k^\star \left(R_{tk} - \mu_k \right), R_{tj} - \mu_j \right) = \frac{\mathrm{Cov}\left(\sum_{k=1}^{N} \omega_k^\star \left(R_{tk} - \mu_k \right), R_{tj} - \mu_j \right)}{\sqrt{\mathrm{Var}\left(\sum_{k=1}^{N} \omega_k^\star \left(R_{tk} - \mu_k \right) \right)} \cdot \sqrt{\mathrm{Var}\left(R_{tj} - \mu_j \right)}}. \tag{B.19}$$

For the numerator,

$$\mathrm{Cov}\left(\sum_{k=1}^{N} \omega_k^\star \left(R_{tk} - \mu_k \right), R_{tj} - \mu_j \right) = \mathrm{Cov}\left(\sum_{k=1}^{N} \omega_k^\star R_{tk}, R_{tj} \right) = \sum_{k=1}^{N} \omega_k^\star \sigma_{jk}. \tag{B.20}$$

For the first term in the denominator, because $R_{tk}, k = 1, \cdots, N$, are i.i.d copies of the true random variable of returns $\boldsymbol{r} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

$$\mathrm{Var}\left(\sum_{k=1}^{N} \omega_k^\star \left(R_{tk} - \mu_k \right) \right) = \left(\boldsymbol{\omega}^\star \right)' \mathrm{Var}\left(\boldsymbol{R}_t \right) \boldsymbol{\omega}^\star = \left(\boldsymbol{\omega}^\star \right)' \boldsymbol{\Sigma} \boldsymbol{\omega}^\star = \sigma^2. \tag{B.21}$$

For the second term in the denominator,

$$\mathrm{Var}\left(R_{tj} - \mu_j \right) = \mathrm{Var}\left(R_{tj} \right) = \sigma_{jj} \qquad \Rightarrow \qquad \sqrt{\mathrm{Var}\left(R_{tj} - \mu_j \right)} = \sigma_j, \tag{B.22}$$

where $\sigma_j \equiv \sqrt{\sigma_{jj}}, j = 1, \cdots, N$. Hence,

$$\rho_j = \frac{\sum_{k=1}^{N} \omega_k^\star \sigma_{jk}}{\sigma \sigma_j}. \tag{B.23}$$

Then

$$D_{j,1} \equiv \sum_{t=1}^{T}\sum_{k=1}^{N} \omega_k^\star \left[\left(R_{tk} - \mu_k \right) \left(R_{tj} - \mu_j \right) - \sigma_{jk} \right]$$

$$= \sum_{t=1}^{T} \sigma\sigma_j \left[\frac{(R_{tj} - \mu_j)}{\sigma_j} \cdot \sum_{k=1}^{N} \left(\frac{\omega_k^\star (R_{tk} - \mu_k)}{\sigma} \right) - \rho_j \right]. \tag{B.24}$$

It is easy to verify that the two random variables in the brackets, $\sum_{k=1}^{N} \omega_k^\star (R_{tk} - \mu_k)/\sigma$ and $(R_{tj} - \mu_j)/\sigma_j$, follow the multivariate normal distribution

$$\begin{pmatrix} \frac{\sum_{k=1}^{N} \omega_k^\star (R_{tk} - \mu_k)}{\sigma} \\ \frac{R_{tj} - \mu_j}{\sigma_j} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_j \\ \rho_j & 1 \end{pmatrix} \right). \tag{B.25}$$

Let $(Z_t)_{t=1}^{T}, (W_{T,j})_{t=1}^{T}, j = 1, \cdots, N$, are i.i.d standard normal random variables, then

$$\begin{aligned} D_{j,1} &\overset{\mathcal{D}}{=} \sum_{t=1}^{T} \sigma\sigma_j \left[\left(\rho_j Z_t + \sqrt{1 - \rho_j^2} W_{t,j} \right) \cdot Z_t - \rho_j \right] \\ &= \sigma\sigma_j \rho_j \sum_{t=1}^{T} \left(Z_t^2 - 1 \right) + \sigma\sigma_j \sqrt{1 - \rho_j^2} \sum_{t=1}^{T} Z_t W_{t,j} \\ &\equiv V_{j,1} + V_{j,2}. \end{aligned} \tag{B.26}$$

For $V_{j,1}$, by Assumption A3, we have

$$\begin{aligned} \mathbb{E}\left(\max_{1 \le j \le N} |V_{j,1}| \right) &= \mathbb{E}\left(\max_{1 \le j \le N} \left| \sigma\sigma_j \rho_j \sum_{t=1}^{T} \left(Z_t^2 - 1 \right) \right| \right) \\ &\le \sigma \mathbb{E}\left(\max_{1 \le j \le N} \left[|\sigma_j| \, |\rho_j| \left| \sum_{t=1}^{T} \left(Z_t^2 - 1 \right) \right| \right] \right) \\ &\le \sigma \mathbb{E}\left(\sqrt{L} \times 1 \times \left| \sum_{t=1}^{T} \left(Z_t^2 - 1 \right) \right| \right) \\ &\le \sigma\sqrt{L} \sqrt{\mathbb{E}\left[\left(\sum_{t=1}^{T} Z_t^2 - T \right)^2 \right]}, \end{aligned}$$

where the third inequality follows from Assumption A3 and the last one from Jensen's inequality.

To calculate the expectation inside the square root, let $Y \equiv \sum_{t=1}^{T} Z_t^2 \sim \chi(T)$, then $\mathbb{E}(Y) = T, \mathrm{Var}(Y) = 2T$, and

$$\mathbb{E}\left[\left(\sum_{t=1}^{T} Z_t^2 - T \right)^2 \right] = \mathbb{E}\left[(Y - T)^2 \right] = \mathrm{Var}\left(Y \right) = 2T. \tag{B.27}$$

Therefore,

$$\mathbb{E}\left(\max_{1 \le j \le N} |V_{j,1}| \right) \le \sigma\sqrt{L} \times \sqrt{2T} = \sigma\sqrt{2LT}. \tag{B.28}$$

For $V_{j,2}$, by Lemma 3,

$$\begin{aligned}
\mathbb{E}\left(\max_{1\le j\le N}|V_{j,2}|\right) &= \sigma\mathbb{E}\left(\max_{1\le j\le N}\left|\sigma_j\sqrt{1-\rho_j^2}\sum_{t=1}^{T}Z_tW_{t,j}\right|\right)\\
&\le \sigma\mathbb{E}\left(\max_{1\le j\le N}\left[|\sigma_j|\left|\sqrt{1-\rho_j^2}\right|\left|\sum_{t=1}^{T}Z_tW_{t,j}\right|\right]\right)\\
&\le \sigma\mathbb{E}\left(\sqrt{L}\times 1\times\max_{1\le j\le N}\left|\sum_{t=1}^{T}Z_tW_{t,j}\right|\right)\\
&\le \sigma\sqrt{L}\cdot 2\sqrt{T\log(2N)}\\
&= 2\sigma\sqrt{LT\log(2N)}, \qquad \text{(B.29)}
\end{aligned}$$

where the third inequality follows from Assumption A3 and the fourth from Lemma 3.

Combining (B.26), (B.28) and (B.29), we have

$$\mathbb{E}\left(\max_{1\le j\le N}|D_{j,1}|\right)\le\mathbb{E}\left(\max_{1\le j\le N}|V_{j,1}|\right)+\mathbb{E}\left(\max_{1\le j\le N}|V_{j,2}|\right)\le\sigma\sqrt{2LT}+2\sigma\sqrt{LT\log(2N)}. \qquad \text{(B.30)}$$

(*ii*) For $D_{j,2}$, note that $D_{j,2}$ is the summation of multiple i.i.d normal random variables, so $D_{j,2}$ is also normally distributed. Thus, we simply calculate the mean and variance of $D_{j,2}$ and then its distribution is obtained.

$$\begin{aligned}
\mathbb{E}(D_{j,2}) &= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{k=1}^{N}\omega_k^\star(R_{tk}-\mu_k)\mu_j\right]=\sum_{t=1}^{T}\sum_{k=1}^{N}\omega_k^\star\mathbb{E}(R_{tk}-\mu_k)\mu_j=0,\\
\mathrm{Var}(D_{j,2}) &= \mathrm{Var}\left[\sum_{t=1}^{T}\sum_{k=1}^{N}\omega_k^\star(R_{tk}-\mu_k)\mu_j\right]=\mu_j^2\mathrm{Var}\left(\sum_{t=1}^{T}\sum_{k=1}^{N}\omega_k^\star(R_{tk}-\mu_k)\right)=T\mu_j^2\sigma^2.
\end{aligned}$$

Therefore, $D_{j,2}\sim N(0,T\mu_j^2\sigma^2)$. By Lemma 1 and Assumption A2, we have

$$\begin{aligned}
\mathbb{E}\left(\max_{1\le j\le N}|D_{j,2}|\right) &\le \max_{1\le j\le N}\left|\sqrt{T}\mu_j\sigma\right|\cdot\sqrt{2\log(2N)}\\
&\le \sigma\sqrt{T}\cdot M\cdot\sqrt{2\log(2N)}\\
&= \sigma M\sqrt{2T\log(2N)}, \qquad \text{(B.31)}
\end{aligned}$$

where the first inequality follows from Lemma 1 and the second from Assumption A2.

(*iii*) For D_3, using the factor that $\boldsymbol{\omega}^\star$ is the theoretical optimal portfolio, we seperate the terms

associated with $\widetilde{\omega}_j$ and $\omega_j^\star$, and rewrite D_3 as

$$\begin{aligned}
D_3 &\equiv \sum_{j=1}^{N}\left(\omega_j^\star - \widetilde{\omega}_j\right)\sum_{t=1}^{T}\sum_{k=1}^{N}\omega_k^\star\sigma_{jk} - \frac{1}{\gamma(1+\zeta)}\sum_{j=1}^{N}\left(\omega_j^\star - \widetilde{\omega}_j\right)\sum_{t=1}^{T}R_{tj} \\
&= \sum_{j=1}^{N}\widetilde{\omega}_j\left[\sum_{t=1}^{T}\left(\frac{1}{\gamma\left(1+\zeta\right)}R_{tj} - \sum_{k=1}^{N}\omega_k^\star\sigma_{jk}\right)\right] + \sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{k=1}^{N}\omega_j^\star\omega_k^\star\sigma_{jk} - \frac{1}{\gamma\left(1+\zeta\right)}\sum_{t=1}^{T}\sum_{j=1}^{N}\omega_j^\star R_{tj} \\
&= \sum_{j=1}^{N}\widetilde{\omega}_j\left[\sum_{t=1}^{T}\left(\frac{1}{\gamma\left(1+\zeta\right)}R_{tj} - \sum_{k=1}^{N}\omega_k^\star\sigma_{jk}\right)\right] + T\sigma^2 - \frac{1}{\gamma\left(1+\zeta\right)}\sum_{t=1}^{T}\sum_{j=1}^{N}\omega_j^\star R_{tj} \qquad \text{(B.32)} \\
&= D_{3a} + D_{3b},
\end{aligned}$$

where

$$D_{3a} \equiv \sum_{j=1}^{N}\widetilde{\omega}_j\left[\sum_{t=1}^{T}\left(\frac{1}{\gamma\left(1+\zeta\right)}R_{tj} - \sum_{k=1}^{N}\omega_k^\star\sigma_{jk}\right)\right], \qquad \text{(B.33)}$$

$$D_{3b} \equiv T\sigma^2 - \frac{1}{\gamma\left(1+\zeta\right)}\sum_{t=1}^{T}\sum_{j=1}^{N}\omega_j^\star R_{tj}. \qquad \text{(B.34)}$$

For D_{3a},

$$\begin{aligned}
D_{3a} &\equiv \sum_{j=1}^{N}\widetilde{\omega}_j\left[\sum_{t=1}^{T}\left(\frac{1}{\gamma\left(1+\zeta\right)}R_{tj} - \sum_{k=1}^{N}\omega_k^\star\sigma_{jk}\right)\right] \\
&= \frac{1}{\gamma\left(1+\zeta\right)}\sum_{j=1}^{N}\widetilde{\omega}_j\sum_{t=1}^{T}R_{tj} - \sum_{t=1}^{T}\left(\boldsymbol{\omega}^\star\right)'\boldsymbol{\Sigma}\widetilde{\boldsymbol{\omega}} \\
&= \frac{1}{\gamma\left(1+\zeta\right)}\sum_{j=1}^{N}\widetilde{\omega}_j\sum_{t=1}^{T}R_{tj} - \sum_{t=1}^{T}\frac{1}{\gamma\left(1+\zeta\right)}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\widetilde{\boldsymbol{\omega}} \\
&= \frac{1}{\gamma\left(1+\zeta\right)}\sum_{j=1}^{N}\widetilde{\omega}_j\sum_{t=1}^{T}R_{tj} - \frac{1}{\gamma\left(1+\zeta\right)}\sum_{t=1}^{T}\sum_{j=1}^{N}\widetilde{\omega}_j\mu_j \\
&= \frac{1}{\gamma\left(1+\zeta\right)}\sum_{j=1}^{N}\widetilde{\omega}_j\sum_{t=1}^{T}\left(R_{tj} - \mu_j\right), \qquad \text{(B.35)}
\end{aligned}$$

where the second equality is obtained by using the definition $\boldsymbol{\omega}^\star = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\gamma(1+\zeta)}$.

Note that the random variable $X_j \equiv \sum_{t=1}^{T}(R_{tj} - \mu_j)$ is normally distributed with mean and variance

$$\mathbb{E}\left(X_j\right) = \mathbb{E}\left(\sum_{t=1}^{T}\left(R_{tj} - \mu_j\right)\right) = \sum_{t=1}^{T}\mathbb{E}\left(R_{tj} - \mu_j\right) = 0, \qquad \text{(B.36)}$$

$$\operatorname{Var}(X_j) = \operatorname{Var}\left(\sum_{t=1}^{T}(R_{tj}-\mu_j)\right) = \sum_{t=1}^{T}\operatorname{Var}(R_{tj}-\mu_j) = \sum_{t=1}^{T}\operatorname{Var}(R_{tj}) = \sum_{t=1}^{T}\sigma_{jj} = T\sigma_{jj}. \quad \text{(B.37)}$$

Then by Lemma 1 and Assumption A3,

$$\begin{aligned}
\mathbb{E}(|D_{3a}|) &= \frac{1}{\gamma(1+\zeta)}\mathbb{E}\left(\left|\sum_{j=1}^{N}\widetilde{\omega}_j X_j\right|\right) \\
&\leq \frac{1}{\gamma(1+\zeta)}\mathbb{E}\left(\sum_{j=1}^{N}|\widetilde{\omega}_j|\,|X_j|\right) \\
&\leq \frac{1}{\gamma(1+\zeta)}\mathbb{E}\left(\sum_{j=1}^{N}|\widetilde{\omega}_j|\max_{1\leq j\leq N}|X_j|\right) \\
&\leq \frac{1}{\gamma(1+\zeta)}\cdot\lambda\cdot\left[\max_{1\leq j\leq N}\left|\sqrt{\operatorname{Var}(X_j)}\right|\cdot\sqrt{2\log(2N)}\right] \\
&\leq \frac{\lambda}{\gamma(1+\zeta)}\cdot\sqrt{L}\cdot\sqrt{2T\log(2N)} \\
&= \frac{\lambda}{\gamma(1+\zeta)}\sqrt{2LT\log(2N)}, \qquad \text{(B.38)}
\end{aligned}$$

where the third inequality follows from the Lasso penalty, and Lemma 1 and the fourth from Assumption A3. Note that, in general, $\tilde{\lambda}$ in the LASSO-regularized regression (12 can be larger than λ in Assumption A4. For simplicity, we assume that λ and

$$\lambda$$

are equal. Nevertheless, the theorem still holds when $\tilde{\lambda} \geq \lambda$.

For D_{3b}, note that the random variable $\sum_{j=1}^{N} R_{tj}\omega_j^\star = (\boldsymbol{\omega}^\star)'\boldsymbol{R}_t$ is also normally distributed with mean and variance

$$\mathbb{E}\left[(\boldsymbol{\omega}^\star)'\boldsymbol{R}_t\right] = (\boldsymbol{\omega}^\star)'\mathbb{E}(\boldsymbol{R}_t) = (\boldsymbol{\omega}^\star)'\mathbb{E}(\boldsymbol{r}) = (\boldsymbol{\omega}^\star)'\boldsymbol{\mu} \equiv r^\star = \frac{\zeta}{\gamma(1+\zeta)}, \quad \text{(B.39)}$$

$$\operatorname{Var}\left[(\boldsymbol{\omega}^\star)'\boldsymbol{R}_t\right] = (\boldsymbol{\omega}^\star)'\operatorname{Var}(\boldsymbol{R}_t)\boldsymbol{\omega}^\star = (\boldsymbol{\omega}^\star)'\operatorname{Var}(\boldsymbol{r})\boldsymbol{\omega}^\star = (\boldsymbol{\omega}^\star)'\boldsymbol{\Sigma}\boldsymbol{\omega}^\star \equiv \sigma^2 = \frac{\zeta}{\gamma^2(1+\zeta)^2}. \quad \text{(B.40)}$$

Let $Z_t \sim N(0,1), t = 1, \cdots, T$, then $\sum_{j=1}^{N}\omega_j^\star R_{tj} = (\boldsymbol{\omega}^\star)'\boldsymbol{R}_t \overset{\mathcal{D}}{=} r^\star + \sigma Z_t$. Thus,

$$\begin{aligned}
D_{3b} &\equiv T\sigma^2 - \frac{1}{\gamma(1+\zeta)}\sum_{t=1}^{T}\sum_{j=1}^{N}\omega_j^\star R_{tj} \\
&\overset{\mathcal{D}}{=} T\sigma^2 - \frac{1}{\gamma(1+\zeta)}\sum_{t=1}^{T}(r^\star + \sigma Z_t)
\end{aligned}$$

$$= -\frac{\sqrt{\zeta}}{\gamma^2 (1+\zeta)^2} \sum_{t=1}^{T} Z_t, \tag{B.41}$$

where the last equality is obtained by plugging the expression of $r^\star$ and σ^2. Because $\sum_{t=1}^{T} Z_T \sim N(0,T)$, then let $\sum_{t=1}^{T} Z_t \overset{\mathcal{D}}{=} \sqrt{T} Z$ where Z is a standard normal random variable.

$$\begin{aligned}
\mathbb{E}\left(|D_{3b}|\right) &= \mathbb{E}\left(\left| -\frac{\sqrt{\zeta}}{\gamma^2 (1+\zeta)^2} \sum_{t=1}^{T} Z_t \right|\right) \\
&= \frac{\sqrt{\zeta}}{\gamma^2 (1+\zeta)^2} \mathbb{E}\left(\left|\sqrt{T} Z\right|\right) \\
&\leq \frac{\sqrt{T\zeta}}{\gamma^2 (1+\zeta)^2},
\end{aligned} \tag{B.42}$$

where the last inequality follows because $\mathbb{E}\left(|Z|\right) = \sqrt{\frac{2}{\pi}} < 1$.

Combining (B.32), (B.38) and (B.42), we have

$$\mathbb{E}\left(|D_3|\right) \leq \mathbb{E}\left(|D_{3a}|\right) + \mathbb{E}\left(|D_{3b}|\right) \leq \frac{\lambda}{\gamma (1+\zeta)} \sqrt{2LT \log(2N)} + \frac{\sqrt{T\zeta}}{\gamma^2 (1+\zeta)^2}. \tag{B.43}$$

Combining (B.15), (B.30), (B.31) and (B.43), we have

$$\begin{aligned}
\mathbb{E}\left(\left\|\boldsymbol{R}^\star - \widetilde{\boldsymbol{R}}\right\|_2^2\right) &\leq \mathbb{E}\left[\sum_{j=1}^{N} \left(\omega_j^\star - \widetilde{\omega}_j\right) D_{j,1} + \sum_{j=1}^{N} \left(\omega_j^\star - \widetilde{\omega}_j\right) D_{j,2} + D_3\right] \\
&\leq \mathbb{E}\left[\sum_{j=1}^{N} \left(\left|\omega_j^\star\right| + \left|\widetilde{\omega}_j\right|\right) \cdot \max_{1\leq j\leq N} |D_{j,1}|\right] + \mathbb{E}\left[\sum_{j=1}^{N} \left(\left|\omega_j^\star\right| + \left|\widetilde{\omega}_j\right|\right) \cdot \max_{1\leq j\leq N} |D_{j,2}|\right] + \mathbb{E}\left(|D_3|\right) \\
&\leq (\lambda+\lambda) \cdot \mathbb{E}\left(\max_{1\leq j\leq N} |D_{j,1}|\right) + (\lambda+\lambda) \cdot \mathbb{E}\left(\max_{1\leq j\leq N} |D_{j,2}|\right) + \mathbb{E}\left(|D_3|\right) \\
&\leq 2\lambda \left[\sigma\sqrt{2LT} + 2\sigma\sqrt{LT\log(2N)}\right] + 2\lambda \left[\sigma M \sqrt{2T\log(2N)}\right] \\
&\quad + \frac{\lambda}{\gamma (1+\zeta)} \sqrt{2LT\log(2N)} + \frac{\sqrt{T\zeta}}{\gamma^2 (1+\zeta)^2}.
\end{aligned} \tag{B.44}$$

(2). We proceed to the second part of the proof where we study the discrepancy of the 1-period future returns by $\boldsymbol{\omega}^\star$ and $\widetilde{\boldsymbol{\omega}}$.

Let $\mathcal{F}$ be the σ-algebra generated by $(R_{tj})_{1\leq t\leq T, 1\leq j\leq N}$. Let $\widetilde{\boldsymbol{r}} \equiv (\widetilde{r}_1, \cdots, \widetilde{r}_N)'$ be 1-period future returns of the N assets. Denote the returns of the portfolios $\boldsymbol{\omega}^\star$ and $\widetilde{\boldsymbol{\omega}}$ by

$$Q^\star = \sum_{j=1}^{N} \omega_j^\star \widetilde{r}_j = \left(\boldsymbol{\omega}^\star\right)' \widetilde{\boldsymbol{r}}, \qquad \widetilde{Q} = \sum_{j=1}^{N} \widetilde{\omega}_j \widetilde{r}_j = \widetilde{\boldsymbol{\omega}}' \widetilde{\boldsymbol{r}}, \tag{B.45}$$

respectively. Because $\widetilde{\boldsymbol{r}}$ is the future return and $\mathcal{F}$ is generated based on the information up to T, then

$$\mathbb{E}\left(\widetilde{\boldsymbol{r}}|\mathcal{F}\right)=\mathbb{E}\left(\widetilde{\boldsymbol{r}}\right)=\mathbb{E}\left(\boldsymbol{r}\right)=\boldsymbol{\mu}, \tag{B.46}$$

$$\mathbb{E}\left(\widetilde{\boldsymbol{r}}\widetilde{\boldsymbol{r}}'|\mathcal{F}\right)=\mathbb{E}\left(\widetilde{\boldsymbol{r}}\widetilde{\boldsymbol{r}}'\right)=\mathbb{E}\left(\boldsymbol{r}\boldsymbol{r}'\right)=\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}'. \tag{B.47}$$

Note that $\widetilde{\boldsymbol{\omega}}$ is independent of $\widetilde{\boldsymbol{r}}$ since it is estimated from the observed data up to T, i.e., $\boldsymbol{R}_t, t=1,\cdots,T$. Note also that $\boldsymbol{\omega}^\star$ is constant with respect to the σ-algebra. Hence, the expected discrepancy between the future returns is

$$\begin{aligned}
\mathbb{E}\left[\left(Q^\star-\widetilde{Q}\right)^2|\mathcal{F}\right] &= \mathbb{E}\left[\left(\left(\boldsymbol{\omega}^\star\right)'\widetilde{\boldsymbol{r}}-\widetilde{\boldsymbol{\omega}}'\widetilde{\boldsymbol{r}}\right)^2|\mathcal{F}\right] \\
&= \mathbb{E}\left[\left(\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right)'\widetilde{\boldsymbol{r}}\widetilde{\boldsymbol{r}}'\left(\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right)|\mathcal{F}\right] \\
&= \left(\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right)'\mathbb{E}\left(\boldsymbol{r}\boldsymbol{r}'\right)\left(\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right) \qquad \text{(B.48)} \\
&= \left(\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right)'\left(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}'\right)\left(\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right) \\
&\geq \left\|\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right\|_{\boldsymbol{\Sigma}}^2. \qquad \text{(B.49)}
\end{aligned}$$

Note that we can rewrite (B.48) as

$$\mathbb{E}\left[\left(Q^\star-\widetilde{Q}\right)^2|\mathcal{F}\right]=\left(\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right)'\mathbb{E}\left(\boldsymbol{r}\boldsymbol{r}'\right)\left(\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right)=\sum_{j=1}^{N}\sum_{k=1}^{N}\left(\omega_j^\star-\widetilde{\omega}_j\right)\left(\omega_k^\star-\widetilde{\omega}_k\right)\mathbb{E}\left(r_jr_k\right). \tag{B.50}$$

Note also that

$$\begin{aligned}
\frac{1}{T}\left\|\boldsymbol{R}^\star-\widetilde{\boldsymbol{R}}\right\|_2^2 &= \frac{1}{T}\left(\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right)'\left(\boldsymbol{R}'\boldsymbol{R}\right)\left(\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right) \\
&= \sum_{j=1}^{N}\sum_{k=1}^{N}\left[\left(\omega_j^\star-\widetilde{\omega}_j\right)\left(\omega_k^\star-\widetilde{\omega}_k\right)\left(\frac{1}{T}\sum_{t=1}^{T}R_{tj}R_{tk}\right)\right]. \qquad \text{(B.51)}
\end{aligned}$$

So

$$\mathbb{E}\left[\left(Q^\star-\widetilde{Q}\right)^2|\mathcal{F}\right]-\frac{1}{T}\left\|\boldsymbol{R}^\star-\widetilde{\boldsymbol{R}}\right\|_2^2=\sum_{j=1}^{N}\sum_{k=1}^{N}\left(\omega_j^\star-\widetilde{\omega}_j\right)\left(\omega_k^\star-\widetilde{\omega}_k\right)\left[\mathbb{E}\left(r_jr_k\right)-\frac{1}{T}\sum_{t=1}^{T}R_{tj}R_{tk}\right]. \tag{B.52}$$

Define

$$Y_{j,k}=\mathbb{E}\left(r_jr_k\right)-\frac{1}{T}\sum_{t=1}^{T}R_{tj}R_{tk}. \tag{B.53}$$

By Assumption A4, we have

$$\begin{aligned}\mathbb{E}\left[\left(Q^\star-\widetilde{Q}\right)^2|\mathcal{F}\right]-\frac{1}{T}\left\|\boldsymbol{R}^\star-\widetilde{\boldsymbol{R}}\right\|_2^2&=\sum_{j=1}^N\sum_{k=1}^N\left(\omega_j^\star-\widetilde{\omega}_j\right)\left(\omega_k^\star-\widetilde{\omega}_k\right)Y_{j,k}\\&\leq\sum_{j=1}^N\sum_{k=1}^N\left(\left|\omega_j^\star\right|+\left|\widetilde{\omega}_j\right|\right)\left(\left|\omega_k^\star\right|+\left|\widetilde{\omega}_k\right|\right)\left(\max_{1\leq j,k\leq N}|Y_{j,k}|\right)\\&\leq 4\lambda^2\left(\max_{1\leq j,k\leq N}|Y_{j,k}|\right).\end{aligned}\tag{B.54}$$

Rearranging terms gives

$$\mathbb{E}\left[\left(Q^\star-\widetilde{Q}\right)^2|\mathcal{F}\right]\leq\frac{1}{T}\left\|\boldsymbol{R}^\star-\widetilde{\boldsymbol{R}}\right\|_2^2+4\lambda^2\left(\max_{1\leq j,k\leq N}|Y_{j,k}|\right).\tag{B.55}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left(\left\|\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right\|_{\boldsymbol{\Sigma}}^2\right)&\leq\mathbb{E}\left\{\mathbb{E}\left[\left(Q^\star-\widetilde{Q}\right)^2|\mathcal{F}\right]\right\}\\&=\mathbb{E}\left[\left(Q^\star-\widetilde{Q}\right)^2\right]\\&\leq\frac{1}{T}\mathbb{E}\left(\left\|\boldsymbol{R}^\star-\widetilde{\boldsymbol{R}}\right\|_2^2\right)+4\lambda^2\mathbb{E}\left(\max_{1\leq j,k\leq N}|Y_{j,k}|\right).\end{aligned}\tag{B.56}$$

By (B.44) and Lemma (4), we know that

$$\frac{1}{T}\mathbb{E}\left(\left\|\boldsymbol{R}^\star-\widetilde{\boldsymbol{R}}\right\|_2^2\right)\leq O\left(\sqrt{\frac{\log(N)}{T}}\right),\tag{B.57}$$

$$\mathbb{E}\left(\max_{1\leq j,k\leq N}|Y_{j,k}|\right)\leq O\left(\sqrt{\frac{\log(N)}{T}}\right).\tag{B.58}$$

Thus,

$$\mathbb{E}\left(\left\|\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right\|_{\boldsymbol{\Sigma}}^2\right)\leq O\left(\sqrt{\frac{\log(N)}{T}}\right).\tag{B.59}$$

By Markov's inequality,

$$\left\|\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right\|_{\boldsymbol{\Sigma}}^2\xrightarrow{P}0,\tag{B.60}$$

as $\log(N)/T\to 0$, so does $\left\|\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right\|_{\boldsymbol{\Sigma}}$ by continuous mapping.

(3). The third part finishes the whole proof. For the discrepancy of the expected returns,

$$\begin{aligned}\left|\boldsymbol{\mu}'\widetilde{\boldsymbol{\omega}}-\boldsymbol{\mu}\boldsymbol{\omega}^\star\right|&=\left|\boldsymbol{\mu}'\left(\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right)\right|\\&=\left|\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-\frac{1}{2}}\cdot\boldsymbol{\Sigma}^{\frac{1}{2}}\left(\boldsymbol{\omega}^\star-\widetilde{\boldsymbol{\omega}}\right)\right|\end{aligned}$$

$$\leq \sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}} \cdot \sqrt{(\boldsymbol{\omega}^\star - \widetilde{\boldsymbol{\omega}})' \boldsymbol{\Sigma} (\boldsymbol{\omega}^\star - \widetilde{\boldsymbol{\omega}})}$$
$$\leq \sqrt{M} \left\| \boldsymbol{\omega}^\star - \widetilde{\boldsymbol{\omega}} \right\|_{\boldsymbol{\Sigma}}, \tag{B.61}$$

where the first and second inequalities follow from Cauchy-Schwarz inequality and Assumption A2, respectively.

As $\log(N)/T \to 0$, by (B.60),

$$\left| \boldsymbol{\mu}'\boldsymbol{\omega}^\star - \boldsymbol{\mu}\widetilde{\boldsymbol{\omega}} \right| \xrightarrow{P} 0. \tag{B.62}$$

For the discrepancy of the volatility, by the triangular inequality for the norm $\|\cdot\|_{\boldsymbol{\Sigma}}$, we have

$$\left| \sqrt{(\widetilde{\boldsymbol{\omega}})' \boldsymbol{\Sigma} \widetilde{\boldsymbol{\omega}}} - \sqrt{(\boldsymbol{\omega}^\star)' \boldsymbol{\Sigma} \boldsymbol{\omega}^\star} \right| = \left| ||\boldsymbol{\omega}^\star||_{\boldsymbol{\Sigma}} - ||\widetilde{\boldsymbol{\omega}}||_{\boldsymbol{\Sigma}} \right| \leq \left\| \boldsymbol{\omega}^\star - \widetilde{\boldsymbol{\omega}} \right\|_{\boldsymbol{\Sigma}}. \tag{B.63}$$

Again by (B.60), we have as $\log(N)/T \to 0$,

$$\left| \sqrt{\widetilde{\boldsymbol{\omega}}' \boldsymbol{\Sigma} \widetilde{\boldsymbol{\omega}}} - \sqrt{(\boldsymbol{\omega}^\star)' \boldsymbol{\Sigma} \boldsymbol{\omega}^\star} \right| \xrightarrow{P} 0. \tag{B.64}$$

□

C.1 Regularized Efficient Portfolio without Risk-free Assets

C.1.1 Formulating the Problem: The Population Case

Suppose the risk-free asset is not available and we want to invest in $N+1$ assets with random non-excess returns $\check{\boldsymbol{r}}_{all}$. The dimension N can be potentially high. The expected returns and the covariance matrix of $\check{\boldsymbol{r}}_{all}$ are denoted by $\check{\boldsymbol{\mu}}_{all}$ and $\check{\boldsymbol{\Sigma}}_{all}$, respectively.

Now an investor with risk-aversion of $\gamma(>0)$ chooses the portfolio $\boldsymbol{\omega}_{all}$ among the $N+1$ assets. Because the risk-free asset is not available, the constraint $\boldsymbol{\omega}_{all}'\mathbf{1}_{N+1}=1$ is imposed. The investor chooses the portfolio $r_p=\boldsymbol{\omega}_{all}'\check{\boldsymbol{r}}_{all}$ to maximize his or her expected quadratic utility, mathematically formulated as follows.

$$\max\ \mathbb{E}\left(r_p-\frac{\gamma}{2}r_p^2\right)\quad\text{subject to}\quad \boldsymbol{\omega}_{all}'\mathbf{1}_{N+1}=1. \tag{C.1}$$

We choose a *basis* asset arbitrarily from the $N+1$ assets and assign an index of "0" to the basis asset and $1,\cdots,N$ to the rest of the N assets without loss of generality. The investment weight in the basis asset (non-basis assets) and its (their) return are denoted by ω_0 ($\boldsymbol{\omega}$) and r_0 ($\boldsymbol{r}_{-0}$) with mean μ_0 ($\boldsymbol{\mu}_{-0}$), respectively. We further define the "excess returns" of the N non-basis assets over the basis asset by $\check{\boldsymbol{r}}\equiv\boldsymbol{r}_{-0}-r_0\mathbf{1}_N$, and their mean and covariance matrix by $\check{\boldsymbol{\mu}}\equiv\boldsymbol{\mu}_{-0}-\mu_0\mathbf{1}_N$ and $\check{\boldsymbol{\Sigma}}\equiv\mathbb{E}\left[(\check{\boldsymbol{r}}-\check{\boldsymbol{\mu}})(\check{\boldsymbol{r}}-\check{\boldsymbol{\mu}})'\right]$.[21]

We find that the optimization problem (C.1) is equivalent to solving the unconstrained regression first

$$\check{\boldsymbol{\omega}}^\star=\arg\min_{\boldsymbol{\omega}}\ \mathbb{E}\left\{\left[\left(\frac{1}{\gamma}-r_0\right)-\boldsymbol{\omega}'\check{\boldsymbol{r}}\right]^2\right\}\quad\Rightarrow\quad\frac{1}{\gamma}-r_0=\boldsymbol{\omega}'\check{\boldsymbol{r}}+\varepsilon, \tag{C.2}$$

and then calculating the weight in the basis asset by $\check{\omega}_0^\star=1-(\check{\boldsymbol{\omega}}^\star)'\mathbf{1}_N$. Hence the final portfolio is $\boldsymbol{\omega}_{all}^\star=(\check{\omega}_0^\star,(\check{\boldsymbol{\omega}}^\star)')'$.

C.1.2 In Reality: The Sample Case

In reality, we observe the T (non-excess) return observations of the basis asset $\boldsymbol{R}_0=(R_{1,0},\cdots,R_{T,0})'$ and of the N non-basis assets $\boldsymbol{R}_{-0}=(\boldsymbol{R}_1,\cdots,\boldsymbol{R}_T)'$ where $\boldsymbol{R}_t=(R_{t1},\cdots,R_{tN})'$. We calculate the

[21] If $\boldsymbol{r}_{all}\sim\boldsymbol{N}(\boldsymbol{\mu}_{all},\boldsymbol{\Sigma}_{all})$, which is the assumption in the following theorem, then the linear transformation $\check{\boldsymbol{r}}$ is multivariate normal distributed as well with mean $\check{\boldsymbol{\mu}}$ and covariance matrix $\check{\boldsymbol{\Sigma}}$.

excess returns of the N non-basis assets and obtain the $T \times N$ *excess return* matrix

$$\boldsymbol{X} \equiv \boldsymbol{R}_{-0} - \boldsymbol{R}_0 \mathbf{1}_N' = \begin{pmatrix} R_{11} - R_{1,0} & \cdots & R_{1N} - R_{1,0} \\ \vdots & \ddots & \vdots \\ R_{T1} - R_{T,0} & \cdots & R_{TN} - R_{T,0} \end{pmatrix} = \begin{pmatrix} \check{R}_{11} & \cdots & \check{R}_{1N} \\ \vdots & \ddots & \vdots \\ \check{R}_{T1} & \cdots & \check{R}_{TN} \end{pmatrix} = \begin{pmatrix} \check{\boldsymbol{R}}_1' \\ \vdots \\ \check{\boldsymbol{R}}_T' \end{pmatrix}. \quad \text{(C.3)}$$

Because the dimension N can be potentially high, we run the sample counterpart of the equivalent regression (C.2) with the LASSO penalty

$$\check{\boldsymbol{\omega}} = \arg\min_{\boldsymbol{\omega}} \left\| \left(\frac{1}{\gamma} \mathbf{1}_T - \boldsymbol{R}_0 \right) - \boldsymbol{X}\boldsymbol{\omega} \right\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\omega}\|_1 \leq \lambda. \quad \text{(C.4)}$$

The estimated weight in the basis asset is then $\check{\omega}_0 = 1 - \check{\boldsymbol{\omega}}' \mathbf{1}_N$, and the estimated overall portfolio is $\check{\boldsymbol{\omega}}_{all} = (\check{\omega}_0, \check{\boldsymbol{\omega}}')'$. We claim in Theorem C.1 that the expected return and volatility of the estimated portfolio $\check{\boldsymbol{\omega}}_{all}$ converge in probability to those yielded by the theoretically optimal portfolio $\check{\boldsymbol{\omega}}_{all}^\star$.

C.2 Theorem of the Convergence Without Risk-free Assets

Theorem C.1. *Given the following six assumptions,*

B1*: The returns of the $N+1$ assets satisfy $\boldsymbol{r}_{all} \sim \boldsymbol{N}(\boldsymbol{\mu}_{all}, \boldsymbol{\Sigma}_{all})$;*

B2*: There exists $M < \infty$ such that $\max\left(\check{\boldsymbol{\mu}}' \check{\boldsymbol{\Sigma}}^{-1} \check{\boldsymbol{\mu}}, \max_{1 \leq j \leq N} |\check{\mu}_j|\right) \leq M$;*

B3*: There exists $L < \infty$ such that $\max_{1 \leq i \leq N} |\check{\sigma}_{ii}| \leq L$;*

B4*: There exists $H < \infty$ such that $\boldsymbol{\sigma}_{-0}' \check{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\sigma}_{-0} \leq H$, where $\boldsymbol{\sigma}_{-0} \equiv Cov(\check{\boldsymbol{r}}, r_0)$;*

B5*: The theoretically efficient portfolio satisfies $\|\check{\boldsymbol{\omega}}^\star\|_1 \leq \lambda$ for some constant λ;*

B6*: The number of assets N and the sample size T go to infinity and satisfy $\sqrt{\log(N)/T} \to 0$,*

then the coefficients of the following LASSO-regularized regression

$$\check{\boldsymbol{\omega}} = \arg\min_{\boldsymbol{\omega}} \left\| \left(\frac{1}{\gamma} \mathbf{1}_T - \boldsymbol{R}_0 \right) - \boldsymbol{X}\boldsymbol{\omega} \right\|_2^2 \quad \textit{subject to} \quad \|\boldsymbol{\omega}\|_1 \leq \tilde{\lambda}, \quad \textit{where} \quad \tilde{\lambda} \geq \lambda, \quad \text{(C.5)}$$

together with $\check{\omega}_0 = 1 - \check{\boldsymbol{\omega}}' \mathbf{1}_N$ form the REP portfolio $\check{\boldsymbol{\omega}}_{all} = (\check{\omega}_0, \check{\boldsymbol{\omega}}')'$ whose expected return and volatility converge in probability to those yielded by the theoretically optimal portfolio $\check{\boldsymbol{\omega}}_{all}^\star$, i.e.,

$$\left| \check{\boldsymbol{\mu}}_{all}' \check{\boldsymbol{\omega}}_{all} - \check{\boldsymbol{\mu}}_{all}' \check{\boldsymbol{\omega}}_{all}^\star \right| \xrightarrow{P} 0, \quad \text{(C.6)}$$

and

$$\left|\sqrt{\check{\boldsymbol{\omega}}_{all}'\check{\boldsymbol{\Sigma}}_{all}\check{\boldsymbol{\omega}}_{all}}-\sqrt{\left(\check{\boldsymbol{\omega}}_{all}^{\star}\right)'\check{\boldsymbol{\Sigma}}_{all}\check{\boldsymbol{\omega}}_{all}^{\star}}\right|\xrightarrow{P}0. \tag{C.7}$$

Note that the quintessence of Theorem C.1 is that the convergence is about all of the $N+1$ assets instead of just the N non-basis assets. The whole proof is divided into two steps, C.3 and C.4.

C.3 Proof Step 1: Convergence of the Non-basis Assets

We first establish that the expected return and the variance of the LASSO-regularied portfolio in the *non-basis* assets $\check{\boldsymbol{\omega}}$ converge in probability to those yielded by the theoretical optimal portfolio in the *non-basis* assets $\check{\boldsymbol{\omega}}^{\star}$. By the property of multivariate normal distribution, it is easy to deduce that the normality assumption of the $N+1$ assets (Assumption B1) in Theorem C.1 implies the normality of the N basis assets, and hence $\check{\boldsymbol{r}}\sim\boldsymbol{N}\left(\check{\boldsymbol{\mu}},\check{\boldsymbol{\Sigma}}\right)$.

The non-basis assets are termed "assets" throughout this part because we do not consider the basis asset temporarily. The theorem is formulated as follows.

C.3.0.1 Theorem of the Convergence of the Non-basis Assets

Theorem C.2. *Given Assumption B1-B6 in Theorem C.1, the coefficients of the following LASSO-penalized regression*

$$\check{\boldsymbol{\omega}}=\arg\min_{\boldsymbol{\omega}}\ \mathbb{E}\left\|\left(\frac{1}{\gamma}\mathbf{1}_T-\boldsymbol{R}_0\right)-\boldsymbol{X}\boldsymbol{\omega}\right\|_2^2\quad \textit{subject to}\quad \|\boldsymbol{\omega}\|_1\leq\lambda,$$

yield the Regularized Efficient Portfolio (REP) for the non-basis assets, whose expected return and variance converge in probability to those yielded by the theoretical efficient portfolio $\check{\boldsymbol{\omega}}^{\star}$, i.e.,

$$\left|\boldsymbol{\mu}'\check{\boldsymbol{\omega}}-\boldsymbol{\mu}'\check{\boldsymbol{\omega}}^{\star}\right|\xrightarrow{P}0, \tag{C.8}$$

and

$$\left|\left(\check{\boldsymbol{\omega}}\right)'\check{\boldsymbol{\Sigma}}\check{\boldsymbol{\omega}}-\left(\check{\boldsymbol{\omega}}^{\star}\right)'\check{\boldsymbol{\Sigma}}\check{\boldsymbol{\omega}}^{\star}\right|\xrightarrow{P}0. \tag{C.9}$$

C.3.1 Lemmas

Before showing Theorem C.1, we give some lemmas that will be frequently used throughout the proof.

Lemma C.1. (Lemma 3 in Chatterjee (2014)). *Suppose that* $\zeta_i \sim N(0, \sigma_i^2)$ *for* $i = 1, \cdots, m$, *which need not be independent. Then*

$$\mathbb{E}\left(\max_{1\leq i\leq m} |\zeta_i|\right) \leq \max_{1\leq i\leq m} \sigma_i \cdot \sqrt{2\log(2m)}. \tag{C.10}$$

Lemma C.2. (Lemma 2 in Ao, Li, and Zheng (2019)). *Suppose that* $\xi_i \sim \chi^2(T)$ *for* $i = 1, \cdots, m$, *which need not be independent. If* $\sqrt{\log(2m)/2T} \leq 1/4$, *then*

$$\mathbb{E}\left(\max_{1\leq i\leq m} |\xi_i - T|\right) \leq 2\sqrt{2T\log(2m)}. \tag{C.11}$$

Lemma C.3. (Lemma 3 in Ao, Li, and Zheng (2019)). *Suppose that* $X_j \sim N(0,1)$ *for* $j = 1, \cdots, p$, *and* $Y_k \sim N(0,1)$ *for* $k = 1, \cdots, q$. *The two sequences* $(X_j)_{j=1}^p$ *and* $(Y_k)_{k=1}^q$ *are independent, but* X_j*'s do not need to be independent, neither do* Y_k*'s. Let* $\left(X_j^t\right)_{t=1}^T$ *and* $\left(Y_k^t\right)_{t=1}^T$ *be i.i.d copies of* $(X_j)_{j=1}^p$ *and* $(Y_k)_{k=1}^q$, *respectively. Further assume that* $\log(2pq)/T \leq 1/2$. *Then*

$$\mathbb{E}\left(\max_{j,k}\left|\sum_{t=1}^T X_j^t Y_k^t\right|\right) \leq 2\sqrt{T\log(2pq)}. \tag{C.12}$$

Lemma C.4. (Lemma 4 in Ao, Li, and Zheng (2019)). *Suppose that* $\boldsymbol{X}_t = (X_{t,1}, \cdots, X_{t,N})', t = 1, \cdots, T$, *are i.i.d random vectors from multivariate normal distribution* $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *with* $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)'$ *and* $\boldsymbol{\Sigma} = (\sigma_{jk})_{1\leq j,k\leq N}$. *For* $j, k = 1, \cdots, N$, *define* $Y_{j,k} \equiv \mathbb{E}(X_{1,j}X_{1,k}) - \frac{1}{T}\sum_{t=1}^T X_{t,j}X_{t,k}$. *If* $\max_{1\leq j\leq N}|\mu_j| \leq M$ *and* $\max_{1\leq j\leq N}|\sigma_{jj}| \leq L$, *then*

$$\mathbb{E}\left(\max_{1\leq j,k\leq N}|Y_{j,k}|\right) \leq 2L\sqrt{\frac{2\log(2N)}{T}} + \left(2L + 2M\sqrt{2L}\right)\sqrt{\frac{\log(2N^2)}{T}}. \tag{C.13}$$

Lemma C.5. *Suppose* $\widetilde{X}_j, j = 1, \cdots, N$, *and* $\widetilde{Y}$ *are multivariate normal distributed from* $\boldsymbol{N}(\boldsymbol{\mu}_{jY}, \boldsymbol{\Sigma}_{jY})$ *with* $\boldsymbol{\mu}_{jY} = (\mu_j, \mu_Y)'$ *and*

$$\boldsymbol{\Sigma}_{XY} = \begin{pmatrix} \sigma_j^2 & \sigma_{jY} \\ \sigma_{jY} & \sigma_Y^2 \end{pmatrix}. \tag{C.14}$$

Further assume that there exists $L \leq 0$ *such that* $\max_{1\leq j\leq N}\left|\sigma_j^2\right| \leq L^2$. *Let* X_{tj} *and* Y_t *be i.i.d*

copies from $\widetilde{X}_j$ *and* $\widetilde{Y}$, $j = 1, \cdots, N, t = 1, \cdots, T$. *Then*

$$\mathbb{E}\left\{\max_{1\leq j\leq N}\left|\sum_{t=1}^{T}\left[\left(X_{tj}-\mu_X\right)\left(Y_t-\mu_Y\right)-\sigma_{jY}\right]\right|\right\} \leq \sigma_Y\sqrt{LT}\left(\sqrt{2}+2\sqrt{\log(2N)}\right). \tag{C.15}$$

Proof. Denote the correlation between $\widetilde{X}_j$ and $\widetilde{Y}$ by $\rho_j \equiv \sigma_{jY}/(\sigma_j\sigma_Y)$. Note that

$$\begin{pmatrix} \frac{\widetilde{X}_j-\mu_j}{\sigma_j} \\ \frac{\widetilde{Y}-\mu_Y}{\sigma_Y} \end{pmatrix} \sim \boldsymbol{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 1 & \rho_j \\ \rho_j & 1 \end{pmatrix}\right). \tag{C.16}$$

Let Z_t and $W_{tj}, j = 1, \cdots, N, t = 1, \cdots, T$, are i.i.d copies from standard normal random variables. Then let

$$\frac{Y_t-\mu_Y}{\sigma_Y} \overset{\mathcal{D}}{=} Z_t, \tag{C.17}$$

$$\frac{X_{tj}-\mu_j}{\sigma_j} \overset{\mathcal{D}}{=} mZ_t + nW_{tj}, \tag{C.18}$$

where m and n are to-be-determined constant so that Z_t and $mZ_t + nW_t$ are multivariate normal distributed as (C.16). The mean of these two random variables are both exactly 0, and the variance of them are.

$$\begin{aligned} \mathrm{Var}\left(Z_t\right) &= 1, \\ \mathrm{Var}\left(mZ_t+nW_t\right) &= m^2+n^2. \end{aligned}$$

The covariance of the two random variables is

$$\begin{aligned} \mathrm{Cov}\left(Z_t, mZ_t+nW_t\right) &= \mathbb{E}\left[Z_t\left(mZ_t+nW_t\right)\right] - \mathbb{E}\left(Z_t\right)\mathbb{E}\left(mZ_t+nW_t\right) \\ &= m\mathbb{E}\left(Z_t^2\right) + n\mathbb{E}\left(Z_tW_t\right) - 0 \\ &= m + n\mathbb{E}\left(Z_t\right)\mathbb{E}\left(W_t\right) \\ &= m. \end{aligned}$$

Put the identities associated with m and n together so as to equate in distribution as (C.16)

$$\begin{cases} m^2+n^2 = 1, \\ m = \rho_j, \end{cases} \quad \Rightarrow \quad \begin{cases} m = \rho_j, \\ n = \sqrt{1-\rho_j^2}. \end{cases} \tag{C.19}$$

Hence,

$$\begin{aligned}
\sum_{t=1}^{T}\left[\left(X_{tj}-\mu_j\right)\left(Y_t-\mu_Y\right)-\sigma_{jY}\right] &= \sigma_j\sigma_Y\sum_{t=1}^{T}\left(\frac{X_{tj}-\mu_j}{\sigma_j}\cdot\frac{Y_t-\mu_Y}{\sigma_Y}-\frac{\sigma_{jY}}{\sigma_j\sigma_Y}\right) \\
&\overset{\mathcal{D}}{=} \sigma_j\sigma_Y\sum_{t=1}^{T}\left[\left(\rho_j Z_t+\sqrt{1-\rho_j^2}W_{tj}\right)Z_t-\rho_j\right] \\
&= \sigma_j\sigma_Y\rho_j\sum_{t=1}^{T}\left(Z_t^2-1\right)+\sigma_j\sigma_Y\sqrt{1-\rho_j^2}\sum_{t=1}^{T}Z_tW_{tj} \\
&\equiv V_{j,1}+V_{j,2}. \qquad \text{(C.20)}
\end{aligned}$$

Then

$$\begin{aligned}
\max_{1\le j\le N}\left|\sum_{t=1}^{T}\left[\left(X_{tj}-\mu_j\right)\left(Y_t-\mu_Y\right)-\sigma_{jY}\right]\right| &\le |V_{j,1}|+|V_{j,2}| \\
&\le \max_{1\le j\le N}|V_{j,1}|+\max_{1\le j\le N}|V_{j,2}|. \qquad \text{(C.21)}
\end{aligned}$$

For $V_{j,1}$,

$$\begin{aligned}
\mathbb{E}\left(\max_{1\le j\le N}|V_{j,1}|\right) &= \mathbb{E}\left(\max_{1\le j\le N}\left|\sigma_j\sigma_Y\rho_j\sum_{t=1}^{T}\left(Z_t^2-1\right)\right|\right) \\
&= \sigma_Y\mathbb{E}\left[\max_{1\le j\le N}\left|\sigma_j\rho_j\sum_{t=1}^{T}\left(Z_t^2-1\right)\right|\right] \\
&\le \sigma_Y\mathbb{E}\left(\max_{1\le j\le N}\left(|\sigma_j|\,|\rho_j|\left|\sum_{t=1}^{T}\left(Z_t^2-1\right)\right|\right)\right) \\
&\le \sigma_Y\mathbb{E}\left(\sqrt{L}\times 1\times\left|\sum_{t=1}^{T}\left(Z_t^2-1\right)\right|\right) \\
&= \sigma_Y\sqrt{L}\mathbb{E}\left(\left|\sum_{t=1}^{T}Z_t^2-T\right|\right) \\
&= \sigma_Y\sqrt{L}\mathbb{E}\left(\sqrt{\left(\sum_{t=1}^{T}Z_t^2-T\right)^2}\right) \\
&\le \sigma_Y\sqrt{L}\sqrt{\mathbb{E}\left[\left(\sum_{t=1}^{T}Z_t^2-T\right)^2\right]}, \qquad \text{(C.22)}
\end{aligned}$$

where the last inequality follows from the Jensen's inequality. To calculate the expectation inside

the square root in the last inequality, let $S \equiv \sum_{t=1}^{T} Z_t^2 \sim \chi(T)$, then $\mathbb{E}(S) = T, \mathrm{Var}(S) = 2T$. Then

$$\mathbb{E}\left[\left(\sum_{t=1}^{T} Z_t^2 - T\right)^2\right] = \mathbb{E}\left[(S-T)^2\right] = \mathrm{Var}\left(S\right) = 2T, \tag{C.23}$$

and thus

$$\mathbb{E}\left(\max_{1\le j\le N} |V_{j,1}|\right) \le \sigma_Y \sqrt{L} \times \sqrt{2T} = \sigma_Y \sqrt{2LT}. \tag{C.24}$$

For $V_{j,2}$,

$$\begin{aligned}
\mathbb{E}\left(\max_{1\le j\le N} |V_{j,2}|\right) &= \sigma_Y \mathbb{E}\left(\max_{1\le j\le N} \left|\sigma_j \sqrt{1-\rho_j^2} \sum_{t=1}^{T} Z_t W_{tj}\right|\right) \\
&\le \sigma_Y \mathbb{E}\left(\max_{1\le j\le N} \left(|\sigma_j| \left|\sqrt{1-\rho_j^2}\right| \left|\sum_{t=1}^{T} Z_t W_{tj}\right|\right)\right) \\
&= \sigma_Y \mathbb{E}\left(\sqrt{L} \times 1 \times \max_{1\le j\le N} \left|\sum_{t=1}^{T} Z_t W_{tj}\right|\right) \\
&= \sigma_Y \sqrt{L} \mathbb{E}\left(\max_{1\le j\le N} \left|\sum_{t=1}^{T} Z_t W_{tj}\right|\right) \\
&\le \sigma_Y \sqrt{L} \cdot 2\sqrt{T \log(2N)} \\
&= 2\sigma_Y \sqrt{LT \log(2N)},
\end{aligned} \tag{C.25}$$

where the last inequality follows from Lemma C.3.

Finally by (C.24) and (C.25), we have

$$\begin{aligned}
\mathbb{E}\left\{\max_{1\le j\le N} \left|\sum_{t=1}^{T} \left[\left(X_{tj} - \mu_j\right)\left(Y_t - \mu_Y\right) - \sigma_{jY}\right]\right|\right\} &\le \mathbb{E}\left(\max_{1\le j\le N} |V_{j,1}| + \max_{1\le j\le N} |V_{j,2}|\right) \\
&= \mathbb{E}\left(\max_{1\le j\le N} |V_{j,1}|\right) + \mathbb{E}\left(\max_{1\le j\le N} |V_{j,2}|\right) \\
&\le \sigma_Y \sqrt{2LT} + 2\sigma_Y \sqrt{LT \log(2N)} \\
&= \sigma_Y \sqrt{LT}\left(\sqrt{2} + 2\sqrt{\log(2N)}\right).
\end{aligned} \tag{C.26}$$

□

C.3.2 Proof of Theorem C.2

Proof. (i) We first solve for the closed form of the theoretical optimal portfolio $\check{\boldsymbol{\omega}}^{\star}$. By (C.2),

$$\check{\boldsymbol{\omega}}^{\star} = \arg\min_{\boldsymbol{\omega}} \ \mathbb{E}\left\{\left[\left(\frac{1}{\gamma} - r_0\right) - \boldsymbol{\omega}'\check{\boldsymbol{r}}\right]^2\right\}. \tag{C.27}$$

Let

$$\begin{aligned}
\mathbb{L} &\equiv \left[\left(\frac{1}{\gamma} - r_0\right) - \boldsymbol{\omega}'\check{\boldsymbol{r}}\right]^2 \\
&= \left(\frac{1}{\gamma} - r_0\right)^2 - 2\left(\frac{1}{\gamma} - r_0\right)\boldsymbol{\omega}'\check{\boldsymbol{r}} + \boldsymbol{\omega}'\check{\boldsymbol{r}}\check{\boldsymbol{r}}'\boldsymbol{\omega} \\
&= \left(\frac{1}{\gamma} - r_0\right)^2 - \frac{2}{\gamma}\boldsymbol{\omega}'\check{\boldsymbol{r}} + 2\boldsymbol{\omega}'\left(r_0\check{\boldsymbol{r}}\right) + \boldsymbol{\omega}'\check{\boldsymbol{r}}\check{\boldsymbol{r}}'\boldsymbol{\omega} \qquad \text{(C.28)} \\
\Rightarrow \quad \mathbb{E}\left(\mathbb{L}\right) &= \mathbb{E}\left[\left(\frac{1}{\gamma} - r_0\right)^2\right] - \frac{2}{\gamma}\boldsymbol{\omega}'\check{\boldsymbol{\mu}} + 2\boldsymbol{\omega}'\mathbb{E}\left(r_0\check{\boldsymbol{r}}\right) + \boldsymbol{\omega}'\left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)\boldsymbol{\omega}. \qquad \text{(C.29)}
\end{aligned}$$

Define

$$\boldsymbol{\sigma}_{-0} = \begin{pmatrix} \sigma_{01} \\ \vdots \\ \sigma_{0N} \end{pmatrix} \equiv \begin{pmatrix} \mathrm{Cov}\left(r_0, \check{r}_1\right) \\ \vdots \\ \mathrm{Cov}\left(r_0, \check{r}_N\right) \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left(r_0\check{r}_1\right) - \mu_0\check{\mu}_1 \\ \vdots \\ \mathbb{E}\left(r_0\check{r}_N\right) - \mu_0\check{\mu}_N \end{pmatrix} = \mathbb{E}\left(r_0\check{\boldsymbol{r}}\right) - \mu_0\check{\boldsymbol{\mu}}. \tag{C.30}$$

So

$$\mathbb{E}\left(r_0\check{\boldsymbol{r}}\right) = \boldsymbol{\sigma}_{-0} + \mu_0\check{\boldsymbol{\mu}}. \tag{C.31}$$

Hence,

$$\begin{aligned}
\mathbb{E}\left(\mathbb{L}\right) &= \mathbb{E}\left[\left(\frac{1}{\gamma} - r_0\right)^2\right] - \frac{2}{\gamma}\boldsymbol{\omega}'\check{\boldsymbol{\mu}} + 2\boldsymbol{\omega}'\mathbb{E}\left(r_0\check{\boldsymbol{r}}\right) + \boldsymbol{\omega}'\left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)\boldsymbol{\omega} \\
&= \mathbb{E}\left[\left(\frac{1}{\gamma} - r_0\right)^2\right] - \frac{2}{\gamma}\boldsymbol{\omega}'\check{\boldsymbol{\mu}} + 2\boldsymbol{\omega}'\left(\boldsymbol{\sigma}_{-0} + \mu_0\check{\boldsymbol{\mu}}\right) + \boldsymbol{\omega}'\left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)\boldsymbol{\omega} \\
&= \mathbb{E}\left[\left(\frac{1}{\gamma} - r_0\right)^2\right] - 2\boldsymbol{\omega}'\left[\left(\frac{1}{\gamma} - \mu_0\right)\check{\boldsymbol{\mu}} - \boldsymbol{\sigma}_{-0}\right] + \boldsymbol{\omega}'\left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)\boldsymbol{\omega}. \qquad \text{(C.32)}
\end{aligned}$$

Then the first-order condition is

$$\begin{aligned}
&\frac{\partial\mathbb{E}\left(\mathbb{L}\right)}{\partial\boldsymbol{\omega}} = -2\left[\left(\frac{1}{\gamma} - \mu_0\right)\boldsymbol{\mu} - \boldsymbol{\sigma}_{-0}\right] + 2\left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)\boldsymbol{\omega} = \mathbf{0} \qquad \text{(C.33)} \\
\Rightarrow \quad &\boldsymbol{\omega}^{\star} = \left(\frac{1}{\gamma} - \mu_0\right)\left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1}\check{\boldsymbol{\mu}} - \left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1}\boldsymbol{\sigma}_{-0}.
\end{aligned}$$

By Shermon-Morrison's formula,

$$\left(\check{\boldsymbol{\Sigma}}+\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1}=\left(\check{\boldsymbol{\Sigma}}^{-1}-\frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{1+\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}\right) \tag{C.34}$$

$$\Rightarrow\left(\check{\boldsymbol{\Sigma}}+\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1}\check{\boldsymbol{\mu}}=\left(\check{\boldsymbol{\Sigma}}^{-1}-\frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{1+\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}\right)\check{\boldsymbol{\mu}}=\frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}{1+\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}. \tag{C.35}$$

So

$$\check{\boldsymbol{\omega}}^{\star}=\left(\frac{1}{\gamma}-\mu_0\right)\frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}{1+\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}-\left(\check{\boldsymbol{\Sigma}}+\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1}\boldsymbol{\sigma}_{-0}, \tag{C.36}$$

$$\Rightarrow\quad(\check{\boldsymbol{\omega}}^{\star})'=\left(\frac{1}{\gamma}-\mu_0\right)\frac{\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{1+\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}-\boldsymbol{\sigma}_{-0}'\left(\check{\boldsymbol{\Sigma}}+\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1}. \tag{C.37}$$

Define three quantities that will be frequently used throughout the proof.

$$a\equiv\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}, \tag{C.38}$$

$$b\equiv\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\sigma}_{-0}\left(=\boldsymbol{\sigma}_{-0}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\right), \tag{C.39}$$

$$c\equiv\boldsymbol{\sigma}_{-0}'\check{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\sigma}_{-0}. \tag{C.40}$$

The expected return of the theoretical optimal portfolio $\check{\boldsymbol{\omega}}^{\star}$ is

$$\begin{aligned}\check{r}^{\star}\equiv(\check{\boldsymbol{\omega}}^{\star})'\check{\boldsymbol{\mu}}&=\left(\frac{1}{\gamma}-\mu_0\right)\frac{\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}{1+\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}-\boldsymbol{\sigma}_{-0}'\left(\check{\boldsymbol{\Sigma}}+\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1}\check{\boldsymbol{\mu}}\\&=\left(\frac{1}{\gamma}-\mu_0\right)\frac{a}{1+a}-\boldsymbol{\sigma}_{-0}'\left(\check{\boldsymbol{\Sigma}}^{-1}-\frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{1+\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}\right)\check{\boldsymbol{\mu}}\\&=\left(\frac{1}{\gamma}-\mu_0\right)\frac{a}{1+a}-\boldsymbol{\sigma}_{-0}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}+\frac{\boldsymbol{\sigma}_{-0}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\boldsymbol{\Sigma}^{-1}\check{\boldsymbol{\mu}}}{1+\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}\\&=\left(\frac{1}{\gamma}-\mu_0\right)\frac{a}{1+a}-b+\frac{ab}{1+a}\\&=\frac{\left(\frac{1}{\gamma}-\mu_0\right)a-b}{1+a}.\end{aligned} \tag{C.41}$$

The variance of the theoretical optimal portfolio $\boldsymbol{\omega}^{\star}$ is

$$\begin{aligned}(\check{\sigma}^{\star})^2\equiv(\check{\boldsymbol{\omega}}^{\star})'\check{\boldsymbol{\Sigma}}\check{\boldsymbol{\omega}}^{\star}&=\left[\left(\frac{1}{\gamma}-\mu_0\right)\frac{\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{1+\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}-\boldsymbol{\sigma}_{-0}'\left(\check{\boldsymbol{\Sigma}}+\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1}\right]\cdot\check{\boldsymbol{\Sigma}}\cdot\\&\qquad\left[\left(\frac{1}{\gamma}-\mu_0\right)\frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}{1+\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}-\left(\check{\boldsymbol{\Sigma}}+\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1}\boldsymbol{\sigma}_{-0}\right]\\&=\left(\frac{1}{\gamma}-\mu_0\right)^2\frac{\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}{\left(1+\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\right)^2}\end{aligned}$$

$$
\begin{aligned}
&+ \boldsymbol{\sigma}_{-0}' \left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1} \check{\boldsymbol{\Sigma}} \left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1} \boldsymbol{\sigma}_{-0} \\
&- 2\left(\frac{1}{\gamma} - \mu_0\right) \frac{\check{\boldsymbol{\mu}}' \left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1} \boldsymbol{\sigma}_{-0}}{1 + \check{\boldsymbol{\mu}}' \check{\boldsymbol{\Sigma}}^{-1} \check{\boldsymbol{\mu}}} \\
= &\left(\frac{1}{\gamma} - \mu_0\right)^2 \frac{a}{(1+a)^2} \\
&+ \boldsymbol{\sigma}_{-0}' \left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1} \check{\boldsymbol{\Sigma}} \left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1} \boldsymbol{\sigma}_{-0} \\
&- 2\left(\frac{1}{\gamma} - \mu_0\right) \frac{\check{\boldsymbol{\mu}}' \left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1} \boldsymbol{\sigma}_{-0}}{1 + a}.
\end{aligned} \tag{C.42}
$$

By Shermon-Morrison's formula (C.34),

$$
\begin{aligned}
\left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1} \check{\boldsymbol{\Sigma}} \left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1} &= \left(\check{\boldsymbol{\Sigma}}^{-1} - \frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{1 + \check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}\right) \check{\boldsymbol{\Sigma}} \left(\check{\boldsymbol{\Sigma}}^{-1} - \frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{1 + \check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}\right) \\
&= \left(\mathbf{I}_N - \frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'}{1+a}\right)\left(\check{\boldsymbol{\Sigma}}^{-1} - \frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{1+a}\right) \\
&= \check{\boldsymbol{\Sigma}}^{-1} - \frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{1+a} - \frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{1+a} + \frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{(1+a)^2} \\
&= \check{\boldsymbol{\Sigma}}^{-1} - \frac{2}{1+a}\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1} + \frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\left(\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}\check{\boldsymbol{\mu}}\right)\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{(1+a)^2} \\
&= \check{\boldsymbol{\Sigma}}^{-1} - \frac{2}{1+a}\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1} + \frac{a}{(1+a)^2}\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1} \\
&= \check{\boldsymbol{\Sigma}}^{-1} + \frac{a - 2(1+a)}{(1+a)^2}\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1} \\
&= \check{\boldsymbol{\Sigma}}^{-1} - \frac{2+a}{(1+a)^2}\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}.
\end{aligned}
$$

$\Rightarrow$

$$
\begin{aligned}
\boldsymbol{\sigma}_{-0}' \left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1} \check{\boldsymbol{\Sigma}} \left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1} \boldsymbol{\sigma}_{-0} &= \boldsymbol{\sigma}_{-0}'\check{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\sigma}_{-0} - \frac{2+a}{(1+a)^2}\boldsymbol{\sigma}_{-0}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\sigma}_{-0} \\
&= c - \frac{(2+a)\,b^2}{(1+a)^2} \\
&= c - \frac{2b^2 + ab^2}{(1+a)^2},
\end{aligned}
$$

and

$$
\begin{aligned}
\check{\boldsymbol{\mu}}' \left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)^{-1} \boldsymbol{\sigma}_{-0} &= \check{\boldsymbol{\mu}}' \left(\check{\boldsymbol{\Sigma}}^{-1} - \frac{\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}}{1+a}\right) \boldsymbol{\sigma}_{-0} \\
&= \check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\sigma}_{-0} - \frac{\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\sigma}_{-0}}{1+a} \\
&= b - \frac{ab}{1+a}
\end{aligned}
$$

$$= \frac{b}{1+a}.$$

Therefore,

$$\begin{aligned}
(\sigma^\star)^2 &= \left(\frac{1}{\gamma} - \mu_0\right)^2 \frac{a}{(1+a)^2} + c - \frac{2b^2 + ab^2}{(1+a)^2} - 2\left(\frac{1}{\gamma} - \mu_0\right)\frac{b}{(1+a)^2} \\
&= c + \frac{1}{(1+a)^2}\left[a\left(\frac{1}{\gamma} - \mu_0\right)^2 - 2b^2 - ab^2 - 2\left(\frac{1}{\gamma} - \mu_0\right)b\right] \\
&= c + \frac{1}{(1+a)^2}\left[a\left(\frac{1}{\gamma^2} - \frac{2\mu_0}{\gamma} + \mu_0^2\right) - 2b^2 - ab^2 - 2\left(\frac{1}{\gamma} - \mu_0\right)b\right] \\
&= c + \frac{1}{(1+a)^2}\left(\frac{a}{\gamma^2} - \frac{2a\mu_0}{\gamma} + a\mu_0^2 - 2b^2 - ab^2 - \frac{2b}{\gamma} + 2b\mu_0\right). \qquad \text{(C.43)}
\end{aligned}$$

(ii) Given the time-series excess returns of the N non-basis assets over the basis asset $\boldsymbol{X}$, the time series of the returns generated by the two portfolios, $\breve{\boldsymbol{\omega}}^\star$ and $\breve{\boldsymbol{\omega}}$, are

$$\breve{\boldsymbol{R}}^\star \equiv \boldsymbol{X}\breve{\boldsymbol{\omega}}^\star, \qquad \breve{\boldsymbol{R}} \equiv \boldsymbol{X}\breve{\boldsymbol{\omega}}, \qquad \text{(C.44)}$$

respectively. In this part, we want to study the dispcrepancy between the time-series returns of the two portfolios, i.e., $\|\breve{\boldsymbol{R}}^\star - \breve{\boldsymbol{R}}\|_2^2$. Define

$$\breve{\boldsymbol{S}}_N = \{\boldsymbol{X}\boldsymbol{\omega} : |\boldsymbol{\omega}_1| + \cdots + |\boldsymbol{\omega}_N| \le \lambda\}. \qquad \text{(C.45)}$$

It is easy to see that $\breve{\boldsymbol{S}}_N$ is a compact convex subset of $\mathbb{R}^T$. Let

$$\boldsymbol{y} \equiv \frac{1}{\gamma}\mathbf{1}_T - \boldsymbol{R}_0. \qquad \text{(C.46)}$$

Note that $\breve{\boldsymbol{R}} \equiv \boldsymbol{X}\breve{\boldsymbol{\omega}}$ is the projection of $\boldsymbol{y}$ onto $\breve{\boldsymbol{S}}_N$. Note also that $\breve{\boldsymbol{R}}^\star \equiv \boldsymbol{X}\boldsymbol{\omega}^\star$ is also in $\breve{\boldsymbol{S}}_N$ by Assumption B5, then $(\breve{\boldsymbol{R}}^\star - \breve{\boldsymbol{R}})$ is an obtuse angle to $(\boldsymbol{y} - \breve{\boldsymbol{R}})$, that is,

$$\left(\breve{\boldsymbol{R}}^\star - \breve{\boldsymbol{R}}\right)'\left(\boldsymbol{y} - \breve{\boldsymbol{R}}\right) \le 0. \qquad \text{(C.47)}$$

Therefore,

$$\begin{aligned}
\|\breve{\boldsymbol{R}}^\star - \breve{\boldsymbol{R}}\|_2^2 &= \left(\breve{\boldsymbol{R}}^\star - \breve{\boldsymbol{R}}\right)'\left(\breve{\boldsymbol{R}}^\star - \boldsymbol{y}\right) + \left(\breve{\boldsymbol{R}}^\star - \breve{\boldsymbol{R}}\right)\left(\boldsymbol{y} - \breve{\boldsymbol{R}}\right) \\
&\le \left(\breve{\boldsymbol{R}}^\star - \breve{\boldsymbol{R}}\right)'\left(\breve{\boldsymbol{R}}^\star - \boldsymbol{y}\right) \\
&= \left(\breve{\boldsymbol{R}}^\star - \breve{\boldsymbol{R}}\right)'\left(\breve{\boldsymbol{R}}^\star - \breve{r}^\star\mathbf{1}_T\right) + \left(\breve{\boldsymbol{R}}^\star - \breve{\boldsymbol{R}}\right)'\left(\breve{r}^\star\mathbf{1}_T - \boldsymbol{y}\right), \qquad \text{(C.48)}
\end{aligned}$$

where by (C.41),

$$\check{r}^{\star}\mathbf{1}_T - \boldsymbol{y} = \frac{\left(\frac{1}{\gamma} - \mu_0\right) a - b}{1+a}\mathbf{1}_T - \frac{1}{\gamma}\mathbf{1}_T + \boldsymbol{R}_0 = \frac{\mu_0 - \frac{1}{\gamma} - b}{1+a}\mathbf{1}_T + \left(\boldsymbol{R}_0 - \mu_0\mathbf{1}_T\right). \tag{C.49}$$

So

$$\|\check{\boldsymbol{R}}^{\star} - \check{\boldsymbol{R}}\|_2^2 \leq \left(\check{\boldsymbol{R}}^{\star} - \check{\boldsymbol{R}}\right)'\left(\check{\boldsymbol{R}}^{\star} - \check{r}^{\star}\mathbf{1}_T\right) + \frac{\mu_0 - \frac{1}{\gamma} - b}{1+a}\left(\check{\boldsymbol{R}}^{\star} - \check{\boldsymbol{R}}\right)'\mathbf{1}_T + \left(\check{\boldsymbol{R}}^{\star} - \check{\boldsymbol{R}}\right)'\left(\boldsymbol{R}_0 - \mu_0\mathbf{1}_T\right). \tag{C.50}$$

We define an expanding set with subscript $i, i \in \mathbb{N}^+$

$$\boldsymbol{\Omega}_i \equiv \left\{V \in \mathbb{N}^+ : \mathbb{E}\left(V\right) \leq O\left(\log(N)/T\right), |\boldsymbol{\Omega}_i| = i\right\}. \tag{C.51}$$

Throughout the proof in this part, we dynamically collect into $\boldsymbol{\Omega}_i$ those terms that will be shown to converge in probability as $\log(N)/T \to 0$. The subscript i is increased by 1 each time a new random variable whose expectation converges in probability is collected.

$$\begin{aligned}
\|\check{\boldsymbol{R}}^{\star} - \check{\boldsymbol{R}}\|_2^2 \leq{}& \left(\check{\boldsymbol{R}}^{\star} - \check{\boldsymbol{R}}\right)'\left(\check{\boldsymbol{R}}^{\star}1 - \mathbf{1}_T\check{\boldsymbol{\mu}}'\check{\boldsymbol{\omega}}^{\star}\right) \\
&+ \frac{\mu_0 - \frac{1}{\gamma} - b}{1+a}\left(\check{\boldsymbol{R}}^{\star} - \check{\boldsymbol{R}}\right)'\mathbf{1}_T \\
&+ \left(\check{\boldsymbol{R}}^{\star} - \check{\boldsymbol{R}}\right)'\left(\boldsymbol{R}_0 - \mu_0\mathbf{1}_T\right) \\
={}& \left(\check{\boldsymbol{\omega}}^{\star} - \check{\boldsymbol{\omega}}\right)'\boldsymbol{X}'\left(\boldsymbol{X} - \mathbf{1}_T\check{\boldsymbol{\mu}}'\right)\check{\boldsymbol{\omega}}^{\star} \\
&+ \frac{\mu_0 - \frac{1}{\gamma} - b}{1+a}\left(\check{\boldsymbol{\omega}}^{\star} - \check{\boldsymbol{\omega}}\right)'\boldsymbol{X}'\mathbf{1}_T \\
&+ \left(\check{\boldsymbol{\omega}}^{\star} - \check{\boldsymbol{\omega}}\right)'\boldsymbol{X}'\left(\boldsymbol{R}_0 - \mu_0\mathbf{1}_T\right) \\
={}& \sum_{j=1}^{N}\left(\check{\omega}_j^{\star} - \check{\omega}_j\right)\sum_{t=1}^{T}\sum_{k=1}^{N}\check{\omega}_k^{\star}\left(\check{R}_{tk} - \check{\mu}_k\right)\check{R}_{tj} \\
&+ \frac{\mu_0 - \frac{1}{\gamma} - b}{1+a}\sum_{j=1}^{N}\left(\check{\omega}_j^{\star} - \check{\omega}_j\right)\sum_{t=1}^{T}\check{R}_{tj} \\
&+ \sum_{j=1}^{N}\left(\check{\omega}_j^{\star} - \check{\omega}_j\right)\sum_{t=1}^{T}\check{R}_{tj}\left(R_{0,t} - \mu_0\right) \\
={}& \boxed{\sum_{j=1}^{N}\left(\check{\omega}_j^{\star} - \check{\omega}_j\right)\sum_{t=1}^{T}\sum_{k=1}^{N}\check{\omega}_k^{\star}\left[\left(\check{R}_{tk} - \check{\mu}_k\right)\left(\check{R}_{tj} - \check{\mu}_j\right) - \check{\sigma}_{jk}\right]}
\end{aligned}$$

$$
\begin{aligned}
&+\boxed{\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\sum_{k=1}^{N}\breve{\omega}_k^{\star}\left(\breve{R}_{tk}-\breve{\mu}_k\right)\breve{\mu}_j}\\
&+\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\sum_{k=1}^{N}\breve{\omega}_k^{\star}\breve{\sigma}_{jk}\\
&+\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\breve{R}_{tj}\\
&+\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\breve{R}_{tj}\left(R_{0,t}-\mu_0\right)\\
=\boldsymbol{\Omega}_2&+\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\sum_{k=1}^{N}\breve{\omega}_k^{\star}\breve{\sigma}_{jk}+\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\breve{R}_{tj}\\
&+\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\breve{R}_{tj}\left(R_{0,t}-\mu_0\right)\\
=\boldsymbol{\Omega}_2&+\sum_{j=1}^{N}\breve{\omega}_j\sum_{t=1}^{T}\left[\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\breve{R}_{tj}-\sum_{k=1}^{N}\breve{\omega}_k^{\star}\breve{\sigma}_{jk}\right]+\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{k=1}^{N}\breve{\omega}_j^{\star}\breve{\omega}_k^{\star}\breve{\sigma}_{jk}\\
&+\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}+\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\breve{R}_{tj}\left(R_{0,t}-\mu_0\right)\\
=\boldsymbol{\Omega}_2&+\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j\breve{R}_{tj}-\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{k=1}^{N}\breve{\omega}_j\breve{\omega}_k^{\star}\breve{\sigma}_{jk}+T\left(\sigma^{\star}\right)^2\\
&+\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}+\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\breve{R}_{tj}\left(R_{0,t}-\mu_0\right)\\
=\boldsymbol{\Omega}_2&+\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j\breve{R}_{tj}-\sum_{t=1}^{T}\left(\breve{\boldsymbol{\omega}}^{\star}\right)'\breve{\boldsymbol{\Sigma}}\breve{\boldsymbol{\omega}}+T\left(\sigma^{\star}\right)^2\\
&+\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}+\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\breve{R}_{tj}\left(R_{0,t}-\mu_0\right)\\
=\boldsymbol{\Omega}_2&+\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j\breve{R}_{tj}+T\left(\sigma^{\star}\right)^2\\
&-\sum_{t=1}^{T}\left[\left(\frac{1}{\gamma}-\mu_0\right)\frac{\breve{\boldsymbol{\mu}}'\breve{\boldsymbol{\Sigma}}^{-1}}{1+\breve{\boldsymbol{\mu}}'\breve{\boldsymbol{\Sigma}}^{-1}\breve{\boldsymbol{\mu}}}-\boldsymbol{\sigma}_{-0}'\left(\boldsymbol{\Sigma}+\breve{\boldsymbol{\mu}}\breve{\boldsymbol{\mu}}'\right)^{-1}\right]\boldsymbol{\Sigma}\breve{\boldsymbol{\omega}}\\
&+\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}+\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\breve{R}_{tj}\left(R_{0,t}-\mu_0\right)\\
=\boldsymbol{\Omega}_2&+\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j\breve{R}_{tj}+T\left(\sigma^{\star}\right)^2\\
&-\left(\frac{1}{\gamma}-\mu_0\right)\sum_{t=1}^{T}\frac{\breve{\boldsymbol{\mu}}'\breve{\boldsymbol{\omega}}}{1+a}+\sum_{t=1}^{T}\boldsymbol{\sigma}_{-0}'\left(\boldsymbol{\Sigma}^{-1}-\frac{\boldsymbol{\Sigma}^{-1}\breve{\boldsymbol{\mu}}\breve{\boldsymbol{\mu}}'\breve{\boldsymbol{\Sigma}}^{-1}}{1+\breve{\boldsymbol{\mu}}'\breve{\boldsymbol{\Sigma}}^{-1}\breve{\boldsymbol{\mu}}}\right)\boldsymbol{\Sigma}\breve{\boldsymbol{\omega}}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}+\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\breve{R}_{tj}\left(R_{0,t}-\mu_0\right)\\
=\boldsymbol{\Omega}_2&+\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j\breve{R}_{tj}+T\left(\sigma^{\star}\right)^2\\
&+\frac{\mu_0-\frac{1}{\gamma}}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j\breve{\mu}_j+\sum_{t=1}^{T}\boldsymbol{\sigma}'_{-0}\breve{\boldsymbol{\omega}}-\sum_{t=1}^{T}\frac{\boldsymbol{\sigma}'_{-0}\breve{\boldsymbol{\Sigma}}^{-1}\breve{\boldsymbol{\mu}}\breve{\boldsymbol{\mu}}'\breve{\boldsymbol{\omega}}}{1+\breve{\boldsymbol{\mu}}'\breve{\boldsymbol{\Sigma}}^{-1}\breve{\boldsymbol{\mu}}}\\
&+\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}+\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\breve{R}_{tj}\left(R_{0,t}-\mu_0\right)\\
=\boldsymbol{\Omega}_2&+\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j\breve{R}_{tj}+T\left(\sigma^{\star}\right)^2\\
&\frac{\mu_0-\frac{1}{\gamma}}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j\breve{\mu}_j+\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j\sigma_{0j}-\frac{b}{1+a}\sum_{t=1}^{T}\breve{\omega}_j\breve{\mu}_j\\
&+\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}+\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\breve{R}_{tj}\left(R_{0,t}-\mu_0\right)\\
=\boldsymbol{\Omega}_2&+\boxed{\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j\left(\breve{R}_{tj}-\breve{\mu}_j\right)}+T\left(\sigma^{\star}\right)^2+\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j\sigma_{0j}\\
&+\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}+\sum_{j=1}^{N}\left(\breve{\omega}_j^{\star}-\breve{\omega}_j\right)\sum_{t=1}^{T}\breve{R}_{tj}\left(R_{0,t}-\mu_0\right)\\
=\boldsymbol{\Omega}_3&+T\left(\sigma^{\star}\right)^2-\sum_{j=1}^{N}\breve{\omega}_j\sum_{t=1}^{T}\left[\breve{R}_{tj}\left(R_{0,t}-\mu_0\right)-\sigma_{0j}\right]\\
&\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}+\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}\left(R_{0,t}-\mu_0\right)\\
=\boldsymbol{\Omega}_3&+T\left(\sigma^{\star}\right)^2-\boxed{\sum_{j=1}^{N}\breve{\omega}_j\sum_{t=1}^{T}\left[\left(\breve{R}_{tj}-\breve{\mu}_j\right)\left(R_{0,t}-\mu_0\right)-\sigma_{0j}\right]}-\boxed{\sum_{j=1}^{N}\breve{\omega}_j\sum_{t=1}^{T}\left(R_{0,t}-\mu_0\right)\breve{\mu}_j}\\
&\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}+\sum_{t=1}^{T}\left(R_{0,t}-\mu_0\right)\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}\\
=\boldsymbol{\Omega}_5&+T\left(\sigma^{\star}\right)^2+\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}+\sum_{t=1}^{T}\left(R_{0,t}-\mu_0\right)\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}.
\end{aligned}
$$

Note that $\sum_{j=1}^{N}\breve{\omega}_j^{\star}\breve{R}_{tj}=\left(\breve{\boldsymbol{\omega}}^{\star}\right)'\breve{\boldsymbol{R}}_t\overset{\mathcal{D}}{=}\breve{r}^{\star}+\breve{\sigma}^{\star}Z_t$ where $Z_t, t=1,\cdots,T$, are i.i.d standard normal random variables. Thus,

$$
\|\breve{\boldsymbol{R}}^{\star}-\breve{\boldsymbol{R}}\|_2^2\overset{\mathcal{D}}{=}\boldsymbol{\Omega}_5+T\left(\breve{\sigma}^{\star}\right)^2+\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}\left(\breve{r}^{\star}+\breve{\sigma}^{\star}Z_t\right)+\sum_{t=1}^{T}\left(R_{0,t}-\mu_0\right)\left(\breve{r}^{\star}+\breve{\sigma}^{\star}Z_t\right)
$$

$$
\begin{aligned}
&= \boldsymbol{\Omega}_5 + T\left(\check{\sigma}^\star\right)^2 + \frac{\mu_0 - \frac{1}{\gamma} - b}{1+a} T\check{r}^\star + \boxed{\check{\sigma}^\star \frac{\mu_0 - \frac{1}{\gamma} - b}{1+a} \sum_{t=1}^{T} Z_t} \\
&\qquad + \boxed{\check{r}^\star \sum_{t=1}^{T} \left(R_{0,t} - \mu_0\right)} + \check{\sigma}^\star \sum_{t=1}^{T} \left(R_{0,t} - \mu_0\right) Z_t \\
&= \boldsymbol{\Omega}_7 + T\left(\check{\sigma}^\star\right)^2 + \frac{\mu_0 - \frac{1}{\gamma} - b}{1+a} T\check{r}^\star + \boxed{\check{\sigma}^\star \sum_{t=1}^{T} \left[\left(R_{0,t} - \mu_0\right)\left(Z_t - 0\right) - \sigma_{0Z}\right]} + \check{\sigma}^\star \sum_{t=1}^{T} \sigma_{0Z} \\
&= \boldsymbol{\Omega}_8 + T\left(\check{\sigma}^\star\right)^2 + \frac{\mu_0 - \frac{1}{\gamma} - b}{1+a} T\check{r}^\star + T\check{\sigma}^\star \sigma_{0Z} \\
&= \boldsymbol{\Omega}_8 + T\left[\left(\check{\sigma}^\star\right)^2 + \frac{\mu_0 - \frac{1}{\gamma} - b}{1+a} \check{r}^\star + \check{\sigma}^\star \sigma_{0Z}\right], \qquad \text{(C.52)}
\end{aligned}
$$

where

$$
\begin{aligned}
\sigma_{0Z} \equiv \operatorname{Cov}\left(Z_t, R_{0,t} - \mu_0\right) &= \frac{1}{\check{\sigma}^\star} \operatorname{Cov}\left(\check{r}^\star + \check{\sigma}^\star Z_t, R_{0,t} - \mu_0\right) \\
&= \frac{1}{\check{\sigma}^\star} \operatorname{Cov}\left(\left(\check{\boldsymbol{\omega}}^\star\right)' \check{\boldsymbol{R}}_t, R_{0,t}\right) \\
&= \frac{1}{\check{\sigma}^\star} \left(\check{\boldsymbol{\omega}}^\star\right)' \operatorname{Cov}\left(\check{\boldsymbol{R}}_t, R_{0,t}\right) \\
&= \frac{1}{\check{\sigma}^\star} \left[\left(\frac{1}{\gamma} - \mu_0\right) \frac{\check{\boldsymbol{\mu}}' \check{\boldsymbol{\Sigma}}^{-1}}{1 + \check{\boldsymbol{\mu}}' \check{\boldsymbol{\Sigma}}^{-1} \check{\boldsymbol{\mu}}} - \boldsymbol{\sigma}_{-0}' \left(\check{\boldsymbol{\Sigma}} + \check{\boldsymbol{\mu}} \check{\boldsymbol{\mu}}'\right)^{-1}\right] \boldsymbol{\sigma}_{-0} \\
&= \frac{1}{\check{\sigma}^\star} \left[\left(\frac{1}{\gamma} - \mu_0\right) \frac{\check{\boldsymbol{\mu}}' \check{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\sigma}_{-0}}{1 + \check{\boldsymbol{\mu}}' \check{\boldsymbol{\Sigma}}^{-1} \check{\boldsymbol{\mu}}} - \boldsymbol{\sigma}_{-0}' \left(\check{\boldsymbol{\Sigma}}^{-1} - \frac{\check{\boldsymbol{\Sigma}}^{-1} \check{\boldsymbol{\mu}} \check{\boldsymbol{\mu}}' \check{\boldsymbol{\Sigma}}^{-1}}{1 + \check{\boldsymbol{\mu}}' \check{\boldsymbol{\Sigma}}^{-1} \check{\boldsymbol{\mu}}}\right) \boldsymbol{\sigma}_{-0}\right] \\
&= \frac{1}{\check{\sigma}^\star} \left[\left(\frac{1}{\gamma} - \mu_0\right) \frac{b}{1+a} - \boldsymbol{\sigma}_{-0}' \check{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\sigma}_{-0} + \frac{\boldsymbol{\sigma}_{-0}' \check{\boldsymbol{\Sigma}}^{-1} \check{\boldsymbol{\mu}} \check{\boldsymbol{\mu}}' \check{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\sigma}_{-0}}{1+a}\right] \\
&= \frac{1}{\check{\sigma}^\star} \left[\left(\frac{1}{\gamma} - \mu_0\right) \frac{b}{1+a} - c + \frac{b^2}{1+a}\right]. \qquad \text{(C.53)}
\end{aligned}
$$

Hence, plugging in the closed form of $\check{r}^\star$ (C.41), $\left(\check{\sigma}^\star\right)^2$ (C.43) and σ_{0Z} (C.53), we obtain the quantity in the bracket in (C.52),

$$
\begin{aligned}
\left(\sigma^\star\right)^2 + \frac{\mu_0 - \frac{1}{\gamma} - b}{1+a} r^\star + \sigma^\star \sigma_{0Z} &= c + \frac{1}{\left(1+a\right)^2} \left(\frac{a}{\gamma^2} - \frac{2a\mu_0}{\gamma} + a\mu_0^2 - 2b^2 - ab^2 - \frac{2b}{\gamma} + 2b\mu_0\right) \\
&\qquad + \frac{\mu_0 - \frac{1}{\gamma} - b}{1+a} \cdot \frac{\left(\frac{1}{\gamma} - \mu_0\right) a - b}{1+a} \\
&\qquad + \sigma^\star \cdot \frac{1}{\sigma^\star} \left[\left(\frac{1}{\gamma} - \mu_0\right) \frac{b}{1+a} - c + \frac{b^2}{1+a}\right] \\
&= c + \frac{1}{\left(1+a\right)^2} \left(\frac{a}{\gamma^2} - \frac{2a\mu_0}{\gamma} + a\mu_0^2 - 2b^2 - ab^2 - \frac{2b}{\gamma} + 2b\mu_0\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{\left(\mu_0-\frac{1}{\gamma}-b\right)\left(\frac{a}{\gamma}-a\mu_0-b\right)}{(1+a)^2}\\
&+\frac{\left(\frac{1}{\gamma}-\mu_0\right)b}{1+a}-c+\frac{b^2}{1+a}\\
&=\frac{1}{(1+a)^2}\left(\frac{a}{\gamma^2}-\frac{2a\mu_0}{\gamma}+a\mu_0^2-2b^2-ab^2-\frac{2b}{\gamma}+2b\mu_0\right)\\
&\quad+\frac{1}{(1+a)^2}\left(\frac{a\mu_0}{\gamma}-a\mu_0^2-b\mu_0-\frac{a}{\gamma^2}+\frac{a\mu_0}{\gamma}+\frac{b}{\gamma}-\frac{ab}{\gamma}+ab\mu_0+b^2\right)\\
&\quad+\frac{1}{(1+a)^2}\left[\left(\frac{1}{\gamma}-\mu_0\right)(b+ab)+b^2+ab^2\right]\\
&=\frac{1}{(1+a)^2}\left(\frac{a}{\gamma^2}-\frac{2a\mu_0}{\gamma}+a\mu_0^2-2b^2-ab^2-\frac{2b}{\gamma}+2b\mu_0\right)\\
&\quad+\frac{1}{(1+a)^2}\left(\frac{a\mu_0}{\gamma}-a\mu_0^2-b\mu_0-\frac{a}{\gamma^2}+\frac{a\mu_0}{\gamma}+\frac{b}{\gamma}-\frac{ab}{\gamma}+ab\mu_0+b^2\right)\\
&\quad+\frac{1}{(1+a)^2}\left[\frac{b}{\gamma}+\frac{ab}{\gamma}-b\mu_0-ab\mu_0+b^2+ab^2\right]\\
&=0. \qquad \text{(C.54)}
\end{aligned}
$$

Therefore,

$$
\|\check{\boldsymbol{R}}^\star-\check{\boldsymbol{R}}\|_2^2\overset{\mathcal{D}}{=}\boldsymbol{\Omega}_8=\sum_{p=1}^{8}\Omega_p, \qquad \text{(C.55)}
$$

where

$$
\Omega_1=\sum_{j=1}^{N}\left(\check{\omega}_j^\star-\check{\omega}_j\right)\sum_{t=1}^{T}\sum_{k=1}^{N}\omega_k^\star\left[\left(\check{R}_{tk}-\check{\mu}_k\right)\left(\check{R}_{tj}-\check{\mu}_j\right)-\check{\sigma}_{jk}\right], \qquad \text{(C.56)}
$$

$$
\Omega_2=\sum_{j=1}^{N}\left(\check{\omega}_j^\star-\check{\omega}_j\right)\sum_{t=1}^{T}\sum_{k=1}^{N}\omega_k^\star\left(\check{R}_{tk}-\check{\mu}_k\right)\check{\mu}_j, \qquad \text{(C.57)}
$$

$$
\Omega_3=\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\check{\omega}_j\left(\check{R}_{tj}-\check{\mu}_j\right), \qquad \text{(C.58)}
$$

$$
\Omega_4=\sum_{j=1}^{N}\check{\omega}_j\sum_{t=1}^{T}\left[\left(\check{R}_{tj}-\check{\mu}_j\right)\left(R_{0,t}-\mu_0\right)-\sigma_{0j}\right], \qquad \text{(C.59)}
$$

$$
\Omega_5=\sum_{j=1}^{N}\check{\omega}_j\sum_{t=1}^{T}\left(R_{0,t}-\mu_0\right)\check{\mu}_j, \qquad \text{(C.60)}
$$

$$
\Omega_6=\check{\sigma}^\star\frac{\mu_0-\frac{1}{\gamma}-b}{1+a}\sum_{t=1}^{T}Z_t, \qquad \text{(C.61)}
$$

$$
\Omega_7=\check{r}^\star\sum_{t=1}^{T}\left(R_{0,t}-\mu_0\right), \qquad \text{(C.62)}
$$

$$\Omega_8 = \sigma^\star \sum_{t=1}^{T} \left[\left(R_{0,t} - \mu_0 \right) \left(Z_t - 0 \right) - \sigma_{0Z} \right]. \tag{C.63}$$

We show now the convergence for these eight terms, respectively. (1) For Ω_1:

$$\begin{aligned} \Omega_1 &= \sum_{j=1}^{N} \left(\breve{\omega}_j^\star - \breve{\omega}_j \right) \sum_{t=1}^{T} \sum_{k=1}^{N} \breve{\omega}_k^\star \left[\left(\breve{R}_{tk} - \breve{\mu}_k \right) \left(\breve{R}_{tj} - \breve{\mu}_j \right) - \breve{\sigma}_{jk} \right] \\ &= \sum_{j=1}^{N} \left(\breve{\omega}_j^\star - \breve{\omega}_j \right) \Omega_{1,j}, \end{aligned} \tag{C.64}$$

where

$$\Omega_{1,j} \equiv \sum_{t=1}^{T} \sum_{k=1}^{N} \breve{\omega}_k^\star \left[\left(\breve{R}_{tk} - \breve{\mu}_k \right) \left(\breve{R}_{tj} - \breve{\mu}_j \right) - \breve{\sigma}_{jk} \right]. \tag{C.65}$$

Define

$$\breve{\rho}_j = \operatorname{Corr} \left(\sum_{k=1}^{N} \breve{\omega}_k^\star \left(\breve{R}_{tk} - \breve{\mu}_k \right), \breve{R}_{tj} - \breve{\mu}_j \right) = \frac{\operatorname{Cov} \left(\sum_{k=1}^{N} \breve{\omega}_k^\star \left(\breve{R}_{tk} - \breve{\mu}_k \right), \breve{R}_{tj} - \breve{\mu}_j \right)}{\sqrt{\operatorname{Var} \left(\sum_{k=1}^{N} \breve{\omega}_k^\star \left(\breve{R}_{tk} - \breve{\mu}_k \right) \right)} \sqrt{\operatorname{Var} \left(\breve{R}_{tj} - \breve{\mu}_j \right)}}. \tag{C.66}$$

For the numerator,

$$\begin{aligned} \operatorname{Cov} \left(\sum_{k=1}^{N} \breve{\omega}_k^\star \left(\breve{R}_{tk} - \breve{\mu}_k \right), \breve{R}_{tj} - \breve{\mu}_j \right) &= \operatorname{Cov} \left(\sum_{k=1}^{N} \breve{\omega}_k^\star \breve{R}_{tk}, \breve{R}_{tj} \right) \\ &= \operatorname{Cov} \left(\left(\boldsymbol{\omega}^\star \right)' \breve{\boldsymbol{R}}_t, \breve{R}_{tj} \right) \\ &= \left(\boldsymbol{\omega}^\star \right)' \operatorname{Cov} \left(\breve{\boldsymbol{R}}_t, \breve{R}_{tj} \right) \\ &= \left(\boldsymbol{\omega}^\star \right)' \begin{pmatrix} \breve{\sigma}_{1j} \\ \vdots \\ \breve{\sigma}_{Nj} \end{pmatrix} \\ &= \sum_{k=1}^{N} \breve{\omega}_k^\star \breve{\sigma}_{jk}. \end{aligned} \tag{C.67}$$

For the first term in the denominator, because $\breve{\boldsymbol{R}}_t, t = 1, \cdots, T$, are i.i.d copies of the true variable of returns $\breve{\boldsymbol{r}} \sim \boldsymbol{N}(\breve{\boldsymbol{\mu}}, \breve{\boldsymbol{\Sigma}})$,

$$\begin{aligned} \operatorname{Var} \left(\sum_{k=1}^{N} \breve{\omega}_k^\star \left(\breve{R}_{tk} - \breve{\mu}_k \right) \right) &= \operatorname{Var} \left(\sum_{k=1}^{N} \breve{\omega}_k^\star \breve{R}_{tk} \right) \\ &= \operatorname{Var} \left(\left(\boldsymbol{\omega}^\star \right)' \boldsymbol{R}_t \right) \\ &= \left(\boldsymbol{\omega}^\star \right)' \operatorname{Var} \left(\boldsymbol{R}_t \right) \boldsymbol{\omega}^\star \end{aligned}$$

$$
\begin{aligned}
&= (\boldsymbol{\omega}^\star)' \operatorname{Var}(\boldsymbol{r})\, \boldsymbol{\omega}^\star \\
&= (\boldsymbol{\omega}^\star)' \check{\boldsymbol{\Sigma}} \boldsymbol{\omega}^\star \\
&= (\check{\sigma}^\star)^2 .
\end{aligned} \tag{C.68}
$$

For the second term in the denomiator,

$$
\operatorname{Var}\left(\check{R}_{tj} - \check{\mu}_j\right) = \operatorname{Var}\left(\check{R}_{tj}\right) = \check{\sigma}_{jj} \quad \Rightarrow \quad \sqrt{\operatorname{Var}\left(\check{R}_{tj} - \check{\mu}_j\right)} = \check{\sigma}_j, \tag{C.69}
$$

where $\check{\sigma}_j \equiv \sqrt{\check{\sigma}_{jj}}, j = 1, \cdots, N$. Hence, (C.66) becomes

$$
\check{\rho}_j = \frac{\sum_{k=1}^N \check{\omega}_k^\star \check{\sigma}_{jk}}{\check{\sigma}^\star \check{\sigma}_j}. \tag{C.70}
$$

Therefore,

$$
\begin{aligned}
\Omega_{1,j} &= \sum_{t=1}^T \sum_{k=1}^N \check{\omega}_k^\star \left[\left(\check{R}_{tk} - \check{\mu}_k\right)\left(\check{R}_{tj} - \check{\mu}_j\right) - \check{\sigma}_{jk}\right] \\
&= \sum_{t=1}^T \left[\sum_{k=1}^N \check{\omega}_k^\star \left(\check{R}_{tk} - \check{\mu}_k\right) \cdot \left(\check{R}_{tj} - \check{\mu}_j\right) - \sum_{k=1}^N \check{\omega}_k^\star \check{\sigma}_{jk}\right] \\
&= \sum_{t=1}^T \check{\sigma}^\star \check{\sigma}_j \left[\frac{\sum_{k=1}^N \check{\omega}_k^\star \left(\check{R}_{tk} - \check{\mu}_k\right)}{\sigma} \cdot \frac{\check{R}_{tj} - \check{\mu}_j}{\check{\sigma}_j} - \frac{\sum_{k=1}^N \check{\omega}_k^\star \check{\sigma}_{jk}}{\sigma \check{\sigma}_j}\right] \\
&= \check{\sigma}^\star \check{\sigma}_j \sum_{t=1}^T \left[\frac{\sum_{k=1}^N \check{\omega}_k^\star \left(\check{R}_{tk} - \check{\mu}_k\right)}{\sigma} \cdot \frac{\check{R}_{tj} - \check{\mu}_j}{\check{\sigma}_j} - \check{\rho}_j\right]
\end{aligned} \tag{C.71}
$$

$$
\Rightarrow \quad \max_{1 \le j \le N} |\Omega_{1,j}| = \check{\sigma}^\star \cdot \max_{1 \le j \le N} \check{\sigma}_j \cdot \max_{1 \le j \le N} \left| \sum_{t=1}^T \left[\frac{\sum_{k=1}^N \check{\omega}_k^\star \left(\check{R}_{tk} - \check{\mu}_k\right)}{\sigma} \cdot \frac{\check{R}_{tj} - \check{\mu}_j}{\check{\sigma}_j} - \check{\rho}_j\right]\right| \tag{C.72}
$$

By Lemma C.5 and Assumption C2,

$$
\mathbb{E}\left(\max_{1 \le j \le N} |\Omega_{1,j}|\right) \check{\sigma}^\star \sqrt{LT}\left(\sqrt{2} + 2\sqrt{\log(2N)}\right). \tag{C.73}
$$

Finally,

$$
\begin{aligned}
|\Omega_1| &= \left|\sum_{j=1}^N \left(\check{\omega}_j^\star - \check{\omega}_j\right) \Omega_{1,j}\right| \\
&\le \sum_{j=1}^N \left(\left|\check{\omega}_j^\star\right| + |\check{\omega}_j|\right) |\Omega_{1,j}|
\end{aligned}
$$

$$\leq \left(\sum_{j=1}^{N}\left|\check{\omega}_j^{\star}\right|+\sum_{j=1}^{N}\left|\check{\omega}_j\right|\right)\cdot\max_{1\leq j\leq N}\left|\Omega_{1,j}\right|$$

$$\leq 2\lambda\cdot\max_{1\leq j\leq N}\left|\Omega_{1,j}\right|$$

$$\Rightarrow \qquad \mathbb{E}\left(\left|\Omega_1\right|\right)\leq 2\lambda\check{\sigma}^{\star}\sqrt{LT}\left(\sqrt{2}+2\sqrt{\log(2N)}\right) \tag{C.74}$$

(2) For Ω_2:

$$\begin{aligned}\Omega_2 &= \sum_{j=1}^{N}\left(\check{\omega}_j^{\star}-\check{\omega}_j\right)\sum_{t=1}^{T}\sum_{k=1}^{N}\check{\omega}_k^{\star}\left(\check{R}_{tk}-\check{\mu}_k\right)\check{\mu}_j\\ &= \sum_{j=1}^{N}\left(\check{\omega}_j^{\star}-\check{\omega}_j\right)\Omega_{2,j},\end{aligned} \tag{C.75}$$

where

$$\Omega_{2,j}\equiv\sum_{t=1}^{T}\sum_{k=1}^{N}\check{\omega}_k^{\star}\left(\check{R}_{tk}-\check{\mu}_k\right)\check{\mu}_j=\check{\mu}_j\sum_{t=1}^{T}\sum_{k=1}^{N}\check{\omega}_k^{\star}\left(\check{R}_{tk}-\check{\mu}_k\right). \tag{C.76}$$

Note that $\Omega_{2,j}$ is the summation of multiple normal random variables, then it is a normal random variable as well and its distribution is obtained once the mean and variance are obtained. Hence,

$$\mathbb{E}\left(\Omega_{2,j}\right)=\check{\mu}_j\mathbb{E}\left[\sum_{t=1}^{T}\sum_{k=1}^{N}\check{\omega}_k^{\star}\left(\check{R}_{tk}-\check{\mu}_k\right)\right]=\check{\mu}_j\sum_{t=1}^{T}\sum_{k=1}^{N}\check{\omega}_k^{\star}\mathbb{E}\left(\check{R}_{tk}-\check{\mu}_k\right)=0, \tag{C.77}$$

$$\begin{aligned}\mathrm{Var}\left(\Omega_{2,j}\right) &= \check{\mu}_j^2\mathrm{Var}\left[\sum_{t=1}^{T}\sum_{k=1}^{N}\check{\omega}_k^{\star}\left(\check{R}_{tk}-\check{\mu}_k\right)\right]\\ &= \check{\mu}_j^2\sum_{t=1}^{T}\mathrm{Var}\left[\left(\boldsymbol{\omega}^{\star}\right)'\boldsymbol{R}_t\right]\\ &= \check{\mu}_j^2\sum_{t=1}^{T}\left(\boldsymbol{\omega}^{\star}\right)'\mathrm{Var}\left(\boldsymbol{R}_t\right)\boldsymbol{\omega}^{\star}\\ &= \check{\mu}_j^2\sum_{t=1}^{T}\left(\boldsymbol{\omega}^{\star}\right)'\boldsymbol{\Sigma}\boldsymbol{\omega}^{\star}\\ &= \check{\mu}_j^2\sum_{t=1}^{T}\left(\check{\sigma}^{\star}\right)^2\\ &= T\check{\mu}_j^2\left(\check{\sigma}^{\star}\right)^2.\end{aligned} \tag{C.78}$$

So $\Omega_{2,j}\sim\boldsymbol{N}(0,T\check{\mu}_j^2\left(\check{\sigma}^{\star}\right)^2)$, and by Lemma 1,

$$\mathbb{E}\left(\left|\Omega_{2,j}\right|\right)\leq\mathbb{E}\left(\max_{1\leq j\leq N}\left|\Omega_{2,j}\right|\right)$$

$$\leq \max_{1\leq j\leq N} \sqrt{T\check{\mu}_j^2 \left(\check{\sigma}^\star\right)^2} \cdot \sqrt{2\log(2N)}$$
$$\leq \check{\sigma}^\star \sqrt{2MT\log(2N)}, \tag{C.79}$$

where the last two inequalities follow from Lemma C.1 and Assumption C2, respectively.

(3) For Ω_3:

$$\begin{aligned}\Omega_3 &= \frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\sum_{t=1}^{T}\sum_{j=1}^{N}\check{\omega}_j\left(\check{R}_{tj}-\check{\mu}_j\right), \\ &= \frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\sum_{j=1}^{N}\check{\omega}_j\sum_{t=1}^{N}\left(\check{R}_{tj}-\check{\mu}_j\right) \\ &= \frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\sum_{j=1}^{N}\check{\omega}_j\Omega_{3,j},\end{aligned} \tag{C.80}$$

where

$$\Omega_{3,j} \equiv \sum_{t=1}^{T}\left(\check{R}_{tj}-\check{\mu}_j\right). \tag{C.81}$$

Note that $\Omega_{3,j}$ is the summation of multiple normal random variables, then it is a normal random variable as well.

$$\mathbb{E}\left(\Omega_{3,j}\right) = \sum_{t=1}^{T}\mathbb{E}\left(\check{R}_{tj}-\check{\mu}_j\right) = 0, \tag{C.82}$$

$$\mathrm{Var}\left(\Omega_{3,j}\right) = \sum_{t=1}^{T}\mathrm{Var}\left(\check{R}_{tj}-\check{\mu}_j\right) = T\check{\sigma}_{jj}. \tag{C.83}$$

So $\Omega_{3,j} \sim \boldsymbol{N}(0, T\check{\sigma}_{jj})$, and by Lemma C.1,

$$\begin{aligned}\mathbb{E}\left(|\Omega_{3,j}|\right) &\leq \mathbb{E}\left(\max_{1\leq j\leq N}|\Omega_{3,j}|\right) \\ &\leq \max_{1\leq j\leq N}\sqrt{T\check{\sigma}_{jj}}\cdot\sqrt{2\log(2N)} \\ &\leq \sqrt{2LT\log(2N)},\end{aligned} \tag{C.84}$$

where the last two inequalities follow from Lemma C.1 and Assumption C3, respectively.

Hence

$$|\Omega_3| = \left|\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\right|\left|\sum_{j=1}^{N}\check{\omega}_j\Omega_{3,j}\right|$$

$$
\begin{aligned}
&\leq \left|\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\right| \sum_{j=1}^{N} |\breve{\omega}_j| \, |\Omega_{3,j}| \\
&\leq \left|\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\right| \sum_{j=1}^{N} |\breve{\omega}_j| \max_{1\leq j\leq N} |\Omega_{3,j}| \\
\Rightarrow \qquad \mathbb{E}\left(|\Omega_3|\right) &\leq \left|\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\right| \sum_{j=1}^{N} |\breve{\omega}_j| \, \mathbb{E}\left(\max_{1\leq j\leq N} |\Omega_{3,j}|\right) \\
&\leq \lambda \left|\frac{b+\frac{1}{\gamma}-\mu_0}{1+a}\right| \sqrt{2LT\log(2N)}. \qquad \text{(C.85)}
\end{aligned}
$$

(4) For Ω_4:

$$
\begin{aligned}
\Omega_4 &= \sum_{j=1}^{N} \breve{\omega}_j \sum_{t=1}^{T} \left[\left(\check{R}_{tj} - \check{\mu}_j\right)\left(R_{0,t} - \mu_0\right) - \sigma_{0j}\right] \\
&= \sum_{j=1}^{N} \breve{\omega}_j \Omega_{4,j}, \qquad \text{(C.86)}
\end{aligned}
$$

where

$$
\Omega_{4,j} \equiv \sum_{t=1}^{T} \left[\left(\check{R}_{tj} - \check{\mu}_j\right)\left(R_{0,t} - \mu_0\right) - \sigma_{0j}\right]. \qquad \text{(C.87)}
$$

By Lemma C.5,

$$
\mathbb{E}\left(\max_{1\leq j\leq N} |\Omega_{4,j}|\right) \leq \sigma_0\sqrt{LT}\left(\sqrt{2} + 2\sqrt{\log(2N)}\right). \qquad \text{(C.88)}
$$

Hence,

$$
\begin{aligned}
|\Omega_4| &\leq \sum_{j=1}^{N} |\breve{\omega}_j| \, |\Omega_{4,j}| \\
&\leq \sum_{j=1}^{N} |\breve{\omega}_j| \max_{1\leq j\leq N} |\Omega_{4,j}| \\
\Rightarrow \qquad \mathbb{E}\left(|\Omega_4|\right) &\leq \sum_{j=1}^{N} |\breve{\omega}_j| \, \mathbb{E}\left(\max_{1\leq j\leq N} |\Omega_{4,j}|\right) \\
&\leq \lambda\sigma_0\sqrt{LT}\left(\sqrt{2} + 2\sqrt{\log(2N)}\right). \qquad \text{(C.89)}
\end{aligned}
$$

(5) For Ω_5:

$$
\Omega_5 = \sum_{j=1}^{N} \breve{\omega}_j \sum_{t=1}^{T} \left(R_{0,t} - \mu_0\right) \check{\mu}_j
$$

$$= \left(\sum_{j=1}^{N} \check{\omega}_j \check{\mu}_j \right) \cdot \sum_{t=1}^{T} (R_{0,t} - \mu_0) .$$

Note that $\sum_{t=1}^{T} (R_{0,t} - \mu_0)$ is a normal random variable and its mean and variance are

$$\mathbb{E}\left[\sum_{t=1}^{T} (R_{0,t} - \mu_0) \right] = \sum_{t=1}^{T} \mathbb{E}\left(R_{0,t} - \mu_0\right) = 0, \tag{C.90}$$

$$\mathrm{Var}\left[\sum_{t=1}^{T} (R_{0,t} - \mu_0) \right] = \sum_{t=1}^{T} \mathrm{Var}\left(R_{0,t} - \mu_0\right) = T\sigma_0^2. \tag{C.91}$$

So $\sum_{t=1}^{T} (R_{0,t} - \mu_0) \sim \boldsymbol{N}(0, T\sigma_0^2)$. Then by Lemma C.1,

$$\mathbb{E}\left[\left| \sum_{t=1}^{T} (R_{0,t} - \mu_0) \right| \right] \leq \sigma_0 \sqrt{2T \log(2)}. \tag{C.92}$$

Hence,

$$\begin{aligned} |\Omega_5| &\leq \sum_{j=1}^{N} |\check{\omega}_j| \, |\check{\mu}_j| \left| \sum_{t=1}^{T} (R_{0,t} - \mu_0) \right| \\ &\leq M \sum_{j=1}^{N} |\check{\omega}_j| \left| \sum_{t=1}^{T} (R_{0,t} - \mu_0) \right| \\ &\leq \lambda M \left| \sum_{t=1}^{T} (R_{0,t} - \mu_0) \right| \\ \Rightarrow \qquad \mathbb{E}\left(|\Omega_5|\right) &\leq \lambda M \sigma_0 \sqrt{2T \log(2)}, \end{aligned} \tag{C.93}$$

where the second inequality follows from Assumption C2.

(6) For Ω_6: note that $\sum_{t=1}^{T} \sim \boldsymbol{N}(0, T)$.

$$\begin{aligned} \Omega_6 &= \check{\sigma}^{\star} \frac{\mu_0 - \frac{1}{\gamma} - b}{1 + a} \sum_{t=1}^{T} Z_t, \\ \Rightarrow \qquad |\Omega_6| &\leq \check{\sigma}^{\star} \frac{\left| \mu_0 - \frac{1}{\gamma} - b \right|}{1 + a} \left| \sum_{t=1}^{T} Z_t \right| \\ \Rightarrow \qquad \mathbb{E}\left(|\Omega_6|\right) &\leq \check{\sigma}^{\star} \frac{\left| \mu_0 - \frac{1}{\gamma} - b \right|}{1 + a} \sqrt{2T \log(2)}, \end{aligned} \tag{C.94}$$

where the last inequality follows from Lemma C.1.

(7) For Ω_7:

$$|\Omega_7| = |\check{r}^\star| \left| \sum_{t=1}^{T} (R_{0,t} - \mu_0) \right|.$$

By (C.92), we have

$$\mathbb{E}\left(|\Omega_7|\right) \leq \sigma_0 \left|\check{r}^\star\right| \sqrt{2T \log(2)}. \tag{C.95}$$

(8) For Ω_8:

$$\Omega_8 = \check{\sigma}^\star \sum_{t=1}^{T} \left[(R_{0,t} - \mu_0) (Z_t - 0) - \sigma_{0Z} \right].$$

By Lemma (C.5), we have

$$\mathbb{E}\left(|\Omega_8|\right) \leq \check{\sigma}^\star \sigma_0 \sqrt{T} \left(\sqrt{2} + 2\sqrt{\log(2)} \right). \tag{C.96}$$

So combining (C.74), (C.79), (C.85), (C.89), (C.93), (C.94), (C.95) and (C.96),

$$\begin{aligned}
\|\check{\boldsymbol{R}}^\star - \check{\boldsymbol{R}}\|_2^2 &\overset{\mathcal{D}}{=} \sum_{p=1}^{8} \Omega_p \\
&\leq \sum_{p=1}^{8} |\Omega_p| , \\
\Rightarrow \qquad \mathbb{E}\left(|\check{\boldsymbol{R}}^\star - \check{\boldsymbol{R}}\|_2^2\right) &\leq \sum_{p=1}^{8} \mathbb{E}\left(|\Omega_p|\right) \\
&= 2\lambda\sigma^\star \sqrt{LT} \left(\sqrt{2} + 2\sqrt{\log(2N)} \right) + \sigma^\star \sqrt{2MT \log(2N)} \\
&\quad + \lambda \left| \frac{b + \frac{1}{\gamma} - \mu_0}{1+a} \right| \sqrt{2LT \log(2N)} + \lambda\sigma_0 \sqrt{LT} \left(\sqrt{2} + 2\sqrt{\log(2N)} \right) \\
&\quad + \lambda M \sigma_0 \sqrt{2T \log(2)} + \sigma^\star \frac{\left| \mu_0 - \frac{1}{\gamma} - b \right|}{1+a} \sqrt{2T \log(2)} \\
&\quad + \sigma_0 \left|r^\star\right| \sqrt{2T \log(2)} + \sigma^\star \sigma_0 \sqrt{T} \left(\sqrt{2} + 2\sqrt{\log(2)} \right).
\end{aligned} \tag{C.97}$$

(iii) We proceed to the third part of the proof where we study the discrepancy of the 1-period future excess returns by $\check{\boldsymbol{\omega}}^\star$ and $\check{\boldsymbol{\omega}}$.

Let $\mathcal{F}$ be the σ-algebra generated by $\left(\check{R}_{tj}\right)_{1 \leq t \leq T, 1 \leq j \leq N}$. Let $\widetilde{\boldsymbol{r}} \equiv (\widetilde{r}_1, \cdots, \widetilde{r}_N)'$ be 1-period

future returns of the N assets. Denote the returns of the portfolios $\check{\boldsymbol{\omega}}^{\star}$ and $\check{\boldsymbol{\omega}}$ by

$$\check{Q}^{\star}=\sum_{j=1}^{N}\check{\omega}_{j}^{\star}\widetilde{\check{r}}_{j}=\left(\check{\boldsymbol{\omega}}^{\star}\right)'\widetilde{\check{\boldsymbol{r}}},\qquad\check{Q}=\sum_{j=1}^{N}\check{\omega}_{j}\widetilde{\check{r}}_{j}=\widetilde{\check{\boldsymbol{\omega}}}'\widetilde{\check{\boldsymbol{r}}},\tag{C.98}$$

respectively. Because $\widetilde{\check{\boldsymbol{r}}}$ is the future return and $\mathcal{F}$ is generated based on the information up to T, then

$$\mathbb{E}\left(\widetilde{\check{\boldsymbol{r}}}|\mathcal{F}\right)=\mathbb{E}\left(\widetilde{\check{\boldsymbol{r}}}\right)=\mathbb{E}\left(\check{\boldsymbol{r}}\right)=\check{\boldsymbol{\mu}},\tag{C.99}$$

$$\mathbb{E}\left(\widetilde{\check{\boldsymbol{r}}}\widetilde{\check{\boldsymbol{r}}}'|\mathcal{F}\right)=\mathbb{E}\left(\widetilde{\check{\boldsymbol{r}}}\widetilde{\check{\boldsymbol{r}}}'\right)=\mathbb{E}\left(\check{\boldsymbol{r}}\check{\boldsymbol{r}}'\right)=\check{\boldsymbol{\Sigma}}+\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'.\tag{C.100}$$

Note that $\check{\boldsymbol{\omega}}$ is independent of $\widetilde{\check{\boldsymbol{r}}}$ since it is estimated from the observed data up to T, i.e., $\check{\boldsymbol{R}}_{t},t=1,\cdots,T$. Note also that $\check{\boldsymbol{\omega}}^{\star}$ is constant with respect to the σ-algebra. Hence, the expected discrepancy between the future returns is

$$\begin{aligned}\mathbb{E}\left[\left(\check{Q}^{\star}-\check{Q}\right)^{2}|\mathcal{F}\right]&=\mathbb{E}\left[\left(\left(\check{\boldsymbol{\omega}}^{\star}\right)'\check{\boldsymbol{r}}-\widetilde{\check{\boldsymbol{\omega}}}'\widetilde{\check{\boldsymbol{r}}}\right)^{2}|\mathcal{F}\right]\\&=\mathbb{E}\left[\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)'\widetilde{\check{\boldsymbol{r}}}\widetilde{\check{\boldsymbol{r}}}'\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)|\mathcal{F}\right]\\&=\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)'\mathbb{E}\left(\check{\boldsymbol{r}}\check{\boldsymbol{r}}'\right)\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)\qquad\text{(C.101)}\\&=\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)'\left(\check{\boldsymbol{\Sigma}}+\check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}'\right)\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)\\&\geq\left\|\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right\|_{\check{\boldsymbol{\Sigma}}}^{2}.\qquad\text{(C.102)}\end{aligned}$$

Note that we can rewrite (C.101) as

$$\mathbb{E}\left[\left(\check{Q}^{\star}-\check{Q}\right)^{2}|\mathcal{F}\right]=\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)'\mathbb{E}\left(\check{\boldsymbol{r}}\check{\boldsymbol{r}}'\right)\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)=\sum_{j=1}^{N}\sum_{k=1}^{N}\left(\check{\omega}_{j}^{\star}-\check{\omega}_{j}\right)\left(\check{\omega}_{k}^{\star}-\check{\omega}_{k}\right)\mathbb{E}\left(\check{r}_{j}\check{r}_{k}\right).\tag{C.103}$$

Note also that

$$\begin{aligned}\frac{1}{T}\left\|\check{\boldsymbol{R}}^{\star}-\check{\boldsymbol{R}}\right\|_{2}^{2}&=\frac{1}{T}\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)'\left(\boldsymbol{X}'\boldsymbol{X}\right)\left(\boldsymbol{\omega}^{\star}-\widetilde{\boldsymbol{\omega}}\right)\\&=\sum_{j=1}^{N}\sum_{k=1}^{N}\left[\left(\check{\omega}_{j}^{\star}-\check{\omega}_{j}\right)\left(\check{\omega}_{k}^{\star}-\check{\omega}_{k}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\check{R}_{tj}\check{R}_{tk}\right)\right].\qquad\text{(C.104)}\end{aligned}$$

So

$$\mathbb{E}\left[\left(\check{Q}^{\star}-\check{Q}\right)^{2}|\mathcal{F}\right]-\frac{1}{T}\left\|b\check{R}^{\star}-\check{\boldsymbol{R}}\right\|_{2}^{2}=\sum_{j=1}^{N}\sum_{k=1}^{N}\left(\check{\omega}_{j}^{\star}-\check{\omega}_{j}\right)\left(\check{\omega}_{k}^{\star}-\check{\omega}_{k}\right)\left[\mathbb{E}\left(\check{r}_{j}\check{r}_{k}\right)-\frac{1}{T}\sum_{t=1}^{T}\check{R}_{tj}\check{R}_{tk}\right].\tag{C.105}$$

Define

$$\check{Y}_{j,k} = \mathbb{E}\left(\check{r}_j \check{r}_k\right) - \frac{1}{T}\sum_{t=1}^{T} \check{R}_{tj}\check{R}_{tk}. \tag{C.106}$$

By Assumption C4, we have

$$\begin{aligned}
\mathbb{E}\left[\left(\check{Q}^\star - \check{Q}\right)^2 |\mathcal{F}\right] - \frac{1}{T}\left\|b\check{R}^\star - \check{\boldsymbol{R}}\right\|_2^2 &= \sum_{j=1}^{N}\sum_{k=1}^{N}\left(\check{\omega}_j^\star - \check{\omega}_j\right)\left(\check{\omega}_k^\star - \check{\omega}_k\right)\check{Y}_{j,k} \\
&\leq \sum_{j=1}^{N}\sum_{k=1}^{N}\left(\left|\check{\omega}_j^\star\right| + \left|\check{\omega}_j\right|\right)\left(\left|\check{\omega}_k^\star\right| + \left|\check{\omega}_k\right|\right)\left(\max_{1\leq j,k\leq N}\left|\check{Y}_{j,k}\right|\right) \\
&\leq 4\lambda^2\left(\max_{1\leq j,k\leq N}\left|\check{Y}_{j,k}\right|\right).
\end{aligned} \tag{C.107}$$

Rearranging terms gives

$$\mathbb{E}\left[\left(\check{Q}^\star - \check{Q}\right)^2 |\mathcal{F}\right] \leq \frac{1}{T}\left\|\check{\boldsymbol{R}}^\star - \check{\boldsymbol{R}}\right\|_2^2 + 4\lambda^2\left(\max_{1\leq j,k\leq N}\left|\check{Y}_{j,k}\right|\right). \tag{C.108}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left(\left\|\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\omega}}\right\|_{\boldsymbol{\Sigma}}^2\right) &\leq \mathbb{E}\left\{\mathbb{E}\left[\left(\check{Q}^\star - \check{Q}\right)^2 |\mathcal{F}\right]\right\} \\
&= \mathbb{E}\left[\left(\check{Q}^\star - \check{Q}\right)^2\right] \\
&\leq \frac{1}{T}\mathbb{E}\left(\left\|\check{\boldsymbol{R}}^\star - \check{\boldsymbol{R}}\right\|_2^2\right) + 4\lambda^2\mathbb{E}\left(\max_{1\leq j,k\leq N}\left|\check{Y}_{j,k}\right|\right).
\end{aligned} \tag{C.109}$$

By (C.97) and Lemma (C.4), we know that

$$\frac{1}{T}\mathbb{E}\left(\left\|\check{\boldsymbol{R}}^\star - \check{\boldsymbol{R}}\right\|_2^2\right) \leq O\left(\sqrt{\frac{\log(N)}{T}}\right), \tag{C.110}$$

$$\mathbb{E}\left(\max_{1\leq j,k\leq N}\left|\check{Y}_{j,k}\right|\right) \leq O\left(\sqrt{\frac{\log(N)}{T}}\right). \tag{C.111}$$

Thus,

$$\mathbb{E}\left(\left\|\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\omega}}\right\|_{\boldsymbol{\Sigma}}^2\right) \leq O\left(\sqrt{\frac{\log(N)}{T}}\right). \tag{C.112}$$

By Markov's inequality,

$$\left\|\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\omega}}\right\|_{\boldsymbol{\Sigma}}^2 \xrightarrow{P} 0, \tag{C.113}$$

as $\log(N)/T \to 0$, so does $\left\|\boldsymbol{\omega}^\star - \widetilde{\boldsymbol{\omega}}\right\|_{\boldsymbol{\Sigma}}$ by continuous mapping.

(iv). The third part finishes the whole proof. For the discrepancy of the expected returns,

$$\begin{aligned}\left|\check{\boldsymbol{\mu}}'\check{\boldsymbol{\omega}}-\check{\boldsymbol{\mu}}\check{\boldsymbol{\omega}}^{\star}\right| &= \left|\check{\boldsymbol{\mu}}'\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)\right| \\ &= \left|\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\cdot\check{\boldsymbol{\Sigma}}^{\frac{1}{2}}\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)\right| \\ &\leq \sqrt{\check{\boldsymbol{\mu}}'\check{\boldsymbol{\Sigma}}^{-1}\check{\boldsymbol{\mu}}}\cdot\sqrt{\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)'\check{\boldsymbol{\Sigma}}\left(\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right)} \\ &\leq \sqrt{M}\left\|\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\right\|_{\check{\boldsymbol{\Sigma}}}, \end{aligned} \tag{C.114}$$

where the last two inequalities follow from the Cauchy-Schwarz inequality and Assumption C2.

As $\log(N)/T\to 0$, by (C.113),

$$\left|\check{\boldsymbol{\mu}}'\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\mu}}\check{\boldsymbol{\omega}}\right|\xrightarrow{P}0. \tag{C.115}$$

For the discrepancy of the volatility, by the triangular inequality for the norm $\|\cdot\|_{\check{\boldsymbol{\Sigma}}}$, we have

$$\left|\check{\boldsymbol{\omega}}'\boldsymbol{\Sigma}\check{\boldsymbol{\omega}}-\left(\check{\boldsymbol{\omega}}^{\star}\right)'\check{\boldsymbol{\Sigma}}\check{\boldsymbol{\omega}}^{\star}\right|=\left|\|\boldsymbol{\omega}^{\star}\|_{\boldsymbol{\Sigma}}^{2}-\|\widetilde{\boldsymbol{\omega}}\|_{\check{\boldsymbol{\Sigma}}}^{2}\right|\leq\|\check{\boldsymbol{\omega}}^{\star}-\check{\boldsymbol{\omega}}\|_{\check{\boldsymbol{\Sigma}}}^{2}\xrightarrow{P}0. \tag{C.116}$$

Again by (C.113), we have as $\log(N)/T\to 0$,

$$\left|\check{\boldsymbol{\omega}}'\boldsymbol{\Sigma}\check{\boldsymbol{\omega}}-\left(\check{\boldsymbol{\omega}}^{\star}\right)'\check{\boldsymbol{\Sigma}}\check{\boldsymbol{\omega}}^{\star}\right|\xrightarrow{P}0. \tag{C.117}$$

□

C.4 Proof Step 2: Finishing the Proof of Theorem C.1

Proof. With the in-probability convergence of the expected return and variance of LASSO-regularized portfolio $\check{\boldsymbol{\omega}}$ for the N non-basis assets, we then turn to prove Theorem C.1 and show that the expected return and volatility of the overall LASSO-regularized portfolio $\check{\boldsymbol{\omega}}_{all}=(\check{\omega}_0,\check{\boldsymbol{\omega}}')'=(1-\check{\boldsymbol{\omega}}'\mathbf{1}_N,\check{\boldsymbol{\omega}}')'$ converge in probability to those yielded by $\boldsymbol{\omega}_{all}^{\star}=(\check{\omega}_0^{\star},(\check{\boldsymbol{\omega}}^{\star})')'=(1-(\check{\boldsymbol{\omega}}^{\star})'\mathbf{1}_N,\check{\boldsymbol{\omega}}^{\star})$. Recall that with random returns on all the assets $\boldsymbol{r}_{all}\equiv(r_0,\boldsymbol{r}_{-0}')'$, the return of a general overall portfolio $\boldsymbol{\omega}_{all}=(1-\boldsymbol{\omega}_N'\mathbf{1}_N,\boldsymbol{\omega}_N')'$ is

$$\boldsymbol{\omega}_{all}'\boldsymbol{r}_{all}=r_0+\boldsymbol{\omega}_N'\left(\boldsymbol{r}_{-0}-r_0\mathbf{1}_N\right)=r_0+\boldsymbol{\omega}_N'\check{\boldsymbol{r}}. \tag{C.118}$$

So the expected return and variance of such general portfolio are

$$\mu_p\equiv\boldsymbol{\mu}_{all}'\boldsymbol{\omega}_{all}$$

$$
\begin{aligned}
&= \mathbb{E}\left(\boldsymbol{\omega}_{all}'\boldsymbol{r}_{all}\right) \\
&= \mathbb{E}\left(r_0 + \boldsymbol{\omega}_N'\check{\boldsymbol{r}}\right) \\
&= \mu_0 + \boldsymbol{\omega}_N'\check{\boldsymbol{\mu}}, \qquad \text{(C.119)} \\
\sigma_p^2 &\equiv \boldsymbol{\omega}_{all}'\boldsymbol{\Sigma}_{all}\boldsymbol{\omega}_{all} \\
&= \mathrm{Var}\left(\boldsymbol{\omega}_{all}'\boldsymbol{r}_{all}\right) \\
&= \mathrm{Var}\left[r_0 + \boldsymbol{\omega}_N'\check{\boldsymbol{r}}\right] \\
&= \sigma_0^2 + 2\boldsymbol{\omega}_N'\mathrm{Cov}\left(\check{\boldsymbol{r}}, r_0\right) + \boldsymbol{\omega}_N'\mathrm{Var}\left(\check{\boldsymbol{r}}\right)\boldsymbol{\omega}_N \\
&= \sigma_0^2 + 2\boldsymbol{\omega}_N'\boldsymbol{\sigma}_{-0} + \boldsymbol{\omega}_N'\boldsymbol{\Sigma}\boldsymbol{\omega}_N. \qquad \text{(C.120)}
\end{aligned}
$$

Replace the general $\boldsymbol{\omega}_N$ above with $\check{\boldsymbol{\omega}}$ and $\check{\boldsymbol{\omega}}^\star$, respectively, and the discrepancy of the expected returns between the two portfolios is then

$$
\begin{aligned}
\left|\boldsymbol{\mu}_{all}'\check{\boldsymbol{\omega}}_{all} - \boldsymbol{\mu}_{all}'\check{\boldsymbol{\omega}}_{all}^\star\right| &= \left|\left(\mu_0 + \check{\boldsymbol{\omega}}'\check{\boldsymbol{\mu}}\right) - \left(\mu_0 + \left(\check{\boldsymbol{\omega}}^\star\right)'\check{\boldsymbol{\mu}}\right)\right| \\
&= \left|\check{\boldsymbol{\mu}}'\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\mu}}'\check{\boldsymbol{\omega}}\right| \\
&\leq \sqrt{M}\left\|\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\omega}}\right\|_{\check{\boldsymbol{\Sigma}}}, \qquad \text{(C.121)}
\end{aligned}
$$

where the last inequality follows from equation (C.114). As $\log(N)/T \to 0$, by (C.113),

$$
\left|\boldsymbol{\mu}_{all}'\check{\boldsymbol{\omega}}_{all} - \boldsymbol{\mu}_{all}'\check{\boldsymbol{\omega}}_{all}^\star\right| \xrightarrow{P} 0. \qquad \text{(C.122)}
$$

As for the discrepancy of variance, by (C.120), we have

$$
\begin{aligned}
\left|\check{\boldsymbol{\omega}}_{all}'\boldsymbol{\Sigma}_{all}\check{\boldsymbol{\omega}}_{all} - \left(\boldsymbol{\omega}_{all}^\star\right)'\boldsymbol{\Sigma}_{all}\boldsymbol{\omega}_{all}^\star\right| &= \left|\left(\sigma_0^2 + 2\check{\boldsymbol{\omega}}'\boldsymbol{\sigma}_{-0} + \check{\boldsymbol{\omega}}'\check{\boldsymbol{\Sigma}}\check{\boldsymbol{\omega}}\right) - \left(\sigma_0^2 + 2\left(\check{\boldsymbol{\omega}}^\star\right)'\boldsymbol{\sigma}_{-0} + \left(\check{\boldsymbol{\omega}}^\star\right)'\check{\boldsymbol{\Sigma}}\check{\boldsymbol{\omega}}^\star\right)\right| \\
&\leq 2\left|\boldsymbol{\sigma}_{-0}'\left(\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\omega}}\right)\right| + \left\|\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\omega}}\right\|_{\check{\boldsymbol{\Sigma}}}^2 \\
&= 2\left|\boldsymbol{\sigma}_{-0}'\check{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}\left(\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\omega}}\right)\right| + \left\|\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\omega}}\right\|_{\check{\boldsymbol{\Sigma}}}^2 \\
&\leq 2\sqrt{\boldsymbol{\sigma}_{-0}'\check{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\sigma}_{-0}}\cdot\left\|\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\omega}}\right\|_{\check{\boldsymbol{\Sigma}}} + \left\|\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\omega}}\right\|_{\check{\boldsymbol{\Sigma}}}^2 \\
&\leq 2\sqrt{H}\left\|\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\omega}}\right\|_{\check{\boldsymbol{\Sigma}}} + \left\|\check{\boldsymbol{\omega}}^\star - \check{\boldsymbol{\omega}}\right\|_{\check{\boldsymbol{\Sigma}}}^2, \qquad \text{(C.123)}
\end{aligned}
$$

where the last inequality holds by Assumption B4. Again by (C.113), as $\log(N)/T \to 0$, we have

$$
\left|\check{\boldsymbol{\omega}}_{all}'\boldsymbol{\Sigma}_{all}\check{\boldsymbol{\omega}}_{all} - \left(\check{\boldsymbol{\omega}}_{all}^\star\right)'\boldsymbol{\Sigma}_{all}\check{\boldsymbol{\omega}}_{all}^\star\right| \xrightarrow{P} 0, \qquad \text{(C.124)}
$$

which completes the proof. □