

# Improving Index Funds via Idiosyncratic Returns

Saerom Lee\* and Tae-Hwy Lee†

September 11, 2026

## Abstract

The rising concentration of major market indices has increasingly undermined the diversification benefits and increased the market risk of index investing. To address this issue, we propose a portfolio strategy designed to minimize variance relative to a given index fund. We decompose each stock's return into an index return and an idiosyncratic return, defined as the excess return over the index. By exploiting the resulting covariance structure—either between the index return and idiosyncratic returns or among the idiosyncratic returns themselves—we construct portfolios that enhance index efficiency by reducing portfolio variance. Using U.S. monthly stock data, we show that incorporating idiosyncratic returns into portfolio construction through this covariance structure consistently reduces out-of-sample variance relative to multiple basis index portfolios, including the S&P 500 index, the Equal-weighted S&P 500 index, and the CRSP value-weighted market portfolio. Crucially, portfolio returns remain comparable to those of the basis portfolios, leading to overall improvements in Sharpe ratios. These findings demonstrate that effective variance reduction can be achieved without sacrificing return performance.

*Keywords:* Return Decomposition, Adaptive Lasso, Boosting, Graphical Lasso, Idiosyncratic Returns, Sharpe Ratio

*JEL classification:* C22, G11

---

\*Department of Economics, University of Toledo, Toledo, OH 43606. E-mail: saerom.lee@utoledo.edu

†Department of Economics, University of California, Riverside, CA 92521. E-mail: taelee@ucr.edu.

# 1 Introduction

Amid growing concerns over market risk driven by the rising concentration of major indices, this paper proposes a portfolio strategy designed to minimize variance relative to an index fund. To this end, we exploit stock-specific information beyond what traditional index investing can capture. We decompose individual stock returns into the index fund return and their respective idiosyncratic returns, defined as the excess return over the index fund. By exploring the resulting covariance structure, we construct portfolios that improve upon the basis portfolio by reducing variance.

The mean-variance framework is a foundational pillar of investment theory, seeking to construct portfolios that optimize the trade-off between expected return and risk (Markowitz (1952)). Within this framework, many investors and researchers prioritize the Global Minimum Variance portfolio, which yields the lowest possible variance for a given set of asset returns. A central insight of modern portfolio theory is that strategic diversification enables investors to mitigate risk without sacrificing performance. This theoretical foundation has fueled a long-standing debate over the relative merits of active stock selection and passive management.

Proponents of active management argue that professional skill is not fully reflected in asset prices, allowing for predictable alpha generation and informed capital allocation (Gruber (1996); Jones and Wermers (2011); Risik (2013)). Empirical evidence suggests that active managers add value specifically through skillful stock selection and tend to outperform during recessionary periods (Moskowitz (2000); Kosowski (2011)). Conversely, proponents of passive management maintain that investment success rests on transparency and cost-efficiency. Passive management—primarily represented by index funds—eliminates manager-specific risks and offers significantly lower fees.

Index funds—passively managed mutual funds or exchange-traded funds that mirror a specific market benchmark such as the S&P 500—have grown in popularity by providing instant diversification and broad market exposure. A growing body of literature supports their widespread adoption, documenting that index funds consistently outperform active funds on an after-expense basis, achieve risk-adjusted performance comparable to that of skilled active man-

agers, and sustain their market power through investor inertia and heterogeneous preferences (Fortin and Michelson (2002); Crane and Crotty (2018); Brown *et al.* (2023)).

While index funds have historically flourished, the rising concentration within major market indices over recent decades has begun to erode the diversification benefits of index investing. The S&P 500 index, historically viewed as a balanced cross-section of the U.S. economy, has transformed into a technology- and AI-dominated index. This structural change, termed the “Great Narrowing”, represents a sharp increase in market concentration that dilutes the efficacy of diversification (RBC Wealth Management (2026)). Between 1990 and 2015, the ten largest companies by market capitalization constituted approximately 19% of the index; by the end of 2025, this figure had surged to nearly 41%. When a handful of mega-cap firms dictate the trajectory of the entire index, the distinction between systematic market risk and idiosyncratic stock risk becomes increasingly blurred. As a result, the index loses much of its diversification benefit, heightening risk for passive investors and leaving portfolios more vulnerable to sector-specific shocks.

Motivated by this structural shift, we aim to improve upon an index fund by reducing portfolio variance, especially when the number of assets,  $N$ , is comparable to or greater than the sample size,  $T$ . Our strategy rests on a decomposition of individual stock returns into an index fund return and idiosyncratic returns—an approach we refer to as the “return decomposition” throughout this paper. This decomposition gives rise to two distinct covariance terms that we exploit to reduce portfolio variance.

First, we consider the covariance between the index fund return and idiosyncratic returns, adjusting individual stock weights toward the minimum variance portfolio through penalized regression (Section 3). Second, we capture the covariance among idiosyncratic returns themselves via the precision matrix, assuming zero correlation between the index fund return and idiosyncratic returns (Section 4). To further highlight the importance of this return decomposition, we also examine a non-decomposed specification where the index fund is directly included as an additional asset alongside individual stocks (Section 5).

The remainder of this paper is organized as follows. Section 2 reviews the minimum variance

portfolio and introduces the return decomposition framework. Sections 3 and 4 develop our proposed strategies by exploring the covariance structure arising from the return decomposition. Section 5 presents the non-decomposition portfolio, which includes the index fund directly alongside individual stocks. Section 6 reports the empirical results using U.S. stock data, and Section 7 concludes.

## 2 Minimum Variance Portfolio with the Return Decomposition

Suppose we consider  $N$  stocks (indexed by  $i$ ) over  $T$  periods (indexed by  $t$ ). Let  $r_{i,t}$  denote the return of individual stock  $i$ , and  $w_i$  denote the corresponding portfolio weight for  $i = 1, \dots, N$ . Throughout this paper, all returns are demeaned by their sample means, so that  $E(r_{i,t}) = 0$ . The portfolio return, denoted as  $r_{p,t}$ , can be represented as:

$$r_{p,t} = \sum_{i=1}^N w_i r_{i,t}, \quad t = 1, \dots, T, \quad (1)$$

where  $\sum_{i=1}^N w_i = 1$  ensures that the portfolio weights sum to one.

The minimum variance portfolio is obtained by solving the following optimization problem:

$$\min_{\{w_i\}_{i=1}^N} E(r_{p,t}^2) = \sigma_p^2 = \sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_{ij} \quad \text{subject to} \quad \sum_{i=1}^N w_i = 1, \quad (2)$$

where  $\sigma_p^2$  represents the portfolio variance, and  $\sigma_{ij}$  is the covariance between the returns of stocks  $i$  and  $j$ . The solution for this minimum variance portfolio is given by  $w_i = \frac{\sum_{j=1}^N \sigma_{ij}}{\sum_{k=1}^N \sum_{j=1}^N \sigma_{kj}}$ .

To reformulate Equation (1) by incorporating the index fund, we define the index fund return as:

$$r_{0,t} = \sum_{i=1}^N a_i r_{i,t}, \quad (3)$$

where  $a_i$  denotes the predetermined weight of stock  $i$  in the index, satisfying  $\sum_{i=1}^N a_i = 1$ . Throughout this paper, the index fund serves as the basis portfolio—the starting point from

which the minimum variance portfolio is constructed by incorporating idiosyncratic return information. Three index portfolios (the S&P 500 index fund, the Equal-weighted S&P 500 portfolio, and the value-weighted portfolio of CRSP firms) are considered in our empirical application (Section 6).

The main goal of this paper is to shift from the basis portfolio weights ( $a_i$ ) toward the risk-optimized minimum variance weights ( $w_i$ ) by exploiting the information contained in idiosyncratic returns. Using Equation (3), the portfolio return  $r_{p,t}$  can be re-expressed as a linear combination of the index return  $r_{0,t}$  and individual stock returns  $r_{i,t}$ , with corresponding weights  $b_0$  and  $b_i$ :

$$r_{p,t} = b_0 r_{0,t} + \sum_{i=1}^N b_i r_{i,t}, \quad (4)$$

where  $b_0 + \sum_{i=1}^N b_i = 1$  ensures that the portfolio weight constraint is satisfied. By adjusting each stock's index weight ( $a_i$ ) using the weight  $b_i$  from Equation (4), the portfolio weight  $w_i$  in Equation (1) can be expressed as:

$$w_i = b_0 a_i + b_i, \quad (5)$$

given that  $r_{0,t} = \sum_{i=1}^N a_i r_{i,t}$ .

Substituting  $b_0 = 1 - \sum_{i=1}^N b_i$  into Equation (4) and rearranging yields:

$$\begin{aligned} r_{p,t} &= r_{0,t} + \sum_{i=1}^N b_i (r_{i,t} - r_{0,t}) \\ &= r_{0,t} + \sum_{i=1}^N b_i d_{i,t}, \end{aligned} \quad (6)$$

where

$$d_{i,t} \equiv r_{i,t} - r_{0,t} \quad (7)$$

is referred to as the “idiosyncratic return” throughout this paper. It captures the portion

of the individual stock return not explained by the index fund return. Equivalently,  $r_{i,t} = r_{0,t} + d_{i,t}$ , which we refer to as the return decomposition, separates each individual stock return into the index fund return and the idiosyncratic return. Accordingly, Equation (6) shows that the portfolio return can be expressed as the index fund return augmented by a weighted sum of idiosyncratic returns, where the weights  $b_i$  determine the extent to which each stock's idiosyncratic return is exploited.

Using the return decomposition, the minimum variance optimization problem in Equation (2) can be rewritten as:

$$\min_{\{b_i\}_{i=1}^N} \sigma_p^2 = \sigma_{r_0}^2 + 2 \sum_{i=1}^N b_i \text{Cov}(r_{0,t}, d_{i,t}) + \sum_{i=1}^N b_i^2 \sigma_{d_i}^2 + \sum_{i \neq j} b_i b_j \text{Cov}(d_{i,t}, d_{j,t}), \quad (8)$$

where  $\sigma_{r_0}^2$  is the variance of the index fund return and  $\sigma_{d_i}^2$  is the variance of the  $i$ -th idiosyncratic return.

Equation (8) reveals two key covariances that arise from the return decomposition: the covariance between the index fund return and idiosyncratic returns,  $\text{Cov}(r_{0,t}, d_{i,t})$ , and the covariance among idiosyncratic returns,  $\text{Cov}(d_{i,t}, d_{j,t})$ , which are explored in Section 3 and Section 4, respectively. Once  $b_i$  is estimated by solving Equation (8), the minimum variance portfolio weights  $w_i$  follow from Equation (5) given the stock's index weight  $a_i$ .

### 3 Exploring Covariance between an Index Fund Return and Idiosyncratic Returns

Building on the return decomposition ( $r_{i,t} = r_{0,t} + d_{i,t}$ ) introduced in Section 2, this section explores the covariance between the index fund return and idiosyncratic returns ( $\text{Cov}(r_{0,t}, d_{i,t})$ ) in Equation (8) to reduce portfolio variance. To do so, we rearrange Equation (6) into a

regression framework:

$$r_{0,t} = \sum_{i=1}^N b_i(-d_{i,t}) + r_{p,t}, \quad (9)$$

where the weights  $b_i$  are estimated by regressing the index fund return  $r_{0,t}$  on the negative idiosyncratic returns  $(d_{i,t})$ , with the regression residuals corresponding to the portfolio returns  $r_{p,t}$ . Minimizing the residual sum of squares is therefore equivalent to finding the minimum variance portfolio weights that exploit the covariance between the index fund return and idiosyncratic returns.

Given that  $N$  may be large relative to  $T$ , we employ three penalized regression methods to estimate the coefficients  $b_i$  in Equation (9). First, we utilize the Adaptive Lasso (ALasso), which applies adaptive weights to the  $L_1$  penalty to penalize coefficients differentially (Zou (2006)). While the Lasso can be inconsistent in variable selection, the ALasso achieves the oracle property.

Second, we employ the Smoothly Clipped Absolute Deviation (SCAD) penalty (Fan and Li (2001)). As a non-concave penalized likelihood estimator, SCAD also achieves the oracle property. Specifically, if the true parameters include zero components, they are estimated as exactly zero with probability approaching one, and the nonzero components are estimated efficiently as if the true model were known in advance.

Third, we use the component-wise  $L_2$ -Boosting procedure (Bühlmann (2006)). This iterative procedure starts from the index fund as the basis portfolio and, at each iteration, incorporates the idiosyncratic return  $d_{i,t}$  that most significantly reduces the portfolio variance. In the empirical application in Section 6, we set the step-size parameter (learning rate) to  $\nu = 0.0001$  and determine the optimal number of iterations—up to a maximum of 5,000—by minimizing the Akaike Information Criterion (Bühlmann (2006)). After estimating  $b_i$  via ALasso, SCAD, or  $L_2$ -Boosting, the final portfolio weights  $w_i$  are constructed according to Equation (5).

## 4 Exploring Covariance among Idiosyncratic Returns

Whereas Section 3 explores the covariance between the index fund return and idiosyncratic returns ( $\text{Cov}(r_{0,t}, d_{i,t})$ ) in Equation (8), this section investigates the covariance among the idiosyncratic returns themselves ( $\text{Cov}(d_{i,t}, d_{j,t})$ ). Throughout this section, we assume that the index fund return and idiosyncratic returns are uncorrelated ( $\text{Cov}(r_{0,t}, d_{i,t}) = 0$ ).<sup>1</sup>

We reformulate the portfolio return in Equation (1) and the minimum variance optimization problem presented in Section 2 into vector form. Let  $\mathbf{r}_t = (r_{1,t}, \dots, r_{N,t})' \in \mathbb{R}^N$  denote the vector of individual stock returns,  $\mathbf{w} = (w_1, \dots, w_N)' \in \mathbb{R}^N$  denote the vector of portfolio weights, and  $\boldsymbol{\iota}_N = (1, \dots, 1)' \in \mathbb{R}^N$  represent a vector of ones. Furthermore, let  $\boldsymbol{\Sigma}_r$  denote the covariance matrix of  $N$  individual stock returns. The portfolio return in Equation (1) is then expressed as  $r_{p,t} = \mathbf{w}'\mathbf{r}_t$ , and the optimization problem in Equation (2) becomes  $\min_{\mathbf{w}} \mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w}$  subject to  $\mathbf{w}'\boldsymbol{\iota}_N = 1$ , with solution  $\mathbf{w} = \frac{\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\iota}_N}{\boldsymbol{\iota}_N'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\iota}_N}$ . Using the return decomposition ( $r_{i,t} = r_{0,t} + d_{i,t}$ ) and under the assumption of zero correlation between  $r_{0,t}$  and  $d_{i,t}$  ( $\text{Cov}(r_{0,t}, d_{i,t}) = 0$ ), the covariance matrix can be expressed as  $\boldsymbol{\Sigma}_r = \sigma_{r_0}^2 \boldsymbol{\iota}_N \boldsymbol{\iota}_N' + \boldsymbol{\Sigma}_d$ , where  $\boldsymbol{\Sigma}_d$  denotes the covariance matrix of the idiosyncratic returns.

When the number of stocks is less than the sample size ( $N < T$ ), the sample covariance matrix can be estimated directly and substituted into the weight solution. However, as the number of stocks grows, the condition number of the sample covariance matrix rises, leading to numerical instability—a phenomenon commonly referred to as “Markowitz’s curse” (DeMiguel *et al.* (2009); López de Prado (2016); López de Prado *et al.* (2025)).

To address this, previous studies have imposed sparsity assumptions on the precision matrix  $\boldsymbol{\Sigma}_r^{-1}$ . For instance, Awoye (2016) propose sparse inverse covariance estimation via the Graphical Lasso (GLasso). Similarly, Cai *et al.* (2020) assume sparsity in the precision matrix and estimate the minimum variance portfolio using constrained  $l_1$  minimization. Lee and Seregina (2024) further develop the Factor Graphical Lasso framework, which integrates graphical models with a factor structure by decomposing the precision matrix into low-rank and sparse components.

---

<sup>1</sup>This amounts to treating the index return  $r_{0,t}$  as a common factor of  $N$  individual stock returns which is orthogonal to the idiosyncratic components, as assumed in the factor model literature, e.g., Lee and Seregina (2024).

We adopt the Factor Graphical Lasso framework of Lee and Seregina (2024) to estimate the precision matrix. In contrast to Lee and Seregina (2024), who consider either latent factors estimated via principal component analysis or observable Fama–French factors, we treat the market index fund as the sole observable common factor in the return decomposition ( $r_{i,t} = r_{0,t} + d_{i,t}$ ). This specification is motivated by Sharpe (1963), who demonstrates that the return of each security is related to the level of a market index, and aligns with our objective of constructing a portfolio that minimizes variance relative to an index fund.

We estimate the precision matrix of idiosyncratic returns ( $\Sigma_d^{-1}$ ) by minimizing a penalized Bregman divergence with a weighted Graphical Lasso penalty, which shrinks selected off-diagonal entries of the precision matrix to zero. For the empirical application in Section 6, the tuning parameter for the shrinkage is selected by minimizing the Bayesian Information Criterion via a grid search (see Equations (3.3) and (3.4) in Lee and Seregina (2024) for details).

After obtaining the estimated precision matrix of idiosyncratic returns ( $\hat{\Sigma}_d^{-1}$ ), the full precision matrix is recovered via the Woodbury matrix identity:

$$\hat{\Sigma}_r^{-1} = \hat{\Sigma}_d^{-1} - \hat{\Sigma}_d^{-1} \boldsymbol{\iota}_N \left[ \frac{1}{\hat{\sigma}_{r_0}^2} + \boldsymbol{\iota}_N' \hat{\Sigma}_d^{-1} \boldsymbol{\iota}_N \right]^{-1} \boldsymbol{\iota}_N' \hat{\Sigma}_d^{-1}. \quad (10)$$

The portfolio weights are then obtained as  $\hat{\mathbf{w}} = \frac{\hat{\Sigma}_r^{-1} \boldsymbol{\iota}_N}{\boldsymbol{\iota}_N' \hat{\Sigma}_r^{-1} \boldsymbol{\iota}_N}$ , and the portfolio is constructed based on Equation (1). For brevity, we label this portfolio as GLasso (with return decomposition) in all tables and figures throughout Section 6 and Appendix A, as its weights are determined using a weighted Graphical Lasso penalty applied to the precision matrix of idiosyncratic returns.

## 5 Minimum Variance Portfolio without the Return Decomposition

In this section, to examine the importance of the return decomposition ( $r_{i,t} = r_{0,t} + d_{i,t}$ ), we consider a direct specification in Equation (4) that foregoes it entirely. Specifically, we treat the index fund as a separate  $(N + 1)$ -th asset alongside the  $N$  individual stocks, constructing

a minimum variance portfolio across  $(N + 1)$  assets. In the empirical application (Section 6), evaluating this portfolio without the return decomposition allows us to illustrate its value by comparing its performance with that of the strategies developed in Sections 3 and 4 based on Equation (6). Our empirical results highlight the merits of applying Equation (6) with the return decomposition, rather than Equation (4) without it.

We rewrite Equation (4) in vector form, including the index fund directly as a separate asset. Let  $\mathbf{R}_t = (r_{0,t}, \mathbf{r}'_t)' \in \mathbb{R}^{N+1}$  denote the vector of all asset returns including the index fund return,  $\mathbf{b} = (b_0, b_1, \dots, b_N)' \in \mathbb{R}^{N+1}$  denote the corresponding portfolio weight vector, and  $\boldsymbol{\iota}_{N+1} = (1, \dots, 1)' \in \mathbb{R}^{N+1}$  represent a vector of ones. Let  $\boldsymbol{\Sigma}_R$  denote the covariance matrix of all  $(N + 1)$  asset returns. The portfolio return in Equation (4) is then expressed as  $r_{p,t} = \mathbf{b}'\mathbf{R}_t$ , and the optimization problem becomes  $\min_{\mathbf{b}} \mathbf{b}'\boldsymbol{\Sigma}_R\mathbf{b}$  subject to  $\mathbf{b}'\boldsymbol{\iota}_{N+1} = 1$ , with solution  $\mathbf{b} = \frac{\boldsymbol{\Sigma}_R^{-1}\boldsymbol{\iota}_{N+1}}{\boldsymbol{\iota}_{N+1}'\boldsymbol{\Sigma}_R^{-1}\boldsymbol{\iota}_{N+1}}$ .

As noted in Section 4, estimating the precision matrix becomes numerically unstable as the number of assets grows. To address this issue, we impose sparsity on the precision matrix of the  $(N + 1)$  asset returns and estimate it by minimizing a penalized Bregman divergence with a weighted Graphical Lasso penalty—equivalent to maximizing the penalized log-likelihood—following the estimation strategy in Section 4. For the empirical application in Section 6, the tuning parameter is selected using the Bayesian Information Criterion. The estimated precision matrix  $\hat{\boldsymbol{\Sigma}}_R^{-1}$  is then used to obtain the portfolio weight vector  $\hat{\mathbf{b}}$  and construct the final portfolio based on Equation (4).

Note that the weights  $\mathbf{b}$  could alternatively be estimated through a regression-based framework under a sparsity assumption, as in Section 3. However, since our primary goal in this section is to examine the role of the return decomposition rather than to compare estimation techniques, we restrict our focus to GLasso. We label this portfolio as GLasso (without return decomposition) in all tables and figures throughout Section 6 and Appendix A.

## 6 Empirical Application

This section evaluates the empirical performance of the portfolios discussed in Sections 3 to 5. Since our primary objective is to minimize portfolio variance relative to an index fund, we focus on the out-of-sample (OOS) variance as our primary performance metric, while also examining the Sharpe ratios (SRs). We consider three basis portfolios for  $r_{0,t}$ : the S&P 500 index fund, the Equal-weighted S&P 500 (EW) portfolio, and the value-weighted portfolio of CRSP firms.

### 6.1 Data and Empirical Setting

We examine whether the portfolios leveraging the return decomposition and idiosyncratic returns (from Sections 3 and 4) achieve lower variance relative to their respective basis portfolios. We also include the portfolio from Section 5, which foregoes the return decomposition, to highlight the importance of exploiting idiosyncratic returns derived from the return decomposition.

We first employ the S&P 500 index fund as the basis portfolio, utilizing monthly excess returns for both the index fund and its constituent stocks. In this exercise, the idiosyncratic returns  $d_{i,t}$  are defined as the excess returns of individual stocks over the S&P 500 index return. Following the data construction procedure of Lee and Seregina (2024), our full sample spans from January 1980 to December 2019 ( $T = 480$ ) and consists of  $N = 355$  stocks.

The total sample is divided into an in-sample period ( $T_1$ ) and an out-of-sample period ( $T_2$ ) such that  $T = T_1 + T_2$ . We consider three distinct sample splits for OOS evaluation: (i) a 30-year in-sample period ( $T_1 = 360$ ) with a 10-year OOS period ( $T_2 = 120$ ); (ii) a 25-year in-sample period ( $T_1 = 300$ ) with a 15-year OOS period ( $T_2 = 180$ ); and (iii) a 20-year in-sample period ( $T_1 = 240$ ) with a 20-year OOS period ( $T_2 = 240$ ). The in-sample period is used to estimate portfolio weights, while the OOS period is reserved for performance evaluation. For all sample splits, the OOS window ends in December 2019, and portfolios are rebalanced on a monthly basis using a rolling-window approach. Additional results with alternative sample splits are provided in Appendix A.

Next, we replace the S&P 500 index fund with the EW portfolio as the basis portfolio,

utilizing the same set of individual stock excess returns. Here, each stock's initial weight is set to  $a_{i,t} = \frac{1}{N}$ , and the idiosyncratic returns  $d_{i,t}$  are defined as the difference between individual stock excess returns and the EW portfolio excess return. This exercise examines whether incorporating idiosyncratic returns further reduces portfolio variance relative to a naive  $\frac{1}{N}$  strategy, motivated by evidence that EW portfolios often outperform market-capitalization-weighted benchmarks over long horizons (e.g., Malladi and Fabozzi (2017); Taljaard and Maré (2021); Swade *et al.* (2023)).

Lastly, we employ the value-weighted portfolio of CRSP firms as the basis portfolio (referred to as CRSP in tables and figures for brevity). This portfolio, corresponding to the market factor in Fama and French (1992) and Fama and French (2015), provides broader market coverage than the S&P 500, which is restricted to large-cap equities. While some S&P 500 constituents may not be present in the CRSP universe due to dual-class shares, listing issues, or timing mismatches, we assume for simplicity that all S&P 500 constituents are contained within the CRSP sample. Under this setting, the value-weighted portfolio of CRSP firms consists of  $M$  stocks (where  $M > N$ ), incorporating equities beyond the S&P 500 universe. Its return at time  $t$  is given by  $r_{0,t} \equiv \sum_{i=1}^M c_i r_{i,t}$ , where  $c_i$  denotes the weight assigned to stock  $i$ , satisfying  $\sum_{i=1}^M c_i = 1$ .

To further characterize the portfolio, these  $M$  stocks are partitioned into two disjoint sets:  $\mathcal{S}$ , the set of S&P 500 constituents with cardinality  $|\mathcal{S}| = N$ , and  $\mathcal{F}$ , the set of CRSP firms not included in the S&P 500 index, where  $|\mathcal{F}| = M - N$ . The portfolio return can thus be decomposed as  $r_{0,t} = \sum_{i=1}^N c_i r_{i,t} + \sum_{i=N+1}^M c_i r_{i,t}$ .

Accordingly, the portfolio return in Equation (1) can be rewritten by partitioning it into these two disjoint sets. Grouping the weights of the S&P 500 constituents and isolating those of the excluded CRSP firms, the portfolio return  $r_{p,t}$  is expressed as:

$$r_{p,t} = \sum_{i=1}^N (b_0 c_i + b_i) r_{i,t} + \sum_{i=N+1}^M b_0 c_i r_{i,t}. \quad (11)$$

Thus, the portfolio weight  $w_i$  depends on set membership: for  $i \in \mathcal{S}$ ,  $w_i = b_0 c_i + b_i$ , and for

$i \in \mathcal{F}$ ,  $w_i = b_0 c_i$ .<sup>2</sup> Given that  $\sum_{i=1}^M c_i = 1$  and  $b_0 = 1 - \sum_{i=1}^N b_i$ , the weights sum to one, satisfying  $\sum_{i=1}^M w_i = 1$ .<sup>3</sup>

By isolating the weights of the S&P 500 constituents in Equation (11), we can restrict the variance minimization exclusively to the idiosyncratic returns of the same S&P 500 constituents ( $N = 355$ ) used in the previous exercises. This demonstrates that exploiting idiosyncratic information from even a subset of constituent stocks can effectively mitigate total portfolio risk. Accordingly, the idiosyncratic returns  $d_{i,t}$  are defined as the difference between the individual excess returns of the S&P 500 constituents and the value-weighted excess return of the CRSP firms.

For all three basis portfolios, we employ an expanding five-fold cross-validation procedure to select the optimal regularization parameters for ALasso and SCAD discussed in Section 3. This approach respects the temporal ordering of time-series data, ensuring that models are trained strictly on historical information. Specifically, we divide the in-sample period into five folds, where the first window is used to evaluate the second window. The training set then expands to include both the first and second windows to evaluate the third, and this process continues sequentially across all folds.

The initial training window covers 40% of the in-sample observations and expands sequentially across folds. For instance, given a 30-year in-sample period ( $T_1 = 360$ ; January 1980 to December 2009), the first training window (spanning 144 months from January 1980 to December 1991) is evaluated on the first test window (43 months, from January 1992 to July 1995). The second training window then expands to include data through July 1995, with the subsequent test window spanning 43 months from August 1995 to February 1999. This process repeats across five folds to identify the penalty parameter that minimizes portfolio variance. As a robustness check, Appendix A provides additional results using an initial training window

---

<sup>2</sup>To derive Equation (11), we start from Equation (4) and express the basis portfolio as  $r_{0,t} = \sum_{i=1}^M c_i r_{i,t}$ . Thus, we have  $r_{p,t} = b_0 r_{0,t} + \sum_{i=1}^N b_i r_{i,t} = b_0 \sum_{i=1}^M c_i r_{i,t} + \sum_{i=1}^N b_i r_{i,t} = b_0 \left( \sum_{i=1}^N c_i r_{i,t} + \sum_{i=N+1}^M c_i r_{i,t} \right) + \sum_{i=1}^N b_i r_{i,t} = \sum_{i=1}^N (b_0 c_i + b_i) r_{i,t} + \sum_{i=N+1}^M b_0 c_i r_{i,t}$ .

<sup>3</sup>The portfolio weights satisfy the sum-to-one constraint, as  $\sum_{i=1}^M w_i = \sum_{i=1}^N (b_0 c_i + b_i) + \sum_{i=N+1}^M b_0 c_i = \sum_{i=1}^N b_0 c_i + \sum_{i=N+1}^M b_0 c_i + \sum_{i=1}^N b_i = b_0 \sum_{i=1}^M c_i + \sum_{i=1}^N b_i = b_0 + (1 - b_0) = 1$  (since  $\sum_{i=1}^M c_i = 1$  and  $b_0 = 1 - \sum_{i=1}^N b_i$ ).

of 30% of in-sample observations.

To evaluate the OOS performance of a portfolio, we report the OOS standard deviation ( $\hat{\sigma}_{\text{oos}}$ ) as our primary portfolio risk measure, alongside the OOS average portfolio return ( $\hat{\mu}_{\text{oos}}$ ), minimum portfolio return ( $\hat{r}_{p,\min}$ ), maximum portfolio return ( $\hat{r}_{p,\max}$ ), and Sharpe ratio (SR). Since standard deviation is the square root of variance, minimizing one is equivalent to minimizing the other. We therefore report standard deviation in all tables, while discussing the results in terms of variance throughout the paper. Transaction costs are excluded for simplicity. These metrics are defined as follows:

$$\begin{aligned}
\hat{\mu}_{\text{oos}} &= \frac{1}{T_2} \sum_{t=T_1}^{T-1} \hat{\mathbf{w}}_t' \mathbf{r}_{t+1}, \\
\hat{\sigma}_{\text{oos}} &= \left( \frac{1}{T_2 - 1} \sum_{t=T_1}^{T-1} (\hat{\mathbf{w}}_t' \mathbf{r}_{t+1} - \hat{\mu}_{\text{oos}})^2 \right)^{\frac{1}{2}}, \\
\hat{r}_{p,\min} &= \min_{T_1 \leq t \leq T-1} (\hat{\mathbf{w}}_t' \mathbf{r}_{t+1}), \\
\hat{r}_{p,\max} &= \max_{T_1 \leq t \leq T-1} (\hat{\mathbf{w}}_t' \mathbf{r}_{t+1}), \\
\text{SR} &= \frac{\hat{\mu}_{\text{oos}}}{\hat{\sigma}_{\text{oos}}}.
\end{aligned} \tag{12}$$

In Tables 1 to 3, these metrics— $\hat{\mu}_{\text{oos}}$ ,  $\hat{\sigma}_{\text{oos}}$ ,  $\hat{r}_{p,\min}$ ,  $\hat{r}_{p,\max}$ , and SR—are referred to as Return, Risk, Min, Max, and SR, respectively.

## 6.2 Empirical Results

Using monthly data, we compare the performance of the portfolios constructed from Sections 3 to 5 against their respective basis portfolios. In particular, by comparing the results of the portfolios in Sections 3 and 4 with that of the non-decomposition portfolio in Section 5, we highlight the importance of employing idiosyncratic returns via the return decomposition to enhance index fund performance.

Tables 1 to 3 report empirical results when the S&P 500 index, the EW portfolio, and the value-weighted portfolio of CRSP firms serve as the basis portfolios, respectively. Figures A1

to A3 illustrate the monthly return trajectories of the portfolios relative to each of the three basis portfolios. Across all tables and figures, portfolios are labeled according to their estimation method: ALasso, SCAD, and  $L_2$ -Boosting from Section 3; GLasso (with return decomposition) from Section 4; and GLasso (without return decomposition) from Section 5. Our main empirical findings are summarized below:

1. Table 1 shows that when the S&P 500 index fund is used as the basis portfolio, all portfolios employing the return decomposition from Sections 3 and 4—estimated via ALasso, SCAD,  $L_2$ -Boosting, and GLasso—consistently reduce portfolio variance relative to the basis portfolio, as evidenced by lower risk across all sample splits. Because the index fund return is retained as the basis and idiosyncratic returns provide incremental adjustments, the average OOS portfolio returns from Sections 3 and 4 are comparable to, or in some cases higher than, those of the S&P 500 index fund, ultimately resulting in higher SRs.
2. Figure A1 further illustrates that the range of OOS returns is narrower for all decomposition-based portfolios (Sections 3 and 4) compared to the S&P 500 index fund. A notable empirical pattern emerges for the ALasso-estimated portfolio from Section 3: its returns closely mirror those of the index fund—meaning the incremental component  $\sum_{i=1}^N \hat{b}_i d_{i,t}$  stays near zero—during the early sub-sample period. One possible explanation is that the idiosyncratic returns  $d_{i,t} = r_{i,t} - r_{0,t}$  are “weak signals” in the early sub-sample, which may cause ALasso to produce unstable coefficient estimates  $\hat{b}_i$  that fail to shrink to zero when they should.<sup>4</sup> This instability, in turn, may cause the portfolio return  $r_{p,t} = r_{0,t} + \sum_{i=1}^N \hat{b}_i d_{i,t}$  from Equation (6) to remain close to zero. As the sample period progresses through the financial crisis, however, the idiosyncratic returns become sufficiently strong signals for ALasso to produce stable coefficient estimates, allowing the portfolio return to deviate from zero.
3. The non-decomposition portfolio discussed in Section 5, which treats the index fund

---

<sup>4</sup>It is well known that Lasso-type regularization works poorly in the presence of weak signals.

as an additional asset alongside individual stocks and is estimated using GLasso, also achieves a variance reduction relative to the S&P 500 index fund. However, it yields the average OOS returns substantially lower than those of the S&P 500 index fund, resulting in a deterioration of OOS SRs. These results underscore the importance of return decomposition and the use of idiosyncratic returns, particularly in preserving the return levels of the basis index.

4. Two factors likely contribute to this deterioration in returns without the return decomposition. First, although the GLasso framework assumes sparsity in the precision matrix of the  $(N + 1)$  asset returns, this assumption is unlikely to hold due to the exposure of the  $(N + 1)$  asset returns to common (fundamental, macroeconomic) factors that affect all assets simultaneously. Second, without the return decomposition, GLasso freely reallocates weights across all  $(N + 1)$  assets based solely on the estimated precision matrix, leading to unconstrained departures from the S&P 500 index fund's target profile.
5. The above advantages of using the return decomposition to improve OOS SRs are not unique to the S&P 500 index fund, but hold robustly across different choices of the basis portfolio. As shown in Tables 2 and 3, the portfolios constructed based on the return decomposition (from Sections 3 and 4) reduce the average OOS variance across all sample splits when the basis portfolio is the EW portfolio or the value-weighted portfolio of CRSP firms. Meanwhile, their average OOS returns do not deviate much from (and in some cases exceed) those of the basis portfolio, thereby improving OOS SRs. Figures A2 and A3 further illustrate that the range of OOS returns is narrower for all portfolios from Sections 3 and 4 relative to their respective basis portfolios. Although the portfolio from Section 5 also reduces average OOS variance, the substantial decrease in mean return leads to a lower SR. This highlights the importance of using return decomposition and idiosyncratic returns to improve the performance of the basis portfolio.

## 7 Conclusion

This paper proposes a portfolio strategy that minimizes variance relative to an index fund by exploiting the return decomposition ( $r_{i,t} = r_{0,t} + d_{i,t}$ ) and the resulting idiosyncratic returns ( $d_{i,t}$ ). By exploring the covariance between the index fund return and idiosyncratic returns ( $\text{Cov}(r_{0,t}, d_{i,t})$  in Section 3) and the covariance among the idiosyncratic returns themselves ( $\text{Cov}(d_{i,t}, d_{j,t})$  in Section 4), we construct portfolios that consistently achieve lower variance than three basis portfolios—the S&P 500 index fund, the Equal-weighted S&P 500 portfolio, and the value-weighted portfolio of CRSP firms—without sacrificing OOS portfolio returns, thereby improving overall OOS SRs.

When the index fund is included directly as an additional asset without applying the return decomposition (Section 5), the resulting non-decomposition portfolio reduces variance, but also substantially reduces the OOS portfolio returns, leading to a deterioration in the SR. This confirms that the primary benefit of our proposed approach stems specifically from exploiting idiosyncratic return structures through this decomposition. These findings suggest that amid growing market concentration in major indices, investors can enhance risk-adjusted portfolio performance by incorporating idiosyncratic return information from the return decomposition, rather than relying solely on traditional index funds or simple reweighting schemes.

## **Disclosure Statement**

The authors report there are no competing interests to declare.

## **Declaration of Generative AI Use**

The authors report generative AI was not used in their research or preparation of this manuscript.

## **Funding Details**

No funding was received for this work.

## **Data Availability Statement**

The data that support the findings of this study are available from the corresponding author, S. Lee, upon reasonable request.

## References

- Awoye, O.A., Markowitz Minimum Variance Portfolio Optimization Using New Machine Learning Methods. PhD thesis, University College London, 2016.
- Brown, Z.Y., Egan, M.L., Jeon, J., Jin, C. and Wu, A.A., Why Do Index Funds Have Market Power? Quantifying Frictions in the Index Fund Market. National Bureau of Economic Research No. w31778, 2023.
- Bühlmann, P., Boosting for high-dimensional linear models. *Annals of Statistics*, 2006, **34**, 559–583.
- Cai, T.T., Hu, J., Li, Y. and Zheng, X., High-dimensional minimum variance portfolio estimation based on high-frequency data. *Journal of Econometrics*, 2020, **214**, 482–494.
- Crane, A.D. and Crotty, K., Passive versus Active Fund Performance: Do Index Funds Have Skill?. *Journal of Financial and Quantitative Analysis*, 2018, **53**, 33–64.
- DeMiguel, V., Garlappi, L. and Uppal, R., Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?. *The Review of Financial Studies*, 2009, **22**, 1915–1953.
- Fama, E.F. and French, K.R., The cross-section of expected stock returns. *The Journal of Finance*, 1992, **47**, 427–465.
- Fama, E.F. and French, K.R., A five-factor asset pricing model. *Journal of Financial Economics*, 2015, **116**, 1–22.
- Fan, J. and Li, R., Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 2001, **96**, 1348–1360.
- Fortin, R. and Michelson, S., Indexing versus active mutual fund management. *Journal of Financial Planning*, 2002, **15**, 82–95.
- Gruber, M.J., Another puzzle: The growth in actively managed mutual funds. *The Journal of Finance*, 1996, **51**, 783–810.

- Jones, R.C. and Wermers, R., Active management in mostly efficient markets. *Financial Analysts Journal*, 2011, **67**, 29–45.
- Kosowski, R., Do mutual funds perform when it matters most to investors? US mutual fund performance and risk in recessions and expansions. *The Quarterly Journal of Finance*, 2011, **1**, 607–664.
- Lee, T.H. and Seregina, E., Optimal portfolio using factor graphical lasso. *Journal of Financial Econometrics*, 2024, **22**, 670–695.
- López de Prado, M., Building diversified portfolios that outperform out-of-sample. *The Journal of Portfolio Management*, 2016, **42**, 59–69.
- López de Prado, M., Simonian, J., Fabozzi, F.A. and Fabozzi, F.J., Enhancing Markowitz’s portfolio selection paradigm with machine learning. *Annals of Operations Research*, 2025, **346**, 319–340.
- Malladi, R. and Fabozzi, F.J., Equal-weighted strategy: Why it outperforms value-weighted strategies? Theory and evidence. *Journal of Asset Management*, 2017, **18**, 188–208.
- Markowitz, H., Modern portfolio theory. *Journal of Finance*, 1952, **7**, 77–91.
- Moskowitz, T.J., Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses: Discussion. *The Journal of Finance*, 2000, **55**, 1695–1703.
- RBC Wealth Management, The ‘Great Narrowing’: S&P 500 Concentration. Available online at: <https://www.rbcwealthmanagement.com/en-us/insights/the-great-narrowing-sp-500-concentration> (accessed 20 March 2026), 2026.
- Risik, E.A., Actively managed mutual fund returns versus the S&P 500 index. *Business and Economics Journal*, 2013, **4**, 1–11.
- Sharpe, W.F., A simplified model for portfolio analysis. *Management Science*, 1963, **9**, 277–293.

- Swade, A., Nolte, S., Shackleton, M. and Lohre, H., Why do equally weighted portfolios beat value-weighted ones?. *Journal of Portfolio Management*, 2023, **49**, 167–187.
- Taljaard, B.H. and Maré, E., Why has the equal weight portfolio underperformed and what can we do about it?. *Quantitative Finance*, 2021, **21**, 1855–1868.
- Zou, H., The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 2006, **101**, 1418–1429.

**Table 1:** Portfolio Performance: S&P 500 as Basis Portfolio

| Out-of-sample period         | 2010M1–2019M12 (10 years) |        |         |        |        | 2005M1–2019M12 (15 years) |        |         |        |        | 2000M1–2019M12 (20 years) |        |         |        |        |
|------------------------------|---------------------------|--------|---------|--------|--------|---------------------------|--------|---------|--------|--------|---------------------------|--------|---------|--------|--------|
|                              | Return                    | Risk   | Min     | Max    | SR     | Return                    | Risk   | Min     | Max    | SR     | Return                    | Risk   | Min     | Max    | SR     |
| S&P 500                      | 0.0086                    | 0.0394 | -0.1159 | 0.1028 | 0.2174 | 0.0052                    | 0.0407 | -0.1836 | 0.1461 | 0.1118 | 0.0032                    | 0.0466 | -0.1836 | 0.1461 | 0.0692 |
| ALasso                       | 0.0099                    | 0.0325 | -0.1202 | 0.1033 | 0.3047 | 0.0054                    | 0.0287 | -0.1006 | 0.2175 | 0.1894 | 0.0023                    | 0.0244 | -0.1280 | 0.1369 | 0.0945 |
| SCAD                         | 0.0065                    | 0.0274 | -0.0943 | 0.0747 | 0.2372 | 0.0042                    | 0.0375 | -0.1602 | 0.1243 | 0.1110 | 0.0016                    | 0.0406 | -0.1470 | 0.1181 | 0.0399 |
| $L_2$ -Boosting              | 0.0049                    | 0.0284 | -0.1092 | 0.0920 | 0.1740 | 0.0033                    | 0.0312 | -0.1114 | 0.0920 | 0.1046 | 0.0021                    | 0.0303 | -0.1114 | 0.0920 | 0.0681 |
| GLasso                       | 0.0100                    | 0.0290 | -0.0844 | 0.0781 | 0.3464 | 0.0083                    | 0.0301 | -0.1016 | 0.0791 | 0.2761 | 0.0065                    | 0.0321 | -0.1489 | 0.0785 | 0.2018 |
| Without Return Decomposition | 0.0003                    | 0.0084 | -0.0244 | 0.0214 | 0.0317 | 0.0008                    | 0.0094 | -0.0251 | 0.0414 | 0.0865 | 0.0030                    | 0.0127 | -0.0483 | 0.0659 | 0.2348 |

<sup>1</sup> Table 1 reports the out-of-sample (OOS) performance of each portfolio across three evaluation periods. “Return,” “Risk,” “Min,” “Max,” and “SR” denote the OOS average return, standard deviation, minimum return, maximum return, and Sharpe ratio, respectively.

<sup>2</sup> “S&P 500” denotes the S&P 500 index fund used as the basis portfolio. The remaining portfolios are constructed as follows. Portfolios estimated using ALasso, SCAD, and  $L_2$ -Boosting (Section 3) explore the covariance between the index fund return and idiosyncratic returns via penalized regression. The portfolio estimated using GLasso with the return decomposition (Section 4) explores the covariance among idiosyncratic returns via the Graphical Lasso. The portfolio estimated using GLasso without the return decomposition (Section 5) treats the index fund as a separate asset in an  $(N + 1)$ -asset framework.

**Table 2:** Portfolio Performance: EW as Basis Portfolio

| Out-of-sample period            | 2010M1–2019M12 (10 years) |        |         |         |        | 2005M1–2019M12 (15 years) |        |         |         |        | 2000M1–2019M12 (20 years) |        |         |         |        |        |
|---------------------------------|---------------------------|--------|---------|---------|--------|---------------------------|--------|---------|---------|--------|---------------------------|--------|---------|---------|--------|--------|
|                                 | Return                    | Risk   | Min     | Max     | SR     | Return                    | Risk   | Min     | Max     | SR     | Return                    | Risk   | Min     | Max     | SR     |        |
| EW                              | 0.0082                    | 0.0472 | -0.1210 | 0.1220  | 0.1729 | 0.0052                    | 0.0599 | -0.2597 | 0.2035  | 0.0864 | 0.0057                    | 0.0564 | -0.2597 | 0.2035  | 0.1005 |        |
| With Return<br>Decomposition    | ALasso                    | 0.0075 | 0.0328  | -0.1374 | 0.1124 | 0.2284                    | 0.0036 | 0.0272  | -0.1314 | 0.1533 | 0.1317                    | 0.0020 | 0.0264  | -0.1207 | 0.1164 | 0.0762 |
|                                 | SCAD                      | 0.0068 | 0.0359  | -0.0941 | 0.0873 | 0.1893                    | 0.0042 | 0.0505  | -0.2243 | 0.1831 | 0.0830                    | 0.0045 | 0.0496  | -0.2064 | 0.1637 | 0.0914 |
|                                 | $L_2$ -Boosting           | 0.0046 | 0.0321  | -0.0836 | 0.0791 | 0.1437                    | 0.0032 | 0.0383  | -0.1575 | 0.1235 | 0.0826                    | 0.0035 | 0.0357  | -0.1575 | 0.1235 | 0.0982 |
|                                 | GLasso                    | 0.0102 | 0.0289  | -0.0754 | 0.0785 | 0.3425                    | 0.0084 | 0.0301  | -0.1016 | 0.0755 | 0.2788                    | 0.0062 | 0.0318  | -0.1434 | 0.0753 | 0.1946 |
| Without Return<br>Decomposition | GLasso                    | 0.0003 | 0.0165  | -0.0575 | 0.0541 | 0.0199                    | 0.0006 | 0.0153  | -0.0581 | 0.0571 | 0.0381                    | 0.0008 | 0.0135  | -0.0588 | 0.0585 | 0.0580 |

<sup>1</sup> Table 2 reports the out-of-sample (OOS) performance of each portfolio across three evaluation periods. “Return,” “Risk,” “Min,” “Max,” and “SR” denote the OOS average return, standard deviation, minimum return, maximum return, and Sharpe ratio, respectively.

<sup>2</sup> “EW” denotes the equally weighted S&P 500 portfolio used as the basis portfolio. The remaining portfolios are constructed as follows. Portfolios estimated using ALasso, SCAD, and  $L_2$ -Boosting (Section 3) explore the covariance between the index fund return and idiosyncratic returns via penalized regression. The portfolio estimated using GLasso with the return decomposition (Section 4) explores the covariance among idiosyncratic returns via the Graphical Lasso. The portfolio estimated using GLasso without the return decomposition (Section 5) treats the index fund as a separate asset in an  $(N + 1)$ -asset framework.

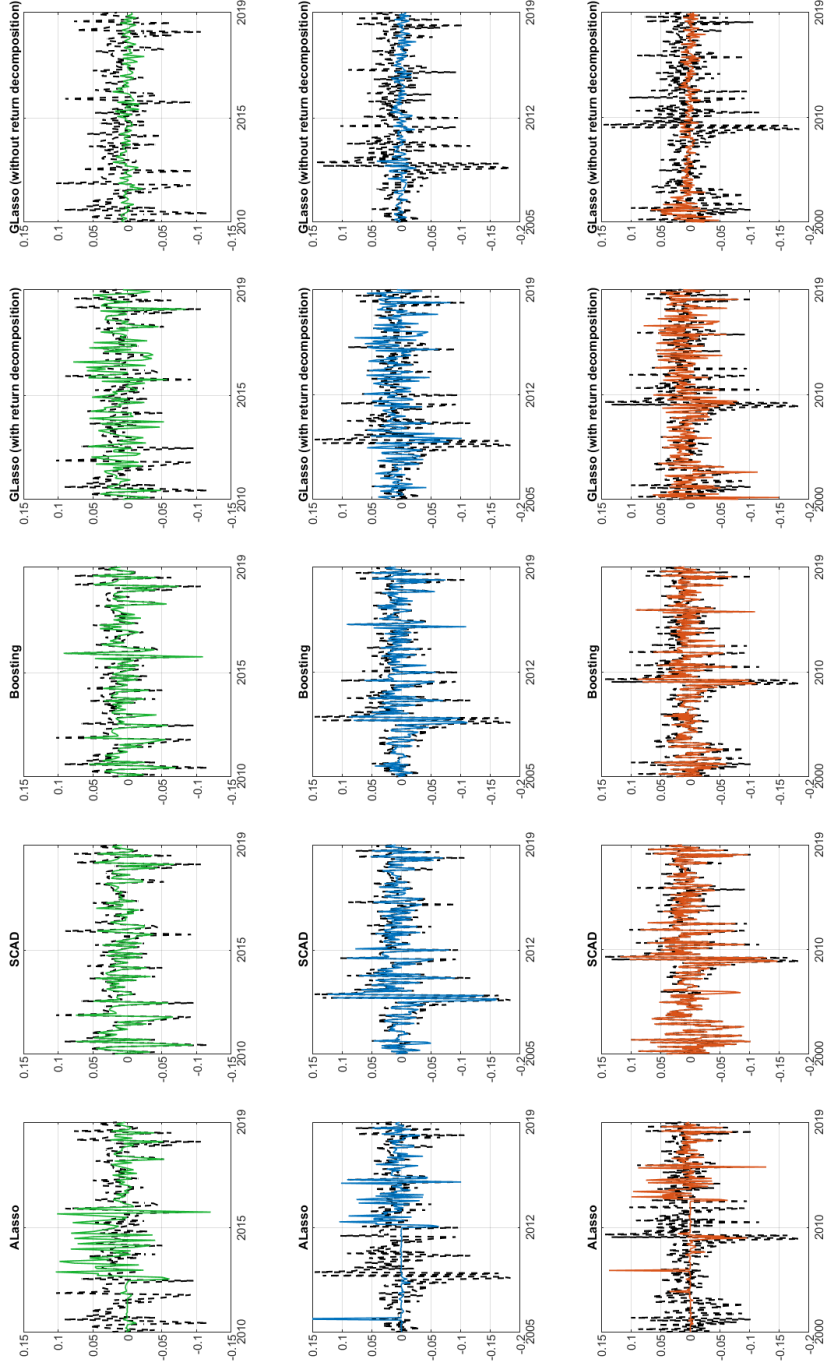
**Table 3:** Portfolio Performance: CRSP as Basis Portfolio

| Out-of-sample period         | 2010M1–2019M12 (10 years) |         |         |         |        | 2005M1–2019M12 (15 years) |         |         |         |        | 2000M1–2019M12 (20 years) |         |         |         |        |
|------------------------------|---------------------------|---------|---------|---------|--------|---------------------------|---------|---------|---------|--------|---------------------------|---------|---------|---------|--------|
|                              | Return                    | Risk    | Min     | Max     | SR     | Return                    | Risk    | Min     | Max     | SR     | Return                    | Risk    | Min     | Max     | SR     |
| CRSP                         | 0.0107                    | 0.0390  | -0.0818 | 0.1078  | 0.2741 | 0.0083                    | 0.0616  | -0.3040 | 0.2072  | 0.1346 | 0.0099                    | 0.1162  | -0.6399 | 0.6076  | 0.0853 |
| ALasso                       | 0.0045                    | 0.0279  | -0.0975 | 0.1022  | 0.1604 | 0.0055                    | 0.0268  | -0.1314 | 0.1400  | 0.2036 | 0.0038                    | 0.0197  | -0.0908 | 0.1072  | 0.1915 |
| SCAD                         | 0.0078                    | 0.0266  | -0.0626 | 0.0781  | 0.2949 | 0.0053                    | 0.0305  | -0.0948 | 0.0849  | 0.1749 | 0.0025                    | 0.0391  | -0.1463 | 0.0896  | 0.0636 |
| $L_2$ -Boosting              | 0.0064                    | 0.0255  | -0.0714 | 0.0688  | 0.2518 | 0.0045                    | 0.0267  | -0.1045 | 0.0688  | 0.1670 | 0.0030                    | 0.0277  | -0.1045 | 0.0688  | 0.1082 |
| GLasso                       | 0.0100                    | 0.0288  | -0.0841 | 0.0716  | 0.3475 | 0.0077                    | 0.0292  | -0.1004 | 0.0662  | 0.2650 | 0.0064                    | 0.0307  | -0.1216 | 0.0661  | 0.2070 |
| Without Return Decomposition | GLasso                    | -0.0028 | 0.0301  | -0.0895 | 0.0670 | -0.0924                   | -0.0016 | 0.0251  | -0.0891 | 0.0689 | -0.00647                  | -0.0003 | 0.0229  | -0.0971 | 0.0646 |

<sup>1</sup> Table 3 reports the out-of-sample (OOS) performance of each portfolio across three evaluation periods. “Return,” “Risk,” “Min,” “Max,” and “SR” denote the OOS average return, standard deviation, minimum return, maximum return, and Sharpe ratio, respectively.

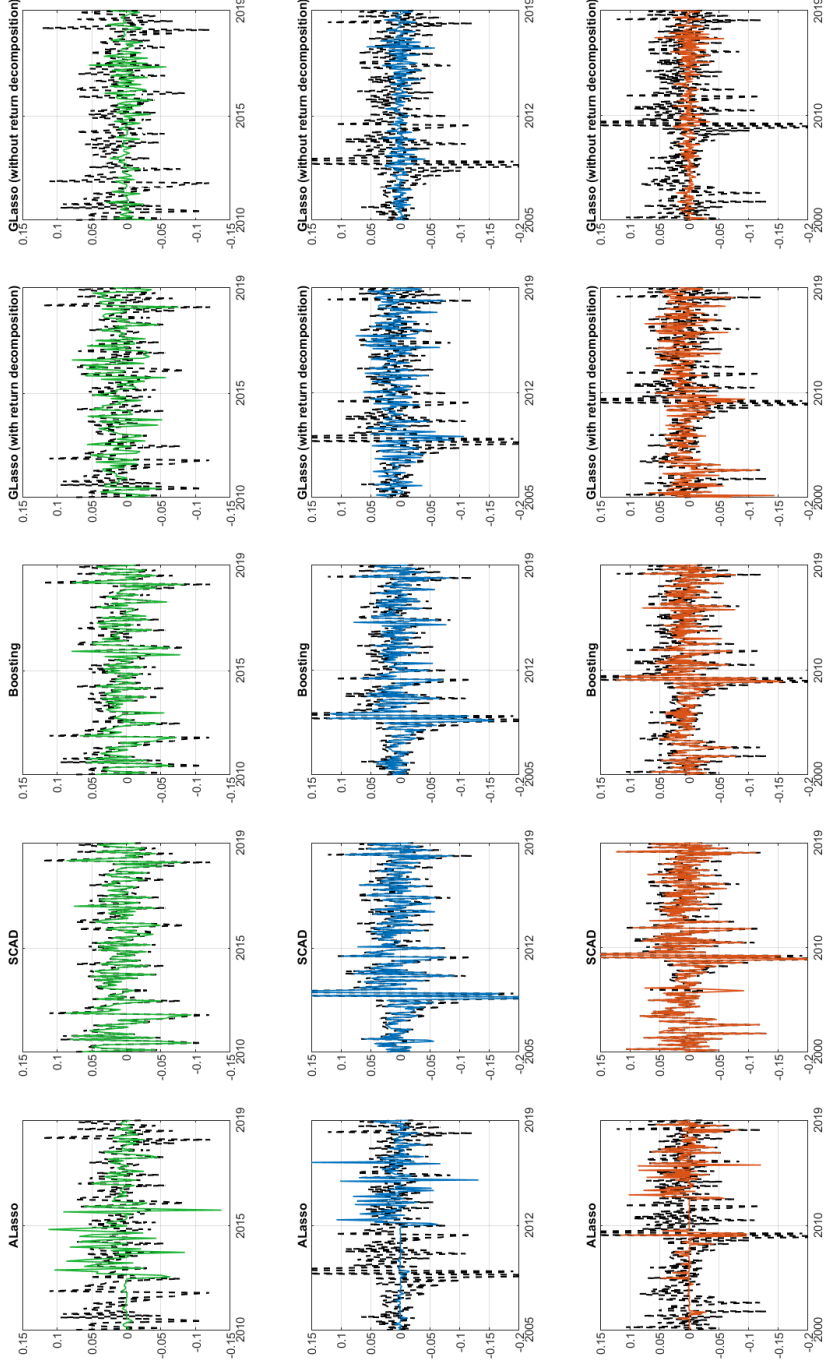
<sup>2</sup> “CRSP” denotes the value-weighted portfolio of CRSP firms used as the basis portfolio. The remaining portfolios are constructed as follows. Portfolios estimated using ALasso, SCAD, and  $L_2$ -Boosting (Section 3) explore the covariance between the index fund return and idiosyncratic returns via penalized regression. The portfolio estimated using GLasso with the return decomposition (Section 4) explores the covariance among idiosyncratic returns via the Graphical Lasso. The portfolio estimated using GLasso without the return decomposition (Section 5) treats the index fund as a separate asset in an  $(N + 1)$ -asset framework.

Figure 1: Monthly Returns: S&P 500 as Basis Portfolio



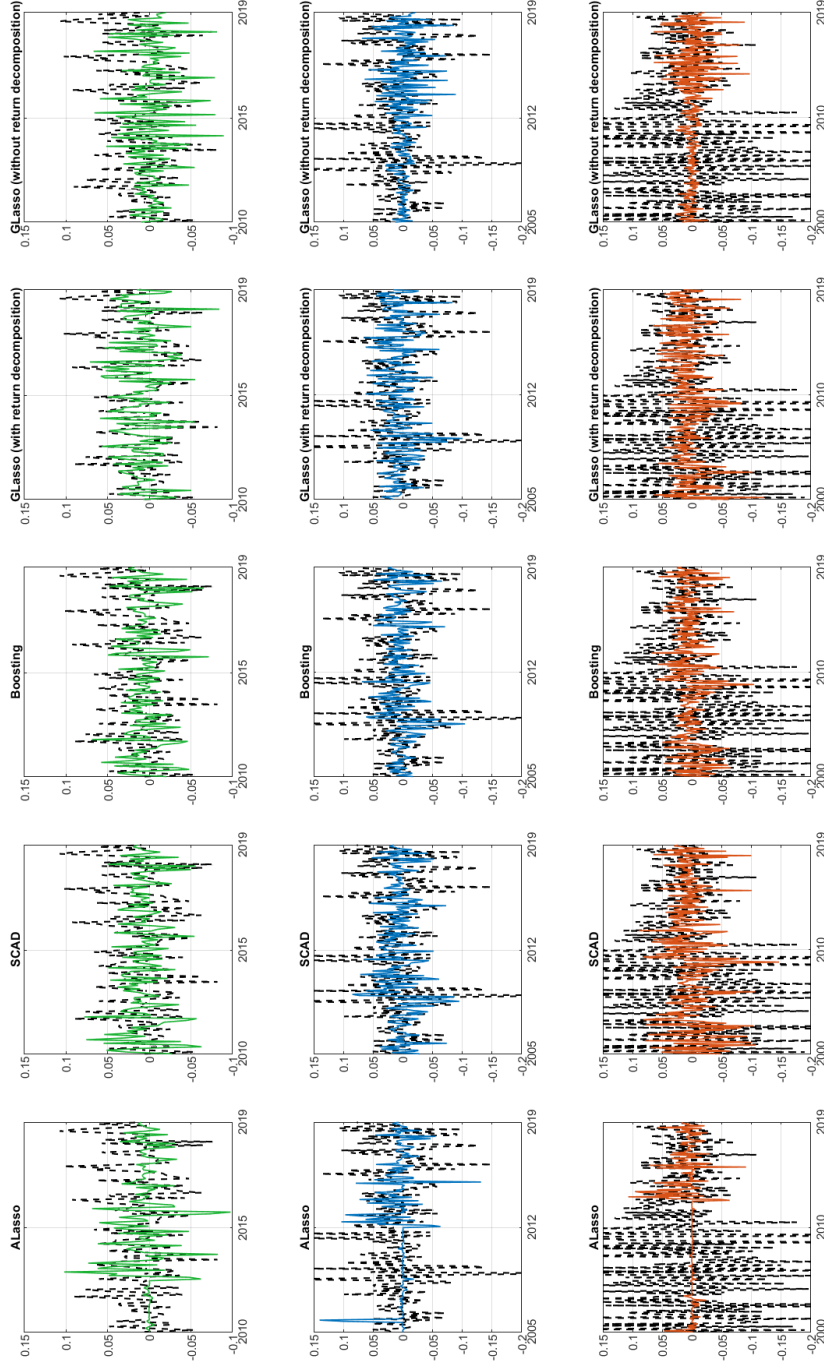
**Notes:** Each panel plots the monthly OOS returns of the portfolios discussed in Sections 3 to 5 against those of the S&P 500 index fund (black dashed line). Solid colored lines correspond to different OOS evaluation periods: green for the 10-year period, blue for the 15-year period, and orange for the 20-year period. See Table 1 for a description of each portfolio and the performance metrics.

Figure 2: Monthly Returns: EW as Basis Portfolio



**Notes:** Each panel plots the monthly OOS returns of the portfolios discussed in Sections 3 to 5 against those of the equally weighted S&P 500 portfolio (black dashed line). Solid colored lines correspond to different OOS evaluation periods: green for the 10-year period, blue for the 15-year period, and orange for the 20-year period. See Table 2 for a description of each portfolio and the performance metrics.

Figure 3: Monthly Returns: CRSP as Basis Portfolio



**Notes:** Each panel plots the monthly OOS returns of the portfolios discussed in Sections 3 to 5 against those of the value-weighted portfolio of CRSP firms (black dashed line). Solid colored lines correspond to different OOS evaluation periods: green for the 10-year period, blue for the 15-year period, and orange for the 20-year period. See Table 3 for a description of each portfolio and the performance metrics.

## Appendix A Additional Empirical Results

This Supplementary Appendix provides additional empirical results complementing the analysis in Section 6.

First, to assess the robustness of the expanding five-fold cross-validation procedure used to select the optimal regularization parameters for ALasso and SCAD, we consider an alternative initial training window covering 30% of the in-sample observations (compared to 40% in the paper). Note that the results for portfolios estimated using  $L_2$ -boosting and GLasso (with and without the return decomposition) are identical to those in the paper, as these methods do not rely on cross-validation for tuning parameter selection. Tables A1 to A3 report the results for the S&P 500 index fund, the EW portfolio, and the value-weighted portfolio of CRSP firms, respectively.

While the magnitude of variance reduction changes slightly compared to the results reported in the paper, the portfolios introduced in Section 3—estimated using ALasso and SCAD with the alternative initial training window—consistently reduce variance relative to their respective basis portfolios, as reflected in the lower portfolio risk. The sole exception is ALasso during the 10-year OOS period when the S&P 500 index fund serves as the basis portfolio.

Next, we examine two additional OOS evaluation periods beyond those considered in Section 6: (i) a 35-year in-sample period ( $T_1 = 420$ ) with a 5-year OOS period ( $T_2 = 60$ ), and (ii) a 15-year in-sample period ( $T_1 = 180$ ) with a 25-year OOS period ( $T_2 = 300$ ). As in the paper, portfolios are rebalanced monthly using a rolling-window approach. Tables A4 to A6 report the results for different sample splits across the three basis portfolios. Figures A1 to A3 depict the monthly returns of each portfolio relative to its basis portfolio.

Consistent with the findings in the paper, all portfolios from Sections 3 and 4—which exploit idiosyncratic returns via the return decomposition—reduce portfolio variance relative to their respective basis portfolios across various sample splits. The returns of these portfolios do not deviate significantly from those of the basis portfolios (and in some cases exceed them), thereby generally improving SRs. Conversely, the portfolio from Section 5, estimated using GLasso without the return decomposition, also achieves variance reduction; however, as noted

in the paper, it yields returns substantially lower than those of the basis portfolios, occasionally turning negative, which consequently lowers its SR.

This divergence likely reflects two compounding factors discussed in the paper: first, the GLasso framework assumes sparsity on the precision matrix of the  $(N + 1)$  asset returns but cannot guarantee sparsity on the returns themselves, given that all assets share common exposure to macroeconomic conditions; second, the absence of the return decomposition allows the GLasso estimator to freely reallocate weights across all  $(N + 1)$  assets based solely on the estimated precision matrix, leading to greater deviations from the basis portfolios.

**Table A1:** Portfolio Performance: S&P 500 as Basis Portfolio

| Out-of-sample period         | 2010M1–2019M12 (10 years) |        |         |        |        | 2005M1–2019M12 (15 years) |        |         |        |        | 2000M1–2019M12 (20 years) |        |         |        |        |
|------------------------------|---------------------------|--------|---------|--------|--------|---------------------------|--------|---------|--------|--------|---------------------------|--------|---------|--------|--------|
|                              | Return                    | Risk   | Min     | Max    | SR     | Return                    | Risk   | Min     | Max    | SR     | Return                    | Risk   | Min     | Max    | SR     |
| S&P 500                      | 0.0086                    | 0.0394 | -0.1159 | 0.1028 | 0.2174 | 0.0052                    | 0.0407 | -0.1836 | 0.1461 | 0.1118 | 0.0032                    | 0.0466 | -0.1836 | 0.1461 | 0.0692 |
| ALasso                       | 0.0105                    | 0.0585 | -0.1235 | 0.5530 | 0.1801 | 0.0075                    | 0.0281 | -0.1072 | 0.1330 | 0.2663 | 0.0048                    | 0.0325 | -0.1280 | 0.2758 | 0.1465 |
| SCAD                         | 0.0059                    | 0.0296 | -0.1044 | 0.0819 | 0.2002 | 0.0046                    | 0.0311 | -0.0963 | 0.0770 | 0.1480 | 0.0014                    | 0.0408 | -0.1660 | 0.1234 | 0.0351 |
| $L_2$ -Boosting              | 0.0049                    | 0.0284 | -0.1092 | 0.0920 | 0.1740 | 0.0033                    | 0.0312 | -0.1114 | 0.0920 | 0.1046 | 0.0021                    | 0.0303 | -0.1114 | 0.0920 | 0.0681 |
| GLasso                       | 0.0100                    | 0.0290 | -0.0844 | 0.0781 | 0.3464 | 0.0083                    | 0.0301 | -0.1016 | 0.0791 | 0.2761 | 0.0065                    | 0.0321 | -0.1489 | 0.0785 | 0.2018 |
| Without Return Decomposition | 0.0003                    | 0.0084 | -0.0244 | 0.0214 | 0.0317 | 0.0008                    | 0.0094 | -0.0251 | 0.0414 | 0.0865 | 0.0030                    | 0.0127 | -0.0483 | 0.0659 | 0.2348 |

<sup>1</sup> Table A1 reports the out-of-sample (OOS) performance of each portfolio across three evaluation periods. “Return,” “Risk,” “Min,” “Max,” and “SR” denote the OOS average return, standard deviation, minimum return, maximum return, and Sharpe ratio, respectively.

<sup>2</sup> “S&P 500” denotes the S&P 500 index fund used as the basis portfolio. The remaining portfolios are constructed as follows. Portfolios estimated using ALasso, SCAD, and  $L_2$ -Boosting (Section 3) explore the covariance between the index fund return and idiosyncratic returns via penalized regression. The portfolio estimated using GLasso with the return decomposition (Section 4) explores the covariance among idiosyncratic returns via the Graphical Lasso. The portfolio estimated using GLasso without the return decomposition (Section 5) treats the index fund as a separate asset in an  $(N + 1)$ -asset framework.

**Table A2:** Portfolio Performance: EW as Basis Portfolio

| Out-of-sample period         | 2010M1–2019M12 (10 years) |        |         |         |        | 2005M1–2019M12 (15 years) |        |         |         |        | 2000M1–2019M12 (20 years) |        |         |        |        |
|------------------------------|---------------------------|--------|---------|---------|--------|---------------------------|--------|---------|---------|--------|---------------------------|--------|---------|--------|--------|
|                              | Return                    | Risk   | Min     | Max     | SR     | Return                    | Risk   | Min     | Max     | SR     | Return                    | Risk   | Min     | Max    | SR     |
| EW                           | 0.0082                    | 0.0472 | -0.1210 | 0.1220  | 0.1729 | 0.0052                    | 0.0599 | -0.2597 | 0.2035  | 0.0864 | 0.0057                    | 0.0564 | -0.2597 | 0.2035 | 0.1005 |
| With Return Decomposition    | ALasso                    | 0.0045 | 0.0312  | -0.1222 | 0.0969 | 0.1447                    | 0.0064 | 0.0469  | -0.1358 | 0.5233 | 0.0063                    | 0.0296 | -0.1216 | 0.1636 | 0.2129 |
|                              | SCAD                      | 0.0052 | 0.0378  | -0.1023 | 0.0954 | 0.1377                    | 0.0050 | 0.0421  | -0.1368 | 0.1728 | 0.0043                    | 0.0499 | -0.2341 | 0.1770 | 0.0868 |
|                              | $L_2$ -Boosting           | 0.0046 | 0.0321  | -0.0836 | 0.0791 | 0.1437                    | 0.0032 | 0.0383  | -0.1575 | 0.1235 | 0.0035                    | 0.0357 | -0.1575 | 0.1235 | 0.0982 |
|                              | GLasso                    | 0.0102 | 0.0289  | -0.0754 | 0.0785 | 0.3425                    | 0.0084 | 0.0301  | -0.1016 | 0.0755 | 0.0062                    | 0.0318 | -0.1434 | 0.0753 | 0.1946 |
| Without Return Decomposition | GLasso                    | 0.0003 | 0.0165  | -0.0575 | 0.0541 | 0.0199                    | 0.0006 | 0.0153  | -0.0581 | 0.0571 | 0.0008                    | 0.0135 | -0.0588 | 0.0585 | 0.0580 |

<sup>1</sup> Table A2 reports the out-of-sample (OOS) performance of each portfolio across three evaluation periods. “Return,” “Risk,” “Min,” “Max,” and “SR” denote the OOS average return, standard deviation, minimum return, maximum return, and Sharpe ratio, respectively.

<sup>2</sup> “EW” denotes the equally weighted S&P 500 portfolio used as the basis portfolio. The remaining portfolios are constructed as follows. Portfolios estimated using ALasso, SCAD, and  $L_2$ -Boosting (Section 3) explore the covariance between the index fund return and idiosyncratic returns via penalized regression. The portfolio estimated using GLasso with the return decomposition (Section 4) explores the covariance among idiosyncratic returns via the Graphical Lasso. The portfolio estimated using GLasso without the return decomposition (Section 5) treats the index fund as a separate asset in an  $(N + 1)$ -asset framework.

**Table A3:** Portfolio Performance: CRSP as Basis Portfolio

| Out-of-sample period         | 2010M1–2019M12 (10 years) |         |         |         |        | 2005M1–2019M12 (15 years) |         |         |         |        | 2000M1–2019M12 (20 years) |         |         |         |        |
|------------------------------|---------------------------|---------|---------|---------|--------|---------------------------|---------|---------|---------|--------|---------------------------|---------|---------|---------|--------|
|                              | Return                    | Risk    | Min     | Max     | SR     | Return                    | Risk    | Min     | Max     | SR     | Return                    | Risk    | Min     | Max     | SR     |
| CRSP                         | 0.0107                    | 0.0390  | -0.0818 | 0.1078  | 0.2741 | 0.0083                    | 0.0616  | -0.3040 | 0.2072  | 0.1346 | 0.0099                    | 0.1162  | -0.6399 | 0.6076  | 0.0853 |
| ALasso                       | 0.0059                    | 0.0278  | -0.0975 | 0.1089  | 0.2124 | 0.0039                    | 0.0221  | -0.1225 | 0.1034  | 0.1755 | 0.0035                    | 0.0245  | -0.1350 | 0.1403  | 0.1413 |
| SCAD                         | 0.0073                    | 0.0287  | -0.0741 | 0.0876  | 0.2538 | 0.0058                    | 0.0282  | -0.0863 | 0.0785  | 0.2057 | 0.0028                    | 0.0376  | -0.1072 | 0.1032  | 0.0743 |
| $L_2$ -Boosting              | 0.0064                    | 0.0255  | -0.0714 | 0.0688  | 0.2518 | 0.0045                    | 0.0267  | -0.1045 | 0.0688  | 0.1670 | 0.0030                    | 0.0277  | -0.1045 | 0.0688  | 0.1082 |
| GLasso                       | 0.0100                    | 0.0288  | -0.0841 | 0.0716  | 0.3475 | 0.0077                    | 0.0292  | -0.1004 | 0.0662  | 0.2650 | 0.0064                    | 0.0307  | -0.1216 | 0.0661  | 0.2070 |
| Without Return Decomposition | GLasso                    | -0.0028 | 0.0301  | -0.0895 | 0.0670 | -0.0924                   | -0.0016 | 0.0251  | -0.0891 | 0.0689 | -0.00647                  | -0.0003 | 0.0229  | -0.0971 | 0.0646 |

<sup>1</sup> Table A3 reports the out-of-sample (OOS) performance of each portfolio across three evaluation periods. “Return,” “Risk,” “Min,” “Max,” and “SR” denote the OOS average return, standard deviation, minimum return, maximum return, and Sharpe ratio, respectively.

<sup>2</sup> “CRSP” denotes the value-weighted portfolio of CRSP firms used as the basis portfolio. The remaining portfolios are constructed as follows. Portfolios estimated using ALasso, SCAD, and  $L_2$ -Boosting (Section 3) explore the covariance between the index fund return and idiosyncratic returns via penalized regression. The portfolio estimated using GLasso with the return decomposition (Section 4) explores the covariance among idiosyncratic returns via the Graphical Lasso. The portfolio estimated using GLasso without the return decomposition (Section 5) treats the index fund as a separate asset in an  $(N + 1)$ -asset framework.

**Table A4:** Portfolio Performance: S&P 500 as Basis Portfolio

| Out-of-sample period         | 2015M1–2019M12 (5 years) |         |         |         |        | 1995M1–2019M12 (25 years) |        |         |         |        |        |
|------------------------------|--------------------------|---------|---------|---------|--------|---------------------------|--------|---------|---------|--------|--------|
|                              | Return                   | Risk    | Min     | Max     | SR     | Return                    | Risk   | Min     | Max     | SR     |        |
| S&P 500                      | 0.0069                   | 0.0365  | -0.1061 | 0.0896  | 0.1881 | 0.0063                    | 0.0453 | -0.1836 | 0.1461  | 0.1389 |        |
| ALasso                       | 0.0034                   | 0.0283  | -0.1334 | 0.0857  | 0.1199 | 0.0039                    | 0.0393 | -0.1067 | 0.6153  | 0.1003 |        |
| SCAD                         | 0.0036                   | 0.0240  | -0.0721 | 0.0536  | 0.1511 | 0.0051                    | 0.0396 | -0.1233 | 0.1192  | 0.1296 |        |
| $L_2$ -Boosting              | 0.0045                   | 0.0307  | -0.1092 | 0.0919  | 0.1455 | 0.0038                    | 0.0284 | -0.1114 | 0.0886  | 0.1333 |        |
| GLasso                       | 0.0076                   | 0.0331  | -0.0849 | 0.0738  | 0.2307 | 0.0077                    | 0.0313 | -0.1161 | 0.0808  | 0.2449 |        |
| Without Return Decomposition | GLasso                   | -0.0018 | 0.0101  | -0.0242 | 0.0227 | -0.1793                   | 0.0023 | 0.0130  | -0.0362 | 0.0680 | 0.1742 |

<sup>1</sup> Table A4 reports the out-of-sample (OOS) performance of each portfolio across two evaluation periods. See Table A1 for a description of each portfolio and the performance metrics.

**Table A5:** Portfolio Performance: EW as Basis Portfolio

| Out-of-sample period            | 2015M1–2019M12 (5 years) |         |         |         |        | 1995M1–2019M12 (25 years) |        |         |         |        |        |
|---------------------------------|--------------------------|---------|---------|---------|--------|---------------------------|--------|---------|---------|--------|--------|
|                                 | Return                   | Risk    | Min     | Max     | SR     | Return                    | Risk   | Min     | Max     | SR     |        |
| EW                              | 0.0052                   | 0.0455  | -0.1210 | 0.1220  | 0.1151 | 0.0081                    | 0.0538 | -0.2597 | 0.2035  | 0.1508 |        |
| With Return<br>Decomposition    | ALasso                   | 0.0067  | 0.0355  | -0.1150 | 0.1193 | 0.1877                    | 0.0026 | 0.0202  | -0.1372 | 0.1333 | 0.1267 |
|                                 | SCAD                     | 0.0008  | 0.0364  | -0.1069 | 0.0909 | 0.0211                    | 0.0072 | 0.0472  | -0.1777 | 0.1585 | 0.1527 |
|                                 | $L_2$ -Boosting          | 0.0034  | 0.0339  | -0.0834 | 0.0791 | 0.0989                    | 0.0049 | 0.0332  | -0.1575 | 0.1235 | 0.1472 |
|                                 | GLasso                   | 0.0077  | 0.0329  | -0.0805 | 0.0743 | 0.2349                    | 0.0074 | 0.0305  | -0.1129 | 0.0703 | 0.2436 |
| Without Return<br>Decomposition | GLasso                   | -0.0008 | 0.0214  | -0.0642 | 0.0545 | -0.0353                   | 0.0009 | 0.0125  | -0.0612 | 0.0590 | 0.0697 |

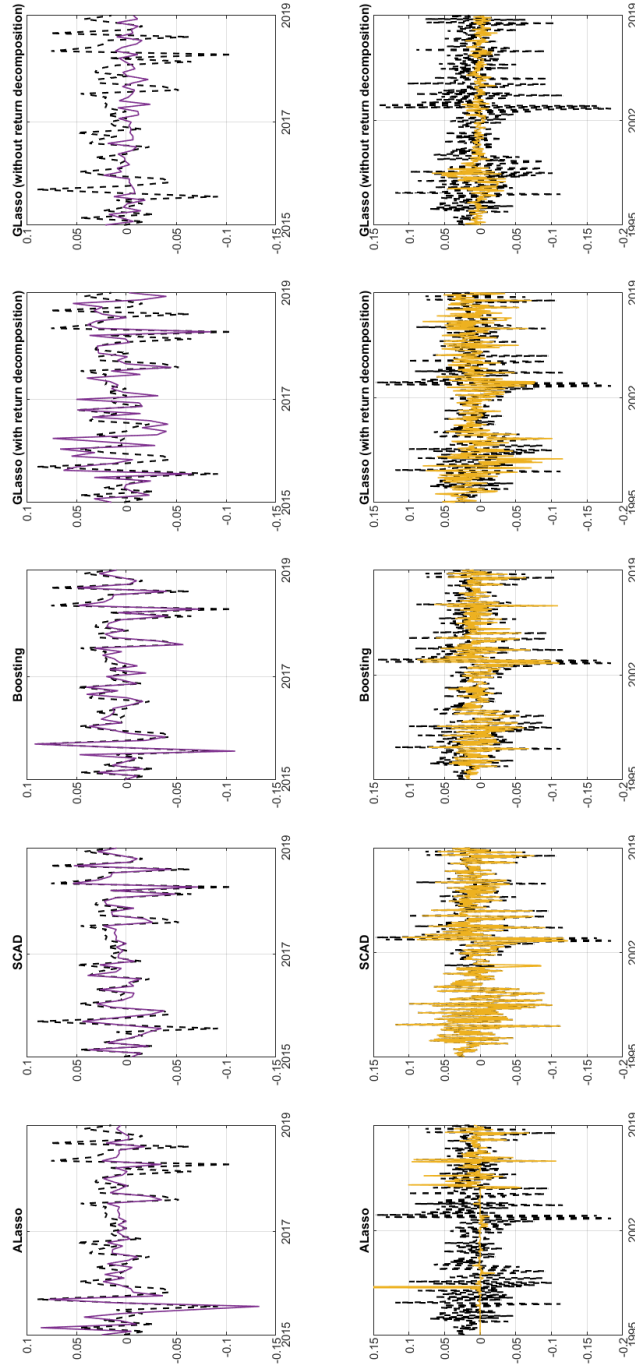
<sup>1</sup> Table A5 reports the out-of-sample (OOS) performance of each portfolio across two evaluation periods. See Table A2 for a description of each portfolio and the performance metrics.

**Table A6:** Portfolio Performance: CRSP as Basis Portfolio

| Out-of-sample period            | 2015M1–2019M12 (5 years) |         |         |         |        | 1995M1–2019M12 (25 years) |         |         |         |        |         |
|---------------------------------|--------------------------|---------|---------|---------|--------|---------------------------|---------|---------|---------|--------|---------|
|                                 | Return                   | Risk    | Min     | Max     | SR     | Return                    | Risk    | Min     | Max     | SR     |         |
| CRSP                            | 0.0092                   | 0.0390  | -0.0972 | 0.1070  | 0.2370 | 0.0043                    | 0.0546  | -0.1669 | 0.2314  | 0.0781 |         |
| With Return<br>Decomposition    | ALasso                   | 0.0043  | 0.0263  | -0.1136 | 0.0715 | 0.1637                    | 0.0033  | 0.0191  | -0.1082 | 0.1070 | 0.1750  |
|                                 | SCAD                     | 0.0051  | 0.0265  | -0.0738 | 0.0689 | 0.1931                    | 0.0056  | 0.0396  | -0.1608 | 0.0866 | 0.1402  |
|                                 | $L_2$ -Boosting          | 0.0058  | 0.0267  | -0.0715 | 0.0496 | 0.2171                    | 0.0044  | 0.0267  | -0.1045 | 0.0688 | 0.1634  |
|                                 | GLasso                   | 0.0075  | 0.0323  | -0.0842 | 0.0664 | 0.2317                    | 0.0073  | 0.0302  | -0.1073 | 0.0753 | 0.2427  |
| Without Return<br>Decomposition | GLasso                   | -0.0033 | 0.0367  | -0.0881 | 0.1028 | -0.0909                   | -0.0004 | 0.0221  | -0.1069 | 0.0679 | -0.0160 |

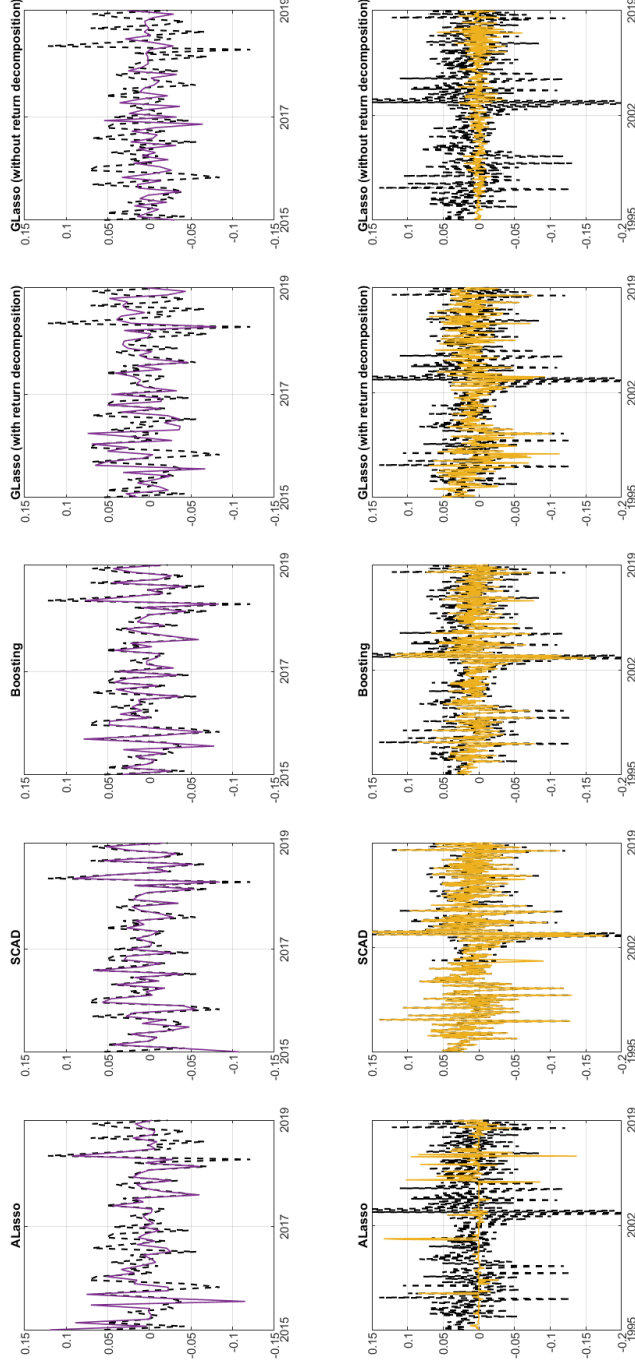
<sup>1</sup> Table A6 reports the out-of-sample (OOS) performance of each portfolio across two evaluation periods. See Table A3 for a description of each portfolio and the performance metrics.

**Figure A1:** Monthly Returns: S&P 500 as Basis Portfolio



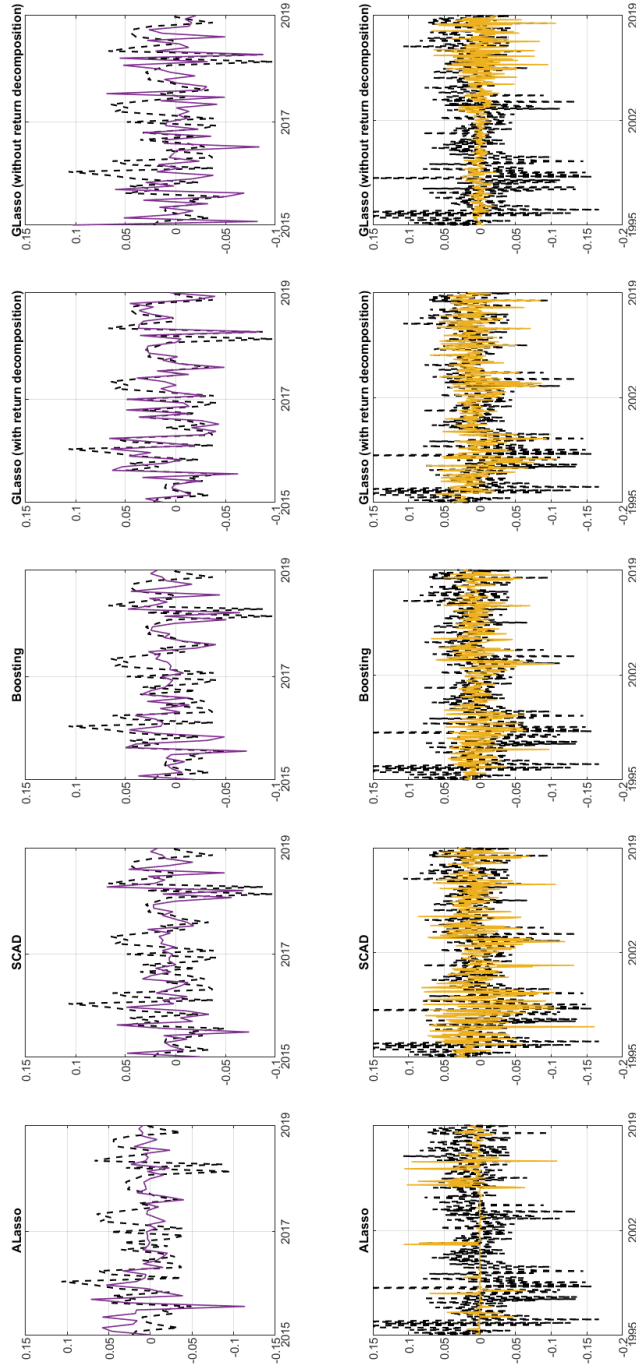
**Notes:** Each panel plots the monthly OOS returns of the portfolios discussed in Sections 3 to 5 against those of the S&P 500 index fund (black dashed line). Solid colored lines correspond to different OOS evaluation periods: purple (first row) for the 5-year period, and yellow (second row) for the 25-year period. See Table A1 for a description of each portfolio and the performance metrics.

**Figure A2:** Monthly Returns: EW as Basis Portfolio



**Notes:** Each panel plots the monthly OOS returns of the portfolios discussed in Sections 3 to 5 against those of the equally weighted S&P 500 portfolio (black dashed line). Solid colored lines correspond to different OOS evaluation periods: purple (first row) for the 5-year period, and yellow (second row) for the 25-year period. See Table A2 for a description of each portfolio and the performance metrics.

**Figure A3:** Monthly Returns: CRSP as Basis Portfolio



**Notes:** Each panel plots the monthly OOS returns of the portfolios discussed in Sections 3 to 5 against those of the value-weighted portfolio of CRSP firms (black dashed line). Solid colored lines correspond to different OOS evaluation periods: purple (first row) for the 5-year period, and yellow (second row) for the 25-year period. See Table A3 for a description of each portfolio and the performance metrics.