

# Aggregating Information for Optimal Portfolio Weights\*

Xiao Li<sup>†</sup>

December 1, 2018

## Abstract

I attempt to address an important issue of the portfolio allocation literature – none of the allocation rules from prior studies consistently delivers good performance. I develop an approach that aggregates information from a wide range of sources to make allocation decisions. Specifically, this approach models the optimal portfolio weights as a function of a broad set of portfolio weights implied by prior allocation rules, and determines the relative contribution from each allocation rule through Elastic Net, a machine-learning technique. Out-of-sample tests suggest that my approach consistently achieves good performance, whereas none of the alternative rules can match the consistency.

JEL Classification: C51, G11

---

\*I thank Scott Cederburg (Committee Chair), Chris Lamoureux, David Brown, and Katherine Barnes for their instructive advice. I thank Sugata Ray for his discussion and other participants at the FMA. I thank the High Performance Computer team at the University of Arizona. All errors are my own.

<sup>†</sup>Xiao Li (lixx2167@email.arizona.edu) is at the Eller College of Management, University of Arizona.

# 1 Introduction

The mean-variance efficiency framework proposed by Markowitz (1952) has been intensively studied by both researchers and practitioners. To implement this framework however, an investor has to estimate the first two moments of asset returns using the observed sample, which often leads to poor out-of-sample performance due to estimation risk. To combat estimation risk, researchers have developed many portfolio allocation rules over the last 60 years. Unfortunately, as documented in DeMiguel, Garlappi, and Uppal (2009), none of these allocation rules can consistently deliver satisfactory performance across different asset samples. Thus, the usefulness of existing allocation rules is open to doubt and the problem of portfolio allocation requires further investigation.

I tackle the problem from a relatively novel angle – the information angle – and I argue that the process of forming portfolio weights is effectively a process of incorporating information. Prior allocation rules are developed under various motivation, but at the very core, they are all trying to incorporate information that researchers believe to be helpful for portfolio allocation. For example, Pástor and Stambaugh (2000) use information from the Fama-French 3 factors to improve moment estimation and form portfolio weights. Kan and Zhou (2007) integrate information from the global minimum variance portfolio into the traditional mean-variance framework to improve out-of-sample performance. From the information angle, the existence of performance inconsistency can be easily understood. This is because the returns in different asset samples exhibit very different profiles (e.g., mean, variance, and covariance), incorporating information from one or two sources (like most, if not all, prior allocation rules) may not sufficiently capture the noisy profiles, which leads to performance inconsistency.

Nevertheless, portfolio weights from prior allocation rules may still contain valuable information for optimal portfolio weights. A potential way to address the performance incon-

sistency issue is to aggregate information from a broad set of allocation rules. This paper develops an approach to achieve this goal. In particular, my approach uses the portfolio weights implied by prior allocation rules (e.g., Pástor and Stambaugh’s (2000) factor-based rule and Kan and Zhou’s (2007) optimal “three fund” rule) as sources of information (instruments, hereafter) and models the optimal portfolio weights as

$$\tilde{w}^* = \phi^0 w^0 + \phi^1 w^1 + \dots + \phi^K w^K, \quad (1)$$

where  $w^0$  through  $w^K$  are instruments implied by  $K + 1$  allocation rules, and  $\phi^0$  through  $\phi^K$  are coefficients that determine the relative importance of each instrument.

Two major issues associated with the instruments warrant special attention. First, some instruments might not be informative about the optimal portfolio. Using these instruments may introduce noise into the estimated portfolio weights. Second, some instruments might be highly correlated with each other, which potentially results in extreme estimates of  $\{\phi^0, \phi^1, \dots, \phi^K\}$  due to multicollinearity. To deal with these issues, I use Elastic Net (Zou and Hastie (2005)), a machine-learning technique, to estimate the coefficients. Elastic Net is designed with two features: (i) the selection effect, which sets the coefficients of uninformative instruments to exactly zero; and (ii) the grouping effect, which deals with the multicollinearity issue by assigning similar weights to highly correlated instruments. These two features speak directly to the two issues, which makes Elastic Net a natural candidate for coefficient estimation.

Previous studies develop an allocation rule and rely solely on this rule for out-of-sample portfolio choice. This practice is equivalent to imposing a constraint of the form  $\{\phi^i = 1, \phi^{-i} = 0\}$  on the coefficients in Equation (1), such that my approach nests using one particular allocation rule as a special case.<sup>1</sup> In contrast, my approach simultaneously incor-

---

<sup>1</sup>The notation  $\phi^{-i}$  means all  $\phi$ ’s except  $\phi^i$ .

**Table I: Table of Asset Samples**

	Datasets	Number of assets	Abbreviation
1	Carhart 4 factors	4	Factor
2	The market factor (Mkt) and the long short legs of SMB, HML, and UMD	7	Factorlegs
3	20 Size and BM portfolios and Carhart 4 factors	24	Size&BM+Factor
4	20 Size and BM portfolios, and Factorlegs	27	Size&BM+Factorlegs
5	20 Size and BM portfolios, 10 momentum portfolios, and Factorlegs	37	Size&BM+Mom+Factorlegs
6	Mkt and 10 industry portfolios	11	Industry
7	Mkt and 10 volatility portfolios	11	Volatility

porates information from several allocation rules to make allocation decisions. If  $w^i$  contains meaningful information for the optimal portfolio weights, then my approach relies more heavily on rule  $i$  (i.e.,  $\hat{\phi}^i$  will be relatively large in magnitude). On the other hand, if  $w^j$  contains only noise, then  $\hat{\phi}^j$  should be set to 0.

Following DeMiguel, Garlappi, and Uppal (2009), I focus on the tangency portfolio (i.e., excess returns of risky assets only) to evaluate the performance of allocation rules. I consider twelve allocation rules from prior studies and conduct out-of-sample tests across the seven asset samples summarized in Table I. Sharpe ratio and certainty-equivalent return (CER) are used as performance measures.

In general, similar to the evidence documented in prior literature, none of the twelve allocation rules can consistently deliver satisfactory out-of-sample performance across all seven asset samples. For example, allocation rules that have positivity constraints achieve monthly Sharpe ratios ranging from 0.257 to 0.303 in Size&BM+Factor, while the performance of other rules ranges from 0.001 to 0.157. However, rules with positivity constraints

achieve only mediocre performance in all other asset samples. Similarly, the top three allocation rules in both Size&BM+Factorlegs and Size&BM+Mom+Factorlegs (sample based mean-variance rule, Bayes-and-Stein rule, and optimal “three-fund” rule.) turn out to be the worst three performers in Volatility and are among the worst in Industry. These observations provide empirical evidence that relying on a particular source of information is not enough to capture various return profiles.

The advantage of aggregating information from multiple instruments is strongly supported by the empirical findings. Using the equally weighted portfolio as a benchmark, my approach achieves statistically better performance in Factorlegs (0.260 vs. 0.134), Size&BM+Factor (0.315 vs. 0.157), Size&BM+Factorlegs (0.347 vs. 0.145), Size&BM+Mom+Factorlegs (0.377 vs. 0.138), and Volatility (0.205 vs. 0.115). In Factor (0.292 vs. 0.279) and Industry (0.156 vs. 0.145), the performance of my approach is higher than but not statistically different from that of the equally weighted portfolio.<sup>2</sup>

Compared with the other eleven rules, my approach also achieves competitive performance. First, among the 77 combinations of allocation rule and asset sample (eleven portfolio allocation rules  $\times$  seven asset samples), my approach achieves significantly better performance in 40% of cases at the 1% level and 45% of cases at the 5% level, while being significantly outperformed only once (0.292 vs. 0.304). Moreover, in three asset samples (Factorlegs, Size&BM+Factor, and Volatility), my approach achieves higher performance than the best performer among all eleven allocation rules.

To provide some evidence that my approach indeed incorporates the most useful information across different asset samples, I calculate the correlation between the out-of-sample returns of my approach and those of each allocation rule. These correlations suggest that my approach is more correlated with instruments that deliver the best performance and less correlated with those that perform poorly for a given asset sample. An analysis of  $\phi$  esti-

---

<sup>2</sup>This situation is potentially caused by the fact that in both asset samples, all allocation rules deliver similar performance and therefore aggregating information does not provide further improvement.

mates provides further evidence. On average, instruments that have better out-of-sample performance tend to receive larger  $\phi$  estimates and those whose out-of-sample performance is poor tend to receive  $\phi$  estimates of zero. Analysis of  $\phi$  also reveals that both the selection effect and the grouping effect contribute to the consistent performance of my approach. In particular, the selection effect appears important in Factorlegs, Size&BM+Factorlegs, Size&BM+Mom+Factorlegs, and Volatility, while the grouping effect seems important in Size&BM+Factor and Industry.

One important question that demands an answer is: does my approach derives its strength by simply following the best performing rules? If this is all that my approach does, an investor can in fact do the same thing more easily, since performance is readily observable. To answer this question, I explore alternative methods for aggregating information that are both intuitive and easy to implement. These methods include putting 100% of one's wealth in the top performing instrument based on historical performance (*Best1*), taking equal positions among the top two (*Best2*) and top three (*Best3*) performing instruments based on historical performance, taking equal positions among all twelve instruments without any discrimination (*Average*), and directly applying Ordinary Least Squares (*OLS*) for coefficient estimation. My evidence shows that none of the five alternative methods is able to deliver good performance with comparable consistency.

To further support the benefit of aggregating information, it is important to compare my approach with the methods proposed by Li (2015) and DeMiguel, Garlappi, Nogales, and Uppal (2009), hereafter, DGNU (2009). All three papers use similar techniques. The key difference between my paper and the other two is that, I apply Elastic Net to deal with issues associated with various instruments, whereas Li (2015) applies Elastic Net and DGNU (2009) apply LASSO and Ridge Regression (two special cases of Elastic Net) directly to asset returns to impose general weight constraints. The evidence suggests that my approach still maintains its competitiveness as it delivers better performance in the majority of cases.

My paper contributes to the literature of portfolio allocation in several dimensions. First, I develop an approach that has a good potential to address the performance inconsistency issue, as evidenced by its consistent performance across a variety of asset samples. Second, my paper closely connects to prior studies that look at combinations of allocation rules. Prior papers focus on deriving the optimal combination theoretically, whereas my paper uses an empirical strategy. There are three advantages of my approach. First, it avoids the theoretical derivation of the optimal combination, which can become a formidable task when the number of rules is large. Second, prior papers (e.g., Tu and Zhou (2011)) focus on combining only two allocation rules, and involve conducting separate derivation for each specific combination. My approach, on the contrary, is able to aggregate information from any number of allocation rules without the need of specific adjustments. Third, instead of combining rules that a researcher believes to perform well based on prior knowledge (e.g.,  $1/N$ ), my approach allows me to be completely agnostic towards the validity of all allocation rules, and systematically decides the relative importance of each allocation rule. Finally, but equally as important, this study reaffirms the usefulness of various sophisticated allocation rules developed in prior literature. Even though these rules cannot consistently deliver satisfactory performance individually, their weights still contain valuable information and serve well as instruments.

This paper also connects the application of machine-learning techniques to finance problems, which has gained considerable popularity in recent years. Bai and Ng (2008) employ Elastic Net to refine the predictors of inflation rate and achieve better prediction accuracy across different forecast horizons. Kozak, Nagel, and Santosh (2017) employ Elastic Net to construct a stochastic discount factor based on the multitude of stock return predictors. Chinco, Clark-Joseph, and Ye (2017) use variable selection techniques (LASSO) to identify short-lived, sparse, and unexpected return predictors for one-minute-ahead stock returns. Gu, Kelly, and Xiu (2018) compare the performance of different machine-learning techniques in

predicting out-of-sample returns based on firm characteristics, and find that most machine-learning techniques outperform traditional methods. Stern, Erel, Tan, and Weisbach (2018) use various machine-learning algorithms to select board directors, and show that directors that are predicted to be good by their algorithms tend to outperform those that are predicted to be bad.

The rest of the paper is organized as follows. Section 2 discusses the classic mean-variance optimization problem, introduces the instrument idea, and develops details of my approach. Section 3 talks about the instruments and the asset samples, and presents evidence on the inconsistency of performance for each allocation rule. Section 4 presents the performance of my approach, conducts analysis of  $\phi$  estimates, and compares my approach with alternative methods of information aggregation. Section 5 conducts robustness tests and Section 6 concludes the paper.

## 2 Methodology

### 2.1 Problem

Consider a mean-variance utility investor who prefers higher expected portfolio returns but dislikes portfolio variance. She attempts to select a set of portfolio weights to maximize her utility such that

$$\max_w w' \mu - \frac{\gamma}{2} w' \Sigma w, \quad (2)$$

where  $\gamma$  is a scalar that represents the level of relative risk aversion,  $w$  is the vector of portfolio weights to be determined,  $\mu$  is the vector of expected excess returns of the underlying assets, and  $\Sigma$  is the covariance matrix among those asset returns. The solution to the problem above is given by  $w = \frac{1}{\gamma} \Sigma^{-1} \mu$ , which implies that the weights of the tangency portfolio are given by



$$w^* = \frac{\Sigma^{-1}\mu}{|\iota'\Sigma^{-1}\mu|}, \quad (3)$$

where  $\iota$  is a vector of ones.<sup>3</sup>

In practice, since the true return moments are unknown, an investor has to replace the true value in Equation (3) with those that are estimated from the sample, which leads to the sample version of the tangency portfolio:

$$\hat{w}_t^* = \frac{\hat{\Sigma}^{-1}\hat{\mu}}{|\iota'\hat{\Sigma}^{-1}\hat{\mu}|}. \quad (4)$$

Unfortunately, estimation risk often leads to poor out-of-sample performance as has been widely documented. A rich literature has emerged to develop various allocation rules to combat estimation risk. Common techniques include imposing moment constraints, employing informative priors, and developing theoretical combination of rules developed in prior studies. Despite all the effort, DeMiguel, Garlappi, and Uppal (2009) show that none of these rules can consistently deliver good performance across a variety of asset samples. Therefore, the estimation risk issue calls for further investigation.

## 2.2 The Idea of Instrument and Estimation Framework

My approach models optimal portfolio weights as a function of instruments (i.e., variables that contain information for allocation decisions). In a general form, the estimated optimal portfolio weights is expressed as

$$\tilde{w}^* = F(\Phi, z^0, z^1, \dots, z^K), \quad (5)$$

---

<sup>3</sup>Following DeMiguel, Garlappi, and Uppal (2009), the absolute value is imposed on the denominator to preserve the sign (i.e., the overall long-short) of the portfolio.

where  $z^0$  through  $z^K$  are instruments that an investor believes to be informative for portfolio allocation, and  $\Phi$  is a vector of parameters that determine the relative importance of each instrument.<sup>4</sup>

In general, instruments can be any variables in an investor's information set and the function  $F(\cdot)$  can take any form. For the choice of instruments, I argue that the portfolio weights from existing allocation rules should be reasonable candidates, since those weights have already incorporated information from different sources that might be useful for portfolio allocation. For example, it has been widely documented that pricing factors (e.g., SMB and HML) play a significant role in explaining asset returns. Therefore, portfolio weights that incorporate information from such factors (e.g., Pástor and Stambaugh (2000)) should be helpful in estimating optimal portfolio weights. For the functional form of  $F(\cdot)$ , I use a linear function not only because of its simplicity, but also because the vector  $\Phi$  can be estimated through a regression framework introduced by Britten-Jones (1999).

In particular, he shows that the weights of the tangency portfolio can be calculated from the following regression:

$$\iota = Xb + u, \tag{6}$$

where  $X = \{x_1, x_2, \dots, x_N\}$  is a  $T \times N$  matrix of excess asset returns,  $\iota$  is a vector of ones, and  $u$  is a vector of error terms. As usual, the solution of the above regression can be obtained by solving the least squared problem:

$$\min_b \frac{1}{2T} \sum_{t=1}^T (1 - X_t b)^2, \tag{7}$$

which yields

$$\hat{b} = (X'X)^{-1}X'\iota. \tag{8}$$

---

<sup>4</sup>Note that, if we only consider two instruments that are the first and second moment estimated from the sample, then we go back to Equation (4), subject to further constraints imposed by the functional form of  $F(\cdot)$ .

Britten-Jones (1999) shows that, the estimated tangency portfolio in Equation (4) can be written as

$$\hat{w} = \frac{\hat{b}}{|\iota'\hat{b}|}, \quad (9)$$

where  $\hat{b}$  is the coefficient estimates in Equation (8). This regression framework is the foundation that my approach builds on.

## 2.3 The Approach

It is helpful to first clarify the timing regarding the implementation of this approach. Given an asset sample with  $T$  months of returns, I first generate a time series of instruments according to each allocation rule (introduced in Section 3) based on a rolling window of  $W_1$  months. This practice generates an instrument sample with  $T - W_1$  observations (first observation corresponds to month  $W_1 + 1$ ). Next, my approach uses these instruments to estimate out-of-sample portfolio weights based on an expanding window with the minimum window length being  $W_2$  months. Specifically, I start out by estimating the out-of-sample portfolio weights for month  $W_1 + W_2 + 1$  (the first out-of-sample portfolio weights) using all available instruments from month  $W_1 + 1$  to month  $W_1 + W_2$  as inputs. Next, I continue to estimate portfolio weights for month  $W_1 + W_2 + 2$  using all available instruments from month  $W_1 + 1$  to month  $W_1 + W_2 + 1$ . The estimation continues until I have estimated the portfolio weights for month  $T$  (the last out-of-sample portfolio weights) using all available instruments from month  $W_1 + 1$  to month  $T - 1$ . This process produces  $T - W_1 - W_2$  out-of-sample portfolio weights and these weights are used to generate the out-of-sample portfolio returns for performance evaluation. In the next few paragraphs, I will develop details of the estimation procedure for an arbitrary month  $\mathcal{T} \in \{W_1 + W_2 + 1, W_1 + W_2 + 2, \dots, T\}$ . For a graphical illustration of the above implementation procedure, please see Figure 1.

To integrate different instruments into the framework, I replace the constant parameter

$b$  in minimization (7) by a dynamic linear function of the instruments:

$$b_t = \phi^0 w_t^0 + \phi^1 w_t^1 + \phi^2 w_t^2 + \dots + \phi^K w_t^K, \quad (10)$$

where  $w_t^0 = \{1/N, 1/N, \dots, 1/N\}$  is the equally weighted portfolio,  $w_t^1$  through  $w_t^K$  are instruments produced by allocation rule 1 through rule  $K$ , and  $\phi^0$  through  $\phi^K$  are coefficients to be estimated. With this modification, the optimization problem in Equation (7) now becomes

$$\min_{\Phi} \frac{1}{2(\mathcal{T} - W_1 - 1)} \sum_{t=W_1+1}^{\mathcal{T}-1} (1 - X_t W_t \Phi)^2. \quad (11)$$

In the optimization above,  $X_t W_t \Phi$  is the instrumented counterpart of  $X_t b$  in equation (7). The row vector  $X_t$  contains the returns of the  $N$  assets in period  $t$ . Matrix  $W_t$  has a dimension of  $N \times (K + 1)$  and takes the form  $\{w_t^0, w_t^1, \dots, w_t^K\}$ . Vector  $\Phi$  contains the coefficients  $\phi^0$  through  $\phi^K$ .

Note that the term  $X_t W_t$  gives a row vector that contains the portfolio returns of  $K + 1$  allocation rules for month  $t$ . Therefore, we can rewrite the optimization in Equation (11) as

$$\min_{\Phi} \frac{1}{2(\mathcal{T} - W_1 - 1)} \sum_{t=W_1+1}^{\mathcal{T}-1} (1 - R_t \Phi)^2, \quad (12)$$

where  $R_t = X_t W_t = \{r_t^0, r_t^1, r_t^2, \dots, r_t^K\}$  is the portfolio returns vector. In other words, the optimization is equivalent to regressing a vector of 1's onto the portfolio returns generated by the instrumented allocation rules.

The most straightforward way to estimate  $\Phi$  is Ordinary Least Squares (OLS). However, two potential problems associated with the instruments make OLS a poor choice. First, some of the instruments might contain only noise and an investor might benefit from ignoring such instruments completely (i.e., assigning a zero coefficient). OLS, however, might assign non-trivial coefficients for these noisy instruments, resulting in poor out-of-sample portfolio

weights. Second, the multicollinearity issue among the instruments might lead OLS to produce extremely large coefficients that tend to result in poor out-of-sample performance. That is, OLS takes extreme positions in an attempt to leverage on highly correlated assets when the out-of-sample correlation might not be as high. My findings in the empirical section further confirm the incapability of OLS.

Instead, I use a machine-learning technique – Elastic Net – to conduct the optimization. It is specifically designed to (i) have the selection effect, which sets the coefficients of instruments that contain only noise to exactly zero, and (ii) encourage the grouping effect, which assigns similar coefficients among highly correlated instruments. Specifically, Elastic Net imposes these two effects by introducing a penalty term on the  $l^1$ -norm and  $l^2$ -norm of the coefficients in optimization (12). With the penalty term, the optimization problem takes the form

$$\min_{\Phi} \frac{1}{2(\mathcal{T} - W_1 - 1)} \sum_{t=W_1+1}^{\mathcal{T}-1} (1 - R_t\Phi)^2 + \lambda[(1 - \alpha)\|\Phi\|_2^2/2 + \alpha\|\Phi\|_1], \quad (13)$$

where  $\lambda[(1 - \alpha)\|\Phi\|_2^2/2 + \alpha\|\Phi\|_1]$  is the penalty term and  $\|\Phi\|_p$  is the  $l^p$ -norm of the vector  $\Phi$ .<sup>5</sup> The  $\lambda$  ( $\lambda \geq 0$ ) parameter controls the intensity of the penalty. When  $\lambda = 0$ , we go back to *OLS*. Larger  $\lambda$  values impose a more intense penalty, which leads to smaller  $\phi$ 's in general and even sets some  $\phi$ 's to zero. When  $\lambda$  surpasses a threshold that depends on the model and the data, all  $\phi$ 's will be set to zero. The  $\alpha$  ( $0 \leq \alpha \leq 1$ ) parameter serves as a tuning parameter that adjusts between the  $l^1$ -norm and  $l^2$ -norm penalties, which balances between the selection effect and the grouping effect. When  $\alpha$  is set to one, only the selection effect is at work, and when  $\alpha$  is set to zero, only the grouping effect is in place. The optimal values for  $\alpha$  and  $\lambda$  are calibrated through cross validation, which will be briefly discussed at the end of this section. For details of the implementation of cross validation, please see the

---

<sup>5</sup>The  $l^p$ -norm of vector  $\Phi$  is given by  $\|\Phi\|_p = (|\phi^0|^p + |\phi^1|^p + \dots + |\phi^K|^p)^{1/p}$ .

Internet Appendix Section D.

The fact that Elastic Net disciplines the  $\phi$  estimates is referred to as coefficient regularization in the machine-learning literature. However, one technical issue brought by regularization is that, *ceteris paribus*, coefficients that are smaller in magnitude are subject to less regularization than coefficients that are larger in magnitude. This issue is particularly pertinent in my setting because different allocation rules generate portfolio returns that have different variance, and therefore the  $\phi$ 's associated with portfolios that have higher variance are subject to less regularization, since the magnitude of these  $\phi$  estimates tends to be small. To deal with this issue, I follow the common practice in the machine-learning literature and standardize the returns of each portfolio by the standard deviation in the observed sample (i.e.,  $\hat{\sigma}^i$  estimated using returns from month  $W_1 + 1$  to month  $\mathcal{T} - 1$  for each rule  $i$ .) and my final optimization problem takes the form

$$\min_{\Phi} \frac{1}{2(\mathcal{T} - W_1 - 1)} \sum_{t=W_1+1}^{\mathcal{T}-1} (1 - R_t^* \Phi)^2 + \lambda[(1 - \alpha)\|\Phi\|_2^2/2 + \alpha\|\Phi\|_1], \quad (14)$$

where  $R_t^* = \{r_t^{*0}, r_t^{*1}, \dots, r_t^{*K}\}$  is the standardized portfolio return with  $r_t^{*i} = r_t^i/\hat{\sigma}^i$ . The  $\phi$  estimates produced by the above optimization,  $\hat{\Phi} = \{\hat{\phi}^0, \hat{\phi}^1, \dots, \hat{\phi}^K\}$ , cannot be directly used to form out-of-sample portfolio weights, since we have to adjust them back to their original magnitude. The adjustment is done by dividing each  $\hat{\phi}^i$  by the associated standard deviation  $\hat{\sigma}^i$ . Finally, the estimated portfolio weights for month  $\mathcal{T}$  are given by

$$\hat{w}_{\mathcal{T}} = \frac{\hat{\phi}^{*0} w_{\mathcal{T}}^0 + \hat{\phi}^{*1} w_{\mathcal{T}}^1 + \dots + \hat{\phi}^{*K} w_{\mathcal{T}}^K}{|\hat{\phi}^{*0} + \hat{\phi}^{*1} + \dots + \hat{\phi}^{*K}|} \quad (15)$$

where  $\hat{\phi}^{*i} = \hat{\phi}^i/\hat{\sigma}^i$  is the adjusted  $\phi$  estimate and the rescaling term in the denominator is to focus on the tangency portfolio. Please see the Internet Appendix Section C for an introduction of numerical estimation of the coefficients. Note that the instruments for month

$\mathcal{T}$ ,  $\{w_{\mathcal{T}}^0, w_{\mathcal{T}}^1, \dots, w_{\mathcal{T}}^K\}$ , in the above equation are ex ante available since they can be calculated by each allocation rule using return data up to month  $\mathcal{T} - 1$ . Therefore, there is no look-ahead bias in Equation (15). Eventually, the out-of-sample portfolio return for month  $\mathcal{T}$  is calculated as

$$R_{\mathcal{T}} = X_{\mathcal{T}} \hat{w}_{\mathcal{T}}. \quad (16)$$

As discussed before, at each time  $\mathcal{T}$ , my approach uses all available instruments to estimate portfolio weights. As  $\mathcal{T}$  iterates through  $\{W_1 + W_2 + 1, W_1 + W_2 + 2, \dots, T\}$ , I generate a time series of out-of-sample portfolio returns that have  $T - W_1 - W_2$  observations for performance evaluation. I use two performance measures, Sharpe ratio and CER, that are given by

$$SR = \bar{R} / \hat{\sigma} \quad (17)$$

$$CER = \bar{R} - \frac{\gamma}{2} \hat{\sigma}^2. \quad (18)$$

In both equations,  $\bar{R}$  and  $\hat{\sigma}$  are the mean and the standard deviation of the out-of-sample portfolio returns, and  $\gamma$  is the coefficient of relative risk aversion. The Sharpe ratio measures how much portfolio return can be expected for each unit of risk ( $\hat{\sigma}$ ) taken. The CER can be interpreted as the constant rate of return that an investor is willing to accept, to avoid holding a risky portfolio. Through out the paper, the investor is assumed to have a risk aversion coefficient of 3 ( $\gamma = 3$ ).

The parameters,  $\lambda$  and  $\alpha$ , are calibrated through cross validation and next comes a brief introduction. First, I select a grid of values for  $\lambda$  and  $\alpha$ . In the baseline result, a grid of 100 values is used for both parameters. Other grids are explored in Section (5). Second, for each parameter pair, I leave one period of instruments out and use all other periods of instruments in the observed sample, to estimate the model in equation (14) and calculate the portfolio weights as in equation (15) for the omitted period. This step is repeated until every period

has been left out once, which generates a time series of portfolio returns for each parameter pair. Finally, I calculate the CER of each time series of returns and pick the parameter pair that achieves the highest CER. To distinguish the CER used in the cross validation and the CER used as out-of-sample performance measure, I denote the CER in cross validation as  $CER_{cv}$  hereafter. At each period  $\mathcal{T}$ , the cross validation process is conducted using only the observed sample, and as  $\mathcal{T}$  iterates through each month, the calibrated values for both parameters are also updated monthly. To account for uncertainty from the out-of-sample, larger risk averse coefficients are adopted during cross validation (Lamoureux and Zhang (2018)). In particular, I use  $\gamma = 6$  (i.e., “twice” as risk averse) in the baseline results and explore the sensitivity of performance using other gamma values ( $\gamma = 4$ ,  $\gamma = 5$ ,  $\gamma = 7$ , and  $\gamma = 8$ ) in Section (5).

## 3 Data and Instruments

### 3.1 Data

I consider seven asset samples that are summarized in Table I. The first six asset samples are from Kenneth French’s data library and I create the Volatility sample according to the instructions in the data library. The Volatility and Industry samples cover a period from July 1926 to December 2016 (1,086 observations). All other asset samples cover a period from January 1927 to December 2016 (1,080 observations) due to the fact that the momentum factor began in January 1927.

The sample that has the fewest number of assets is Factor, which only includes the size factor (SMB), the value factor (HML), the momentum factor (UMD), and the market factor (Mkt). Since some of the portfolio allocation rules involve positivity constraints that are not compatible with the embedded short positions in SMB, HML, and UMD, I split the long and short legs of these factors and combine these factor legs with Mkt to form a new



asset sample, Factorlegs. Following DeMiguel, Garlappi, and Uppal (2009), I combine Factor and Factorlegs with the 20 size and book-to-market portfolios (25 Size and B/M portfolios without the five portfolios in the largest size quintile) respectively, to form Size&BM+Factor and Size&BM+Factorlegs.

Prior literature has shown that both the number of assets and the squared Sharpe ratio of the tangency portfolio are important features that influence the performance of portfolio allocation rules. Therefore, I combine Size&BM+Factorlegs with 10 momentum portfolios and form Size&BM+Mom+Factorlegs to further increase the number of assets and the squared Sharpe ratio. I also include 10 industry portfolios (plus market factor) and 10 volatility portfolios (plus market factor) to enrich the variety of asset samples. Finally, across the seven asset samples, the number of assets ranges from 4 to 37 and the squared Sharpe ratio ranges from 0.040 (Volatility) to 0.167 (Size&BM+Mom+Factorlegs).<sup>6</sup>

## 3.2 Instruments

I consider instruments implied by the equally weighted portfolio and eleven portfolio allocation rules from DeMiguel, Garlappi, and Uppal (2009): sample based mean-variance rule (“mv”), optimal “three fund” rule (“mv-min”, Kan and Zhou (2007)), Bayes-Stein shrinkage rule (“bs”, James and Stein (1961)), Bayesian “data and model” rule (“dm(0.01)”, Pástor and Stambaugh (2000)), sample based minimum variance rule (“min”), mixture of naive and minimum variance rule (“ew-min”, DeMiguel, Garlappi, and Uppal (2009)), unobservable factor model (“mp”, MacKinlay and Pástor (2000)), sample based mean-variance rule with positivity constraint (“mv-c”), sample based minimum variance rule with positivity constraint (“min-c”), Bayes-Stein shrinkage rule with positivity constraint (“bs-c”), and

---

<sup>6</sup>The squared Sharpe ratio of each asset sample is calculated as  $\mu'\Sigma^{-1}\mu$ , where  $\mu$  and  $\Sigma$  are the mean and covariance matrix calculated using the entire sample. For the seven asset samples I consider in this paper, the squared Sharpe ratios are 0.040, 0.046, 0.081, 0.104, 0.136, 0.139, and 0.167 for Volatility, Industry, Factor, Factorlegs, Size&BM+Factor, Size&BM+Factorlegs, and Size&BM+Mom+Factorlegs respectively. These squared Sharpe ratios are largely comparable with those in DeMiguel, Garlappi, and Uppal (2009).

combination of naive rule and minimum variance rule with positivity constraint (“g-min-c”, DeMiguel, Garlappi, and Uppal (2009)). These rules employ a variety of techniques including Bayesian methods, methods that impose constraints, and mixture of methods developed in prior studies. Using simulation, Kan and Zhou (2007) show that, a window of roughly 250 months of data is needed for the sample based mean-variance rule to deliver reasonable performance.<sup>7</sup> Therefore, I calculate the instruments using a rolling window of 240 months ( $W_1 = 240$ ) to ensure reasonable information quality. In the Internet Appendix Section A, I provide further evidence that demonstrates that instruments estimated using 240 months of data contain better information, compared with shorter estimation window.

Table II presents the Sharpe ratios for each rule across the seven asset samples. The equally weighted portfolio is used as a benchmark and the  $p$ -value is from a test for whether the difference between the Sharpe ratio of a particular allocation rule and that of the equally weighted portfolio is zero. These  $p$ -values are calculated following Jobson and Korkie (1981) after making the correction pointed out by Memmel (2003).

Several performance patterns deserve highlights. Let us first focus on rules without positivity constraints. Four of these rules are minimum variance rules (“1/N”, “min”, “mp”, and “ew-min”) that focus on minimizing portfolio variance.<sup>8</sup> The other four are mean-variance rules (“mv”, “bs”, “dm(0.01)”, and “mv-min”) that conduct mean-variance optimization. From an information perspective, minimum variance rules only take information from the covariance matrix, whereas mean-variance rules also consider the information in the mean.

In Factorlegs, Size&BM+Factorlegs, and Size&BM+Mom+Factorlegs samples, mean-variance rules tend to generate much higher Sharpe ratios than minimum variance rules. In particular, in Size&BM+Mom+Factorlegs sample, mean-variance rules have Sharpe ra-

---

<sup>7</sup>Note that, since their tests are based on simulated data, there is no look ahead bias for selection of estimation window.

<sup>8</sup>The “mp” rule is considered as minimum variance rule because it mimics the 1/N portfolio most of the time. The time series of portfolio return of these two rules has a correlation of more than 0.91 in all asset samples except Size&BM+Mom+Factor.

tios of 0.323 (“dm(0.01)”), 0.396 (“mv”), 0.411 (“bs”), and 0.416 (“mv-min”), whereas the highest Sharpe ratio realized by minimum variance rules is 0.217 (“ew-min”). On the other hand, in the Industry sample, the highest Sharpe ratio (0.178) is achieved by “ew-min” and “min”, both of which are minimum variance rules. More strikingly, the three rules “bs”, “mv-min”, and “mv” that have the top three Sharpe ratios in both Size&BM+Factor and Size&BM+Mom+Factorlegs turn out to have the worst Sharpe ratios in the Volatility sample and are among the lowest in the Industry sample.

As argued in Kirby and Ostdiek (2012), the spread of the mean in industry-sorted portfolios is not different from zero (noisy information), whereas characteristic-sorted portfolios have more persistent spread in the mean vector (meaningful information). Therefore, completely ignoring the noisy information in the mean leads to better Sharpe ratios in the Industry sample, whereas considering the mean leads to better results in, for example, Size&BM+Mom+Factorlegs. The trade-off between mean-variance and minimum variance can also be seen in the four rules with positivity constraints, “mv-c”, “min-c”, “bs-c”, and “g-min-c”. Both “mv-c” and “bs-c” are mean-variance rules and “min-c” and “g-min-c” are minimum variance rules. Again, we can observe alternations of performance across different asset samples. In particular, among Size&BM+Factorlegs, Size&BM+Mom+Factorlegs, and Factorlegs, “mv-c” and “bs-c” tend to perform better, whereas in Industry and Volatility, “min-c” and “g-min-c” tend to perform better.

A different pattern emerges when asset samples involve factors – assets that have embedded long-short positions. In Size&BM+Factor, Sharpe ratios of rules with positivity constraints are 0.303 (“mv-c”), 0.296 (“min-c”), 0.303 (“bs-c”), and 0.257 (“g-min-c”), whereas the highest Sharpe ratio achieved by rules without positivity constraints is 0.157 (“1/N”). These different patterns can be potentially driven by the factors themselves. Due to the embedded long-short positions, factors tend to see occasional crashes (e.g., the momentum crash), which distort the estimated correlation structure – the high correlation estimated

from the observed sample may not be as high, or even reversed in the out-of-sample. Positivity constraints, as shown in Jagannathan and Ma (2003), can be more suitable in this situation by using lower correlation (than estimated) to form portfolios, which explains the good performance of rules with positivity constraints to some extent. However, in almost all other asset samples, rules with positivity constraints achieve only mediocre performance.

There are several important takeaways from the above discussion. First, incorporating information from limited sources cannot deliver consistent performance as evidenced by the performance variation of individual rules. Second, the set of instruments from all twelve rules together, could potentially provide a wide range of information that captures different return profiles across different asset samples, as some rules deliver good performance in some asset samples. Third, to successfully aggregate information, it is important to filter away noisy instruments. Fourth, the occasional similar performances among certain instruments and the common component (e.g., both “min” and “ew-min” have “min” in common) in the formation of those instruments also suggest a potential multicollinearity issue.<sup>9</sup>

## 4 Empirical

### 4.1 Baseline Results

This section demonstrates the out-of-sample performance of my approach (hereafter, *EN*) in comparison with the twelve instrumented allocation rules. *EN* is implemented based on an expanding window with the minimum window length being 120 months ( $W_2=120$ ). Please see the Internet Appendix Section B for how to determine the baseline minimum window length. Performance of *EN* using alternative minimum window length is explored in Section (5).

Table III presents the monthly Sharpe ratio for the twelve portfolio allocation rules

---

<sup>9</sup>More evidence for the fourth point comes in subsequent sections.

and  $EN$ . Each column contains the Sharpe ratios for a particular asset sample among all allocation rules and each row contains the Sharpe ratios for an allocation rule across seven asset samples. The parentheses contain the  $p$ -values for testing whether the difference between the Sharpe ratio of  $EN$  and that of a particular allocation rule is equal to zero.

The merit of aggregating information from multiple instruments is strongly supported by the results. In Factorlegs, Size&BM+Factor, Size&BM+Factorlegs, Size&BM+Mom+Factorlegs, and Volatility,  $EN$  delivers Sharpe ratios that are statistically and economically higher than those of the equally weighted portfolio. In Industry and Factor,  $EN$  has Sharpe ratios that are higher than but not statistically different from those of the equally weighted portfolio.

Compared with other allocation rules, the performance of my approach is also consistently competitive. For all 77 combinations of allocation rule and asset sample (eleven allocation rules  $\times$  seven asset samples), my approach delivers a statistically higher Sharpe ratio 40% of cases at the 1% level, and 45% of cases at the 5% level. Second, among all cases, my approach is statistically outperformed only once (by “bs” in Factor), though the economic difference is small (0.292 vs. 0.304). Finally, in Factorlegs, Size&BM+Factor, and Volatility, my approach delivers higher Sharpe ratios than the top performers among all eleven rules.

The limitation of the Sharpe ratio is that it is invariant to proportional changes of the mean and standard deviation. That is, when both the mean and standard deviation become twice as large, the Sharpe ratio remains constant. As a consequence, a portfolio that looks scary in the eye of a risk averse investor due to high variance, might still maintain a decent Sharpe ratio by offering a high enough mean. Therefore, it is relevant to use CER to evaluate the performance of allocation rules from the perspective of a risk averse investor.

Table IV contains the monthly CERs (in percent). To test whether the difference between the CER of my approach and that of an allocation rule is zero, I follow Greene (2002) and report the  $p$ -values in parentheses. The next example demonstrates that risk aversion indeed impacts how an investor perceives the performance of allocation rules. Note that

“mv” has a Sharpe ratio of 0.131 in Size&BM+Factor. It is certainly worse than the Sharpe ratio achieved by the top performer, but it is still far better than 0.001 achieved by “min”. However, once we incorporate risk aversion, things change dramatically. The monthly CER of “mv” in Size&BM+Factor is -3.452%, which means an investor is willing to give away 41.42% of his wealth to avoid taking the “mv” portfolio on an annual basis. At the same time, “min” has a monthly CER of -0.2% which is far better than that of “mv”.

My approach withstands the alternative performance measure. Specifically, in the same five asset samples as mentioned before, my approach achieves statistically higher CERs than the equally weighted portfolio. Among the 77 rule and asset sample combinations, my approach delivers statistically better CERs 48% of cases at the 1% level (55% of cases at the 5% level), while being outperformed five times.

The evidence presented so far only looks at absolute performance. To provide further evidence regarding how each allocation rule performs in relation to others, I next examine relative performance ( $RP$ ). For allocation rule  $i$ , the  $RP$  in asset sample  $j$  is defined as:

$$RP_j^i = \frac{SR_j^i - SRmin_j}{SRmax_j - SRmin_j}, \quad (19)$$

where  $SR_j^i$  is the Sharpe ratio achieved by rule  $i$  in sample  $j$ ,  $SRmin_j$  ( $SRmax_j$ ) is the lowest (highest) Sharpe ratio achieved among all thirteen allocation rules in sample  $j$ . The  $RP$ 's in terms of CER can be similarly defined. Equation (19) implies that the best performing allocation rule has a  $RP = 100\%$ , the worst performing allocation rule has a  $RP = 0\%$ , and all other allocation rules have  $RP$ 's between  $0\%$  and  $100\%$ . Since there are seven asset samples, each allocation rule has seven  $RP_j$ 's ( $j = 1, \dots, 7$ ).

To provide a summary of the seven  $RP$ 's, I conduct a Box-Whisker plot for each allocation rule as shown in Figure 2 (Sharpe ratio) and Figure 3 (CER). Each box depicts the lowest (lower bar), second to lowest (lower end of the rectangle), median (middle bar), second to

highest (higher end of the rectangle), and the highest (upper bar) values of the seven  $RP_j$ 's. These box plots provide more direct evidence of the consistent performance achieved by  $EN$ , as the max-min distance is relatively short and the mass of the box is concentrated right below the 100% bar. Other rules tend to have wider min-max distances and boxes, which indicates more volatile performance.

Another advantage of  $RP$  is that, given an allocation rule, I can take the average of the  $RP$ 's across the seven asset samples to reflect an average performance, whereas taking the average of Sharpe ratio or CER might lead to a biased reflection.<sup>10</sup> The average  $RP$  for each allocation rule is listed on the right hand side.  $EN$  has an average  $RP$  of 90% and 89% in terms of Sharpe ratio and CER, respectively, whereas the highest average  $RP$  achieved by other rules are 69% for Sharpe ratio and 74% for CER.<sup>11</sup>

Recall that the goal of my approach is to incorporate the most useful information into portfolios. If this approach fulfills this intention, we should expect to see the portfolio formed by my approach to be more correlated to those portfolios implied by the most informative instruments. Therefore, I compute the correlation between the out-of-sample portfolio returns of my approach and those of each allocation rule. Panel A of Table V reports the correlations and Panel B provides the Sharpe ratios (repeated from Table II) for the convenience of comparison. Two observations warrant attention. First, when all allocation rules deliver similar performance, my approach has similar correlations with all allocation rules (e.g., in Factor and Industry). This observation might also explain why in Factor and Industry, my approach achieves similar performance to that of the equally weighted portfolio. That is, since all allocation rules deliver similar performance, aggregating information does not

---

<sup>10</sup>Here is an example. Suppose we have four asset samples whose true (highest possible) Sharpe ratios are 1.0, 0.25, 0.25, and 0.25, respectively. We have two allocation rules whose realized Sharpe ratios are 0.95, 0.05, 0.04, and 0.04 for the first rule, and 0.25, 0.25, 0.25, and 0.25 for the second. The average Sharpe ratio of the first rule is higher than that of the second, which suggests that the first rule is consistently better. However, the second rule achieves the highest possible Sharpe ratio in three of the four asset samples, and therefore can also be rightfully considered as a better allocation rule.

<sup>11</sup>Contrary to the findings in DeMiguel, Garlappi, and Uppal (2009), "1/N" underperforms most other allocation rules. Longer estimation window ( $W_1 = 240$ ) for instruments calculation should contribute to this observation.

grant further performance improvement. Second, when there is a sizable variation among the performance of allocation rules, my approach tends to be highly correlated with rules that deliver top performance and less correlated with poorly performing rules (e.g., in Factorlegs, Size&BM+Factor, Size&BM+Factorlegs, Size&BM+Mom+Factorlegs, and Volatility).

As discussed in Section 2.3, Elastic Net grants both the selection and the grouping effects. The selection effect assigns zero  $\phi$  estimates to instruments that contain only noise. The grouping effect grants the freedom of assigning similar coefficients among highly correlated instruments, which deals with the issue of extreme coefficient estimates due to multicollinearity. For a detailed introduction to the selection effect, please see the Internet Appendix Section C. For the grouping effect, please see Zou and Hastie (2005) Theorem 1. I next focus on empirical analyses of how both effects regularize the  $\phi$  estimates and how each effect contributes to the performance of my approach.

To demonstrate the presence of the selection effect, I plot the time series of  $\phi$  estimates of the instruments. For a clear demonstration, given an asset sample, I only plot the time series for the three instruments whose  $\hat{\phi}$  is set to zero most often.<sup>12</sup> Figure 4 presents these plots. Several observations need to be emphasized. First, the selection effect indeed sets the coefficients of some instruments to zero. Second, consistent with the performance inconsistency documented before, the instruments whose  $\hat{\phi}$ 's are set to zero most often, vary greatly across different asset samples. This observation provides additional evidence for the limitation of relying on a single source of information for portfolio allocation. Third, combined with the ex post performance shown in Table III and Table IV, instruments that have  $\hat{\phi} = 0$  most often, tend to be the ones that realize poor out-of-sample performance. Together, these observations demonstrate that, in general, selection effect filters away noisy information.

To demonstrate the presence of the grouping effect, I calculate the time series average of  $\phi$  estimates under *EN* (Panel A of Table VI) and compare them with the time series average

---

<sup>12</sup>These instruments are determined by, first calculating the percentage of the months that receive a zero  $\phi$  estimate for each instrument, and next picking the three instruments that have the highest percentage in each sample.



generated by OLS (Panel B of Table VI). I focus on the  $\phi$  estimates in Size&BM+Factor for a brief comparison.<sup>13</sup> Due to the common component “min”, the “min” rule and the “ew-min” rule tend to be highly correlated. This high correlation leads OLS to produce large positive  $\phi$  estimates for “min” (24.63, on average) and large negative  $\phi$  estimates for “ew-min” (-25.76, on average). However, combined with the ex post performance, we know that both rules realized poor out-of-sample performances that are almost identical (Sharpe ratio being 0.001 and 0.003), which makes taking advantage of the correlation a potentially dangerous practice. On the contrary, the grouping effect assigns similar negative  $\phi$  estimates for both rules (-1.51 for “min” and -1.50 for “ew-min”). The intuition is that, since both rules are performing poorly, an investor might benefit from shorting both instruments simultaneously. The performance of my approach in this asset sample justifies this intuition, as it delivers the highest Sharpe ratio and CER among all allocation rules. It is important to point out that Elastic Net always has the freedom not to impose the grouping effect and assign positive-negative positions to leverage on correlations among certain instruments. For example, in Size&BM+Mom+Factorlegs, “mv-c” and “bs-c” are highly correlated instruments and each receives an average coefficient estimate of 3.20 and -2.45, respectively.

The above discussion has two implications. First, the strength of my approach comes not only from assigning relatively large  $\hat{\phi}$  to instruments that have realized good performance, but also from systematically considering the correlation among all instruments and the plausibility of leveraging on such correlation. Second, it is important to compare the performance of my approach with that of pure performance-chasing strategies (i.e., taking big positions among the best performing instruments, which mimics assigning large  $\hat{\phi}$ .) as the latter can be done very easily by an investor. Therefore, I explore performance chasing strategies in Section 4.2, together with other methods of information aggregation.

---

<sup>13</sup>Due to the complicated relationship among the instruments, it is difficult to draw a general conclusion regarding how the  $\phi$ 's for different instruments are related to each other. Therefore, I need to focus on cases in which the grouping effect is more pronounced.

The evidence presented so far only shows how  $\phi$  estimates behave differently under the influence of the selection and grouping effects. It does not show how each effect contributes to out-of-sample performance. To capture the benefits introduced by the two effects separately, I next compare the performance of  $EN$  with that of two special versions of this approach where the parameter  $\alpha$  is fixed at 0 (denoted as *Group*) and 1 (denoted as *Select*). The first version only imposes the grouping effect and the second only imposes the selection effect. Panel A of Table VII shows the Sharpe ratios and Panel B shows the CERs. My evidence suggests that both the selection effect and the grouping effect possess unique advantages. Specifically, *Group* delivers good performance in Size&BM+Factor and Industry, while *Select* realizes competitive performance in Size&BM+Factorlegs, Size&BM+Mom+Factorlegs, and Volatility. However, to achieve better consistency, combining both effects simultaneously is essential.

## 4.2 Alternative Methods of Information Aggregation

This section focuses on alternative methods of information aggregation that are both intuitive and easy to implement. Had these “easier” methods consistently delivered good performance, the effort for coefficient estimation in my approach seems pointless. The most straightforward way to aggregate information is to simply take equal positions among all twelve instruments without imposing any discrimination conditions (*Average*). The second method is to pick instruments based on observed performance as mentioned in Section 4.1. That is, at each period, a risk averse investor can simply distribute all of his wealth among the instruments that achieved the highest CER and hope the good performance will persist into the future. Therefore, I consider three such methods including: evenly distributing one’s wealth among the top one (*Best1*), two (*Best2*), and three (*Best3*) instruments that have

achieved the highest CER in the observed sample.<sup>14</sup> Exploring these three methods can also help finding out whether the strength of my approach comes entirely as a consequence of performance chasing. The fifth method I consider is directly applying OLS for the coefficient estimation (*OLS*). In Section 2.3, I only demonstrate that the  $\hat{\phi}$ 's under OLS tend to be volatile and often incur extreme values. It is still important to show how those coefficient estimates translate into performance.

Panel A of Table VIII presents the Sharpe ratios and Panel B presents the CERs for *Average*, *Best1*, *Best2*, *Best3*, and *OLS*. The  $p$ -value comes from a test of whether the difference between the Sharpe ratio or CER realized by *EN* and those realized by a certain method is zero. Among all five methods, *OLS* performs the worst. It realizes significantly lower Sharpe ratios in four of the asset samples and often sees negative CERs. Picking instruments based on historical performance does not deliver satisfactory results either. In the best cases, methods based on observed performance achieve similar performance as *EN*, but in most cases these methods are significantly outperformed by *EN*, especially in terms of CER. These findings also show that, beyond performance chasing, systematically considering correlation among instruments also make important contribution to the strength of my approach. *Average* delivers fairly similar Sharpe ratios as *EN* in Factorlegs, Size&BM+Factorlegs, Size&BM+Mom+Factorlegs, and even higher Sharpe ratio in Industry. However, it performs significantly worse in Size&BM+Factor and Volatility. In particular, the CER realized by *Average* in Volatility is -10.31%, the lowest CER realized among all five alternative methods in all seven asset samples. Overall, the evidence suggests that none of the alternative methods for information aggregation delivers a comparable performance.

---

<sup>14</sup>Due to a similar conservative concern, larger risk aversion coefficient ( $\gamma = 6$ , same as the baseline case of *EN*) is used for the calculation of the in-sample CER. Smaller  $\gamma$  values produce almost identical out-of-sample performance and therefore are not reported.

### 4.3 Elastic Net, LASSO, and Ridge Regression Directly to Underlying Assets

This paper is not the first in the portfolio allocation literature to use Elastic Net, LASSO, or Ridge Regression to construct out-of-sample portfolio. The key difference between my paper and previous studies is that my paper uses Elastic Net to deal with the issues – noise and multicollinearity – associated with the instruments, whereas prior papers use these techniques to motivate sparse and stable portfolio (e.g., Li (2015)) or impose general constraints on portfolio weights (e.g., DGNU (2009)). Another way to look at the difference is that previous studies apply these techniques directly to the returns of underlying assets, whereas my paper applies Elastic Net to various instruments. Therefore, I next compare the performance of my approach with that of directly applying Elastic Net (*ElasticNet*), LASSO (*LASSO*), and Ridge Regression (*Ridge*) to the underlying assets.

Panel A of Table IX presents the Sharpe ratio and Panel B presents the CER. Among the three methods, *Ridge* seems to be the worst method as it realizes similar performance as “1/N”. This observation is consistent with DGNU (2009) as they document that *Ridge* often generates portfolio weights that are close to “1/N”. Compared with *Ridge*, *LASSO* sees much better performance. It delivers better performance than *EN* in Factor and Industry. However, in most of the other five asset samples, *LASSO* statistically underperforms *EN* in terms of Sharpe ratio and/or CER. The most interesting comparison is between *EN* and *ElasticNet* as both approaches employ the same technique. Indeed, in many cases *ElasticNet* delivers a performance that is close to *EN*. However, *EN* still maintains its advantage as it achieves a higher performance than *ElasticNet* in all 14 cases (two measure  $\times$  seven asset samples) except one (Factor in terms of CER). Moreover, *EN* achieves significantly better Sharpe ratios in Size&BM+Mom+Factorlegs, and significantly better CER in Size&BM+Factor and Volatility.

## 5 Robustness

The robustness of this approach is explored in three dimensions. First, I test whether the performance of this approach is sensitive to the fineness of the parameter grids. Second, I explore whether the performance of this approach is sensitive to the minimum window length, which would also help to test whether the performance is driven by a particular historical period. Lastly, I explore whether different values of risk aversion coefficient ( $\gamma = 4, 5, 7,$  and  $8$ ), used to calculate  $\text{CER}_{cv}$  in cross validation, affect the out-of-sample performance of this approach.

### 5.1 Alternative Parameter Grids

In the baseline result, I use a grid of 100 candidate values for both  $\alpha$  and  $\lambda$ . In this section, I test whether the performance of this approach is sensitive to the fineness of the grids. Both finer and coarser grids are tested. For coarser grids, I include 50 candidate values for  $\alpha$  with 50 candidate values for  $\lambda$ , 50 candidate values for  $\alpha$  with 100 candidate values for  $\lambda$ , and 100 candidate values for  $\alpha$  with 50 candidate values for  $\lambda$ . For finer grids, I include 150 candidate values for  $\alpha$  with 100 candidate values for  $\lambda$ , 100 candidate values for  $\alpha$  with 150 candidate values for  $\lambda$ , and 150 candidate values for  $\alpha$  with 150 candidate values for  $\lambda$ . I still keep the same notation  $EN$  for the baseline case and denote other cases as  $EN_{X\&Y}$ , where  $X$  denotes the number of candidate values for  $\alpha$  and  $Y$  for  $\lambda$ . Panel A of Table X shows the Sharpe ratios and Panel B shows the CERs. As the evidence suggests, both Sharpe ratio and CER are highly stable across different grids. Therefore, the performance of my approach is not sensitive to the fineness of parameter grids as long as the fineness is above a reasonable level.

## 5.2 Various Minimum Windows

In the baseline case, my approach is conducted using a minimum window length of 120 months ( $W_2=120$ ). To alleviate the concern that the performance of  $EN$  is driven by a particular minimum window length, I next present graphs that plot the Sharpe ratios against a series of minimum window lengths for each asset sample. In each graph, I also plot the Sharpe ratio of the equally weighted portfolio as a benchmark. Equally important, changing the minimum window length can also alleviate the concern that the performance is entirely driven by a particular historical period, since a different minimum window length is associated with a different length of out-of-sample. Figure 5 presents these graphs. The shortest minimum window length is 120 months (point 0 on the horizontal axis), which is associated with an out-of-sample of roughly 720 months and the longest minimum window length is 480 months (point 360 on the horizontal axis), which is associated with an out-of-sample of roughly 360 months. Across all seven asset samples,  $EN$  realizes highly persistent performance for various minimum window lengths, which also shows that the performance is not entirely driven by a particular historical period.

## 5.3 Alternative Risk Aversion in Cross Validation

In the baseline case,  $CER_{cv}$  is calculated using  $\gamma = 6$ . This practice is motivated by the conservative concern. That is, due to the uncertainty in the out-of-sample, an investor could benefit from adopting higher levels of risk aversion when estimating the model using the observed sample. Therefore, I next explore how different levels of risk aversion in calculating  $CER_{cv}$  influence the out-of-sample performance, if at all. In particular, I test whether there is a significant difference in out-of-sample performance between the baseline case ( $\gamma = 6$ ) and alternative cases ( $\gamma=4, 5, 7, \text{ and } 8$ ). All values for  $\gamma$  are reasonably large risk aversion coefficients that are often used in the literature.

Ex ante, I form two expectations that are motivated by the performance difference between mean-variance and minimum variance allocation rules (Section 3.2). That is, if the mean of the asset sample contains meaningful information (e.g., Size&BM+Mom+Factorlegs), mean-variance allocation rules perform better, and if the mean contains only noisy information (e.g., Industry), minimum variance rules deliver better results. Therefore, I can potentially see that higher risk aversion leads to better performance in asset samples where the mean does not contain meaningful information (e.g., Industry). I also expect that higher risk aversion leads to lower performance in asset samples where the mean does contain useful information (e.g., Size&BM+Mom+Factorlegs).

Panel A of Table XI presents the Sharpe ratios. In general, different  $\gamma$  values do not have meaningful impact on the performance of my approach. Even though the performance is significantly higher in Factor when  $\gamma = 8$ , the economic difference is tiny. Consistent with the first expectation, the Sharpe ratio monotonically increases in Industry as  $\gamma$  goes from 4 to 8. However, there is no significant difference compared with the baseline case ( $\gamma = 6$ ).

Panel B presents the CER. In all asset samples, except Size&BM+Mom+Factorlegs and Size&BM +Factorlegs, different levels of risk aversion in  $CER_{cv}$  calculation do not have meaningful impact on out-of-sample performance. In Size&BM+Mom+Factorlegs and Size&BM+Factorlegs, consistent with the second expectation, I observe that higher levels of risk aversion ( $\gamma = 7$  and 8) lead to statistically lower CER. However, compared with all other allocation rules, my approach still maintains its competitiveness. Overall, the evidence from both the Sharpe ratio and the CER suggest that my approach is robust to alternative values of risk aversion coefficient.

## 6 Conclusion

I attempt to address the performance inconsistency issue in the portfolio allocation literature. I develop an approach that aggregates information from multiple instruments (portfolio weights from existing allocation rules) to estimate optimal portfolio weights. This approach obtains its strength from a machine-learning technique – Elastic Net – which grants the selection effect that filters noise contained in those instruments and the grouping effect which combats the multicollinearity issue. Out-of-sample performance shows that my approach consistently delivers satisfactory Sharpe ratios and certainty-equivalent returns across seven asset samples, whereas none of the instrumented portfolio allocation rules can match the consistency of performance. Moreover, alternative ways of aggregating information that are intuitive and easy-to-implement cannot deliver performance with similar consistency.



## References

- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2001, The Distribution of Realized Exchange Rate Volatility, *Journal of American Statistical Association* 96, 42–55.
- Bai, Jushan, and Serena Ng, 2008, Forecasting Economic Time Series Using Targeted Predictors, *Journal of Econometrics* 146, 304–317.
- Bawa, Vijay S., S. Brown, and R Klein, 1979, *Estimation Risk and Optimal Portfolio Choice*, North-Holland Publishing Company: Amsterdam, the Netherlands.
- Brandt, Michael W., Pedro Santa-Clara, and Rossen Valkanov, 2009, Parametric Portfolio Policies: Exploiting Characteristics in the Cross-section of Equity Returns, *The Review of Financial Studies* 22, 3411–3447.
- Britten-Jones, Mark, 1999, The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights, *The Journal of Finance* 54, 655–671.
- Bruder, Benjamin, Nicolas Gaussel, Jean-Charles Richard, and Thierry Roncalli, 2013, Regularization of Portfolio Allocation, Working paper, Paris Diderot University.
- Chinco, Alexander M, Adam D Clark-Joseph, and Mao Ye, 2017, Sparse Signals in the Cross-Section of Returns, Forthcoming, *Journal of Finance*.
- DeMiguel, Victor, Lorenzo Garlappi, Francisco J Nogales, and Raman Uppal, 2009, A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms, *Management Science* 55, 798–812.
- DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009, Optimal Versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?, *The Review of Financial Studies* 22, 1915–1953.

- DeMiguel, Victor, Alberto Martin-Utrera, Francisco J Nogales, and Raman Uppal, 2017, A Portfolio Perspective on the Multitude of Firm Characteristics, Working paper, London Business School.
- Fama, Eugene F, and Kenneth R French, 1992, The Cross-Section of Expected Stock Returns, *The Journal of Finance* 47, 427–465.
- Fan, Jianqing, Jingjin Zhang, and Ke Yu, 2012, Vast Portfolio Selection with Gross-Exposure Constraints, *Journal of the American Statistical Association* 107, 592–606.
- Friedman, Jerome, Trevor Hastie, and Rob Tibshirani, 2010, Regularization Paths for Generalized Linear Models via Coordinate Descent, *Journal of Statistical Software* 33, 1–22.
- Green, Richard C, and Burton Hollifield, 1992, When Will Mean-Variance Efficient Portfolios Be Well Diversified?, *The Journal of Finance* 47, 1785–1809.
- Greene, W. H., 2002, *Econometric Analysis*: New York: Prentice-Hall.
- Gu, Shihao, Bryan T Kelly, and Dacheng Xiu, 2018, Empirical Asset Pricing Via Machine Learning, Working paper, University of Chicago.
- Guo, Danqiao, Phelim P. Boyle, Chengguo Weng, and Tong S. Wirjanto, 2018, Eigen Portfolio Selection: A Robust Approach to Sharpe Ratio Maximization, Working paper, University of Waterloo.
- Hautsch, Nikolaus, Lada M Kyj, and Peter Malec, 2015, Do High-Frequency Data Improve High-Dimensional Portfolio Allocations?, *Journal of Applied Econometrics* 30, 263–290.
- Jacob, Laurent, Guillaume Obozinski, and Jean-Philippe Vert, 2009, Group LASSO with Overlap and Graph LASSO, in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 433–440.

- Jagannathan, Ravi, and Tongshu Ma, 2003, Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps, *The Journal of Finance* 58, 1651–1683.
- James, William, and Charles Stein, 1961, Estimation with Quadratic Loss, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 361–379.
- Jobson, J. David, and Bob M. Korkie, 1980, Estimation for Markowitz Efficient Portfolios, *Journal of the American Statistical Association* 75, 544–554.
- , 1981, Performance Hypothesis Testing with the Sharpe and Treynor Measures, *The Journal of Finance* 36, 889–908.
- Jorion, Philippe, 1985, International Portfolio Diversification with Estimation Risk, *Journal of Business* 58, 259–278.
- , 1986, Bayes-Stein Estimation for Portfolio Analysis, *Journal of Financial and Quantitative Analysis* 21, 279–292.
- Kan, Raymond, and Guofu Zhou, 2007, Optimal Portfolio Choice with Parameter Uncertainty, *Journal of Financial and Quantitative Analysis* 42, 621–656.
- Kirby, Chris, and Barbara Ostdiek, 2012, It’s All in the Timing: Simple Active Portfolio Strategies that Outperform Naive Diversification, *Journal of Financial and Quantitative Analysis* 47, 437–467.
- Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2017, Shrinking the Cross Section, Working paper, University of Chicago.
- Kyj, Lada, Barbara Ostdiek, and Katherine Ensor, 2009, Realized Covariance Estimation in Dynamic Portfolio Optimization, Working paper, Humboldt-Universit.

- Lamoureux, Christopher, and huacheng Zhang, 2018, An Empirical Assessment of Characteristics and Optimal Portfolios, Working paper, University of Arizona.
- Ledoit, Olivier, and Michael Wolf, 2004a, Honey, I Shrunk the Sample Covariance Matrix, *The Journal of Portfolio Management* 30, 110–119.
- , 2004b, A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices, *Journal of Multivariate Analysis* 88, 365–411.
- Li, Jiahua, 2015, Sparse and Stable Portfolio Selection with Parameter Uncertainty, *Journal of Business & Economic Statistics* 33, 381–392.
- Li, Jiahua, and Weiye Chen, 2014, Forecasting Macroeconomic Time Series: LASSO-Based Approaches and Their Forecast Combinations with Dynamic Factor Models, *International Journal of Forecasting* 30, 996–1015.
- MacKinlay, Craig A., and Lúboš Pástor, 2000, Asset Pricing Models: Implications for Expected Returns and Portfolio Selection, *The Review of Financial Studies* 13, 883–916.
- Markowitz, Harry M., 1952, Portfolio Selection., *Journal of Finance* 7, 77–91.
- Memmel, Christoph, 2003, Performance Hypothesis Testing with the Sharpe Ratio, *Finance Letters* 1, 21–23.
- Pástor, Lúboš, and Robert F Stambaugh, 2000, Comparing Asset Pricing Models: An Investment Perspective, *Journal of Financial Economics* 56, 335–381.
- Pooter, Michiel de, Martin Martens, and Dick van Dijk, 2008, Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data—But Which Frequency to Use?, *Econometric Reviews* 27, 199–229.

- Sharpe, William F, 1963, A Simplified Model for Portfolio Analysis, *Management Science* 9, 277–293.
- Stein, Charles, 1956, Inadmissibility of the Usual Estimator for the Mean of A Multivariate Normal Distribution, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pp. 197–206.
- Stern, Léa H, Isil Erel, Chenhao Tan, and Michael S Weisbach, 2018, Selecting Directors Using Machine Learning, Working paper, University of Washington.
- Tibshirani, Robert, 1996, Regression Shrinkage and Selection via the LASSO, *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 267–288.
- Tu, Jun, and Guofu Zhou, 2011, Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies, *Journal of Financial Economics* 99, 204–215.
- Yuan, Ming, and Yi Lin, 2006, Model Selection and Estimation in Regression with Grouped Variables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68, 49–67.
- Zou, Hui, and Trevor Hastie, 2005, Regularization and Variable Selection via the Elastic Net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 301–320.

**Table II: Monthly Sharpe Ratio for Instrument**

This table presents the Sharpe ratio for each allocation rule across 7 asset samples including: Factor (N=4), Factorlegs (N=7), Size&BM+Factor (N=24), Size&BM+Factorlegs (N=27), Size&BM+Mom+Factorlegs (N=37), Industry (N=11), and Volatility (N=11). Portfolio allocation rules included are: equally weighted portfolio, sample based mean-variance rule, Bayes-Stein rule, Data-and-Model rule, sample based minimum variance rule, missing-factor (MacKinlay and Pastor (2000)), optimal “three fund” rule (Kan and Zhou (2007)), mixture of minimum variance and naive rule, sample based mean-variance with positivity constraint, Bayes-Stein with positivity constraint, minimum variance with positivity constraint, and minimum variance rule with generalized constraints. Instruments are calculated based on 20-year rolling window ( $W_1 = 240$ ).

Strategy	Factor (N=4)	Factorlegs (N=7)	Size&BM +Factor (N=24)	Size&BM +Factorlegs (N=27)	Size&BM+Mom +Factorlegs (N=37)	Industry (N=11)	Volatility (N=11)
1/N	0.279	0.134	0.157	0.145	0.138	0.145	0.115
mv	0.297 (0.48)	0.220 (0.10)	0.131 (0.62)	0.344 (0.00)	0.396 (0.00)	0.128 (0.81)	0.066 (0.35)
mv-min	0.303 (0.31)	0.224 (0.08)	0.127 (0.57)	0.370 (0.00)	0.416 (0.00)	0.159 (0.64)	0.070 (0.38)
bs	0.304 (0.29)	0.224 (0.07)	0.129 (0.59)	0.362 (0.00)	0.411 (0.00)	0.153 (0.76)	0.069 (0.37)
dm(0.01)	0.210 (0.01)	0.257 (0.00)	0.129 (0.57)	0.274 (0.01)	0.323 (0.00)	0.148 (0.79)	0.194 (0.00)
min	0.295 (0.51)	0.144 (0.73)	0.001 (0.00)	0.209 (0.04)	0.122 (0.01)	0.178 (0.16)	0.145 (0.43)
ew-min	0.298 (0.37)	0.146 (0.63)	0.003 (0.00)	0.207 (0.01)	0.217 (0.00)	0.178 (0.11)	0.150 (0.33)
mp	-0.005 (0.00)	0.134 (0.34)	0.141 (0.00)	0.141 (0.00)	0.135 (0.01)	0.138 (0.00)	0.101 (0.00)
mv-c	0.298 (0.46)	0.189 (0.00)	0.303 (0.00)	0.191 (0.00)	0.191 (0.00)	0.118 (0.25)	0.143 (0.17)
min-c	0.295 (0.49)	0.142 (0.47)	0.296 (0.00)	0.146 (0.94)	0.143 (0.70)	0.164 (0.28)	0.169 (0.04)
bs-c	0.304 (0.29)	0.188 (0.00)	0.303 (0.00)	0.186 (0.00)	0.182 (0.00)	0.127 (0.48)	0.146 (0.15)
g-min-c	0.298 (0.39)	0.139 (0.29)	0.257 (0.00)	0.147 (0.76)	0.142 (0.60)	0.161 (0.22)	0.145 (0.00)

**Table III: Monthly Sharpe Ratio (Baseline)**

This table presents the Sharpe ratio for 7 asset samples including: Factor (N=4), Factorlegs (N=7), Size&BM+Factor (N=24), Size&BM+Factorlegs (N=27), Size&BM+Mom+Factorlegs (N=37), Industry (N=11), and Volatility (N=11). *EN* denotes the method proposed in this paper with the risk aversion coefficient  $\gamma = 6$  in cross validation. Portfolio allocation rules included are: equally weighted portfolio, sample based mean-variance rule, Bayes-Stein rule, Data-and-Model rule, sample based minimum variance rule, missing-factor (MacKinlay and Pastor (2000)), optimal “three fund” rule (Kan and Zhou (2007)), mixture of minimum variance and naive rule, sample based mean-variance with positivity constraint, Bayes-Stein with positivity constraint, minimum variance with positivity constraint, and minimum variance rule with generalized constraints. Instruments are calculated based on 20-year rolling window ( $W_1 = 240$ ). *EN* is implemented based on expanding window with the minimum window being 10 years ( $W_2 = 120$ ).

Strategy	Factor (N=4)	Factorlegs (N=7)	Size&BM +Factor (N=24)	Size&BM +Factorlegs (N=27)	Size&BM+Mom +Factorlegs (N=37)	Industry (N=11)	Volatility (N=11)
<i>EN</i>	0.292	0.260	0.315	0.347	0.377	0.156	0.205
mv	0.297 (0.53)	0.220 (0.13)	0.131 (0.00)	0.344 (0.93)	0.396 (0.53)	0.128 (0.47)	0.066 (0.00)
mv-min	0.303 (0.14)	0.224 (0.16)	0.127 (0.00)	0.370 (0.33)	0.416 (0.13)	0.159 (0.78)	0.069 (0.00)
bs	0.304 (0.10)	0.224 (0.16)	0.129 (0.00)	0.362 (0.55)	0.411 (0.22)	0.153 (0.98)	0.069 (0.00)
dm(0.01)	0.210 (0.02)	0.257 (0.86)	0.129 (0.00)	0.274 (0.04)	0.323 (0.06)	0.148 (0.64)	0.194 (0.74)
1/N	0.279 (0.62)	0.134 (0.00)	0.157 (0.00)	0.145 (0.00)	0.138 (0.00)	0.145 (0.62)	0.115 (0.04)
min	0.295 (0.85)	0.144 (0.00)	0.001 (0.00)	0.209 (0.00)	0.122 (0.00)	0.178 (0.26)	0.145 (0.13)
ew-min	0.298 (0.70)	0.146 (0.00)	0.003 (0.00)	0.207 (0.00)	0.217 (0.00)	0.178 (0.24)	0.150 (0.16)
mp	-0.005 (0.00)	0.133 (0.00)	0.141 (0.00)	0.141 (0.00)	0.135 (0.00)	0.138 (0.45)	0.101 (0.02)
mv-c	0.298 (0.49)	0.189 (0.02)	0.303 (0.67)	0.191 (0.00)	0.191 (0.00)	0.118 (0.19)	0.143 (0.11)
min-c	0.295 (0.83)	0.142 (0.00)	0.296 (0.55)	0.146 (0.00)	0.143 (0.00)	0.164 (0.71)	0.169 (0.35)
bs-c	0.303 (0.11)	0.188 (0.02)	0.303 (0.67)	0.186 (0.00)	0.182 (0.00)	0.127 (0.31)	0.146 (0.13)
g-min-c	0.298 (0.71)	0.139 (0.00)	0.257 (0.09)	0.147 (0.00)	0.142 (0.00)	0.161 (0.84)	0.145 (0.15)

**Table IV: Monthly Certainty-Equivalent Return (Baseline)**

This table presents the CER (in percent) for 7 asset samples including: Factor (N=4), Factorlegs (N=7), Size&BM+Factor (N=24), Size&BM+Factorlegs (N=27), Size&BM+Mom+Factorlegs (N=37), Industry (N=11), Volatility (N=11). *EN* denotes the method proposed in this paper with the risk aversion coefficient  $\gamma = 6$  in the cross validation process. The CER for out-of-sample performance is evaluated using  $\gamma = 3$ . Portfolio allocation rules included are: equally weighted portfolio, sample based mean-variance rule, Bayes-Stein rule, Data-and-Model rule, sample based minimum variance rule, missing-factor (MacKinlay and Pastor (2000)), optimal “three fund” rule (Kan and Zhou (2007)), mixture of minimum variance and naive rule, sample based mean-variance with positivity constraint, Bayes-Stein with positivity constraint, minimum variance with positivity constraint, and minimum variance rule with generalized constraints. Instruments are calculated based on 20-year rolling window ( $W_1 = 240$ ). *EN* is implemented based on expanding window with the minimum window being 10 years ( $W_2 = 120$ ).

Strategy	Factor (N=4)	Factorlegs (N=7)	Size&BM +Factor (N=24)	Size&BM +Factorlegs (N=27)	Size&BM+Mom +Factorlegs (N=37)	Industry (N=11)	Volatility (N=11)
<i>EN</i>	0.421	1.072	1.647	1.608	1.816	0.383	0.678
mv	0.434 (0.29)	-1.375 (0.00)	-3.452 (0.00)	1.974 (0.25)	2.610 (0.03)	0.070 (0.24)	-300.11 (0.00)
mv-min	0.417 (0.72)	0.021 (0.01)	-1.726 (0.00)	1.918 (0.06)	2.468 (0.00)	0.422 (0.64)	-136.07 (0.00)
bs	0.418 (0.74)	0.067 (0.02)	-1.887 (0.00)	1.979 (0.06)	2.566 (0.00)	0.382 (0.88)	-153.46 (0.00)
dm(0.01)	0.422 (0.99)	0.908 (0.17)	-2.028 (0.00)	0.864 (0.11)	1.630 (0.63)	0.355 (0.67)	0.553 (0.47)
1/N	0.407 (0.71)	0.298 (0.00)	0.404 (0.00)	0.346 (0.00)	0.315 (0.00)	0.343 (0.66)	0.208 (0.04)
min	0.394 (0.29)	0.327 (0.00)	-0.002 (0.00)	0.558 (0.00)	0.635 (0.00)	0.446 (0.37)	0.244 (0.04)
ew-min	0.397 (0.33)	0.336 (0.00)	-0.001 (0.00)	0.553 (0.00)	0.596 (0.00)	0.443 (0.38)	0.312 (0.05)
mp	-0.999 (0.00)	0.293 (0.00)	0.326 (0.00)	0.324 (0.00)	0.296 (0.00)	0.315 (0.49)	0.118 (0.02)
mv-c	0.435 (0.27)	0.567 (0.00)	0.541 (0.00)	0.582 (0.00)	0.587 (0.00)	0.229 (0.20)	0.323 (0.07)
min-c	0.395 (0.30)	0.331 (0.00)	0.397 (0.00)	0.349 (0.00)	0.335 (0.00)	0.397 (0.83)	0.373 (0.09)
bs-c	0.418 (0.74)	0.560 (0.00)	0.544 (0.00)	0.553 (0.00)	0.532 (0.00)	0.270 (0.33)	0.333 (0.07)
g-min-c	0.397 (0.30)	0.323 (0.00)	0.524 (0.00)	0.358 (0.00)	0.333 (0.00)	0.389 (0.94)	0.328 (0.08)



**Table V: Correlation of Out-of-Sample Return**

Panel A presents the correlation between the out-of-sample returns of  $EN$  and each of the twelve portfolio allocation rule. These portfolio allocation rules include: equally weighted portfolio, sample based mean-variance rule, Bayes-Stein rule, Data-and-Model rule, sample based minimum variance rule, missing-factor (MacKinlay and Pastor (2000)), optimal “three fund” rule (Kan and Zhou (2007)), mixture of minimum variance and naive rule, sample based mean-variance with positivity constraint, Bayes-Stein with positivity constraint, minimum variance with positivity constraint, and minimum variance rule with generalized constraints. To facilitate comparison, Panel B presents the Sharpe ratios of the out-of-sample returns for each allocation rule that are directly copied from Table II. Instruments are calculated based on 20-year rolling window ( $W_1 = 240$ ) and  $EN$  is implemented based on expanding window with the minimum window being 10 years ( $W_2 = 120$ ).

Strategy	Factor (N=4)	Factorlegs (N=7)	Size&BM +Factor (N=24)	Size&BM +Factorlegs (N=27)	Size&BM+Mom +Factorlegs (N=37)	Industry (N=11)	Volatility (N=11)
<b>Panel A: Correlation between <math>EN</math> and Instruments</b>							
mv	0.98	0.75	0.54	0.73	0.73	0.61	0.40
mv-min	0.98	0.76	0.53	0.81	0.79	0.77	0.41
bs	0.98	0.76	0.53	0.78	0.77	0.73	0.41
dm(0.01)	0.61	0.90	0.31	0.58	0.74	0.87	0.61
1/N	0.80	0.50	0.36	0.43	0.40	0.77	0.31
min	0.90	0.51	0.02	0.49	0.57	0.81	0.46
ew-min	0.90	0.52	0.03	0.51	0.56	0.83	0.47
mp	0.02	0.50	0.32	0.42	0.39	0.75	0.27
mv-c	0.98	0.68	0.74	0.59	0.60	0.73	0.46
min-c	0.90	0.51	0.67	0.41	0.39	0.76	0.51
bs-c	0.98	0.68	0.74	0.57	0.57	0.74	0.47
g-min-c	0.91	0.51	0.59	0.42	0.39	0.79	0.41
<b>Panel B: Monthly Sharpe Ratio</b>							
mv	0.297	0.220	0.131	0.344	0.396	0.128	0.066
mv-min	0.303	0.224	0.127	0.370	0.416	0.159	0.070
bs	0.304	0.224	0.129	0.362	0.411	0.153	0.069
dm(0.01)	0.210	0.257	0.129	0.274	0.323	0.148	0.194
1/N	0.279	0.134	0.157	0.145	0.138	0.145	0.115
min	0.295	0.144	0.001	0.209	0.122	0.178	0.145
ew-min	0.298	0.146	0.003	0.207	0.217	0.178	0.150
mp	-0.005	0.134	0.141	0.141	0.135	0.138	0.101
mv-c	0.298	0.189	0.303	0.191	0.191	0.118	0.143
min-c	0.295	0.142	0.296	0.146	0.143	0.164	0.169
bs-c	0.304	0.188	0.303	0.186	0.182	0.127	0.146
g-min-c	0.298	0.139	0.257	0.147	0.142	0.161	0.145

**Table VI: Average of Time Series of  $\phi$ 's**

Panel A of this table presents the time series average of  $\phi$  estimates for *EN* and Panel B, for *OLS*. The time series of  $\phi$ 's are estimated based on an expanding window with the minimum window length being 120 months. In each period, the  $\phi$  estimates are rescaled so that the absolute value of the summation of the rescaled  $\phi$ 's equals to one.

Strategy	Factor (N=4)	Factorlegs (N=7)	Size&BM +Factor (N=24)	Size&BM +Factorlegs (N=27)	Size&BM+Mom +Factorlegs (N=37)	Industry (N=11)	Volatility (N=11)
<b>Panel A: Average of the Time Series of <math>\phi</math>'s for <i>EN</i></b>							
mv	0.16	-0.00	0.26	-0.13	-0.02	0.22	-0.01
mv-min	0.12	0.08	0.39	0.73	0.38	0.02	-0.16
bs	0.13	0.03	0.35	-0.03	-0.08	0.22	-0.04
dm(0.01)	0.05	0.76	0.72	0.04	0.07	-0.96	1.37
1/N	0.14	-0.33	0.05	0.02	0.37	1.09	0.24
min	-0.01	0.08	-1.51	0.11	0.31	1.42	-0.27
ew-min	-0.01	0.09	-1.50	0.21	0.58	-1.51	0.11
mp	-0.06	-0.29	0.03	-0.22	-0.77	0.28	-1.38
mv-c	0.17	0.63	0.52	1.21	3.20	0.67	-0.33
min-c	0.02	-0.36	0.79	-0.76	-0.94	0.07	0.59
bs-c	0.13	0.63	0.51	0.45	-2.45	-0.46	0.50
g-min-c	0.17	-0.32	0.39	-0.64	0.35	-0.06	0.38
<b>Panel B: Average of the Time Series of <math>\phi</math>'s for <i>OLS</i></b>							
mv	-0.89	4.02	-0.56	1.67	5.58	-12.11	0.08
mv-min	-23.09	12.38	-4.20	3.45	7.47	-24.54	0.27
bs	-42.76	-17.46	4.78	-4.76	-13.80	42.42	-0.59
dm(0.01)	-0.00	0.99	0.64	0.05	0.07	-4.21	0.85
1/N	0.72	-61.17	0.86	12.82	13.92	9.78	-11.83
min	21.68	-16.83	24.63	0.07	3.09	8.39	-6.72
ew-min	-0.12	19.72	-25.76	0.62	-1.70	-13.39	6.44
mp	-0.12	32.77	-0.80	-10.41	-13.56	-2.93	-9.30
mv-c	-0.95	-43.62	-1.82	5.53	7.26	1.77	-0.68
min-c	-23.23	-21.01	0.37	0.45	-2.95	1.44	-20.56
bs-c	69.68	47.03	1.24	-4.01	-7.31	-1.89	1.13
g-min-c	0.18	44.20	0.35	-4.48	2.94	-2.72	41.92

**Table VII: Selection and Grouping Effect**

Panel A of this table presents the Sharpe ratios and Panel B presents the CERs for *EN*, and two special cases of *EN* where only the selection effect (*Select*) and only the grouping effect is granted (*Group*). The *p*-value for Sharpe ratio is calculated following Jobson and Korkie (1981) after making correction pointed out by Memmel (2003) and the *p*-value for CER is calculated following to Greene (2002).

Strategy	Factor (N=4)	Factorlegs (N=7)	Size&BM +Factor (N=24)	Size&BM +Factorlegs (N=27)	Size&BM+Mom +Factorlegs (N=37)	Industry (N=11)	Volatility (N=11)
<b>Panel A: Monthly Sharpe Ratio</b>							
<i>EN</i>	0.292	0.260	0.315	0.347	0.377	0.156	0.205
<i>Select</i>	0.301 (0.19)	0.245 (0.09)	0.242 (0.00)	0.343 (0.56)	0.380 (0.66)	0.152 (0.18)	0.192 (0.66)
<i>Group</i>	0.306 (0.16)	0.217 (0.08)	0.341 (0.00)	0.261 (0.01)	0.246 (0.00)	0.179 (0.10)	0.166 (0.09)
<b>Panel B: Monthly CER</b>							
<i>EN</i>	0.421	1.072	1.647	1.608	1.816	0.383	0.678
<i>Select</i>	0.437 (0.11)	0.990 (0.27)	0.972 (0.00)	1.682 (0.06)	1.853 (0.38)	0.374 (0.85)	0.594 (0.58)
<i>Group</i>	0.412 (0.56)	0.663 (0.01)	0.827 (0.00)	0.828 (0.00)	0.757 (0.00)	0.452 (0.20)	0.436 (0.09)

**Table VIII: Alternative Ways for Information Aggregation**

Panel A of this table presents the Sharpe ratios and Panel B presents the CERs for *EN*, taking equal positions among the best one (*Best1*), two (*Best2*) and three (*Best3*) instruments based on observed performance, taking equal positions among all instruments (*Average*), and directly applying *OLS* for coefficient estimation. At each period, the best performing rules are those that achieve the highest CER with  $\gamma = 6$ . The *p*-value for Sharpe ratio is calculated following Jobson and Korkie (1981) after making correction pointed out by Memmel (2003) and the *p*-value for CER is calculated following to Greene (2002).

Strategy	Factor (N=4)	Factorlegs (N=7)	Size&BM +Factor (N=24)	Size&BM +Factorlegs (N=27)	Size&BM+Mom +Factorlegs (N=37)	Industry (N=11)	Volatility (N=11)
<b>Panel A: Monthly Sharpe Ratio</b>							
<i>EN</i>	0.292	0.260	0.315	0.347	0.377	0.156	0.205
<i>Average</i>	0.267 (0.25)	0.257 (0.83)	0.212 (0.00)	0.339 (0.74)	0.386 (0.68)	0.185 (0.04)	0.081 (0.00)
<i>Best1</i>	0.244 (0.02)	0.254 (0.71)	0.217 (0.00)	0.259 (0.00)	0.216 (0.00)	0.149 (0.90)	0.184 (0.54)
<i>Best2</i>	0.275 (0.23)	0.231 (0.20)	0.260 (0.04)	0.241 (0.00)	0.217 (0.00)	0.152 (0.93)	0.197 (0.81)
<i>Best3</i>	0.286 (0.64)	0.220 (0.11)	0.287 (0.27)	0.239 (0.00)	0.217 (0.00)	0.154 (0.96)	0.190 (0.64)
<i>OLS</i>	0.050 (0.00)	0.090 (0.00)	0.105 (0.00)	0.336 (0.61)	0.382 (0.81)	0.135 (0.47)	0.118 (0.03)
<b>Panel B: Annualized CER</b>							
<i>EN</i>	0.421	1.072	1.647	1.608	1.816	0.383	0.678
<i>Average</i>	0.374 (0.14)	0.968 (0.36)	0.682 (0.00)	1.294 (0.03)	1.557 (0.04)	0.477 (0.09)	-10.310 (0.00)
<i>Best1</i>	0.398 (0.52)	0.910 (0.17)	0.700 (0.00)	0.960 (0.00)	0.683 (0.00)	0.357 (0.91)	0.493 (0.29)
<i>Best2</i>	0.418 (0.88)	0.766 (0.04)	0.734 (0.00)	0.805 (0.00)	0.677 (0.00)	0.362 (0.92)	0.513 (0.30)
<i>Best3</i>	0.415 (0.77)	0.704 (0.02)	0.702 (0.00)	0.744 (0.00)	0.648 (0.00)	0.368 (0.94)	0.456 (0.17)
<i>OLS</i>	-0.250 (0.00)	-3.363 (0.00)	0.088 (0.00)	1.832 (0.24)	2.114 (0.07)	0.293 (0.53)	-1.524 (0.00)

**Table IX: LASSO, Ridge Regression and Elastic Net**

Panel A of this table presents the Sharpe ratios and Panel B presents the CERs for *EN*, *LASSO*, *Ridge*, and *ElasticNet*, where *LASSO*, *Ridge*, and *ElasticNet* denote applying LASSO, Ridge regression, and Elastic Net to the returns of the underlying assets, respectively. The  $p$ -value for Sharpe ratio is calculated following Jobson and Korkie (1981) after making correction pointed out by Memmel (2003) and the  $p$ -value for CER is calculated following Greene (2002).

Strategy	Factor (N=4)	Factorlegs (N=7)	Size&BM +Factor (N=24)	Size&BM +Factorlegs (N=27)	Size&BM+Mom +Factorlegs (N=37)	Industry (N=11)	Volatility (N=11)
<b>Panel A: Monthly Sharpe Ratio</b>							
<i>EN</i>	0.292	0.261	0.315	0.347	0.377	0.156	0.205
<i>LASSO</i>	0.282 (0.65)	0.238 (0.38)	0.307 (0.81)	0.281 (0.02)	0.310 (0.02)	0.189 (0.25)	0.151 (0.14)
<i>Ridge</i>	0.280 (0.57)	0.147 (0.00)	0.175 (0.00)	0.153 (0.00)	0.155 (0.00)	0.149 (0.77)	0.136 (0.11)
<i>ElasticNet</i>	0.284 (0.70)	0.258 (0.90)	0.300 (0.63)	0.308 (0.17)	0.302 (0.02)	0.155 (0.88)	0.174 (0.38)
<b>Panel B: Monthly CER</b>							
<i>EN</i>	0.421	1.072	1.647	1.608	1.816	0.383	0.678
<i>LASSO</i>	0.470 (0.16)	0.921 (0.38)	0.993 (0.03)	1.214 (0.03)	1.557 (0.27)	0.566 (0.16)	0.317 (0.04)
<i>Ridge</i>	0.403 (0.54)	0.361 (0.00)	0.461 (0.00)	0.391 (0.00)	0.400 (0.00)	0.356 (0.79)	0.307 (0.08)
<i>ElasticNet</i>	0.461 (0.22)	1.029 (0.77)	1.026 (0.04)	1.423 (0.32)	1.475 (0.15)	0.373 (0.92)	0.393 (0.10)

**Table X: Alternative Grids**

This table documents the performance of my approach with different grids for parameter  $\alpha$  and  $\lambda$ . These alternative grids include, 50 candidate values for  $\alpha$  with 50 candidate values for  $\lambda$ , 50 candidate values for  $\alpha$  with 100 candidate values for  $\lambda$ , 100 candidate values for  $\alpha$  with 50 candidate values for  $\lambda$ , 100 candidate values for  $\alpha$  with 150 candidate values for  $\lambda$ , 150 candidate values for  $\alpha$  with 100 candidate values for  $\lambda$ , and 150 candidate values for  $\alpha$  with 150 candidate values for  $\lambda$ . Same notation,  $EN$ , is used for the baseline case (i.e., 100 candidate values for both parameters) and I denote other cases as  $EN_{X\&Y}$ , where  $X$  denotes the number of candidate values for  $\alpha$  and  $Y$  for  $\lambda$ . Panel A of this table presents the Sharpe ratios and Panel B presents the CERs. The  $p$ -values is for testing the difference between performance in the baseline results and those under alternative grids. The  $p$ -value for Sharpe ratio is calculated following to Jobson and Korkie (1981) after making correction pointed out by Memmel (2003) and the  $p$ -value for CER is calculated following Greene (2002).

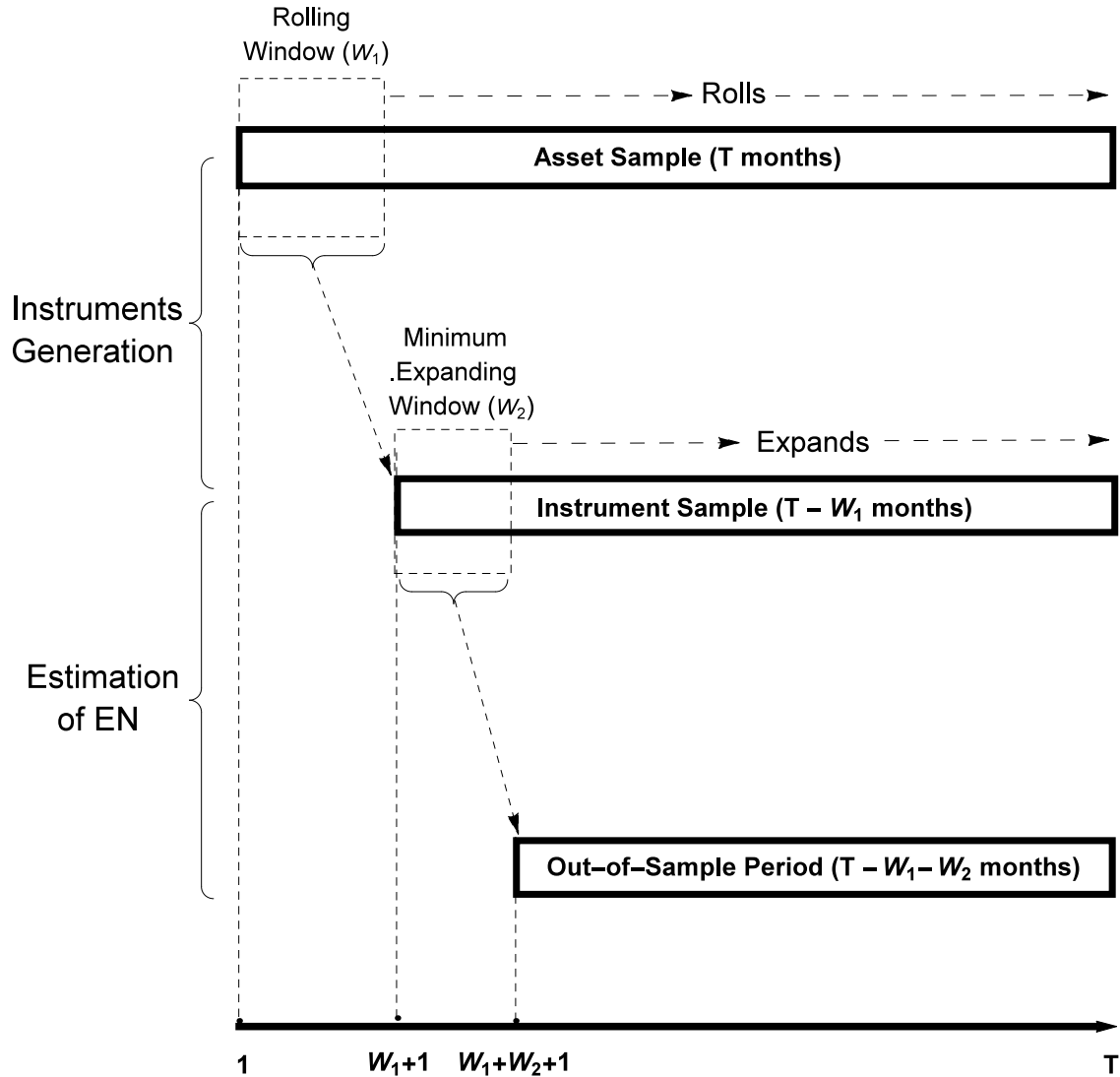
Strategy	Factor (N=4)	Factorlegs (N=7)	Size&BM +Factor (N=24)	Size&BM +Factorlegs (N=27)	Size&BM+Mom +Factorlegs (N=37)	Industry (N=11)	Volatility (N=11)
<b>Panel A: Monthly Sharpe Ratio</b>							
$EN$	0.292	0.260	0.315	0.347	0.377	0.156	0.205
$EN_{50\&50}$	0.289 (0.14)	0.260 (0.68)	0.311 (0.02)	0.347 (0.82)	0.377 (0.56)	0.156 (0.93)	0.206 (0.51)
$EN_{50\&100}$	0.291 (0.34)	0.260 (0.34)	0.314 (0.61)	0.346 (0.23)	0.377 (0.53)	0.156 (0.77)	0.205 (0.78)
$EN_{100\&50}$	0.292 (0.30)	0.261 (0.03)	0.314 (0.55)	0.347 (0.69)	0.377 (0.76)	0.156 (0.92)	0.205 (0.76)
$EN_{100\&150}$	0.290 (0.36)	0.261 (0.53)	0.313 (0.06)	0.347 (0.23)	0.377 (0.46)	0.158 (0.70)	0.207 (0.40)
$EN_{150\&100}$	0.292 (0.97)	0.261 (0.55)	0.315 (0.84)	0.347 (0.29)	0.377 (0.85)	0.156 (0.45)	0.205 (0.93)
$EN_{150\&150}$	0.289 (0.32)	0.261 (0.94)	0.314 (0.42)	0.347 (0.34)	0.377 (0.72)	0.160 (0.41)	0.206 (0.49)
<b>Panel B: Monthly CER</b>							
$EN$	0.421	1.072	1.647	1.608	1.816	0.383	0.678
$EN_{50\&50}$	0.417 (0.24)	1.068 (0.59)	1.613 (0.03)	1.609 (0.92)	1.817 (0.33)	0.380 (0.86)	0.684 (0.51)
$EN_{50\&100}$	0.421 (0.36)	1.069 (0.29)	1.644 (0.59)	1.608 (0.51)	1.816 (0.77)	0.383 (0.78)	0.678 (0.82)
$EN_{100\&50}$	0.421 (0.61)	1.078 (0.06)	1.640 (0.51)	1.610 (0.44)	1.817 (0.51)	0.383 (0.99)	0.681 (0.77)
$EN_{100\&150}$	0.419 (0.66)	1.073 (0.61)	1.631 (0.06)	1.608 (0.42)	1.815 (0.52)	0.392 (0.66)	0.689 (0.42)
$EN_{150\&100}$	0.421 (0.82)	1.073 (0.57)	1.648 (0.86)	1.608 (0.46)	1.816 (0.63)	0.382 (0.48)	0.678 (0.93)
$EN_{150\&150}$	0.419 (0.60)	1.072 (0.97)	1.639 (0.40)	1.608 (0.55)	1.816 (0.79)	0.401 (0.37)	0.687 (0.51)

**Table XI: Alternative Performance Measures for Cross Validation**

This table documents the performance of my approach under different risk aversion coefficients ( $\gamma = 4$ ,  $\gamma = 5$ ,  $\gamma = 7$ , and  $\gamma = 8$ ) for  $CER_{cv}$  calculation in cross validation. In all cases, the grid for both parameter  $\alpha$  and  $\lambda$  has 100 candidate values. Panel A of this table presents the Sharpe ratios and Panel B presents the CERs. The  $p$ -values is for testing the difference between performance in the baseline results and those in the alternative cases. The  $p$ -value for Sharpe ratio is calculated following to Jobson and Korkie (1981) after making correction pointed out by Memmel (2003) and the  $p$ -value for CER is calculated following Greene (2002).

Strategy	Factor (N=4)	Factorlegs (N=7)	Size&BM +Factor (N=24)	Size&BM +Factorlegs (N=27)	Size&BM+Mom +Factorlegs (N=37)	Industry (N=11)	Volatility (N=11)
<b>Panel A: Monthly Sharpe Ratio</b>							
$EN_{\gamma=6}$	0.292	0.260	0.315	0.347	0.377	0.156	0.205
$EN_{\gamma=4}$	0.290 (0.11)	0.254 (0.47)	0.308 (0.05)	0.338 (0.55)	0.378 (0.90)	0.144 (0.11)	0.209 (0.29)
$EN_{\gamma=5}$	0.291 (0.34)	0.258 (0.64)	0.312 (0.19)	0.340 (0.55)	0.378 (0.78)	0.148 (0.20)	0.208 (0.09)
$EN_{\gamma=7}$	0.292 (0.43)	0.261 (0.64)	0.315 (0.91)	0.332 (0.29)	0.373 (0.57)	0.160 (0.58)	0.199 (0.42)
$EN_{\gamma=8}$	0.293 (0.09)	0.255 (0.66)	0.315 (0.99)	0.328 (0.30)	0.370 (0.53)	0.170 (0.13)	0.210 (0.79)
<b>Panel B: Monthly CER</b>							
$EN_{\gamma=6}$	0.421	1.072	1.647	1.608	1.816	0.383	0.678
$EN_{\gamma=4}$	0.417 (0.08)	1.077 (0.95)	1.382 (0.11)	1.707 (0.38)	1.820 (0.84)	0.337 (0.15)	0.704 (0.26)
$EN_{\gamma=5}$	0.419 (0.25)	1.088 (0.73)	1.596 (0.49)	1.662 (0.51)	1.822 (0.74)	0.352 (0.25)	0.700 (0.08)
$EN_{\gamma=7}$	0.421 (0.33)	1.022 (0.42)	1.605 (0.45)	1.445 (0.05)	1.730 (0.08)	0.393 (0.69)	0.632 (0.26)
$EN_{\gamma=8}$	0.423 (0.04)	0.950 (0.15)	1.528 (0.23)	1.349 (0.02)	1.629 (0.01)	0.432 (0.20)	0.628 (0.63)

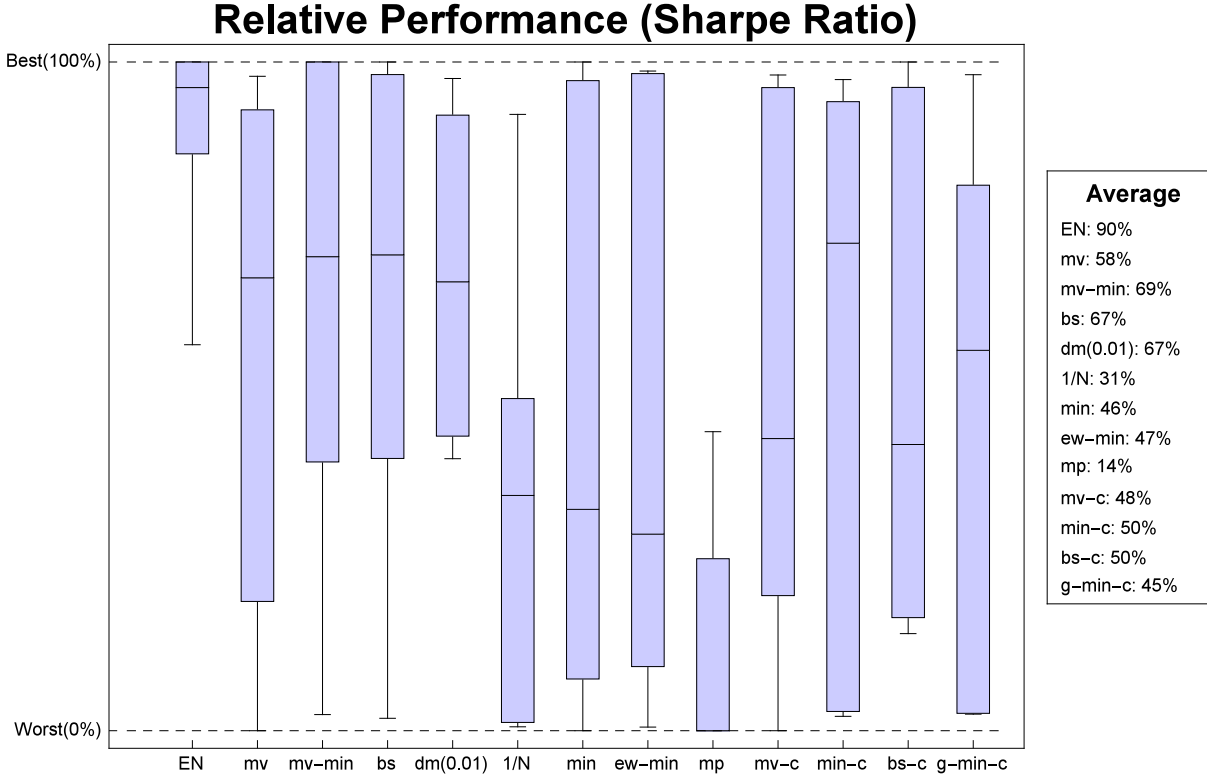
# Time Horizon



**Figure 1: Time Horizon of Implementation**

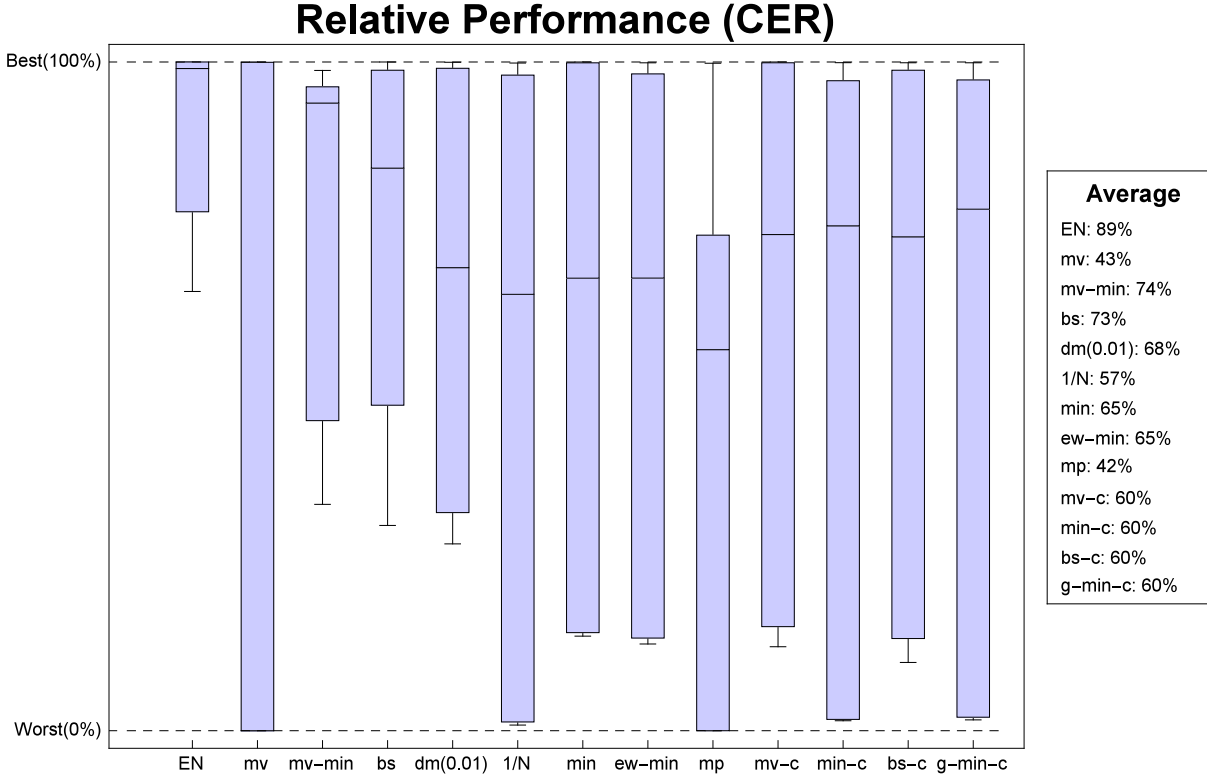
This figure illustrates the implementation of  $EN$  for a given asset sample. First, a sample of instruments ( $T - W_1$  months) is generated from the asset sample based on a rolling window of  $W_1$  months. Second, using the instrument sample,  $EN$  is implemented based on an expanding window with a minimum window length of  $W_2$  months. This procedure generates an out-of-sample period that has  $T - W_1 - W_2$  months of portfolio returns for performance evaluation.





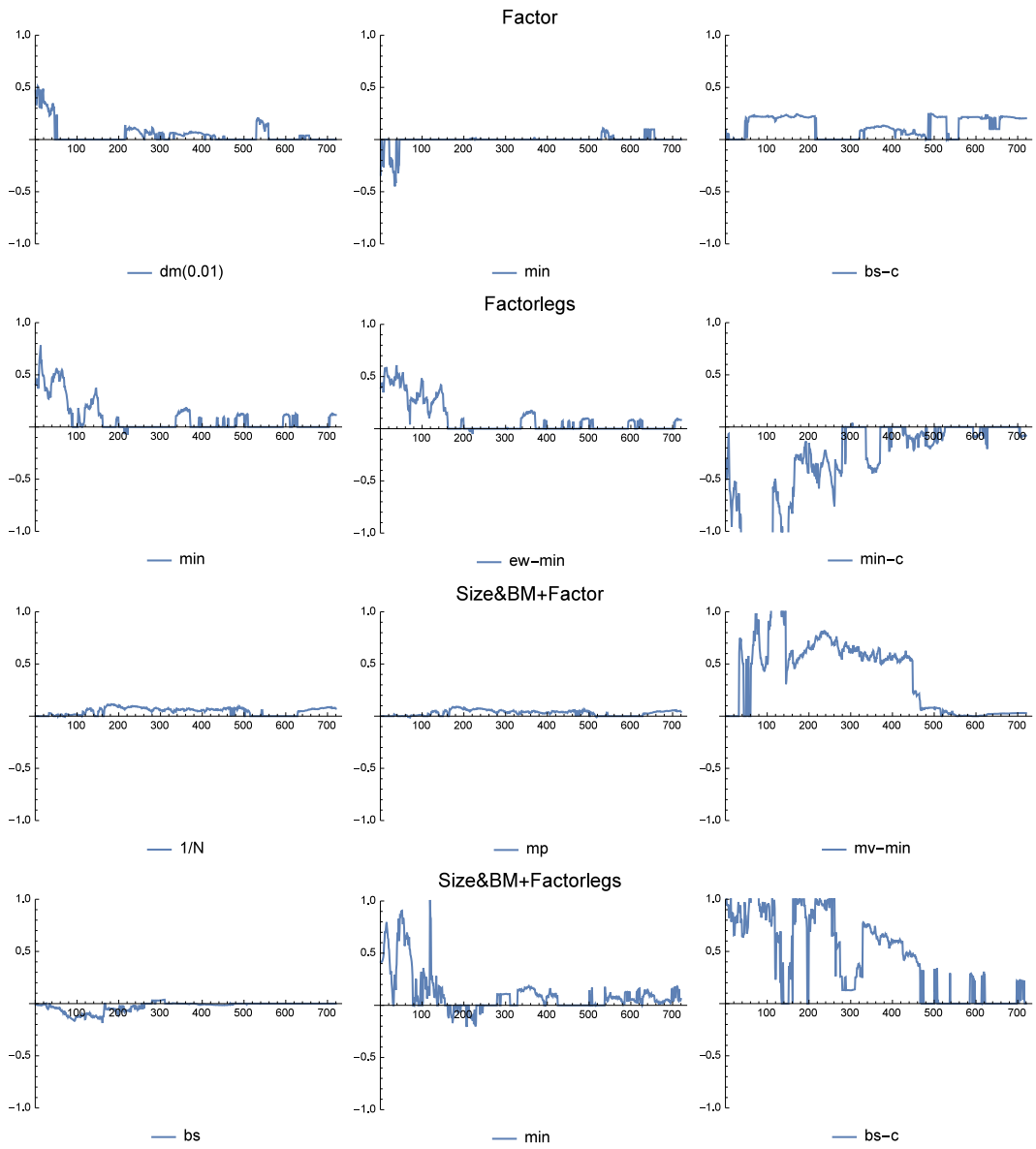
**Figure 2: Relative Performance (Sharpe Ratio)**

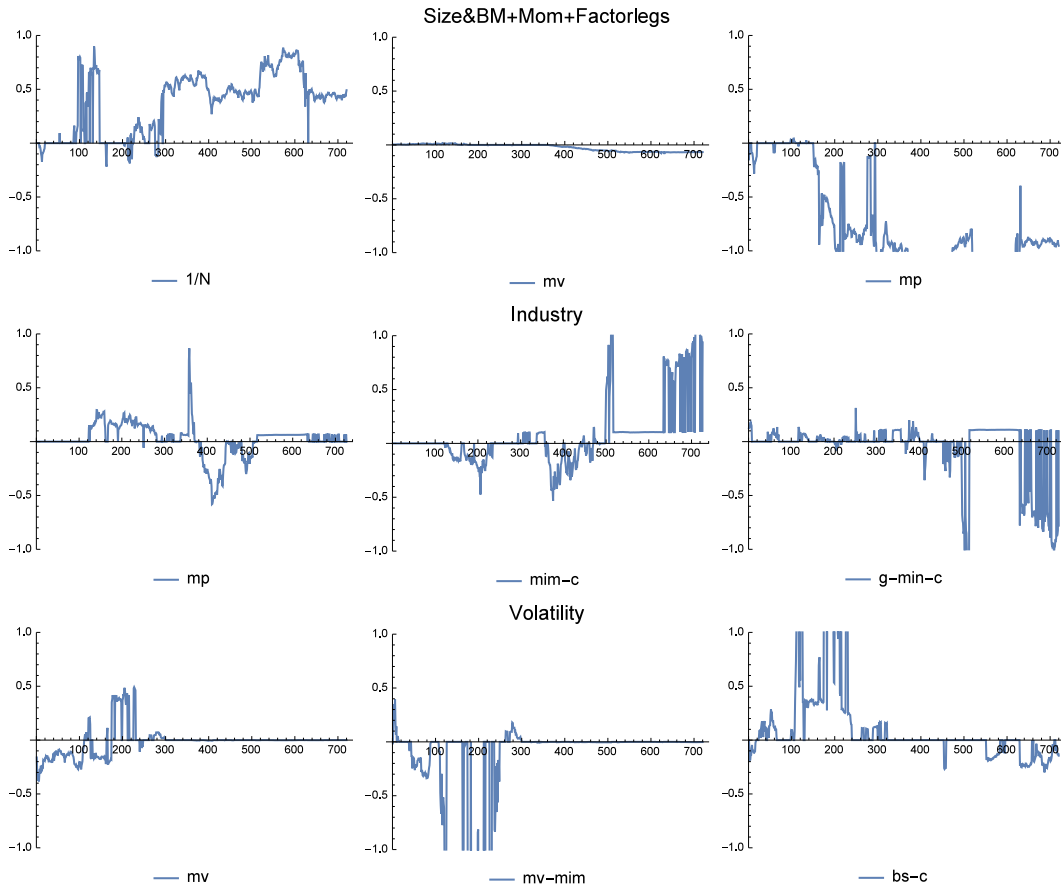
This figure presents the Box-Whisker plot of relative performance ( $RP$ ) for each allocation rule. Given an asset sample  $j$ , the relative performance of allocation rule  $i$  is calculated as:  $RP_j^i = (SR_j^i - SRmin_j)/(SRmax_j - SRmin_j)$ , where  $SR_j^i$  is the Sharpe ratio achieved by rule  $i$  in sample  $j$ ,  $SRmin_j$  ( $SRmax_j$ ) is the lowest (highest) Sharpe ratio achieved among all allocation rules in sample  $j$ . Each allocation rule has seven  $RP_j$ 's ( $j = 1, \dots, 7$ ) (since there are seven asset samples) and each box depicts the the lowest (lower bar), second to lowest (lower end of the rectangle), median (middle bar), second to highest (higher end of the rectangle), and the highest (upper bar) among the seven  $RP_j$ 's. The average  $RP$  for each allocation rule is listed on the right hand side.



**Figure 3: Relative Performance (CER)**

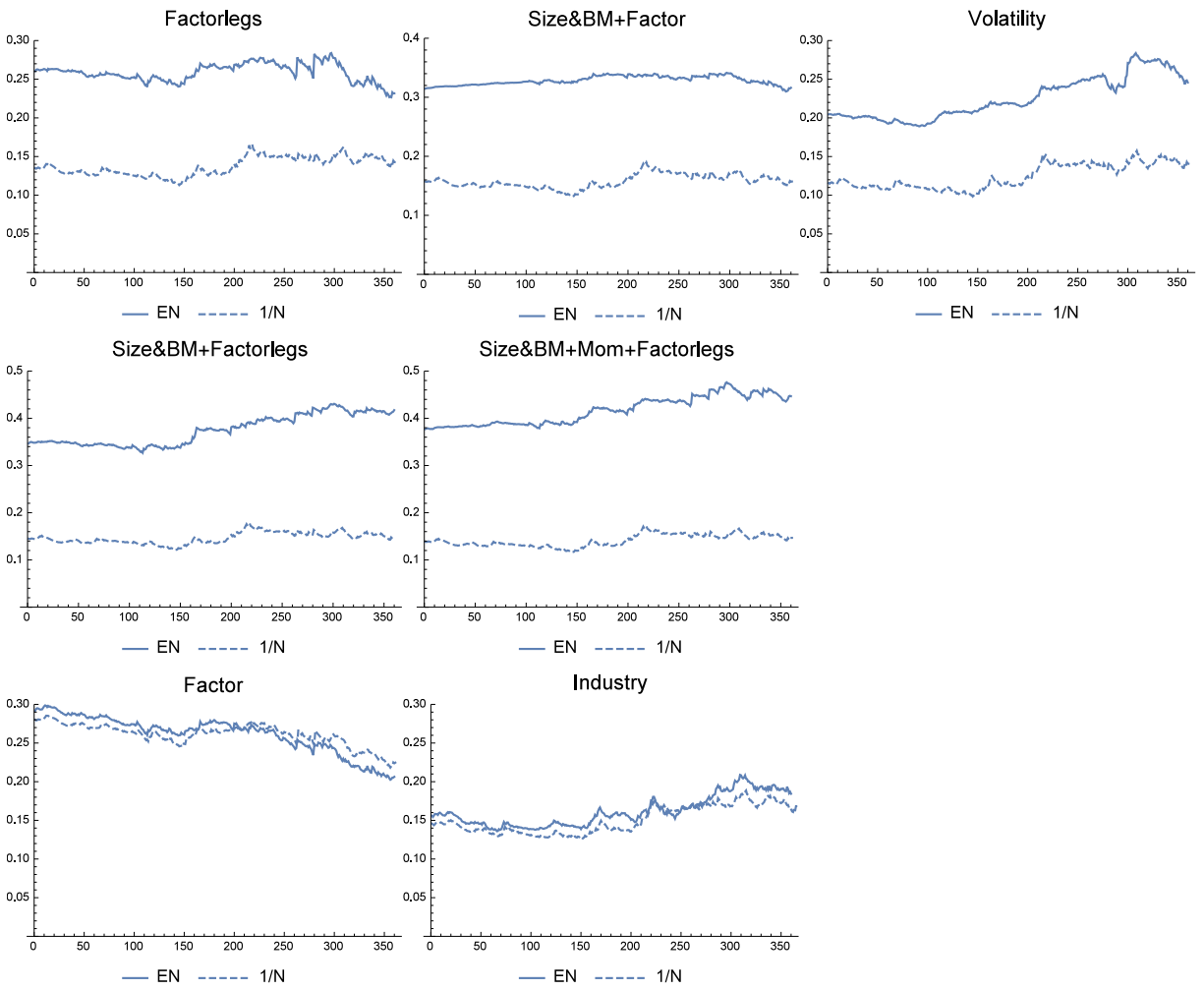
This figure presents the Box-Whisker plot of relative performance ( $RP$ ) for each allocation rule. Given an asset sample  $j$ , the relative performance of allocation rule  $i$  is calculated as:  $RP_j^i = (CER_j^i - CERmin_j)/(CERmax_j - CERmin_j)$ , where  $CER_j^i$  is the CER achieved by rule  $i$  in sample  $j$ ,  $CERmin_j$  ( $CERmax_j$ ) is the lowest (highest) CER achieved among all allocation rules in sample  $j$ . Each allocation rule has seven  $RP_j$ 's ( $j = 1, \dots, 7$ ) (since there are seven asset samples) and each box depicts the the lowest (lower bar), second to lowest (lower end of the rectangle), median (middle bar), second to highest (higher end of the rectangle), and the highest (upper bar) among the seven  $RP_j$ 's. The average  $RP$  for each allocation rule is listed on the right hand side.





**Figure 4: Time Series of  $\phi$  estimates**

This figure plots the time series of  $\phi$  estimates for all seven asset samples to show that the selection effect can indeed set the coefficient for some of the instruments to 0. For the purpose of clear demonstration, for each asset sample, I only plot the three instruments whose coefficient estimates are set to 0 most often.



**Figure 5: Sharpe Ratio v.s. Various Minimum Window**

This figure plots Sharpe ratio of  $EN$  and equally weighted portfolio against various minimum windows for all seven asset samples. The shortest minimum window is 120 months, which is associated with a out-of-sample of roughly 720 months and the longest is 480 months, which is associated with an out-of-sample of roughly 360 months.

**Internet Appendix**  
for  
**Aggregating Information for Optimal Portfolio  
Weights**

December 1 2018

Xiao Li  
University of Arizona

## A Instrument Calculation

The instruments (portfolio weights implied by each allocation rule) are calculated based on a rolling window following DeMiguel, Garlappi, and Uppal (2009) with correction of the optimal “three-fund” rule. One thing that has great influence on the quality of the instruments yet is still subject to change is the length of the rolling window. Since the purpose of my approach is to aggregate information, it is important to use a window that generates instruments with reasonable information quality. Based on simulation results, Kan and Zhou (2007) suggest a window of 240 months can be a reasonable candidate. In addition, I next will provide some empirical evidence.

I start out with the most commonly used window lengths in prior literature: 60 months, 120 months, and 240 months. Among the three window length candidates, I do not consider 60 month because prior literature in general has shown that it tends to generate poor performance and hence instruments with bad quality. For the other two candidates, I formally compare their validity based on how much wealth would an mean-variance efficient investor assign to a certain allocation rule vis-à-vis equally weighted portfolio, had the investor observed the performance of that rule in the entire sample. Intuitively, we might expect the investor to put more wealth on the portfolio implied by a certain rule (for all rules) when the estimation window is 240 months since longer estimation window tends to generate more accurate moment estimates as evidenced in prior literature. Here is a description of the procedure.

First, for each window length  $W_1 = 120$  and  $240$ , generate a time series of portfolio weights for each allocation rule. Next, for each rule ( $i = 1, 2, \dots, 11$ ), estimate  $\hat{\alpha}^i$  and  $\hat{\beta}^i$  by

solving the minimization problem

$$\underset{(\alpha^i, \beta^i)}{\text{Min}} \sum_{t=1}^T (1 - X_t(\alpha^i + \beta^i w_t^i))^2 \quad (\text{A1})$$

using the entire time series of weights generated in the first step. Third, calculate the relative weight assigned to allocation rule  $i$  as  $w_r^i = \frac{\hat{\beta}^i}{|\hat{\beta}^i + \hat{\alpha}^i|}$ . Finally, for each allocation rule, compare the relative weights across the two window length candidates and pick the window length that has, on average, higher relative weight as the window for instrument calculation.

Asset samples that include the UMD factor (or long short legs) have 960 and 840 periods of portfolio weights for window length of 120 months and 240 months and those that do not include the UMD factor (or long short legs), have 966 and 846. Table A.I reports the relative weights  $w_r^i$  for all 11 portfolio allocation rules across 7 asset samples. Consistent with our expectation, we observe a roughly monotonic increase in relative weights for almost all allocation rules as the window length extends, which is likely to result from more accurate moments estimation. Therefore, to obtain instruments that are likely to contain better information, I pick  $W_1 = 240$  as the window length for instrument calculation. One glaring exception is the “mp” method, whose weight becomes extremely negative as the window length increases. This is largely due to the fact that, for the majority of time, “mp” method mimics the equal weighted portfolio fairly closely (the time series of portfolio returns for the “mp” rule and the equally weighted portfolio have a correlation of over 0.9) except for two months where the equal weighted portfolio realized a big positive return but the mp method took a short position on similar weights and therefore realized a negative return of comparable size. In this situation, since the big negative return comes at a relatively late position in the time series, more observations are cut as one extends the estimation window so that the impact of those two extreme observations becomes larger and hence the ever more extreme negative weights assignment for window length of 240 months. This observation also



highlights the pertinence of regularization on those coefficients as discussed in Section 2.3.

**Table A.I: Relative Weights of Allocation Rule**

This table presents the relative weights assigned to the each allocation rule by a mean-variance efficient investor had he observed the entire time series of portfolio returns of each allocation rule across 2 rolling window  $W_1 = 120$  and 240. Seven asset samples are included that are Factor (N=4), Factorlegs (N=7), Size&BM+Factor (N=24), Size&BM+Factorlegs (N=27), Size&BM+Mom+Factorlegs (N=37), Industry (N=11), Volatility (N=11). Portfolio allocation rules included here are: equally weighted portfolio, sample based mean-variance rule, Bayes-Stein rule, Data-and-Model rule, sample based minimum variance rule, missing-factor (MacKinlay and Pastor (2000)), optimal “three fund” rule (Kan and Zhou (2007)), mixture of minimum variance and naive rule, sample based mean-variance with positivity constraint, Bayes-Stein with positivity constraint, minimum variance with positivity constraint, and minimum variance rule with generalized constraints.

Method	Window	Factor (N=4)	Factorlegs (N=7)	Size&BM +Factor (N=24)	Size&BM +Factorlegs (N=27)	Size&BM+Mom +Factorlegs (N=37)	Industry (N=11)	Volatility (N=11)
mv	120	0.015	-0.001	0.002	-0.006	0.031	0.019	0.001
	240	0.661	0.195	0.149	0.220	0.084	0.259	0.015
bs	120	0.148	0.000	0.003	-0.011	0.068	0.028	0.002
	240	0.740	0.295	0.175	0.413	0.183	0.370	0.021
dm	120	0.013	0.188	0.631	0.014	0.002	0.034	0.193
	240	0.050	1.537	0.163	0.247	0.252	0.658	2.193
min	120	0.397	0.556	-1.226	0.726	0.696	0.572	0.693
	240	0.570	0.725	-1.570	0.938	1.002	0.571	0.744
mp	120	0.000	-0.104	-0.976	-0.654	-0.352	-0.237	-0.507
	240	-0.063	-15.876	-5.087	-15.949	-14.03	-5.430	-5.701
mv-min	120	0.193	-0.001	0.003	-0.006	0.148	0.027	0.005
	240	0.727	0.315	0.177	0.663	0.433	0.407	0.023
ew-min	120	0.467	0.616	-1.227	0.858	0.883	0.657	0.756
	240	0.627	0.787	-1.566	1.091	1.248	0.602	0.791
mv-c	120	0.561	3.727	0.973	2.679	2.167	0.022	0.662
	240	0.669	4.137	0.964	3.720	2.644	0.041	1.135
min-c	120	0.436	1.238	0.955	1.045	0.938	0.698	1.178
	240	0.574	1.406	0.977	0.934	1.281	0.617	1.121
bs-c	120	0.645	4.215	0.976	2.103	1.759	0.179	1.081
	240	0.740	4.130	0.966	3.499	2.410	0.069	1.121
g-min-c	120	0.608	2.142	1.313	1.865	1.129	1.057	2.300
	240	0.711	2.856	1.430	1.457	2.080	0.918	2.216

## B Determining the Minimum Expending Window

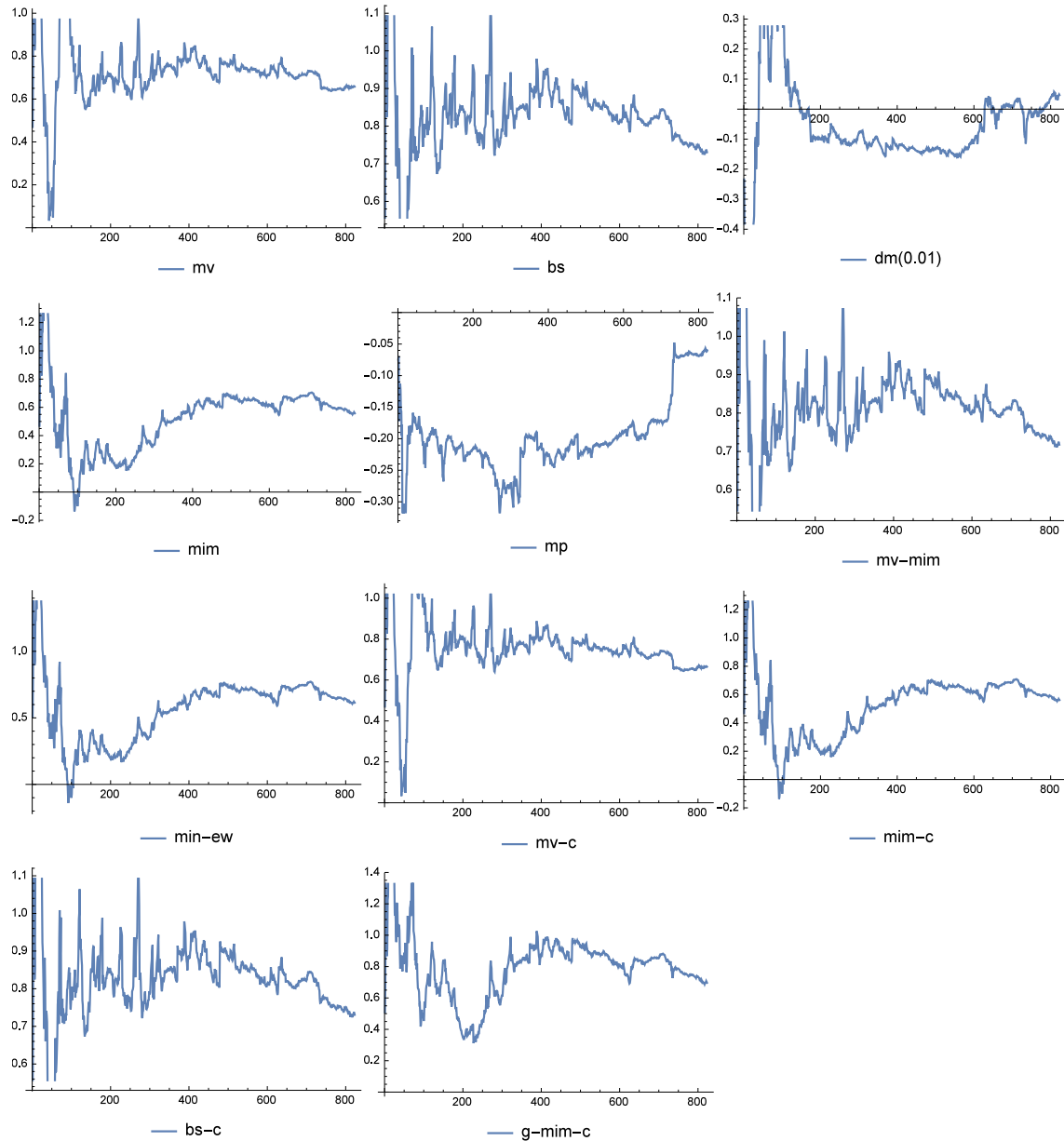
Recall that  $EN$  is implemented using an expanding window (i.e., at each period, all observed instruments are used.), therefore we need to determine the minimum window length. On the one hand, we need more data and hence a longer minimum window to conduct meaningful and reliable estimation. On the other hand however, we want shorter minimum window since it is also important to keep a longer out-of-sample time series in order to increase the reliability of the out-of-sample tests. To keep a good balance between those two goals, the minimum window is picked as the shortest window after which the relative weight  $w_r$  starts to stabilize.<sup>15</sup> That is, for each rule, we calculate a series of  $w_r$  using ever expanding windows, plot those  $w_r$ 's against the corresponding window length, and pick the shortest window where the plot starts to stabilize. Figure B.1 through Figure B.7 illustrate those plots for the seven asset sample, respectively.

For all seven asset samples, the shortest estimation window has 12 months of instruments and the longest has 834 for Industry and Volatility and 828 for the other five asset samples. There are usually some extreme  $w_r$ 's in all plots, especially when the estimation window is short. Since plotting these extreme numbers tends to make the entire graph look almost constant while featuring a large spike, all plots are cut off at some point on the vertical axis. In general, when estimation window is short, relative weights tend to be highly volatile. Stabilization of  $w_r$ 's starts to appear differently for different allocation rules, across different asset samples. Some allocation rules (e.g., bs-c and mv-c) only start to see relatively stable  $w_r$  in the far later part of the asset sample. However, a good portion of the allocation rules start to see relatively stable  $w_r$  after around the first 100 months. Therefore, I choose  $W_2 = 120$  as the baseline minimum window as it is a commonly window length in the literature. For robustness concern, the performance of  $EN$  using alternative minimum window lengths is

---

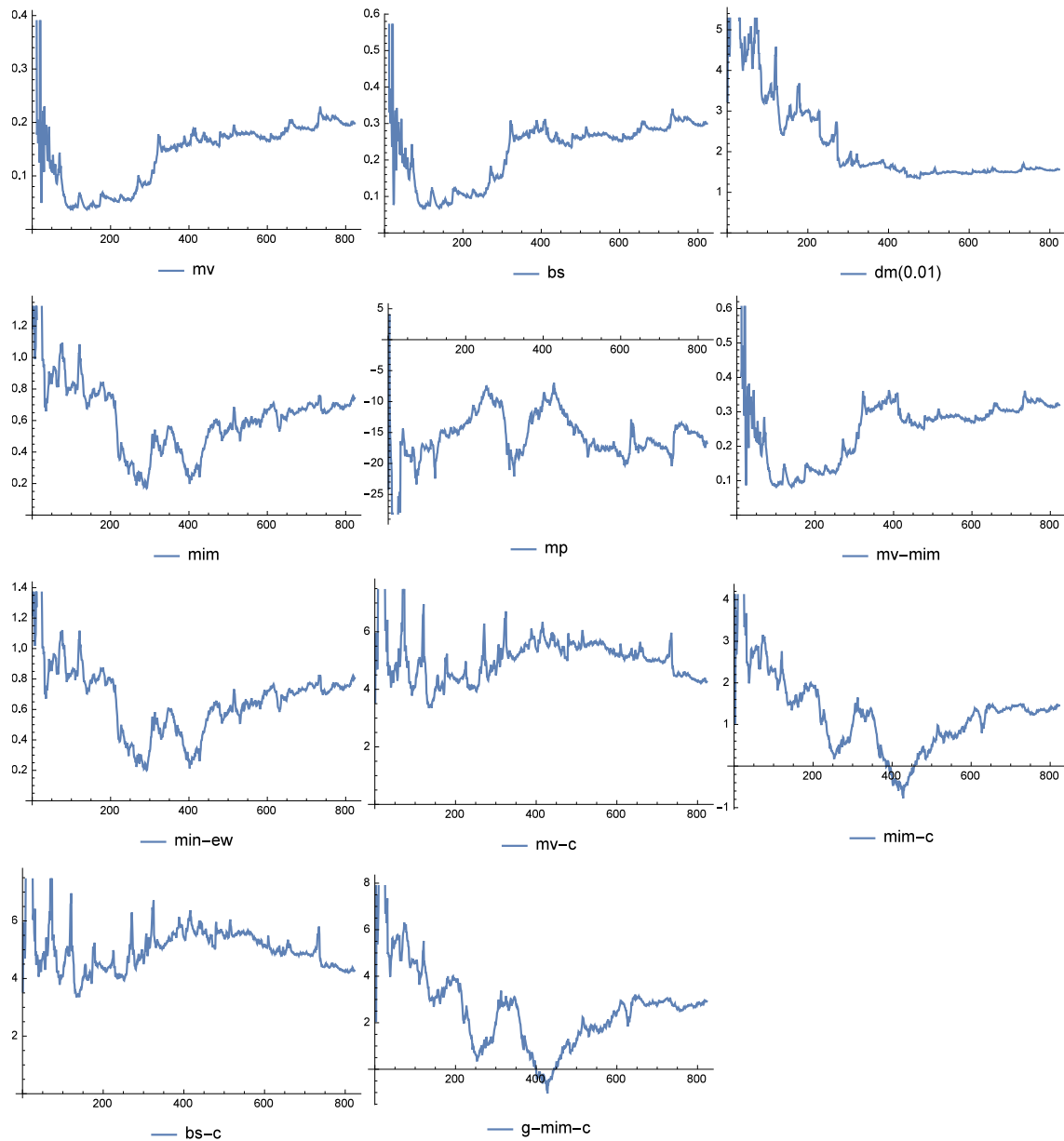
<sup>15</sup>In Appendix A,  $w_r$  is calculated using the entire time series of portfolio return, whereas in this section,  $w_r$  is calculated using various window lengths.

explored in Section (5).



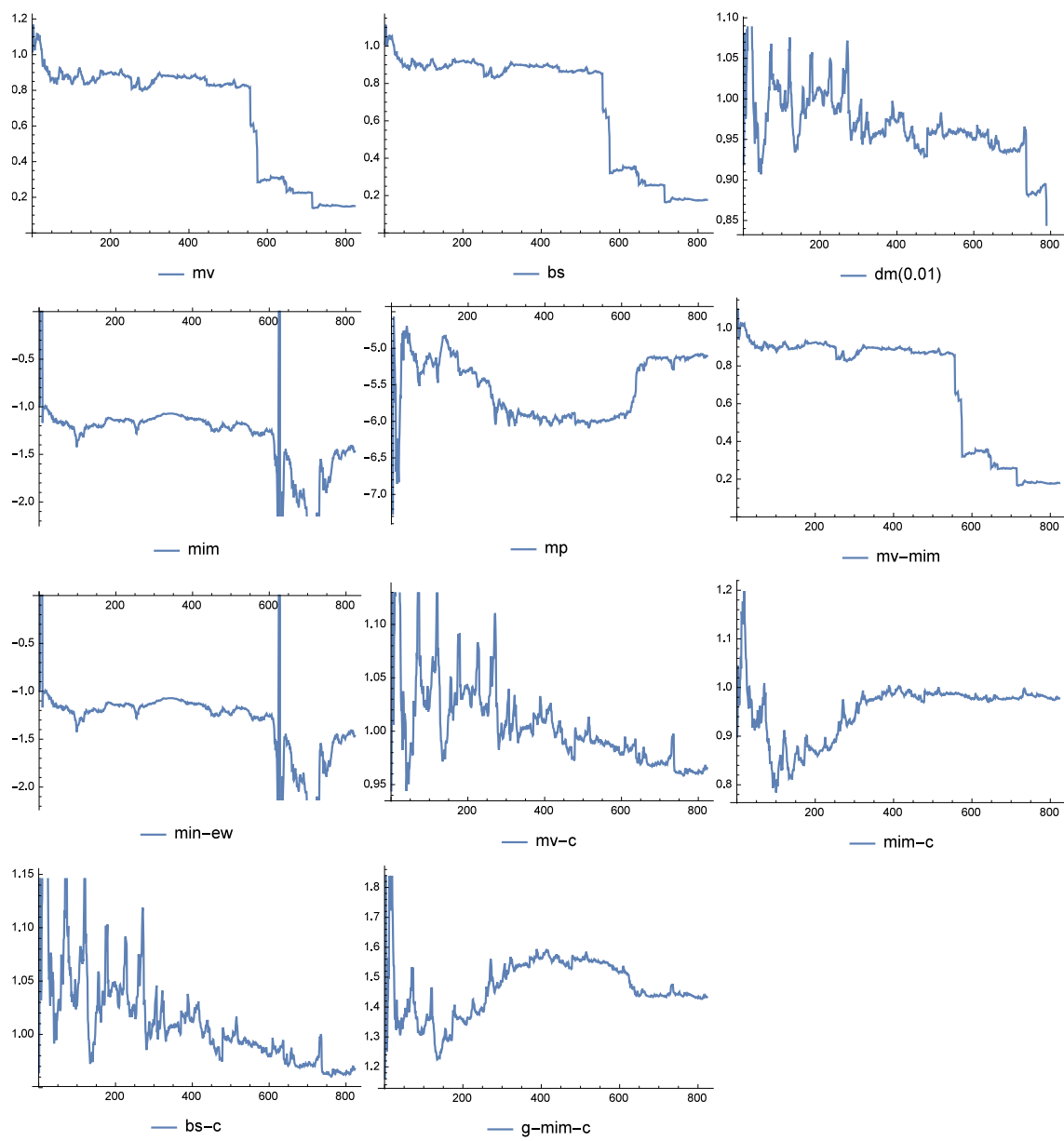
**Figure B.1: Relative Weights v.s. Estimation Window (Factor)**

This figure plots relative weights against various estimation window lengths for the Factor sample. The shortest window length has 12 months and the longest has 828 months.



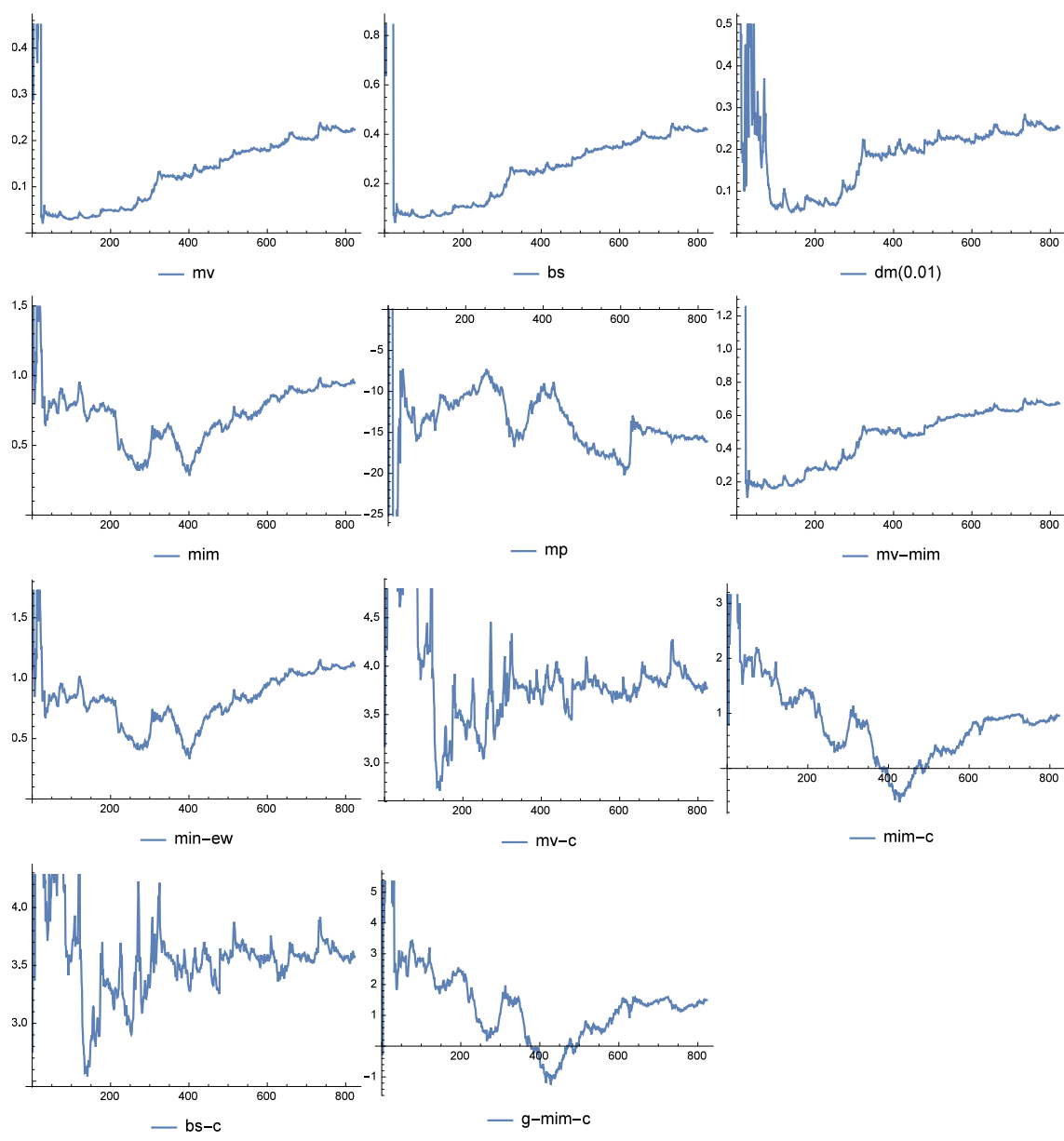
**Figure B.2: Relative Weights v.s. Estimation Window (Factorlegs)**

This figure plots relative weights against various estimation window lengths for the Factorlegs sample. The shortest window length has 12 months and the longest has 828 months.



**Figure B.3: Relative Weights v.s. Estimation Window (Size&BM+Factor)**

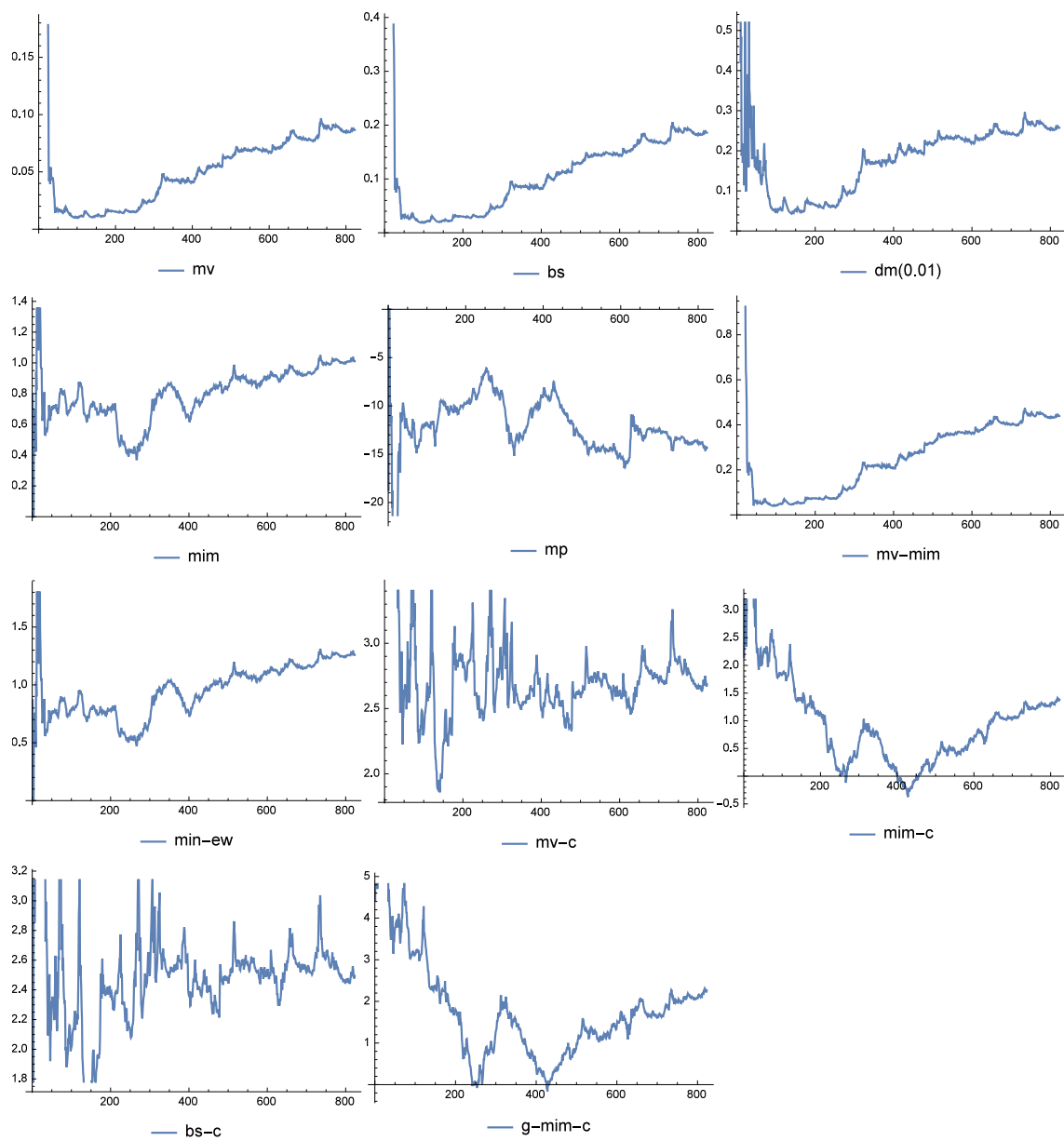
This figure plots relative weights against various estimation window lengths for the Size&BM+Factor sample. The shortest window length has 12 months and the longest has 828 months.



**Figure B.4: Relative Weights v.s. Estimation Window (Size&BM+Factorlegs)**

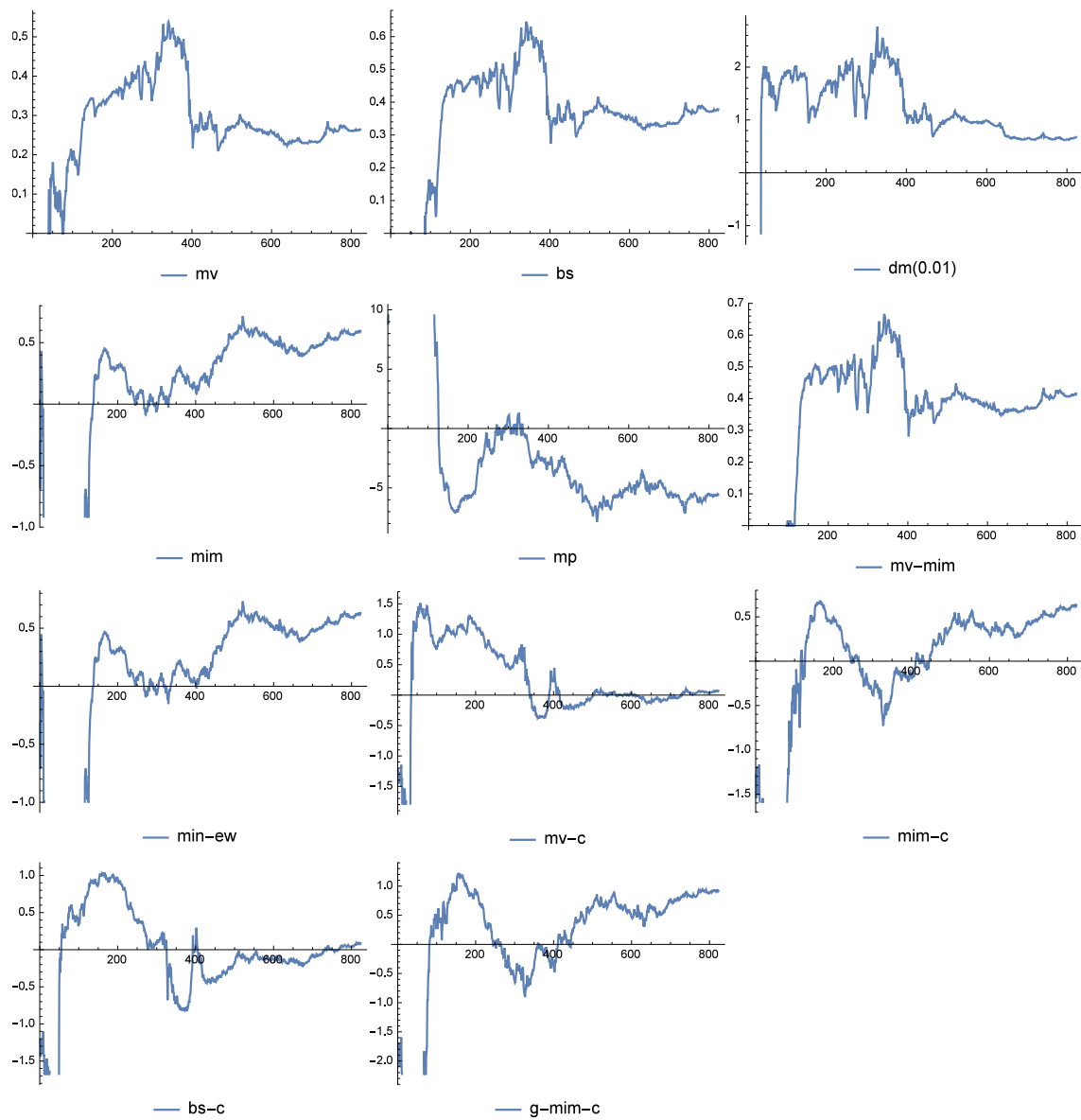
This figure plots relative weights against various estimation window lengths for the Size&BM+Factorlegs sample. The shortest window length has 12 months and the longest has 828 months.





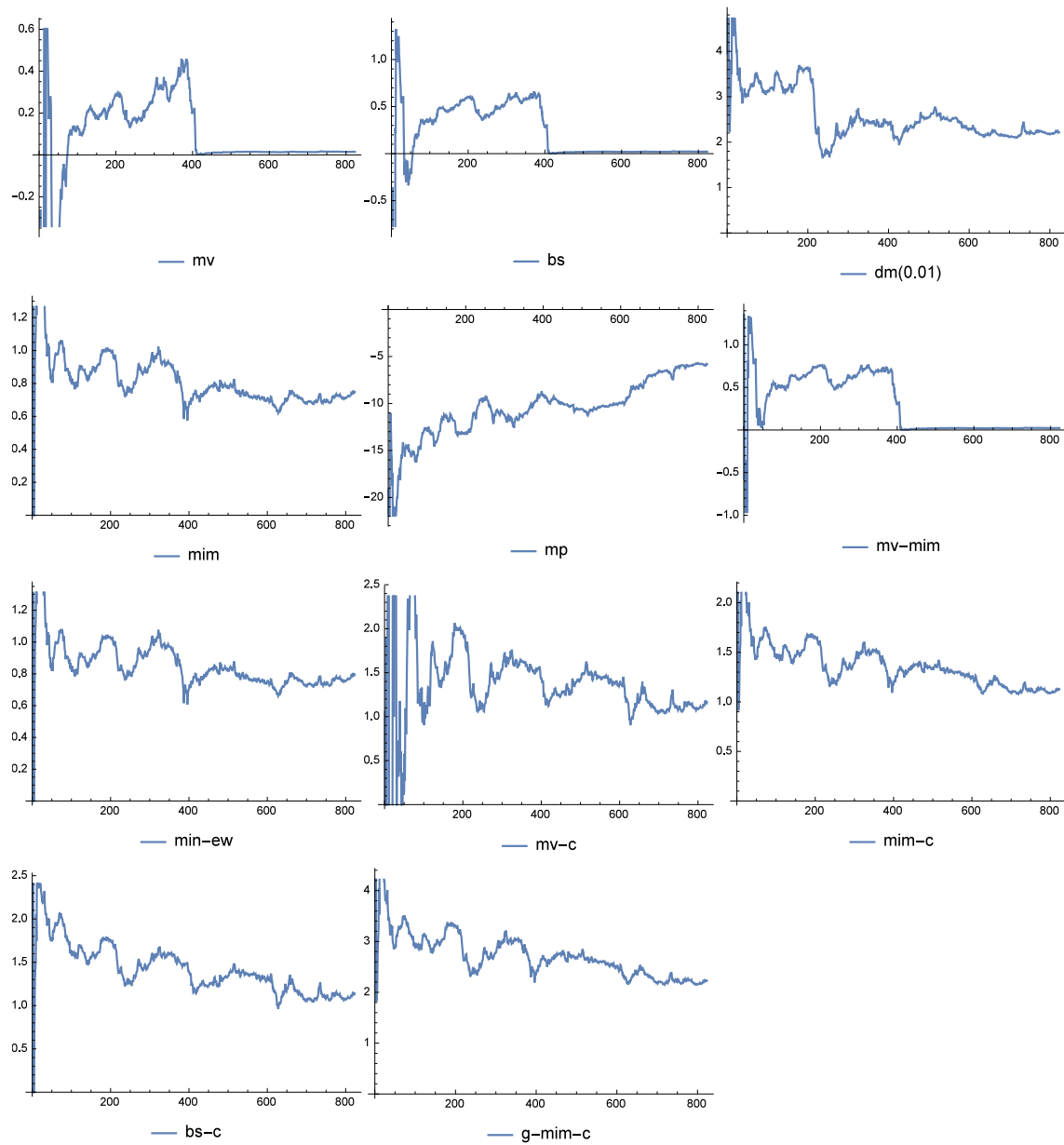
**Figure B.5: Relative Weights v.s. Estimation Window (Size&BM+Mom+Factorlegs)**

This figure plots relative weights against various estimation window lengths for the Size&BM+Mom+Factorlegs sample. The shortest window length has 12 months and the longest has 828 months.



**Figure B.6: Relative Weights v.s. Estimation Window (Industry)**

This figure plots relative weights against various estimation window lengths for the Industry sample. The shortest window length has 12 months and the longest has 834 months.



**Figure B.7: Relative Weights v.s. Estimation Window (Volatility)**

This figure plots relative weights against various estimation window lengths for the Volatility sample. The shortest window length has 12 months and the longest has 834 months.

## C Coefficient Estimation and the Selection Effect

There is no analytic solution for a regression with Elastic Net penalty. The algorithm used in this paper for calculating the numerical solution is cyclical coordinate descent, which successively optimizes the objective function over each coefficient with others fixed, and cycles repeatedly until convergence. Specifically, at each iteration, the update of coefficient for instrument  $j$  is given by the following equation:

$$\hat{\phi}^j \leftarrow \frac{S(\frac{1}{T^*-1} \sum_{t=1}^{T^*-1} r_{t,j}^*(y_t - \hat{y}_t^j), \lambda\alpha)}{1 + \lambda(1 - \alpha)} \quad (\text{C2})$$

where  $r_{t,j}^*$  is the standardized portfolio return of instrument  $j$ ,  $\hat{y}_t^j = \hat{\phi}^0 r_{t,0}^* + \sum_{k \neq j} \hat{\phi}^k r_{t,k}^*$  and  $S(z, \gamma)$  is the soft-thresholding operator with value  $\text{sign}(z)(|z| - \gamma)_+$ . The function  $\text{sign}(z)$  is equal to 1 if  $z$  is positive and -1 if  $z$  is negative, and the function  $(x)_+$  equals to  $x$  if  $x$  is positive and 0 otherwise.

It is straightforward to see how parameters  $\lambda$  and  $\alpha$  grant the selection effect by serving as the threshold variable for function  $S(\cdot, \cdot)$  in Equation (C2). Inside this function, the first variable  $\sum_{t=1}^{T^*-1} r_{t,j}^*(y_t - \hat{y}_t^j)$  measures the information content of each instrument and the second variable  $\lambda\alpha$  is the threshold that discriminates the information. By the definition of the soft-thresholding function, only instruments whose information content is higher than the threshold can have nonzero coefficients. As the value of  $\lambda\alpha$  varies from small to large, the number of instruments that enter the model decreases which corresponds to an ever selective attitude towards information content.

Unfortunately, it is difficult to generate any general conclusion regarding which instrument enters the model (i.e., having a non-zero coefficient) since the coefficient of instruments are interdependent. To provide some intuition, I will focus on simple situation where the parameter of  $\lambda$  is set high enough so that only one instrument enters the model. In such case, the information content variable  $\sum_{t=1}^{T^*-1} r_{t,j}^*(y_t - \hat{y}_t^j)$  degenerates to  $\sum_{t=1}^{T^*-1} r_{t,j}^* y_t$  since

all other coefficients are 0. Also note that  $y_t = 1$  for all observations, therefore  $\sum_{t=1}^{T^*-1} r_{t,j}^* y_t$  gives the Sharpe ratio (up to a scalar  $\frac{1}{T^*-1}$ ) for allocation  $j$ . As a result, when only one instrument is allowed into the model, we pick the rule that realizes the highest Sharpe ratio.

## D Cross Validation

There are two parameter  $\alpha$  and  $\lambda$  that need to be calibrated before calculating out-of-sample weights. However, before we actually implement the cross validation process, we need to first determine a grid of  $\alpha$  and  $\lambda$  candidate values from which we can select the optimal pair. Therefore, in the next few paragraphs I will first explain how I construct the grids and next how to implement the cross validation process.

Constructing a grid of candidate values involves i) pinning down a maximum and a minimum value for the parameter and ii) selecting a series of values between the maximum and minimum inclusively as parameter candidates for the cross validation. The range for  $\alpha$  is always between 0 and 1 ( $\alpha \in (0, 1)$ ) and the maximum and minimum value for  $\alpha$  are set to be 0.99999 and 0.00001 in order to preserve the effect from both the  $l^1$  and  $l^2$  regularization terms. The grid of  $\alpha$  is constructed by joining two pieces of grids above and below 0.5. For the piece below 0.5, I follow the common practice in the machine learning literature and select Q evenly-spaced  $\alpha$ 's between 0.00001 and 0.5 on the log scale. That is, I take Q evenly-spaced  $\log(\alpha)$ 's between  $\log(0.00001)$  and  $\log(0.5)$  and transform them back to the original scale. This practice will leave a series of  $\alpha$  values that are more densely populated towards 0. The above 0.5 grid is conducted in a similar fashion but with the more densely populated part happening towards 1. The final grid is the union of the  $\alpha$  values from both pieces with 2Q  $\alpha$ 's that are more densely populated towards the two ends. This density structure is to accommodate the fact that empirical evidence suggests that, for the majority of the time, the truly useful  $\alpha$  are either around 0 or around 1.

The range of  $\lambda$  is data-dependent. In particular, at each period, I follow Friedman, Hastie, and Tibshirani (2010) and set the largest  $\lambda_{max}$  as the smallest  $\lambda$  such that the coefficients of all independent variables are zeros. This practice implies that  $T\alpha\lambda_{max} = \max_l |x_l \cdot y|$ , where  $a \cdot b$  is the inner product of the two vectors. In other words,  $\lambda_{max}$  is the absolute value

of the largest inner products among the independent variables  $x_l$  and  $y$ , adjusted by the product of the number of observations  $T$  and  $\alpha$ . The minimum value  $\lambda_{min}$  is set to be  $\epsilon\lambda_{max}$ , where  $\epsilon$  is a small positive value. As suggested in Friedman, Hastie, and Tibshirani (2010),  $\epsilon = 0.001$  through out the entire cross validation.  $\epsilon = 0.001$  is also small enough so that when  $\lambda = \lambda_{min}$ , all independent variables have non-zero coefficients through the entire time across all seven asset samples. The grid between  $\lambda_{min}$  and  $\lambda_{max}$  is constructed by taking  $P$  evenly spaced values in the log scale. That is, we take  $P$  evenly spaced value between  $\log(\lambda_{min})$  and  $\log(\lambda_{max})$  and transform these values back to the original scale. So at each period, there are  $2Q \times P$  pairs of  $\alpha$  and  $\lambda$  that need to be determined by the cross validation process which I will explain next.

Since there are two parameters to calibrate, the cross validation is conducted in a two round fashion, inner round and outer round. In the inner round, the  $\alpha$  is fixed, we optimally select a  $\lambda$  through a hold-one-out process as follows. Suppose we have  $T$  periods of instruments. First, for the first  $\lambda$ , we sequentially leave the first through the last period of instruments out and used the rest of the  $T - 1$  periods of instruments to do the estimation as in equation (14) and calculate the portfolio weights as given in equation (15). These weights are then used to calculate the portfolio return for each left-out period, which generates a time series of portfolio returns for the first  $\lambda$ . Second, we repeat the first step for all  $P$   $\lambda$ 's and generate  $P$  time series of portfolio returns. Third, for each time series, calculate the CER and pick the lambda that generated highest CER as the optimal  $\lambda$  for the fixed  $\alpha$ , which finishes the inner round. In the outer round, we select an optimal  $\alpha$  according to the performance (CER) of the  $\lambda$  associated with it and thus we have a pair of  $\alpha$  and  $\lambda$  to estimate the out-of-sample weights. Note that, since we are using an expanding window and the optimal parameter pair is updated monthly, the cross validation process is taking longer time in later periods than earlier periods. In the baseline model,  $Q$  is set to be 50 and  $P$  is set to be 100, and therefore there are 100  $\alpha$ 's and  $\lambda$ 's respectively. In robustness tests,

alternative value of  $Q$  and  $P$  are used.