Optimal Portfolio Choice with Estimation Risk: No Risk-free Asset Case

Raymond Kan, Xiaolu Wang and Guofu Zhou*

JEL classification: G11; G12; C11

Keywords: Portfolio choice; estimation risk; mean-variance optimization; optimal combining.

Current version: November, 2020

*Kan is from the University of Toronto, Wang is from Iowa State University and Zhou is from Washington University in St. Louis and CAFR. Corresponding author: Raymond Kan, Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6; Tel: (416) 978-4291; Fax: (416) 978-5433; email: raymond.kan@rotman.utoronto.ca. We thank Victor DeMiguel, Lorenzo Garlappi, Steven Kou, Thomas Maurer, Tim Simin, Thomas Tang, Ghoshi Tong, Raman Uppal, Michael Wolf, Chu Zhang, and seminar participants at Australian National University, Chinese University of Hong Kong, Durham University, East China Normal University, Hong Kong University of Science and Technology, Lancaster University, Nankai University, Shanghai Advanced Institute of Finance, Southwestern University of Finance and Economics, Tianjin University, Tsinghua University School of Economics and Management, University of Texas at Dallas, Washington University in St. Louis and participants at the 2nd Financial Econometrics and Risk Management Conference at Western University, the 9th Annual SoFiE Conference, the 2016 China International Conference in Finance, the 2016 China Finance Review International Conference, the 10th Annual Risk Management Conference at National University of Singapore, the 2016 Northern Finance Association Conference, and the 29th Australasian Finance and Banking Conference for helpful comments. We are especially grateful to Tyler Shumway (the Department Editor), an Associate Editor and two anonymous referees for their insightful and detailed comments that have helped improve the paper substantially. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada.

Optimal Portfolio Choice with Estimation Risk: No Risk-free Asset Case

Abstract

We propose an optimal combining strategy to mitigate estimation risk for the popular mean-variance portfolio choice problem in the case without a risk-free asset. We find that our strategy performs well in general, and it can be applied to known estimated rules and the resulting new rules outperform the original ones. We further obtain the exact distribution of the out-of-sample returns and explicit expressions of the expected out-of-sample utilities of the combining strategy, providing not only a fast and accurate way of evaluating the performance but also analytical insights into the portfolio construction.

1. Introduction

Since Markowitz's (1952) seminal paper, the mean-variance framework has been the major model used in practice in asset allocation and active portfolio management.[1] One main reason is that many implementation issues, such as factor exposures and trading constraints, can be easily accommodated within this framework, which allows for analytical insights and fast numerical solutions. Another reason is that the intertemporal hedging demand is typically found to be small. However, to implement the mean-variance optimal portfolio, we need to know the true values of asset means, variances, and covariances, which are unavailable in practice and must be estimated from historical data. Treating estimated parameters as true parameters, the plug-in method typically results in a substantial deterioration in out-of-sample portfolio performance. This is the well-known estimation risk problem.

Various strategies have been proposed in the literature to address the estimation risk problem. Early works by Brown (1976) and Bawa, Brown, and Klein (1979) show that the plug-in method is generally outperformed by the Bayesian decision rule under a diffuse prior. Jorion (1986, 1991) proposes the use of a Bayes-Stein estimator instead of the sample mean. Ledoit and Wolf (2004, 2017) recommend the use of shrinkage covariance matrix estimators in place of the sample covariance matrix. MacKinlay and Pástor (2000) assume a single factor structure and exploit the implications of an asset pricing model to parameter estimation. Jagannathan and Ma (2003) suggest imposing no-short-sale constraints in estimating optimal portfolios. Due to the poor performance of the estimated optimal portfolios, some studies focus on the global minimum variance (GMV) portfolio, which ignores the means.[2] DeMiguel, Garlappi and Uppal (2009) show that the $1/N$ rule, which does not use any sample information thus without estimation errors, can outperform various estimated optimal portfolios in many empirical datasets. Kirby and Ostdiek (2012) propose two timing strategies, which use sample variances and sample means (but not sample covariances) in a way different from the usual optimization framework.

Theoretically, the optimal strategy is to design a portfolio rule to minimize the utility loss

[1]See, e.g., Grinold and Kahn (1999), Litterman (2003), Meucci (2005), Qian, Hua, and Sorensen (2007) for practical applications of the mean-variance framework.

[2]For example, Jobson, Korkie and Ratti (1979), Jagannathan and Ma (2003), Kempf and Memmel (2006), Basak, Jaganathan and Ma (2009), and Bodnar, Parolya and Schmid (2018).

resulted from estimation risk. Along this line, Kan and Zhou (2007) develop an optimal three-fund rule that maximizes the expected out-of-sample utility in the case where a risk-free asset is available. Following similar theoretical optimization framework, Tu and Zhou (2011) further improve the three-fund rule of Kan and Zhou (2007) by combining it with the $1/N$ rule.

For the case without a risk-free asset, similar portfolio rules are not available in the literature due perhaps to the difficulties of this problem.[3] Note that simply normalizing the optimal portfolios derived in the case with a risk-free asset (e.g., the three-fund rule of Kan and Zhou (2007)) so that weights on risky assets add up to one does not generate optimal portfolios for the no risk-free asset case.[4] The no risk-free asset case is, however, important. For example, most equity funds are required to be fully invested in risky assets, which is a portfolio selection problem in the no risk-free asset case.[5]

In this paper, we focus on the case without a risk-free asset, and make several contributions. First, for the no risk-free asset case, there is a lack of studies on optimal portfolio rules that explicitly take into account estimation risk, and we provide the first such a rule here. We develop new analytical methods which enables us to obtain the exact distribution of the out-of-sample returns as well as the explicit expression of the expected out-of-sample utility of the plug-in rule. Given these, we are able to derive an optimal portfolio rule that minimizes the utility loss under estimation risk, and the rule takes the form of optimally combining the sample GMV portfolio with a sample zero-investment portfolio.

In addition to the plug-in rule and the newly developed optimal combining rule, we show that other optimal portfolios such as the one based on the Bayes-Stein shrinkage estimator and the three-fund rule of Kan and Zhou (2007) can also be expressed as a combination of the sample GMV and the sample zero-investment portfolio. Our earlier analytical methods can be used to obtain the exact distribution of all the portfolio rules in the group.[6] For the rules more sophisticated than

[3]In the case without a risk-free asset, the plug-in rule can also be expressed as a combining portfolio similar to Kan and Zhou (2007) but with the combining coefficients involving random variables. For such case, the results in Kan and Zhou (2007) are not enough to obtain the explicit expression of the expected out-of-sample utility.

[4]In Section 3.5.3 and 4, we show, respectively, that the out-of-sample returns of the normalized three-fund rule of Kan and Zhou (2007) do not have finite moments and its performance is generally very poor.

[5]Many practitioners books, such as Michaud and Michaud (2008, p. 17), state that fully invested in risky assets is the case of interest. Bodie, Kane, and Marcus (2011, p. 97) point out that equity funds invest primarily in stocks, but may hold 4-5% in cash to meet redemption needs.

[6]Without such new analytical results, the only existing method is to use simulation, which is computationally

the plug-in rule, additional analytical methods are developed to obtain the expected out-of-sample utilities, which have remained intractable for a long time. The new analytical methods in the paper provide a fast and accurate way of evaluating the portfolios and offer analytical insights in portfolio construction and performance evaluation.

Both theoretically and empirically, we find that the newly developed optimal combining portfolio performs the best in this group of estimated portfolio rules. Theoretically, the optimal combining portfolio has the highest expected out-of-sample utility. In all the empirical datasets examined, the new portfolio generates the highest certainty equivalent returns (CER), the highest Sharpe ratio, and the lowest turnover in the group.

Second, we extend the results of prior studies to the no risk-free asset case and show that the newly derived optimal combining strategy can be readily combined with the shrinkage covariance matrix estimators of Ledoit and Wolf (2004, 2017) or the single factor structure of MacKinlay and Pástor (2000) to form new optimal portfolios. In both cases, the portfolios using the new optimal combining coefficient outperform those not using the coefficient in terms of CER or Sharpe ratio, and using the optimal coefficient also leads to lower portfolio turnover.

Relative to the portfolio using the optimal coefficient alone, we find that portfolios adopting both the optimal coefficient and the shrinkage estimators perform better overall; and the portfolio adopting both the optimal coefficient and the single factor structure tends to perform better in datasets with relatively large number of risky assets (e.g., 100 individual stocks).

Third, we compare the newly obtained optimal combining portfolios with other portfolio strategies proposed in the literature, including imposing no-short-sale constraints, ignoring sample mean (GMV), and non-optimization based rules such as the $1/N$ rule and the two timing strategies of Kirby and Ostdiek (2012). Without transaction costs, the optimal combining portfolios compare favorably with other rules in general. However, it is noted that these alternative rules generally have low turnover compared to the optimal combining portfolios. As a result, when transaction costs are taken into account, the performance gaps become smaller, sometimes even negative. Assuming transaction costs of 20 basis points, we notice that the portfolio using the optimal combining coefficient alone is less likely to outperform these other rules; but the portfolios adopting both the

expensive and inaccurate (e.g., in the tails).

optimal coefficient and the shrinkage estimators continue to perform well against other rules. In dataset with relatively large number of risky assets (e.g., 100 individual stocks), we continue to observe the outperformance of the portfolio adopting both the optimal coefficient and the single factor structure given transaction costs.

2. Theoretical Results

In this section, we present the theoretical results. In particular, the new optimal combining coefficient is derived, which facilitates the proposed optimal combining strategy.

2.1. The Setup

Consider a portfolio choice problem in which an investor chooses his optimal portfolio among N risky assets. Denote the returns of the N risky assets at time t by r_t, with mean μ and covariance matrix Σ. Let w be the weights of a portfolio on the N risky assets.[7] The investor chooses his portfolio weights w to maximize the mean-variance utility function

$$U(w) = w'\mu - \frac{\gamma}{2} w'\Sigma w, \tag{1}$$

where γ is the risk aversion coefficient, with an additional constraint of $1_N' w = 1$ where 1_N stands for the $N \times 1$ vector of ones so that the investor has 100% weights invested in the risky assets.

When both μ and Σ are known, it is well known that the optimal portfolio p^* must be on the efficient frontier, and we can show that portfolio p^* can be expressed as a combination of the GMV portfolio and another efficient portfolio. Specifically, the weights of the optimal portfolio p^* are given by

$$w^* = w_g + \frac{1}{\gamma} w_z, \tag{2}$$

where

$$w_g = \frac{\Sigma^{-1} 1_N}{1_N' \Sigma^{-1} 1_N}, \qquad w_z = \Sigma^{-1} \left(\mu - 1_N \mu_g\right), \tag{3}$$

[7]The analysis holds whether r_t is defined as raw or excess returns. Our calibration and empirical tests are based on the latter for easy comparison with those studies using excess returns.

and w_g is the weights of the GMV portfolio, $\mu_g = 1_N'\Sigma^{-1}\mu/(1_N'\Sigma^{-1}1_N)$ is the expected return of the GMV portfolio, and w_z is the weights of an efficient zero-investment portfolio (i.e., $1_N'w_z = 0$). Equation (2) suggests that investors always hold 100% of the GMV portfolio, and depending on their degrees of risk aversion, their exposures to the efficient zero-investment portfolio vary. As the risk aversion varies, the optimal portfolio from (2) will trace out the upper half of the minimum-variance frontier.

Let $r_{p^*,t+1} = w^{*\prime}r_{t+1}$ be the return of portfolio p^* at time $t+1$. The mean and variance of $r_{p^*,t+1}$ are given by

$$\mu_{p^*} = \mu_g + \frac{\psi^2}{\gamma}, \tag{4}$$

$$\sigma_{p^*}^2 = \sigma_g^2 + \frac{\psi^2}{\gamma^2}, \tag{5}$$

where $\sigma_g^2 = 1/(1_N'\Sigma^{-1}1_N)$ is the variance of the GMV portfolio and $\psi^2 = \mu'\Sigma^{-1}\mu - \mu_g^2/\sigma_g^2$ is the squared slope of the asymptote to the *ex ante* minimum-variance frontier. It follows that the utility from holding the optimal portfolio is

$$U(w^*) = \mu_g - \frac{\gamma}{2}\sigma_g^2 + \frac{\psi^2}{2\gamma}. \tag{6}$$

This equation shows that w^* outperforms w_g by a certainty equivalent return of $\psi^2/(2\gamma)$, which is coming from the exposure to w_z. Its magnitude is determined by the slope of the asymptote to the *ex ante* minimum-variance frontier (ψ) and the risk aversion coefficient (γ).

2.2. Estimation Risk

In practice, however, the optimal portfolio weights, w^*, are not computable because μ and Σ are unknown, and they need to be estimated. We assume an investor estimates μ and Σ using an estimation window of h periods of historical return data with $h > N$. For analytical tractability, we make the usual assumption that r_t is independent and identically distributed over time, and has a multivariate normal distribution. Under this assumption, the maximum likelihood estimators of μ and Σ at time t are given by

$$\hat{\mu}_t = \frac{1}{h}\sum_{s=t-h+1}^{t} r_s, \tag{7}$$

$$\hat{\Sigma}_t = \frac{1}{h} \sum_{s=t-h+1}^{t} (r_s - \hat{\mu}_t)(r_s - \hat{\mu}_t)'. \tag{8}$$

Replacing μ and Σ in (2) by $\hat{\mu}_t$ and $\hat{\Sigma}_t$, we can obtain an implementable portfolio p (which is termed the plug-in rule hereafter)

$$\hat{w}_{p,t} = \hat{w}_{g,t} + \frac{1}{\gamma} \hat{w}_{z,t}, \tag{9}$$

where

$$\hat{w}_{g,t} = \frac{\hat{\Sigma}_t^{-1} 1_N}{1_N' \hat{\Sigma}_t^{-1} 1_N}, \qquad \hat{w}_{z,t} = \hat{\Sigma}_t^{-1} (\hat{\mu}_t - 1_N \hat{\mu}_{g,t}) \tag{10}$$

represent the weights of the sample GMV portfolio and those of the sample zero-investment portfolio and $\hat{\mu}_{g,t} = (1_N' \hat{\Sigma}_t^{-1} \hat{\mu}_t)/(1_N' \hat{\Sigma}_t^{-1} 1_N)$.

When estimated parameters instead of true parameters are used, the plug-in portfolio p underperforms the true optimal portfolio p^* due to estimation errors. In this paper, we focus on one strategy to deal with the estimation risk, that is, to adjust the exposure to the sample zero-investment portfolio $\hat{w}_{z,t}$. Specifically, we consider the class of portfolios with weights:

$$\hat{w}_t(\tilde{c}) = \hat{w}_{g,t} + \frac{\tilde{c}}{\gamma} \hat{w}_{z,t}, \tag{11}$$

where $\tilde{c}$ is a scalar combining coefficient. Note that the weight on the sample GMV portfolio is always 100% so that the investor remains fully invested in risky assets. When $\tilde{c} = 1$, we have the plug-in portfolio p. When $0 < \tilde{c} < 1$, the effect of estimation risk is reduced due to a smaller exposure to $\hat{w}_{z,t}$. Note that $\hat{w}_{g,t}$ involves smaller estimation errors than $\hat{w}_{z,t}$ because $\hat{w}_{g,t}$ depends only on $\hat{\Sigma}_t$ while $\hat{w}_{z,t}$ depends on both $\hat{\mu}_t$ and $\hat{\Sigma}_t$, and it is well known that the sample mean $\hat{\mu}_t$ is a very noisy estimator of μ. This motivates many practitioners and researchers to focus only on the GMV portfolio (i.e., $\tilde{c} = 0$). This practice is appropriate only if the cost associated with the estimation risk in $\hat{w}_{z,t}$ outweighs the utility gain from the exposure to $\hat{w}_{z,t}$. This, however, is not typically the case. Instead of completely ignoring $\hat{w}_{z,t}$, we show that optimally adjusting the exposure to $\hat{w}_{z,t}$ is a better strategy.

Let $\mathcal{N}(\mu_0, v_0)$ stand for a random variable that is normally distributed with mean μ_0 and variance v_0, and χ^2_ν stand for a random variable that follows a chi-squared distribution with ν degrees of freedom. The following Proposition expresses the exact distribution of the out-of-sample returns of portfolios in the class specified in (11) in terms of a set of independent univariate random variables.

PROPOSITION **1:** *Suppose $N > 3$. Let $z_2 \sim \mathcal{N}(\sqrt{h}\psi, 1)$, $u_0 \sim \chi^2_{N-2}$, $v_2 \sim \chi^2_{h-N+1}$, $w_1 \sim \chi^2_{h-N+3}$, $w_2 \sim \chi^2_{h-N+2}$, $s_1 \sim \chi^2_{N-4}$, $s_2 \sim \chi^2_{N-3}$, $x_{11} \sim \mathcal{N}(0,1)$, $x_{21} \sim \mathcal{N}(0,1)$, $a \sim \mathcal{N}(0,1)$, $b \sim \mathcal{N}(0,1)$, $c \sim \mathcal{N}(0,1)$, and they are independent of each other.*[8] *Then, the distribution of $\hat{\psi}_t^2$ is given by*

$$\hat{\psi}_t^2 = \hat{\mu}_t'\hat{\Sigma}_t^{-1}\hat{\mu}_t - \frac{\left(1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t\right)^2}{\left(1_N'\hat{\Sigma}_t^{-1}1_N\right)} = \frac{z_2^2 + u_0}{v_2}. \tag{12}$$

Define

$$y_1 = \frac{x_{11}}{\sqrt{w_1}} + \frac{bx_{21}}{\sqrt{w_1 w_2}} + \frac{ax_{21}}{\sqrt{v_2 w_2}}, \tag{13}$$

$$y_2 = \frac{c}{\sqrt{w_1}} + \frac{b\sqrt{s_2}}{\sqrt{w_1 w_2}} + \frac{a\sqrt{s_2}}{\sqrt{v_2 w_2}}. \tag{14}$$

The out-of-sample return of portfolio $\hat{w}_t(\tilde{c})$, $r_{t+1}(\tilde{c}) = \hat{w}_t(\tilde{c})'r_{t+1}$, is conditionally normally distributed with conditional mean and variance given by

$$\mu_t(\tilde{c}) = \hat{w}_t(\tilde{c})'\mu = \mu_g + \frac{\sigma_g \psi}{\hat{\psi}_t}\left(\frac{\sqrt{u_0}y_1}{\sqrt{v_2}} + \frac{az_2}{v_2}\right) + \frac{\tilde{c}\sqrt{h}\psi}{\gamma v_2}\left(\frac{x_{21}\sqrt{u_0}}{\sqrt{w_2}} + z_2\right), \tag{15}$$

$$\begin{aligned}\sigma_t^2(\tilde{c}) = \hat{w}_t(\tilde{c})'\Sigma\hat{w}_t(\tilde{c}) &= \sigma_g^2\left(y_1^2 + y_2^2 + 1 + \frac{s_1}{w_1} + \frac{a^2}{v_2}\right) + \frac{\tilde{c}^2 h\hat{\psi}_t^2}{\gamma^2 v_2}\left(1 + \frac{x_{21}^2 + s_2}{w_2}\right) + \\ &\frac{2\tilde{c}\sqrt{h}\sigma_g\hat{\psi}_t}{\gamma\sqrt{v_2}}\left(\frac{a}{\sqrt{v_2}} + \frac{x_{21}y_1}{\sqrt{w_2}} + \frac{\sqrt{s_2}y_2}{\sqrt{w_2}}\right).\end{aligned} \tag{16}$$

Proposition 1 shows that the out-of-sample return of portfolio $\hat{w}_t(\tilde{c})$ depends only on three unknown scalar parameters: μ_g, σ_g, and ψ. Therefore, any performance measures constructed based on the out-of-sample portfolio returns (e.g., expected out-of-sample utility, out-of-sample Sharpe ratio) also depend only on these three parameters.

In addition, Proposition 1 helps us better understand the effect of estimation risk. From (4) and (5) in Subsection 2.1, we know that without estimation risk, the return distribution of the GMV portfolio (w_g) depends on μ_g and σ_g, the return distribution of the zero-investment portfolio (w_z) depends only on ψ, and the two portfolios are uncorrelated (i.e., $w_g'\Sigma w_z = 0$). From (15) and (16) in Proposition 1, we see that other than μ_g and σ_g, the out-of-sample return of $\hat{w}_{g,t}$ (i.e., the terms not involving $\tilde{c}$) also depends on ψ, and ψ influences the conditional mean but not the conditional

[8]If we set $u_0 = 0$, $x_{11} = 0$, $x_{21} = 0$, $s_1 = 0$, $s_2 = 0$, and $c = 0$ when $N = 2$, and set $s_1 = 0$, $s_2 = 0$, and $c = 0$ when $N = 3$, then the results in Proposition 1 also hold for the cases of $N = 2$ or $N = 3$.

variance. The last term in (16) suggests that due to estimation errors, the returns of these two portfolios ($\hat{w}_{g,t}$ and $\hat{w}_{z,t}$) are no longer uncorrelated, and the conditional covariance depends on both σ_g and ψ. The term involving $\tilde{c}$ in (15) and the term involving $\tilde{c}^2$ in (16) indicate that the return distribution of the zero-investment portfolio ($\hat{w}_{z,t}$) continues to just depend on ψ even when there is estimation risk.

Other than offering analytical insights in the portfolio performance, results from Proposition 1, in particular that expressing the out-of-sample returns of sample optimal portfolios in terms of a set of independent univariate random variables, enable a much faster and accurate way of computing the exact distribution. In addition, with the expressions in Proposition 1, we can easily obtain explicit expressions of the moments and the expected out-of-sample utilities of the portfolios so that we can evaluate the portfolios more efficiently.

The following two lemmas present the results for portfolio $\hat{w}_t(\tilde{c})$ when $\tilde{c}$ is a constant scalar.

LEMMA **1:** *When $h > N+3$, the expected out-of-sample utility of portfolio $\hat{w}_t(\tilde{c}) = \hat{w}_{g,t} + \tilde{c}\hat{w}_{z,t}/\gamma$, where $\tilde{c}$ is a constant scalar, is given by*

$$E[U(\hat{w}_t(\tilde{c}))] = \mu_g - \frac{\gamma(h-2)\sigma_g^2}{2(h-N-1)} + \frac{h}{\gamma(h-N-1)}\left[\tilde{c}\psi^2 - \frac{\tilde{c}^2(h-2)(h\psi^2+N-1)}{2(h-N)(h-N-3)}\right]. \tag{17}$$

Economically, the expected out-of-sample utility represents the utility level that an investor can achieve on average by applying the portfolio rule repeatedly, and this is because the conditional mean and variance of the portfolio, $\hat{\mu}_t(\tilde{c})$ and $\sigma_t^2(\tilde{c})$, are both random variables, so is $U(\hat{w}_t(\tilde{c}))$. The expected out-of-sample utility is often used as a performance measure to evaluate portfolio theoretically. Lemma 1 tells us what determines this performance measure. The first two terms in (17) capture the expected out-of-sample utility of $\hat{w}_{g,t}$, which depends on two unknown parameters μ_g and σ_g^2. For given μ_g and σ_g^2, the expected out-of-sample utility of $\hat{w}_{g,t}$ increases with h and decreases with N. The last term in (17) represents the expected out-of-sample utility due to the exposure to $\hat{w}_{z,t}$ (i.e., $\tilde{c}\hat{w}_{z,t}/\gamma$), and it depends only on one unknown parameter ψ^2.

When there is no estimation risk, it is always beneficial to allocate weight to the zero-investment portfolio as long as $\psi^2 > 0$, as shown in (6). However, when there is estimation risk, we need a sufficiently large ψ^2 to realize the benefit of the allocation to the zero-investment portfolio. Notice that when $\psi^2 = 0$, the last term in (17) is negative. This suggests that when ψ^2 is not large enough, the cost associated with estimation risk can outweigh the benefit of a positive ψ^2, and as a result, the exposure to $\hat{w}_{z,t}$ decreases, instead of increases, portfolio performance. We know that when the expected returns of the risky assets are similar to each other, the *ex ante* minimum-variance frontier tends to be flat (i.e., small ψ^2). This explains why we typically see better performance for the optimal portfolios when anomaly portfolios are used as test assets, and the portfolio performance tends to be poor when industry portfolios are used as test assets, for example.

Also note that the coefficient associated with ψ^2 in (17) is positive only if $0 < \tilde{c} < 2(h-N)(h-N-3)/h/(h-2)$. That is, even for a large enough ψ^2, the allocation to $\hat{w}_{z,t}$ must be in this range so that the allocation is beneficial. This range becomes narrow when N is large relative to h. For example, when $h = 120$ and $N = 100$, we need $0 < \tilde{c} < 0.048$. This explains why the plug-in rule (i.e., $\tilde{c} = 1$) typically performs poorly in datasets with a large number of risky assets.

Finally, note that the last term in (17) is a quadratic function of $\tilde{c}$. The term involving $\tilde{c}$ is positive, representing the benefit of including $\hat{w}_{z,t}$ in terms of increased expected portfolio return; and the term involving $\tilde{c}^2$ is negative, representing the cost of including $\hat{w}_{z,t}$ in terms of increased portfolio variance. The quadratic function suggests that there exists an optimal $\tilde{c}$, and therefore, ignoring $\hat{w}_{z,t}$ and focusing only on the GMV portfolio is not the best strategy.

LEMMA **2:** *The unconditional mean and variance of a portfolio with weights $\hat{w}_t(\tilde{c}) = \hat{w}_{g,t} + \tilde{c}\hat{w}_{z,t}/\gamma$, where $\tilde{c}$ is a constant scalar, are given by*

$$\mu(\tilde{c}) = E[\mu_t(\tilde{c})] = \mu_g + \frac{\tilde{c}h\psi^2}{\gamma(h-N-1)} \quad \textit{for } h > N+1, \tag{18}$$

$$\begin{aligned}\sigma^2(\tilde{c}) &= E[\sigma_t^2(\tilde{c})] + E[\mu_t^2(\tilde{c})] - E[\mu_t(\tilde{c})]^2 \\ &= \frac{\sigma_g^2(h-2+\psi^2)}{h-N-1} + \frac{\tilde{c}^2 h(h-2)[N-1+(h+1)\psi^2]}{\gamma^2(h-N)(h-N-1)(h-N-3)} \\ &\quad + \frac{2\tilde{c}^2h^2\psi^4}{\gamma^2(h-N-1)^2(h-N-3)} \quad \textit{for } h > N+3.\end{aligned} \tag{19}$$

From (18), we can see that $\mu(\tilde{c} = 1) > \mu_{p^*}$, suggesting that the plug-in rule $\hat{w}_{p,t}$ is not an unbiased estimator of w^*, i.e., $E[\hat{w}_{p,t}] \neq w^*$. The unbiased rule can be obtained by setting $\tilde{c} = (h-N-1)/h$ so that $E[\hat{w}_{p,t}] = w^*$ and $\mu(\tilde{c}) = \mu_{p^*}$. When $\tilde{c} = 1$ or $\tilde{c} = (h-N-1)/h$, we can show from (19) that $\sigma^2(\tilde{c}) > \sigma^2_{p^*}$. This is expected because estimation risk leads to mean-variance utility loss. Given the same or higher mean, the estimated optimal portfolios must have higher variances compared to that of the true optimal portfolio. It can also be verified from (18) and (19) that both $\mu(\tilde{c})$ and $\sigma^2(\tilde{c})$ are decreasing functions of h and increasing functions of N. As $h \to \infty$, $\mu(\tilde{c})$ converges to μ_{p^*} and $\sigma^2(\tilde{c})$ converges to $\sigma^2_{p^*}$ if $\tilde{c} \to 1$.

2.3. Optimal Combining Coefficient

Given that (17) is a quadratic function in $\tilde{c}$, we can obtain the optimal value of $\tilde{c}$ that maximizes the expected out-of-sample utility:

$$\tilde{c}^* = \frac{k\psi^2}{\psi^2 + \frac{N-1}{h}}, \tag{20}$$

where

$$k = \frac{(h-N)(h-N-3)}{h(h-2)}. \tag{21}$$

Note that $\tilde{c}^*$ is derived by explicitly taking into account the effect of estimation errors in $\hat{w}_{z,t}$. For $h > N+3$, it is easy to show that $0 < \tilde{c}^* < 1$. That is, the optimal combining portfolio addresses the estimation risk by lowering the exposure to $\hat{w}_{z,t}$.

Because $\tilde{c}^*$ depends on ψ^2 which is unknown to investors in practice, $\hat{w}_t(\tilde{c}^*)$ is not implementable. Adopting the adjusted estimator of ψ^2 in Kan and Zhou (2007), an implementable version of $\tilde{c}^*$ can be obtained as

$$\hat{c}_t = \frac{k\hat{\psi}^2_{a,t}}{\hat{\psi}^2_{a,t} + \frac{N-1}{h}}, \tag{22}$$

where

$$\hat{\psi}^2_{a,t} = \frac{(h-N-1)\hat{\psi}^2_t - (N-1)}{h} + \frac{2(\hat{\psi}^2_t)^{\frac{N-1}{2}}(1+\hat{\psi}^2_t)^{-\frac{h-2}{2}}}{h\mathrm{B}_{\hat{\psi}^2_t/(1+\hat{\psi}^2_t)}((N-1)/2,(h-N+1)/2)} \tag{23}$$

and

$$B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}\mathrm{d}y \tag{24}$$

is the incomplete beta function. Note that $\hat{c}_t$ is a function of $\hat{\psi}^2_t$ and we use g_3 to denote it,

$$\hat{c}_t = g_3(\hat{\psi}^2_t) = \frac{k\hat{\psi}^2_{a,t}}{\hat{\psi}^2_{a,t} + \frac{N-1}{h}}. \tag{25}$$

Given $\hat{c}_t$, we have an implementable optimal combining portfolio, denoted as portfolio q,

$$\hat{w}_{q,t} = \hat{w}_t(\hat{c}_t) = \hat{w}_{g,t} + \frac{g_3(\hat{\psi}_t^2)}{\gamma}\hat{w}_{z,t}. \tag{26}$$

Because $\hat{c}_t = g_3(\hat{\psi}_t^2)$ is no longer a constant scalar, Lemmas 1 and 2 are not applicable. Proposition 2 and Lemma 3 provide explicit expressions of the expected out-of-sample utility and the unconditional mean and variance of a combining portfolio with the combining coefficient being a function of $\hat{\psi}_t^2$. Let $\chi_m^2(\delta)$ stand for a random variable that follows a noncentral chi-squared distribution with m degrees of freedom and a noncentrality parameter δ. To facilitate our presentation, we use $\mathscr{G}_{m,n}^{\delta}$ to stand for a random variable $y = x_1/x_2$ where $x_1 \sim \chi_m^2(\delta)$ and $x_2 \sim \chi_n^2$, independent of each other.

PROPOSITION **2:** *When $h > N+3$, the expected out-of-sample utility of portfolio $\hat{w}_t(c_t) = \hat{w}_{g,t} + c_t\hat{w}_{z,t}/\gamma$ where $c_t = g(\hat{\psi}_t^2)$ is a function of $\hat{\psi}_t^2$, is given by*

$$E[U(\hat{w}_t(c_t))] = \mu_g - \frac{\gamma(h-2)\sigma_g^2}{2(h-N-1)} + \frac{h\psi^2 E[g(q_3)]}{\gamma(h-N-1)} - \frac{h(h-2)E[g^2(q_4)q_4]}{2\gamma(h-N)(h-N-1)}, \tag{27}$$

where $q_3 \sim \mathscr{G}_{N+1,h-N-1}^{h\psi^2}$, and $q_4 \sim \mathscr{G}_{N-1,h-N-1}^{h\psi^2}$.

Proposition 2 suggests that the expected out-of-sample utility of the implementable optimal combining portfolio can be expressed in terms of a one-dimensional integral, which provides a speedy and accurate way to obtain the expected out-of-sample utility. The first two terms in (27) are the same as those in (17), reflecting the expected out-of-sample utility of $\hat{w}_{g,t}$. The remaining two terms capture the utility due to the exposure to $\hat{w}_{z,t}$, with the first term being the benefit of the exposure to $\hat{w}_{z,t}$ in terms of increased portfolio expected return and the second term being the cost of the exposure to $\hat{w}_{z,t}$ in terms of higher portfolio variance.

LEMMA **3:** *The unconditional mean and variance of portfolio $\hat{w}_t(c_t) = \hat{w}_{g,t} + c_t\hat{w}_{z,t}/\gamma$ where $c_t = g(\hat{\psi}_t^2)$ is a function of $\hat{\psi}_t^2$, are given by*

$$\mu(c_t) = E[\mu_t(c_t)] = \mu_g + \frac{h\psi^2}{\gamma(h-N-1)}E[g(q_3)] \quad \textit{for } h > N+1, \tag{28}$$
$$\sigma^2(c_t) = E[\sigma_t^2(c_t)] + E[\mu_t(c_t)^2] - E[\mu_t(c_t)]^2$$

$$= \frac{\sigma_g^2(h-2+\psi^2)}{h-N-1} + \frac{h(h-2+\psi^2)E[g^2(q_4)q_4]}{\gamma^2(h-N)(h-N-1)} - \frac{h^2\psi^4(E[g(q_3)])^2}{\gamma^2(h-N-1)^2} + \frac{h\psi^2\left(E[g^2(q_5)]+h\psi^2E[g^2(q_6)]\right)}{\gamma^2(h-N)(h-N-3)} \quad \text{for } h > N+3, \tag{29}$$

where $q_3 \sim \mathscr{G}^{h\psi^2}_{N+1,h-N-1}$, $q_4 \sim \mathscr{G}^{h\psi^2}_{N-1,h-N-1}$, $q_5 \sim \mathscr{G}^{h\psi^2}_{N+1,h-N-3}$ *and* $q_6 \sim \mathscr{G}^{h\psi^2}_{N+3,h-N-3}$.

Lemma 3 presents results similar to those in Lemma 2, but allowing the combining coefficient to be any function of $\hat{\psi}_t^2$. This results in slightly more complicated expressions of the first two moments of the estimated optimal portfolios. Nevertheless, we can continue to show that as $h \to \infty$, both $\mu(c_t)$ and $\sigma^2(c_t)$ converge to the mean and the variance of the true optimal portfolio if $c_t \to 1$.

3. Portfolio Rules

This section introduces the portfolio rules that we will study in this paper. Subsection 3.1 presents a set of portfolios in the class specified in (11), and the theoretical results obtained in the previous section can be readily applied to these portfolios. Other than adjusting the exposure to the zero-investment portfolio $\hat{w}_{z,t}$, Subsection 3.2 to 3.4 present some alternative strategies that can be used to deal with the estimation risk when forming optimal portfolios, including the use of the shrinkage covariance matrix estimators of Ledoit and Wolf (2004, 2017), imposing the single factor structure of MacKinlay and Pástor (2000), and imposing no-short-sale constraints. In particular, we show that our newly derived optimal combining strategy can be applied together with the use of the shrinkage estimators or the single factor structure to form new optimal portfolios. Subsection 3.5 studies some other rules that are derived from the optimization framework but do not maximize the expected out-of-sample utility. Subsection 3.6 introduces a set of *ad hoc* portfolio rules not derived from the optimization framework.

3.1. Invariant Optimal Portfolio Rules

We consider four portfolio rules in this subsection, and they are all in the class specified in (11). These portfolio rules have some nice properties. First, they are all *invariant* to asset repackaging. That is, let A be an $N \times N$ non-singular matrix with $A'1_N = 1_N$. Portfolio returns remain the same whether constructing portfolios based on the original N assets with returns r_t or on the N linear combinations of the original assets with returns $y_t = A'r_t$. In addition, these portfolio rules all converge to the true optimal rule p^* as the estimation window goes to infinity (i.e., $h \to \infty$). Because of these properties, we name this set of portfolios as invariant optimal portfolios. The distribution, moments, and expected out-of-sample utility of these portfolios can be obtained using the results derived in the previous section. Lemma 1 and Proposition 2 suggest that the relative ranking of these portfolios in terms of expected out-of-sample utility is *invariant* to the value of the risk aversion coefficient γ.

Following are the four portfolio rules considered, and they differ in their exposures to the zero-investment portfolio $\hat{w}_{z,t}$.

- The first rule is the plug-in rule p as specified in (9), i.e.,

$$\hat{w}_{p,t} = \hat{w}_{g,t} + \frac{1}{\gamma}\hat{w}_{z,t},$$

  with the out-of-sample portfolio return $r_{p,t+1} = \hat{w}'_{p,t} r_{t+1}$.

- The second rule is the unbiased rule, obtained by setting $\tilde{c} = (h-N-1)/h$

$$\hat{w}_{u,t} = \hat{w}_{g,t} + \frac{(h-N-1)}{\gamma h}\hat{w}_{z,t}. \tag{30}$$

  We denote this unbiased portfolio as portfolio u, and its out-of-sample portfolio return is $r_{u,t+1} = \hat{w}'_{u,t} r_{t+1}$.

- The third rule is the implementable optimal combining rule q as specified in (26), i.e.,

$$\hat{w}_{q,t} = \hat{w}_{g,t} + \frac{g_3(\hat{\psi}_t^2)}{\gamma}\hat{w}_{z,t},$$

  with the out-of-sample portfolio return $r_{q,t+1} = \hat{w}'_{q,t} r_{t+1}$.

- The last portfolio rule is based on a Bayes-Stein estimator developed in Jorion (1986, 1991). We term this portfolio rule as the BS rule[9]

$$\hat{w}_{BS,t} = \hat{w}_{g,t} + \frac{g_4(\hat{\psi}_t^2)}{\gamma}\hat{w}_{z,t}, \tag{31}$$

[9]The derivation of this expression is provided in the Online Appendix.

where

$$g_4(\hat{\psi}_t^2) = \frac{(h-N-2)^2\hat{\psi}_t^2}{(h+1)(h-N-2)\hat{\psi}_t^2 + h(N+2)}. \tag{32}$$

The out-of-sample portfolio return is $r_{BS,t+1} = \hat{w}'_{BS,t} r_{t+1}$.

Note that relative to the plug-in rule $\hat{w}_{p,t}$, the other three rules (i.e., $\hat{w}_{u,t}$, $\hat{w}_{q,t}$ and $\hat{w}_{BS,t}$) have lower exposure to $\hat{w}_{z,t}$, suggesting that these three portfolios contain less estimation risk than the plug-in rule. In addition, it is easy to verify that $0 < g_3(\hat{\psi}_t^2) < (h-N-1)/h < 1$ and $0 < g_4(\hat{\psi}_t^2) < (h-N-1)/h < 1$ when $h > N+3$. Thus, both the implementable optimal combining rule and the BS rule have lower estimation errors than the unbiased rule.

3.2. Rules with Shrinkage Covariance Matrix Estimators

To implement the mean-variance optimal portfolio, we need to invert the estimated covariance matrix. When N is large relative to h, the sample covariance matrix is typically not well-conditioned.[10] To address the issue, Ledoit and Wolf (2004) introduce a shrinkage estimator which is a linear combination of the sample covariance matrix and the identity matrix,

$$\hat{\Sigma}_t^{LW2004} = (1-\rho_t)\hat{\Sigma}_t + \rho_t \nu_t I_N, \tag{33}$$

where I_N is an $N \times N$ identity matrix, ν_t is the shrinkage target which equals to the average of the eigenvalues of $\hat{\Sigma}_t$, and ρ_t is the shrinkage intensity

$$\rho_t = \frac{\text{Min}\left[\frac{1}{h^2}\sum_{s=t-h+1}^{t} ||(r_s - \hat{\mu}_t)(r_s - \hat{\mu}_t)' - \hat{\Sigma}_t||^2, ||\hat{\Sigma}_t - \nu_t I_N||^2\right]}{||\hat{\Sigma}_t - \nu_t I_N||^2} \tag{34}$$

with $||A|| = \sqrt{\text{tr}(AA')/N}$ being the Frobenius norm. When $\hat{\Sigma}_t^{LW2004}$ instead of $\hat{\Sigma}_t$ is plug-in, we obtain a plug-in portfolio with the shrinkage covariance matrix

$$\hat{w}_{p,t}^{LW2004} = \hat{w}_{g,t}^{LW2004} + \frac{1}{\gamma}\hat{w}_{z,t}^{LW2004}, \tag{35}$$

where

$$\hat{w}_{g,t}^{LW2004} = \frac{(\hat{\Sigma}_t^{LW2004})^{-1} 1_N}{1'_N (\hat{\Sigma}_t^{LW2004})^{-1} 1_N}, \tag{36}$$

[10] In this paper, we assume $h > N+3$, so the sample covariance matrix is invertible.

$$\hat{w}_{z,t}^{LW2004} = (\hat{\Sigma}_t^{LW2004})^{-1} \left[\hat{\mu}_t - 1_N \left(\hat{\mu}_t' \hat{w}_{g,t}^{LW2004}\right)\right]. \tag{37}$$

In their recent paper, Ledoit and Wolf (2017) propose an improved nonlinear shrinkage estimator of the covariance matrix which is more flexible than the previous shrinkage estimator. The new estimator can be computed as

$$\hat{\Sigma}_t^{LW2017} = \hat{U}_t \hat{D}_t \hat{U}_t', \tag{38}$$

where $\hat{U}_t$ is the orthogonal matrix obtained from a spectral decomposition of $\hat{\Sigma}_t$, and $\hat{D}_t$ is a diagonal matrix $\hat{D}_t = \text{Diag}(\hat{d}_1(\lambda_1), \cdots, \hat{d}_N(\lambda_N))$ with $\lambda_1, \cdots, \lambda_N$ being the eigenvalues of $\hat{\Sigma}_t$. For $i = 1, \cdots, N$,

$$\hat{d}_i(\lambda_i) = \frac{1}{\lambda_i |s(\lambda_i)|^2}. \tag{39}$$

where $s(\lambda_i)$ is an estimator of the Stieltjes (1894) transform of the limiting empirical distribution of sample eigenvalues. The plug-in portfolio with the nonlinear shrinkage covariance matrix can be obtained as

$$\hat{w}_{p,t}^{LW2017} = \hat{w}_{g,t}^{LW2017} + \frac{1}{\gamma} \hat{w}_{z,t}^{LW2017}, \tag{40}$$

where

$$\hat{w}_{g,t}^{LW2017} = \frac{(\hat{\Sigma}_t^{LW2017})^{-1} 1_N}{1_N' (\hat{\Sigma}_t^{LW2017})^{-1} 1_N}, \tag{41}$$

$$\hat{w}_{z,t}^{LW2017} = (\hat{\Sigma}_t^{LW2017})^{-1} \left[\hat{\mu}_t - 1_N \left(\hat{\mu}_t' \hat{w}_{g,t}^{LW2017}\right)\right]. \tag{42}$$

Note that explicit expressions of the expected out-of-sample utilities of $\hat{w}_{p,t}^{LW2004}$ or $\hat{w}_{p,t}^{LW2017}$ are not available, and we need to use simulation to evaluate them theoretically.

In addition to the use of the shrinkage covariance matrix estimators, we can further adjust the exposure to the zero-investment portfolio when constructing the optimal portfolios. Without explicit expressions of the expected out-of-sample utilities, we are unable to derive the exact forms of the optimal combining coefficients when $\hat{\Sigma}_t^{LW2004}$ or $\hat{\Sigma}_t^{LW2017}$ are used. As an alternative, we directly apply the implementable optimal combining coefficient $\hat{c}_t = g_3(\hat{\psi}_t^2)$ from Subsection 2.3. Specifically, we study the following two portfolios:

$$\hat{w}_{q,t}^{LW2004} = \hat{w}_{g,t}^{LW2004} + \frac{g_3(\hat{\psi}_t^2)}{\gamma} \hat{w}_{z,t}^{LW2004}, \tag{43}$$

$$\hat{w}_{q,t}^{LW2017} = \hat{w}_{g,t}^{LW2017} + \frac{g_3(\hat{\psi}_t^2)}{\gamma}\hat{w}_{z,t}^{LW2017}. \tag{44}$$

We want to examine whether adjusting the exposure to the zero-investment portfolio can further improve portfolio performance when the shrinkage covariance matrix estimators are used.

3.3. Rules with MacKinlay-Pástor Single Factor Structure

MacKinlay and Pástor (2000) adopt a different strategy to deal with the estimation risk. They exploit the implications of an asset pricing model with a single risk factor in estimation of expected returns. Specifically, they assume the covariance matrix to take the following form:

$$\Sigma = \sigma^2 I_N + a\mu\mu', \tag{45}$$

where a and σ^2 are positive scalars. By imposing such a single factor structure, the estimation risk is reduced because fewer parameters need to be estimated (i.e., instead of μ and Σ, we only need to estimate μ and two scalar parameters a and σ^2).

Under the assumption of (45), we have

$$\Sigma^{-1} = \frac{1}{\sigma^2}\left(I_N - \frac{a\mu\mu'}{\sigma^2 + a\mu'\mu}\right). \tag{46}$$

Therefore, the weights of the optimal portfolio for the no risk-free asset case are given by

$$w_{MP} = \frac{(\sigma^2 + a\mu'\mu)1_N - a1_N'\mu\mu}{N(\sigma^2 + a\mu'\mu) - a(1_N'\mu)^2} + \frac{1}{\gamma}\frac{N\mu - (1_N'\mu)1_N}{N(\sigma^2 + a\mu'\mu) - a(1_N'\mu)^2}. \tag{47}$$

Note that the first term in the above equation is the GMV portfolio and the second term captures the zero-investment portfolio in the setup of MacKinlay and Pástor (2000).

At time t, a, σ^2, and μ are obtained by maximizing the log-likelihood function. The closed-form solution is not available, and we use the semi-analytical solution in Tu and Zhou (2011). Plugging the estimates $\hat{a}_{MP,t}$, $\hat{\sigma}_{MP,t}^2$, and $\hat{\mu}_{MP,t}$ into (47), we obtain the implementable version of the optimal portfolio rule based on the MacKinlay-Pástor estimators

$$\hat{w}_{p,t}^{MP} = \hat{w}_{g,t}^{MP} + \frac{1}{\gamma}\hat{w}_{z,t}^{MP}, \tag{48}$$

where

$$\hat{w}_{g,t}^{MP} = \frac{(\hat{\sigma}_{MP,t}^2 + \hat{a}_{MP,t}\hat{\mu}_{MP,t}'\hat{\mu}_{MP,t})1_N - \hat{a}_{MP,t}1_N'\hat{\mu}_{MP,t}\hat{\mu}_{MP,t}}{N(\hat{\sigma}_{MP,t}^2 + \hat{a}_{MP,t}\hat{\mu}_{MP,t}'\hat{\mu}_{MP,t}) - \hat{a}_{MP,t}(1_N'\hat{\mu}_{MP,t})^2}, \tag{49}$$

$$\hat{w}_{z,t}^{MP} = \frac{N\hat{\mu}_{MP,t} - (1_N'\hat{\mu}_{MP,t})1_N}{N(\hat{\sigma}_{MP,t}^2 + \hat{a}_{MP,t}\hat{\mu}_{MP,t}'\hat{\mu}_{MP,t}) - \hat{a}_{MP,t}(1_N'\hat{\mu}_{MP,t})^2}. \tag{50}$$

We term this rule as the MP rule. Note that the MP rule is not consistent (i.e., it does not converge to w^* when $h \to \infty$) unless the assumption of Σ in (45) is true. By imposing such a single factor structure, the estimation risk is reduced, but at the same time, a bias is introduced into the optimal portfolio. So the MP rule provides a tradeoff between estimation errors and bias.

The expression in (48) suggests that we can also adjust the exposure to the zero-investment portfolio when the factor structure such as (45) is imposed. Therefore, we also examine whether applying $\hat{c}_t = g_3(\hat{\psi}_t^2)$ can further improve the performance of the MP rule. Specifically, we study the following portfolio

$$\hat{w}_{q,t}^{MP} = \hat{w}_{g,t}^{MP} + \frac{g_3(\hat{\psi}_t^2)}{\gamma}\hat{w}_{z,t}^{MP}. \tag{51}$$

3.4. Rule with No-Short-Sale Constraints

Empirically, it has been documented that imposing nonnegative portfolio weights can improve the out-of-sample performance of optimal portfolios (e.g., Frost and Savarino, 1988). Jagannathan and Ma (2003) explain why constraining portfolio weights to be nonnegative can reduce the risk in estimated optimal portfolios even when the constraints are wrong. They show that "with no-short-sale constraints in place, the sample covariance matrix performs as well as covariance matrix estimates based on factor models, shrinkage estimators, and daily data."

Given the sample mean and the sample covariance matrix, the optimal portfolio with no-short-sale constraints is the solution to the following optimization problem:

$$\begin{aligned} \max_{w} \quad & w'\hat{\mu}_t - \frac{\gamma}{2}w'\hat{\Sigma}_t w \\ \text{s.t.} \quad & w'1_N = 1, \quad w \geq 0_N \end{aligned}$$

where 0_N is an $N \times 1$ vector of zeros. This optimization problem can be readily solved using quadratic programming. We term this portfolio rule as the NS rule, and denote it as $\hat{w}_{p,t}^{NS}$.

3.5. Other Rules from Portfolio Optimization

This subsection presents three portfolio rules that are also derived from the optimization framework but do not maximize the expected out-of-sample utility.

3.5.1. Sample Global Minimum Variance (GMV) Portfolio

It is known that the sample mean is an imprecise estimator of the population mean. Some even argue that nothing much is lost in ignoring the mean altogether because the estimation error in the sample mean is so large. As a result, instead of the optimal portfolio, it might be better focusing on the sample GMV portfolio:

$$\hat{w}_{g,t} = \frac{\hat{\Sigma}_t^{-1} 1_N}{1_N' \hat{\Sigma}_t^{-1} 1_N}.$$

3.5.2. GMV with No-Short-Sale Constraints

Jagannathan and Ma (2003) show that the GMV portfolio, with the no-short-sale constraints in place, performs well. Such portfolio is the solution to the following problem:

$$\begin{aligned} \min_{w} \quad & w' \hat{\Sigma}_t w \\ \text{s.t.} \quad & w' 1_N = 1, \quad w \geq 0_N. \end{aligned}$$

Similarly, we can solve this problem using quadratic programming, and we denote the resulting portfolio as $\hat{w}_{g,t}^{NS}$.

3.5.3. Normalized Kan-Zhou (2007) Three-fund Rule

In the case in which a risk-free asset is available, Kan and Zhou (2007) derive an optimal three-fund rule under the assumption that the returns of the risky assets are i.i.d. multivariate normally distributed. The portfolio optimally combines the risk-free asset, the sample tangency portfolio, and the sample GMV portfolio.

In the case without a risk-free asset, one may think that we can obtain a similar optimal portfolio by normalizing the weights such that the investor is fully invested in risky assets. It can be shown that the normalized optimal three-fund rule takes the following form

$$\hat{w}_{KZ3,t} = \hat{w}_{g,t} + \left(\frac{1}{1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t}\right)\left(\frac{\hat{\psi}_{a,t}^2}{\hat{\psi}_{a,t}^2 + N/h}\right)\hat{w}_{z,t}, \tag{52}$$

which is also a combination of $\hat{w}_{g,t}$ and $\hat{w}_{z,t}$ and belongs to the class as specified in (11). We term this portfolio as the normalized KZ 3-fund rule. We can apply the results in Proposition 1 to obtain the distribution of its out-of-sample returns. However, the out-of-sample return of this portfolio does not have finite moments, so its expected out-of-sample utility does not exist. The problem is caused by the normalization process. The normalization induces the term $1_N'\hat{\Sigma}_t^{-1}\hat{\mu}_t$ into the denominator which has non-negligible density at zero, and a zero denominator will lead to extreme positions in the risky assets.[11] This explains why we often find that the normalized optimal portfolios have poor out-of-sample performance empirically (e.g., DeMiguel, Garlappi, and Uppal, 2009). Therefore, an optimal portfolio derived in the risk-free asset case cannot be transferred into an optimal portfolio in the case without a risk-free asset. Normalization typically generates a portfolio that is no longer optimal.

3.6. Non-Optimization Rules

To implement the optimal portfolio rules, we need to estimate the parameters based on historical return data. Some argue that the impact of estimation errors can be so large that the optimization may no longer have value. Following this argument, some alternative portfolio rules that are not based on optimization are proposed in the literature. The alternative rules either completely ignore the information contained in historical return data, or use only partial information from the sample and in a way different from the optimization framework. This subsection presents some of these alternative rules.

3.6.1. The $1/N$ Rule

The $1/N$ rule refers to the portfolio strategy with equal weights in the N risky assets, i.e., $w_{ew} = 1_N/N$. DeMiguel, Garlappi, and Uppal (2009) compare the $1/N$ rule with various optimal rules

[11] Under the normality assumption, Okhrin and Schmid (2006) show that the expectation of the weights of the sample tangency portfolio do not exist. Note that the normality assumption is not critical, and this is generally true based on a Lemma of Sargan (1976).

derived in the case where a risk-free asset is available, and find that the $1/N$ rule outperforms the optimal rules in many empirical datasets. Note that when a risk-free asset is not available, the $1/N$ rule is completely free of estimation errors. Under the i.i.d. assumption, the mean and variance of the out-of-sample portfolio return of the $1/N$ rule are given by

$$\mu_{ew} = \frac{1_N'\mu}{N}, \quad \sigma_{ew}^2 = \frac{1_N'\Sigma 1_N}{N^2}. \tag{53}$$

The out-of-sample utility of the $1/N$ rule is therefore

$$U_{ew} = \mu_{ew} - \frac{\gamma}{2}\sigma_{ew}^2. \tag{54}$$

3.6.2. Volatility Timing

While the $1/N$ rule completely ignores the information contained in historical return data, Kirby and Ostdiek (2012) propose two portfolio strategies that use partial sample information and show that they outperform the $1/N$ rule.

The first strategy is the volatility timing strategy (denoted as KO_{VT}). Specifically, the weights of this portfolio are given by

$$\hat{w}_{it} = \frac{(1/\hat{\sigma}_{it}^2)}{\sum_{j=1}^N (1/\hat{\sigma}_{jt}^2)}, \quad i = 1, 2, \ldots, N, \tag{55}$$

where $\hat{\sigma}_{it}^2$ is the estimated conditional variance on the ith risky asset at time t. Note that this strategy uses only the information about the conditional variance from the sample and ignores the means and the covariances.

Kirby and Ostdiek (2012) generalize the above volatility timing strategy to the following:

$$\hat{w}_{it} = \frac{(1/\hat{\sigma}_{it}^2)^\eta}{\sum_{j=1}^N (1/\hat{\sigma}_{jt}^2)^\eta}, \quad i = 1, 2, \ldots, N, \tag{56}$$

where $\eta \geq 0$ is a tuning parameter that measures the timing aggressiveness. Kirby and Ostdiek (2012) do not specify how to choose η, but evaluate portfolio performance with different values of η, i.e., $\eta = 1$, 2, and 4.

3.6.3. Reward-to-risk Timing

The second strategy proposed by Kirby and Ostdiek (2012) is the reward-to-risk timing strategy. In addition to the conditional variances, this timing strategy also incorporates the conditional means from the sample. The weights of this portfolio are given by

$$\hat{w}_{it} = \frac{(\hat{\mu}_{it}^{+}/\hat{\sigma}_{it}^{2})}{\sum_{j=1}^{N}(\hat{\mu}_{jt}^{+}/\hat{\sigma}_{jt}^{2})}, \quad i = 1, 2, \ldots, N, \tag{57}$$

where $\hat{\mu}_{it}^{+} = \text{Max}(\hat{\mu}_{i,t}, 0)$ with $\hat{\mu}_{it}$ being the maximum likelihood estimator of the conditional mean of the ith asset at time t. Using only the positive value of the estimated conditional mean avoids short sales. We denote this portfolio as KO_{RT}.

Similarly, Kirby and Ostdiek (2012) define a generalized version of the reward-to-risk timing strategy:

$$\hat{w}_{it} = \frac{(\hat{\mu}_{it}^{+}/\hat{\sigma}_{it}^{2})^{\eta}}{\sum_{j=1}^{N}(\hat{\mu}_{jt}^{+}/\hat{\sigma}_{jt}^{2})^{\eta}}, \quad i = 1, 2, \ldots, N, \tag{58}$$

where $\eta \geq 0$ is a tuning parameter. Again, they evaluate the portfolio performance for $\eta = 1$, 2, and 4.

In addition to the maximum likelihood estimator of the conditional mean, Kirby and Ostdiek propose an alternative approach to estimate the conditional mean of risky assets. The second estimator exploits the relation between the first and second moments of returns imposed by linear asset pricing models. For a K-factor model, assuming the factors have identical risk premiums, then the portfolio weights are given by

$$\hat{w}_{it} = \frac{(\bar{\beta}_{it}^{+}/\hat{\sigma}_{it}^{2})^{\eta}}{\sum_{j=1}^{N}(\bar{\beta}_{jt}^{+}/\hat{\sigma}_{jt}^{2})^{\eta}}, \quad i = 1, 2, \ldots, N, \tag{59}$$

where $\bar{\beta}_{it}^{+} = \text{Max}(\bar{\beta}_{it}, 0)$ and $\bar{\beta}_{it} = (1/K)\sum_{k=1}^{K}\hat{\beta}_{ikt}$ with $\hat{\beta}_{ikt}$ being the estimated conditional beta of asset i with respect to factor k at time t. We use KO_{BT} to denote the reward-to-risk timing strategy based on $\bar{\beta}_{it}^{+}$. In our empirical results, we evaluate the performance of KO_{BT} with $\hat{\beta}_{ikt}$ obtained with respect to the Carhart four-factor model.

4. Empirical Results

In this section, we empirically compare the performance of various portfolio rules across eight datasets containing monthly excess returns.[12] The first four datasets are obtained from Ken French's website. They are: (i) 10 momentum portfolios ("Momentum"); (ii) Fama-French 5×5 size and book-to-market ranked portfolios ("Size-B/M"); (iii) 25 portfolios formed on operating profitability and investment ("OP-Inv"); (iv) 49 industry portfolios ("Industry"). The first two datasets cover the period from January 1927 to December 2018, the third one covers the period from July 1963 to December 2018, and the last one covers the period from July 1969 to December 2018. All portfolios are value-weighted.

The next two datasets are obtained from Robert Novy-Marx's website. Novy-Marx and Velikov (2016) analyze the performance of twenty-three of the best known, and strongest performing, anomaly strategies. Each anomaly strategy involves a long-side and a short-side value-weighted portfolio. Instead of focusing on the long-short zero-investment portfolios, we include both the long-side and the short-side portfolios in our datasets because we are interested in portfolio strategies with 100% invested in risky assets. Our first dataset involves the portfolios from the eight low turnover anomaly strategies ("NM-V (LT)") specified in Novy-Marx and Velikov (2016), and covers the period from July 1963 to December 2013. The second dataset contains the portfolios from all twenty-three anomaly strategies ("NM-V (All)"), and the sample period is from July 1973 to December 2013. These two datasets contain portfolios that are constructed based on different anomaly findings, and the optimal portfolio rules provide a way to jointly realize the value of the various findings.

The remaining two datasets are constructed using the CRSP data. The first one contains the excess returns of 10 portfolios sorted by idiosyncratic volatility. At the beginning of each month from January 1927 to December 2018, idiosyncratic volatility relative to the Fama-French three-factor model is obtained for each stock using daily data in the previous three months. Stocks are assigned to 10 portfolios based on the idiosyncratic volatility using the NYSE breakpoints. Value-weighted portfolios are held for one month. Stocks with fewer than 20 non-missing daily data in

[12]The Online Appendix provides the comparison based on the expected out-of-sample utility using either the analytical results in Section 2 or simulations.

the three-month period are excluded from the portfolios.

The last dataset contains monthly excess returns of 100 individual stocks selected from the S&P 500 index over the period of March 1957 to December 2018. The investment universe is updated annually. At the end of February of each year t, the largest 100 stocks (in terms of market value) from the S&P 500 index with non-missing monthly returns in the previous h months constitute the investment universe for the period from March of year t to February of year $t+1$. It is possible that some of the selected stocks do not survive the whole year before updating the investment universe. As a result, the number of assets used to construct portfolios can be smaller than 100 for some months.

We use a rolling estimation window approach with $h = 120$ (The Online Appendix provides results with $h = 240$). At the beginning of a given month t, we use the data in the most recent h months (i.e., month $t-h$ to month $t-1$) to compute the weights of various portfolio rules, and obtain the out-of-sample portfolio returns in month t. This practice generates $T-h$ out-of-sample portfolio returns where T stands for the number of months in the sample period. Based on these $T-h$ returns, we compute the sample mean ($\hat{\mu}$) and the sample variance ($\hat{\sigma}^2$) for a given portfolio rule. The certainty equivalent return ($CER = \hat{\mu} - \frac{\gamma}{2}\hat{\sigma}^2$) is typically used as an empirical proxy for the expected out-of-sample utility of a portfolio, and we use this as the first empirical measure to evaluate portfolios. In this section, we assume $\gamma = 3$ (see the Online Appendix for results with $\gamma = 5$). In addition to CER, we also report portfolio Sharpe ratio ($SR = \hat{\mu}/\hat{\sigma}$), which is another widely used measure to evaluate portfolio.

When implementing a portfolio rule in practice, portfolio turnover and the associated trading costs can be a non-trivial aspect of portfolio performance (e.g., Novy-Marx and Velikov, 2016). In this section, we also report the average turnover of a given portfolio, and examine portfolio performance after transaction costs. Let $w_t = [w_{1,t}, \ldots, w_{N,t}]'$ denote the weights determined by a given portfolio rule at the beginning of month t (i.e., using information over months $t-h$ to $t-1$), and $R_t = [R_{1,t}, \ldots, R_{N,t}]'$ be the gross asset returns in month t. The turnover of this portfolio rule in month t is computed as

$$Turn_t = \sum_{i=1}^{N} \left| w_{i,t} - \frac{w_{i,t-1} R_{i,t-1}}{\sum_{i=1}^{N} w_{i,t-1} R_{i,t-1}} \right|. \tag{60}$$

The average turnover across the $T-h$ months in the sample period is reported.

Using live equity trading data from a large institutional money manager over the period 1998 to 2016, Frazzini, Isreal, and Moskowitz (2018) find that actual trading costs are much smaller than previous studies suggest. The mean market impact of the trades in their sample is about 10 basis points. Based on orders executed by Mongan Stanley in 2004, Engle, Ferstenberg, and Russell (2012) show similar level of transaction costs (i.e., 10 basis points) for large institutional investors. Eaton, Irvine, and Liu (2020) examine order-level data provided by Abel Noser over the period from 1999 to 2011, and show the mean price impact of institutional trades is 22 basis points. Given the above findings, we set the transaction cost to be 20 basis points of the amount traded in our base case calculation.[13] We also report the results based on 10 or 50 basis points in the Online Appendix.

4.1. Certainty Equivalent Return (CER)

Table 1 reports the CER results of the portfolios for $h = 120$ and $\gamma = 3$. The first row, w^*, is the performance of the optimal portfolio based on the in-sample estimates of mean and covariance matrix. This portfolio generates the highest CER; but it is not attainable to investors because it requires look-ahead information. The rest of the portfolios are grouped into six categories, following the same order as they are introduced in Section 3.

Note that the newly derived optimal combining coefficient $\hat{c}_t$ can be used to generate four new optimal combining portfolios: $\hat{w}_{q,t}$, $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, and $\hat{w}_{q,t}^{MP}$; and these four portfolios belong to the first three categories, i.e., invariant optimal portfolio rules, rules with shrinkage covariance matrix estimators, and rules with MacKinlay-Pástor single factor structure. Portfolios in the same category adopt similar strategy to deal with estimation risk, and the only difference across portfolios is the combining coefficient used. We compare the portfolios using $\hat{c}_t$ with those not using $\hat{c}_t$ in each of the three categories, and report the one-sided p-value in *italics*. Such comparison offers direct insight in the effect of adopting the optimal combining coefficient $\hat{c}_t$. The CER results of the portfolios in the first three categories in Table 1 show that portfolios adopting the optimal combining coefficient $\hat{c}_t$ always outperform those not using $\hat{c}_t$, and the performance

[13] Some studies (e.g., Gârleanu and Pedersen (2013), Olivares-Nadal and DeMiguel (2018)) use the quadratic transaction costs. As the weights are independent of the portfolio scale, it seems difficult to calibrate such costs in our setting. This is an interesting problem for future research.

improvement is mostly significant.

Comparing across the three categories, we see that $\hat{w}_{p,t}^{LW2004}$, $\hat{w}_{p,t}^{LW2017}$, and $\hat{w}_{p,t}^{MP}$ all outperform $\hat{w}_{p,t}$ but underperform $\hat{w}_{q,t}$ in general. This suggests that the use of shrinkage covariance matrix estimators or imposing the MacKinlay-Pástor single factor structure are both effective ways to address the estimation risk in the plug-in rule, but they are less effective than the optimal combining strategy. In addition, we notice that $\hat{w}_{q,t}^{LW2004}$ and $\hat{w}_{q,t}^{LW2017}$ outperform $\hat{w}_{q,t}$, indicating that adopting the optimal combining coefficient $\hat{c}_t$ together with the shrinkage estimators generates a better performing portfolio than using $\hat{c}_t$ alone to deal with estimation risk. This is, however, not always the case when the single factor structure is used together with $\hat{c}_t$: $\hat{w}_{q,t}^{MP}$ only outperforms $\hat{w}_{q,t}$ in datasets with relatively large number of risky assets (e.g., 49 industry portfolios or 100 individual stocks). This is because in most cases, the single factor structure is not true in population, and imposing such a structure introduces a bias. When $\hat{c}_t$ is used, the effect of estimation risk is already reduced significantly, and it is only beneficial to further impose the factor structure if the gain from further reduced estimation risk outweighs the cost coming from the introduced bias. When more risky assets are involved, there are more estimation errors; and it is more likely for $\hat{w}_{q,t}^{MP}$ to outperform $\hat{w}_{q,t}$.

Next, we consider the portfolios in the remaining three categories in the table, i.e., rule with no-short-sale constraints, other rules from portfolio optimization, and non-optimization rules.[14] Note that the optimal combining coefficient $\hat{c}_t$ is not applicable to these portfolios. Therefore, instead of testing the effect of using $\hat{c}_t$, we compare the performance of the portfolios in these three categories with the four newly obtained optimal portfolios, i.e., $\hat{w}_{q,t}$, $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, and $\hat{w}_{q,t}^{MP}$. The corresponding one-sided p-values are reported in *italics* in the four rows below the portfolio CER.

Table 1 shows that $\hat{w}_{p,t}^{NS}$ outperforms $\hat{w}_{p,t}$ in every case, suggesting that imposing no-short-sale constraints effectively reduces the estimation risk in the plug-in rule. However, it is in general less effective compared to the optimal combining strategy, noting that $\hat{w}_{q,t}$ outperforms $\hat{w}_{p,t}^{NS}$ in all cases except for the dataset containing the largest number of risky assets (i.e., 100 individual stocks). For 100 individual stocks, when we apply the optimal combining coefficient $\hat{c}_t$ together with the

[14] We find that the performance of the timing strategies, i.e., KO_{VT}, KO_{RT}, and KO_{BT}, is generally better with $\eta = 4$ than with $\eta = 1$ or 2. For brevity, only the results with $\eta = 4$ are reported in the table.

shrinkage covariance matrix estimators or the single factor structure, the resulting portfolios (i.e., $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, $\hat{w}_{q,t}^{MP}$) all outperform $\hat{w}_{p,t}^{NS}$.

In six out of the eight datasets examined, the sample GMV $\hat{w}_{g,t}$ underperforms $\hat{w}_{q,t}$. For the remaining two datasets (i.e., 49 industry portfolios and 100 individual stocks), when the optimal combining coefficient $\hat{c}_t$ is used together with the shrinkage covariance matrix estimators or the single factor structure, the resulting portfolios (i.e., $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, $\hat{w}_{q,t}^{MP}$) outperform $\hat{w}_{g,t}$. When no-short-sale constraints are imposed on the sample GMV, portfolio performance does not always improve, unlike the case of the plug-in rule. Similar to $\hat{w}_{g,t}$, we see that $\hat{w}_{g,t}^{NS}$ underperforms $\hat{w}_{q,t}$ in the first six datasets. In dataset with 49 industry portfolios, $\hat{w}_{g,t}^{NS}$ performs similarly as $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, and $\hat{w}_{q,t}^{MP}$. In dataset with 100 individual stocks, $\hat{w}_{g,t}^{NS}$ performs similarly as $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$ but underperforms $\hat{w}_{q,t}^{MP}$. Together, these results indicate that completely ignoring information in the sample mean and focusing only on GMV may not be an ideal strategy.

Next in the table, the performance of the normalized KZ 3-fund rule $\hat{w}_{KZ3,t}$ is reported. As discussed previously, normalizing an optimal portfolio derived in the case with a risk-free asset does not generate an optimal portfolio for the case without a risk-free asset, and the returns of the normalized portfolio do not have finite moments. Consistent with our prediction, $\hat{w}_{KZ3,t}$ performs poorly in general. In many cases, it even underperforms the plug-in rule $\hat{w}_{p,t}$.

Lastly, the performance of the non-optimization rules (i.e., the $1/N$ rule, and the two timing strategies, KO_{VT}, KO_{RT}, and KO_{BT}) are reported. We notice that among these four portfolios, the $1/N$ rule never performs the best and often performs the worst. This is consistent with the conclusion in Kirby and Ostdiek (2012) that there are substantial benefits of using sample information to guide portfolio selection. Comparing the non-optimization rules with the four newly obtained optimal portfolios, we find that $\hat{w}_{q,t}$, $\hat{w}_{q,t}^{LW2004}$, and $\hat{w}_{q,t}^{LW2017}$ perform well in the first six datasets. For 49 industry portfolios, $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, and $\hat{w}_{q,t}^{MP}$ perform similarly as the non-optimization rules. For the dataset including the most assets (i.e., 100 individual stocks), $\hat{w}_{q,t}^{MP}$ performs the best.

Across the eight datasets examined, we notice that adopting only the optimal combining coefficient $\hat{c}_t$, $\hat{w}_{q,t}$ performs well in the first six datasets but less so in the last two datasets (i.e., 49 industry portfolios and 100 individual stocks) when relatively large number of risky assets are involved. More risky assets result in a higher level of estimation risk. In addition, unlike the anomaly portfolios, no significant spread in expected returns has been documented for e.g., industry portfolios. A smaller spread indicates that the mean-variance frontier is flatter (i.e., smaller ψ^2), which suggests a lower value of using $\hat{c}_t$ as discussed in Subsection 2.2. For those cases, Table 1 shows that adopting $\hat{c}_t$ together with other measures to address estimation risk (such as the shrinkage covariance matrix or the single factor structure) can significantly improve portfolio performance.

4.2. Sharpe Ratio

Table 2 reports the Sharpe ratios of the portfolio rules using the same eight datasets for $h = 120$ and $\gamma = 3$. The results are presented in the same way as in Table 1.

For the portfolios in the first three categories, i.e., invariant optimal portfolio rules, rules with shrinkage covariance matrix estimators, and rules with MacKinlay-Pástor single factor structure, we see patterns similar to the CER results in Table 1. Even though the optimal combining coefficient $\hat{c}_t$ is derived to maximize the expected out-of-sample utility, we continue to observe the portfolios using $\hat{c}_t$ to outperform those not using $\hat{c}_t$ in each of the three categories when Sharpe ratio is used to evaluate portfolios. Adopting $\hat{c}_t$ together with the shrinkage covariance matrix estimators generates better portfolio Sharpe ratio than using $\hat{c}_t$ alone (noticing that $\hat{w}_{q,t}^{LW2004}$ and $\hat{w}_{q,t}^{LW2017}$ outperform $\hat{w}_{q,t}$). However, it is not always the case when $\hat{c}_t$ is used together with the single factor structure: $\hat{w}_{q,t}^{MP}$ is more likely to outperform $\hat{w}_{q,t}$ when more risky assets are involved (e.g., 49 industry portfolios and 100 individual stocks).

Unlike the CER results in Table 1, imposing no-short-sale constraints does not seem to be effective in improving the Sharpe ratio of the plug-in rule. Table 2 shows that $\hat{w}_{p,t}^{NS}$ underperforms $\hat{w}_{p,t}$ in five out of the eight datasets examined. Relative to the optimal combining portfolio using only $\hat{c}_t$ (i.e., $\hat{w}_{q,t}$), $\hat{w}_{p,t}^{NS}$ underperforms in all cases except for the dataset with 100 individual stocks. For that dataset, when $\hat{c}_t$ is used together with the shrinkage estimators or the single factor structure, the optimal combining rules (i.e., $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, $\hat{w}_{q,t}^{MP}$) outperform $\hat{w}_{p,t}^{NS}$.

Completely ignoring the information in sample mean, both $\hat{w}_{g,t}$ and $\hat{w}_{g,t}^{NS}$ underperform $\hat{w}_{q,t}$ in six out of the eight datasets examined. When $\hat{c}_t$ is applied together with the shrinkage estimators or the single factor structure, the optimal combining portfolios (i.e., $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, $\hat{w}_{q,t}^{MP}$) outperform $\hat{w}_{g,t}$ in the remaining two datasets (i.e., 49 industry portfolios and 100 individual stocks), and generate similar Sharpe ratios as $\hat{w}_{g,t}^{NS}$ in 49 industry portfolios. For 100 individual stocks, only $\hat{w}_{q,t}^{MP}$ outperforms $\hat{w}_{g,t}^{NS}$.

Similar to the CER results in Table 1, the Sharpe ratio of the normalized KZ 3-fund rule is also poor, further supporting the argument that normalizing an optimal portfolio obtained in the case with a risk-free asset does not generate an optimal portfolio for the case without a risk-free asset.

Finally, the Sharpe ratio results of the non-optimization rules are also similar to the CER results in Table 1. The optimal combining portfolios $\hat{w}_{q,t}$, $\hat{w}_{q,t}^{LW2004}$, and $\hat{w}_{q,t}^{LW2017}$ outperform the non-optimization rules in the first six datasets. For 49 industry portfolios, $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, and $\hat{w}_{q,t}^{MP}$ perform similarly as the non-optimization rules. For 100 individual stocks, $\hat{w}_{q,t}^{MP}$ outperforms the non-optimization rules.

4.3. Turnover and Performance Net of Trading Costs

Table 3 reports the average turnover of the portfolios with $h = 120$ and $\gamma = 3$ based on the same eight datasets. Portfolios are grouped in the same way as in Tables 1 and 2.

For portfolios in the first three categories, i.e., invariant optimal portfolio rules, rules with shrinkage covariance matrix estimators, and rules with MacKinlay-Pástor single factor structure, we have some interesting observations. In each of the three categories, the portfolios using $\hat{c}_t$ have lower average turnover than those not using $\hat{c}_t$. Comparing across the three categories, we find that using the shrinkage covariance matrix estimators further reduces portfolio turnover; and imposing the single factor structure results in even lower turnover. As a result, among the four newly derived optimal combining portfolios, $\hat{w}_{q,t}$ has the highest turnover, and $\hat{w}_{q,t}^{MP}$ has the lowest turnover.

For the portfolios in the remaining three categories, we can see that imposing no-short-sale constraints or ignoring the sample mean both help to reduce portfolio turnover. The normalized KZ 3-fund rule, on the other hand, has a high level of turnover. Finally, the non-optimization rules involve low turnover in general.

Tables 4 and 5 report portfolio CER and Sharpe ratio with transaction cost of 20 basis points. For portfolios in the first three categories, we continue to see the portfolios using $\hat{c}_t$ outperform those not using $\hat{c}_t$ in each category after transaction costs. In addition, portfolios using both $\hat{c}_t$ and the shrinkage covariance matrix estimators continue to outperform the one using $\hat{c}_t$ alone. Due to the low turnover of $\hat{w}_{q,t}^{MP}$, the performance of $\hat{w}_{q,t}^{MP}$ relative to other portfolios in the first three categories improves after the transaction costs are considered. Note that in terms of CER, it is more likely for $\hat{w}_{q,t}^{MP}$ to outperform $\hat{w}_{q,t}$ after transaction costs.

Except for the normalized KZ 3-fund rule, the turnover of the portfolios in the remaining three categories are, in general, lower than those in the first three categories. As a result, we expect the performance of the four newly derived optimal combining portfolios (i.e., $\hat{w}_{q,t}$, $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, and $\hat{w}_{q,t}^{MP}$) relative to the portfolios in the remaining three categories to decrease. Nevertheless, we continue to observe that the optimal combining portfolios using both $\hat{c}_t$ and the shrinkage estimators to outperform in most cases after transaction costs, and $\hat{w}_{q,t}^{MP}$ continues to perform well in the dataset with the largest number of risky assets (i.e., 100 individual stocks).

5. Conclusion

In this paper, we analyze optimal portfolio choice for the case without a risk-free asset under estimation risk. We propose an optimal combining strategy to mitigate the impact of estimation risk, which is the first such a rule in the no risk-free asset case. The new strategy can be directly applied to the plug-in rule, or it can be applied together with some other estimated strategies, including the use of the shrinkage covariance matrix estimators of Ledoit and Wolf (2004, 2017) and the single factor structure of MacKinlay-Pástor (2000). We show that the portfolios adopting the new combining strategy outperform those without using it in terms of higher CER, higher Sharpe ratio, and lower turnover.

In addition, we compare the new strategy with other portfolio strategies from the literature, such as the rule with no-short-sale constraints, the GMV, and the non-optimization based portfolio rules (i.e., the $1/N$ rule and the two timing strategies of Kirby and Ostdiek (2012)). We show that the optimal combining strategy, in particular the one applied together with the shrinkage estimators, performs well against these alternative portfolio rules; and the optimal combining strategy applied together with the single factor structure tends to perform well in dataset with relatively large number of risky assets (e.g., 100 individual stocks).

Unlike earlier studies, such as Kan and Zhou (2007) and Tu and Zhou (2011), which rely on simulations to compute the expected out-of-sample utility of the invariant optimal portfolios, we develop new analytical methods which enable us to obtain the exact distribution of the out-of-sample returns and to derive the explicit expressions of the expected out-of-sample utility of these portfolios. Besides allowing for speedy computation, the explicit expressions provide analytical insights into what drives the out-of-sample performance of the rules.

For future research, it will be of interest to examine the gains of our proposed strategy in other asset markets, such as bonds, currencies and commodities. It will also be of interest to explore the implications of estimation errors in some other settings such as the out-of-sample performance of estimated stochastic discount factors (e.g., Kozak, Nagel, and Santosh, 2020).

Appendix

This appendix contains an outline of the proof of Proposition 1 and all the other proofs. The detailed proof of Proposition 1 is in the Online Appendix. We begin by citing two lemmas from Kan and Wang (2017). Suppose $z \sim \mathcal{N}(\mu, 1)$, $w \sim \chi^2_{m-1}$, $u \sim \chi^2_n$, and they are independent of each other. It follows that $v = z^2 + w \sim \chi^2_m(\delta)$, where $\delta = \mu^2$.

Lemma A1: Let $g(v)$ be a function of v. When the expectations exist, we have

$$E[g(v)z] = \mu E[g(v_1)], \tag{A1}$$

$$E[g(v)z^2] = E[g(v_1)] + \delta E[g(v_2)], \tag{A2}$$

where $v_1 \sim \chi^2_{m+2}(\delta)$ and $v_2 \sim \chi^2_{m+4}(\delta)$.

Lemma A2: Let $g(y)$ be a function of $y = v/u \sim \mathscr{G}^{\delta}_{m,n}$. When the expectation exists, we have

$$E\left[\frac{g(y)}{u^k}\right] = \frac{E[g(y_1)]}{2^k\left(\frac{n}{2}-k\right)_k} \quad \text{for } k < \tfrac{n}{2}, \tag{A3}$$

where $y_1 \sim \mathscr{G}^{\delta}_{m,n-2k}$ and $(a)_k = a(a+1)\cdots(a+k-1)$.

PROOF OF PROPOSITION 1: The key to the proof is to define two $N \times N$ orthonormal matrices P and Q. With P, we can transform $\hat{\mu}_t$ and $\hat{\Sigma}_t$ into z and W, which are independent of each other and their distributions, instead of μ and Σ, depend only on the scalar parameters θ_g and ψ. With the definition of z and W, we show that the exact distribution of $r_{t+1}(\tilde{c}) = \hat{w}'_t(\tilde{c})' r_{t+1}$ are determined by eight univariate random terms constructed based on z and W. Next, define $A = (Q'W^{-1}Q)^{-1}$. We show that the distributions of the eight terms are closely linked to the distributions of different elements in A^{-1}. Applying Theorem 3.2.10 of Muirhead (1982), the results in Dickey (1967), and the Bartlett decomposition, we are able to obtain the distributions of those elements. Q.E.D.

PROOF OF LEMMA 1: From the detailed proof of Proposition 1, we obtain

$$\mu_{z,t} = \frac{\sqrt{h}\psi}{v_2}\left(\frac{x_{21}\sqrt{u_0}}{\sqrt{w_2}} + z_2\right), \tag{A4}$$

$$\mu_{g,t} = \mu_g + \frac{\sigma_g \psi}{\sqrt{z_2^2 + u_0}}\left(y_1\sqrt{u_0} + \frac{a z_2}{\sqrt{v_2}}\right), \tag{A5}$$

$$\sigma_{g,t}^2 = \sigma_g^2\left(y_1^2+y_2^2+1+\frac{s_1}{w_1}+\frac{a^2}{v_2}\right), \tag{A6}$$

$$\sigma_{z,t}^2 = \frac{h\hat{\psi}_t^2}{v_2}\left(1+\frac{x_{21}^2+s_2}{w_2}\right), \tag{A7}$$

$$\sigma_{gz,t} = \frac{\sqrt{h}\sigma_g\hat{\psi}_t}{\sqrt{v_2}}\left(\frac{a}{\sqrt{v_2}}+\frac{x_{21}}{\sqrt{w_2}}y_1+\frac{\sqrt{s_2}}{\sqrt{w_2}}y_2\right). \tag{A8}$$

Taking expectations and using the fact that $E[y_1]=E[y_2]=0$, $E[y_1^2]=1/(h-N-1)$, and $E[y_2^2]=(h+N-7)/[(h-N-1)(h-N+1)]$, we get

$$E[\mu_{g,t}] = \mu_g, \tag{A9}$$

$$E[\mu_{z,t}] = \frac{h\psi^2}{h-N-1}, \tag{A10}$$

$$E[\sigma_{g,t}^2] = \frac{(h-2)\sigma_g^2}{h-N-1}, \tag{A11}$$

$$E[\sigma_{z,t}^2] = \frac{h(h-2)(h\psi^2+N-1)}{(h-N)(h-N-1)(h-N-3)}, \tag{A12}$$

$$E[\sigma_{gz,t}] = 0. \tag{A13}$$

With the above expressions, we obtain $E[U(\hat{w}_t(\tilde{c}))]$. Q.E.D.

PROOF OF LEMMA 2: Using (A9) to (A13), it is easy to obtain $\mu(\tilde{c})=E[\mu_t(\tilde{c})]$ and $E[\sigma_t^2(\tilde{c})]$. To obtain the expression of $\sigma^2(\tilde{c})$, the only term that needs to be derived is

$$E[\mu_t(\tilde{c})^2] = E\left[\mu_{g,t}^2+\frac{\tilde{c}^2}{\gamma^2}\mu_{z,t}^2+\frac{2\tilde{c}}{\gamma}\mu_{g,t}\mu_{z,t}\right]. \tag{A14}$$

Using (A4) and (A5), we get

$$E[\mu_{g,t}^2] = \mu_g^2+\frac{\sigma_g^2\psi^2}{h-N-1}, \tag{A15}$$

$$E[\mu_{z,t}^2] = \frac{h\psi^2}{(h-N-1)(h-N-3)}\left(\frac{h-2}{h-N}+h\psi^2\right), \tag{A16}$$

$$E[\mu_{g,t}\mu_{z,t}] = \frac{h\psi^2\mu_g}{h-N-1}. \tag{A17}$$

With these expressions, we can obtain $\sigma^2(\tilde{c})$. Q.E.D.

PROOF OF PROPOSITION 2: The expressions of $\mu_t(c_t)$ and $\sigma_t^2(c_t)$ are given by

$$\mu_t(c_t) = \mu_g+\frac{\sigma_g\psi}{\hat{\psi}_t}\left(\frac{\sqrt{u_0}y_1}{\sqrt{v_2}}+\frac{az_2}{v_2}\right)+\frac{\sqrt{h}\psi g(\hat{\psi}_t^2)}{\gamma v_2}\left(\frac{x_{21}\sqrt{u_0}}{\sqrt{w_2}}+z_2\right), \tag{A18}$$

$$\sigma_t^2(c_t) = \sigma_g^2\left(y_1^2+y_2^2+1+\frac{s_1}{w_1}+\frac{a^2}{v_2}\right)+\frac{hg^2(\hat{\psi}_t^2)\hat{\psi}_t^2}{\gamma^2 v_2}\left(1+\frac{x_{21}^2+s_2}{w_2}\right)$$
$$+\frac{2\sqrt{h}\sigma_g g(\hat{\psi}_t^2)\hat{\psi}_t}{\gamma\sqrt{v_2}}\left(\frac{a}{\sqrt{v_2}}+\frac{x_{21}y_1}{\sqrt{w_2}}+\frac{\sqrt{s_2}y_2}{\sqrt{w_2}}\right). \tag{A19}$$

Take expectations and applying Lemmas A1 and A2, we get

$$E[\mu_t(c_t)] = \mu_g+\frac{\sqrt{h}\psi}{\gamma}E\left[\frac{g(\hat{\psi}_t^2)z_2}{v_2}\right] = \mu_g+\frac{h\psi^2}{\gamma(h-N-1)}E[g(q_3)], \tag{A20}$$
$$E[\sigma_t^2(c_t)] = \frac{(h-2)\sigma_g^2}{h-N-1}+\frac{h(h-2)}{\gamma^2(h-N)}E\left[\frac{g^2(\hat{\psi}_t^2)\hat{\psi}_t^2}{v_2}\right] = \frac{(h-2)\sigma_g^2}{h-N-1}+\frac{h(h-2)E[g^2(q_4)q_4]}{\gamma^2(h-N)(h-N-1)}, \tag{A21}$$

where $q_3 \sim \mathscr{G}^{h\psi^2}_{N+1,h-N-1}$ and $q_4 \sim \mathscr{G}^{h\psi^2}_{N-1,h-N-1}$. Using these expressions, we obtain the expected out-of-sample utility of portfolio $\hat{w}_t(c_t)$. Q.E.D.

PROOF OF LEMMA 3: The expression for $\mu(c_t) = E[\mu_t(c_t)]$ has already been derived in (A20). For $\sigma^2(c_t) = E[\sigma_t^2(c_t)]+E[\mu_t(c_t)^2]-E[\mu_t(c_t)]^2$, the only term that we need to derive is $E[\mu_t(c_t)^2]$, which can be expressed as

$$E[\mu_t(c_t)^2] = E[\mu_{g,t}^2]+\frac{E[g^2(\hat{\psi}_t^2)\mu_{z,t}^2]}{\gamma^2}+\frac{2E[g(\hat{\psi}_t^2)\mu_{g,t}\mu_{z,t}]}{\gamma}. \tag{A22}$$

Expression of $E[\mu_{g,t}^2]$ is in (A15). Using (A4), (A5), and the fact that $\hat{\psi}_t^2 = (z_2^2+u_0)/v_2$, take expectations and apply Lemmas A1 and A2, we obtain

$$\begin{aligned}
E[g^2(\hat{\psi}_t^2)\mu_{z,t}^2] &= h\psi^2 E\left[g^2(\hat{\psi}_t^2)\left(\frac{u_0}{(h-N)v_2^2}+\frac{z_2^2}{v_2^2}\right)\right] \\
&= \frac{h\psi^2}{h-N}E\left[\frac{g^2(\hat{\psi}_t^2)\hat{\psi}_t^2}{v_2}\right]+\frac{h\psi^2(h-N-1)}{h-N}E\left[\frac{g^2(\hat{\psi}_t^2)z_2^2}{v_2^2}\right] \\
&= \frac{h\psi^2 E[g^2(q_4)q_4]}{(h-N)(h-N-1)}+\frac{h\psi^2(E[g^2(q_5)+h\psi^2 E[g^2(q_6)])}{(h-N)(h-N-3)}, &\text{(A23)} \\
E[g(\hat{\psi}_t^2)\mu_{g,t}\mu_{z,t}] &= \mu_g\sqrt{h}\psi E\left[g(\hat{\psi}_t)\frac{z_2}{v_2}\right] \\
&= \frac{\mu_g h\psi^2 E[g(q_3)]}{h-N-1}, &\text{(A24)}
\end{aligned}$$

where $q_3 \sim \mathscr{G}^{h\psi^2}_{N+1,h-N-1}$, $q_4 \sim \mathscr{G}^{h\psi^2}_{N-1,h-N-1}$, $q_5 \sim \mathscr{G}^{h\psi^2}_{N+1,h-N-3}$ and $q_6 \sim \mathscr{G}^{h\psi^2}_{N+3,h-N-3}$. Using these expressions and after simplification, we obtain (29). Q.E.D.

REFERENCES

Basak, Gopal K., Ravi Jagannathan, Tongshu Ma, 2009, Jackknife Estimator for Tracking Error Variance of Optimal Portfolios, Management Science 55, 990–1002.

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

Bodie, Zvi, Alex Kane, and Alan Marcus, 2011, *Investments, Ninth Edition* (McGraw-Hill Education).

Bodnar, Taras, Nestor Parolya, and Wolfgang Schmid, 2018, Estimation of the global minimum variance portfolio in high dimensions. *European Journal of Operational Research* 266, 371–390.

Brown, Stephen J., 1976, Optimal portfolio choice under uncertainty: A Bayesian approach, Ph.D. dissertation, University of Chicago.

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

Dickey, James M., 1967, Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution, *Annals of Mathematical Statistics* 38, 511–518.

Eaton, Gregory W., Paul Irvine, and Tingting Liu, 2020, Measuring institutional trading costs and the implications for finance research: The case of tick size reductions. *Journal of Financial Economics*, forthcoming.

Engle, Robert, Robert Ferstenberg, and Jeffrey Russell, 2012, Measuring and modeling execution cost and risk, *Journal of Portfolio Management* 38, 14–29.

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

Frazzini, Andrea, Ronen Isreal, and Tobias J. Moskowitz, 2018, Trading costs, working Paper, Yale University.

Frost, Peter A., and James E. Savarino, 1988, For better performance: Constrain portfolio weights, *Journal of Portfolio Management* 15, 29–34.

Gârleanu, Nicolae, and Lasse H. Pedersen, 2013, Dynamic trading with predictable returns and transaction costs, *Journal of Finance* 68, 2309–2340.

Grinold, Richard C., and Ronald N. Kahn, 1999, *Active Portfolio Management: Quantitative Theory and Applications* (McGraw-Hill, New York).

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

Jobson, J. D., B. Korkie, and V. Ratti, 1979, Improved estimation for Markowitz portfolios using James-Stein type estimators. *Proceedings of the American Statistical Association, Business and Economics Statistics Section*, 41 (1979), 279–284.

Jorion, Philippe, 1991, Bayesian and CAPM estimators of the means: Implications for portfolio selection, *Journal of Banking and Finance* 15, 717–727.

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

Kan, Raymond, and Xiaolu Wang, 2017, On the economic value of alphas, working paper, University of Toronto.

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

Kempf, Alexander, and Christoph Memmel, 2006, Estimating the global minimum variance portfolio. *Schmalenbach Business Review* 58, 332–348.

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

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

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

Ledoit, Olivier, and Michael Wolf, 2017, Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks, *Review of Financial Studies*, 30, 4349–4388.

Litterman, Bob, 2003, *Modern Investment Management: An Equilibrium Approach* (Wiley, New York).

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

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

Meucci, Attilio, 2005, *Risk and Asset Allocation* (Springer-Verlag, New York).

Michaud, Richard O., and Robert O. Michaud, 2008, *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation, Second Edition* (Oxford University Press).

Muirhead, Robb J., 1982, *Aspects of Multivariate Statistical Theory* (Wiley, New York).

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

Okhrin, Yarema, and Wolfgang Schmid, 2006, Distributional properties of portfolio weights, *Journal of Econometrics* 134, 235–256.

Olivares-Nadal, Alba V., and Victor DeMiguel, 2018, Technical note – a robust perspective on transaction costs in portfolio optimization, *Operations Research* 66, 733–739.

Politis, Dimitris N., and Joseph P. Romano, 1994, The stationary bootstrap, *Journal of the American Statistical Association* 89, 1303–1313.

Qian, Edward, Ronald Hua, and Eric Sorensen, 2007, *Quantitative Equity Portfolio Management: Modern Techniques and Applications* (Chapman & Hall, New York).

Sargan, John D., 1976, The existence of the moments of estimated reduced form coefficients, Discussion Paper A6, London School of Economics Econometrics Program.

Tu, Jun, and Guofu Zhou, 2004, Data-generating process uncertainty: What difference does it make in portfolio decisions? *Journal of Financial Economics* 72, 385–421.

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

Table 1: CER Comparison

This table reports the certainty equivalent returns of the portfolios studied in this paper with $h = 120$ and $\gamma = 3$, based on the eight datasets containing excess monthly returns. The four newly obtained optimal combining portfolios, i.e., $\hat{w}_{q,t}$, $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, and $\hat{w}_{q,t}^{MP}$, are highlighted with a box around. In the three categories to which the four new portfolios belong, one-sided tests are conducted to assess the value of using the newly derived optimal combining coefficient $\hat{c}_t$, and the p-values are reported in *italics*. For the portfolios in the remaining three categories, one-sided tests are conducted to compare them with the four newly obtained portfolios, and the corresponding p-values are reported in the four rows below the portfolio CER. We set $\eta = 4$ for the timing strategies KO_{VT}, KO_{RT}, and KO_{BT}.

	Momentum $N = 10$	Size-B/M $N = 25$	IVOL $N = 10$	OP-Inv $N = 25$	NM-V (LT) $N = 16$	NM-V (All) $N = 46$	Industry $N = 49$	Stocks $N = 100$
w^*	0.0131	0.0195	0.0107	0.0245	0.0258	0.1167	0.0205	0.0293
				Invariant Optimal Portfolio Rules				
$\boxed{\hat{w}_{q,t}}$	0.0098	0.0102	0.0064	0.0060	0.0081	0.0259	0.0022	−0.0068
$\hat{w}_{p,t}$	−0.0063	−0.0635	−0.0141	−0.0705	−0.0270	−1.2294	−0.3998	-27.0937
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
$\hat{w}_{u,t}$	−0.0013	−0.0284	−0.0075	−0.0361	−0.0147	−0.3039	−0.1278	−0.7464
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
$\hat{w}_{BS,t}$	0.0088	0.0029	0.0048	−0.0017	0.0048	−0.1404	−0.0210	−0.1201
	0.02	*0.00*	*0.00*	*0.00*	*0.03*	*0.00*	*0.00*	*0.00*
				Rules with Shrinkage Covariance Matrix Estimators				
$\boxed{\hat{w}_{q,t}^{LW2004}}$	0.0120	0.0112	0.0100	0.0083	0.0100	0.0695	0.0042	0.0020
$\hat{w}_{p,t}^{LW2004}$	0.0070	−0.0009	0.0048	−0.0219	0.0013	−0.0444	−0.1243	−0.5255
	0.01	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.00*
$\boxed{\hat{w}_{q,t}^{LW2017}}$	0.0113	0.0123	0.0082	0.0088	0.0093	0.0769	0.0045	0.0030
$\hat{w}_{p,t}^{LW2017}$	−0.0006	−0.0204	−0.0066	−0.0275	−0.0135	−0.2046	−0.1205	−0.2643
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
				Rules with MacKinlay-Pástor Single Factor Structure				
$\boxed{\hat{w}_{q,t}^{MP}}$	0.0047	0.0036	0.0036	0.0063	0.0030	0.0066	0.0043	0.0050
$\hat{w}_{p,t}^{MP}$	0.0039	−0.0003	−0.0010	−0.0014	−0.0017	−0.0012	0.0013	−0.0028
	0.15	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.09*	*0.00*
				Rule with No-Short-Sale Constraints				
$\hat{w}_{p,t}^{NS}$	0.0060	0.0050	0.0028	0.0045	0.0048	0.0090	0.0003	0.0000
	0.11	*0.04*	*0.18*	*0.31*	*0.20*	*0.21*	*0.24*	*0.97*
	0.00	*0.00*	*0.00*	*0.05*	*0.02*	*0.00*	*0.05*	*0.19*
	0.03	*0.00*	*0.07*	*0.04*	*0.09*	*0.00*	*0.04*	*0.09*
	0.85	*0.83*	*0.22*	*0.14*	*0.86*	*0.86*	*0.08*	*0.03*

Table 1: CER Comparison (Cont'd)

	Momentum $N=10$	Size-B/M $N=25$	IVOL $N=10$	OP-Inv $N=25$	NM-V (LT) $N=16$	NM-V (All) $N=46$	Industry $N=49$	Stocks $N=100$
				Other Rules from Portfolio Optimization				
$\hat{w}_{g,t}$	0.0051	0.0060	0.0020	0.0050	0.0050	0.0035	0.0025	−0.0056
	0.07	*0.06*	*0.15*	*0.35*	*0.19*	*0.15*	*0.59*	*0.98*
	0.00	*0.00*	*0.00*	*0.03*	*0.01*	*0.00*	*0.05*	*0.00*
	0.02	*0.00*	*0.05*	*0.03*	*0.07*	*0.00*	*0.04*	*0.00*
	0.66	*0.99*	*0.04*	*0.07*	*0.96*	*0.05*	*0.11*	*0.00*
$\hat{w}_{g,t}^{NS}$	0.0035	0.0041	0.0041	0.0049	0.0025	0.0059	0.0042	0.0027
	0.03	*0.02*	*0.28*	*0.36*	*0.07*	*0.17*	*0.85*	*1.00*
	0.00	*0.00*	*0.01*	*0.06*	*0.00*	*0.00*	*0.50*	*0.77*
	0.01	*0.00*	*0.13*	*0.05*	*0.02*	*0.00*	*0.39*	*0.33*
	0.08	*0.69*	*0.73*	*0.11*	*0.30*	*0.29*	*0.44*	*0.02*
$\hat{w}_{KZ3,t}$	−0.7926	−0.0174	−1.2156	−0.0387	−0.6193	−1.4438	−0.1179	−0.0426
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
				Non-Optimization Rules				
$1/N$	0.0026	0.0036	0.0022	0.0034	0.0022	0.0013	0.0039	0.0024
	0.02	*0.02*	*0.16*	*0.21*	*0.08*	*0.13*	*0.75*	*1.00*
	0.00	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.43*	*0.61*
	0.00	*0.00*	*0.06*	*0.02*	*0.02*	*0.00*	*0.37*	*0.32*
	0.04	*0.51*	*0.16*	*0.03*	*0.31*	*0.01*	*0.40*	*0.05*
KO_{VT}	0.0036	0.0043	0.0041	0.0045	0.0030	0.0046	0.0049	0.0037
	0.03	*0.03*	*0.28*	*0.32*	*0.10*	*0.16*	*0.91*	*1.00*
	0.00	*0.00*	*0.01*	*0.05*	*0.00*	*0.00*	*0.71*	*0.95*
	0.01	*0.00*	*0.13*	*0.04*	*0.03*	*0.00*	*0.61*	*0.80*
	0.10	*0.75*	*0.73*	*0.07*	*0.49*	*0.10*	*0.66*	*0.10*
KO_{RT}	0.0047	0.0045	0.0034	0.0049	0.0046	0.0061	0.0038	0.0019
	0.06	*0.03*	*0.23*	*0.36*	*0.18*	*0.18*	*0.78*	*1.00*
	0.00	*0.00*	*0.00*	*0.06*	*0.01*	*0.00*	*0.40*	*0.46*
	0.01	*0.00*	*0.10*	*0.05*	*0.08*	*0.00*	*0.31*	*0.18*
	0.47	*0.74*	*0.46*	*0.14*	*0.85*	*0.39*	*0.38*	*0.02*
KO_{BT}	0.0049	0.0046	0.0026	0.0046	0.0050	0.0053	0.0041	0.0027
	0.07	*0.04*	*0.18*	*0.34*	*0.21*	*0.17*	*0.78*	*1.00*
	0.00	*0.00*	*0.00*	*0.06*	*0.02*	*0.00*	*0.48*	*0.69*
	0.02	*0.00*	*0.07*	*0.05*	*0.10*	*0.00*	*0.41*	*0.42*
	0.59	*0.75*	*0.22*	*0.11*	*0.91*	*0.23*	*0.45*	*0.07*

Table 2: Sharpe Ratio Comparison

This table reports the Sharpe ratio of the portfolios studied in this paper with $h = 120$ and $\gamma = 3$, based on the eight datasets containing excess monthly returns. The four newly obtained optimal combining portfolios, i.e., $\hat{w}_{q,t}$, $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, and $\hat{w}_{q,t}^{MP}$, are highlighted with a box around. In the three categories to which the four new portfolios belong, one-sided tests are conducted to assess the value of using the newly derived optimal combining coefficient $\hat{c}_t$, and the p-values are reported in *italics*. For the portfolios in the remaining three categories, one-sided tests are conducted to compare them with the four newly obtained portfolios, and the corresponding p-values are reported in the four rows below the portfolio Sharpe ratio. We set $\eta = 4$ for the timing strategies KO_{VT}, KO_{RT}, and KO_{BT}.

	Momentum $N = 10$	Size-B/M $N = 25$	IVOL $N = 10$	OP-Inv $N = 25$	NM-V (LT) $N = 16$	NM-V (All) $N = 46$	Industry $N = 49$	Stocks $N = 100$
w^*	0.2827	0.3445	0.2547	0.3843	0.3947	0.8369	0.3519	0.4193
				Invariant Optimal Portfolio Rules				
$\boxed{\hat{w}_{q,t}}$	0.2521	0.2479	0.2452	0.1916	0.2238	0.5778	0.1194	0.0234
$\hat{w}_{p,t}$	0.2375	0.1999	0.2387	0.1470	0.1896	0.5614	0.0610	−0.0482
	0.06	*0.00*	*0.23*	*0.03*	*0.04*	*0.07*	*0.08*	*0.09*
$\hat{w}_{u,t}$	0.2405	0.2069	0.2400	0.1534	0.1940	0.5648	0.0671	−0.0428
	0.10	*0.00*	*0.27*	*0.04*	*0.06*	*0.12*	*0.09*	*0.09*
$\hat{w}_{BS,t}$	0.2511	0.2284	0.2453	0.1718	0.2135	0.5715	0.0809	−0.0434
	0.36	*0.01*	*0.51*	*0.07*	*0.14*	*0.11*	*0.11*	*0.06*
				Rules with Shrinkage Covariance Matrix Estimators				
$\boxed{\hat{w}_{q,t}^{LW2004}}$	0.2693	0.2838	0.2456	0.2312	0.2591	0.6925	0.1661	0.1093
$\hat{w}_{p,t}^{LW2004}$	0.2545	0.2393	0.2383	0.1792	0.2118	0.6558	0.0728	−0.0138
	0.07	*0.01*	*0.23*	*0.02*	*0.03*	*0.01*	*0.02*	*0.01*
$\boxed{\hat{w}_{q,t}^{LW2017}}$	0.2629	0.2784	0.2478	0.2357	0.2368	0.6866	0.1746	0.1382
$\hat{w}_{p,t}^{LW2017}$	0.2515	0.2350	0.2437	0.1885	0.1970	0.6618	0.0767	−0.0094
	0.11	*0.00*	*0.33*	*0.03*	*0.03*	*0.02*	*0.02*	*0.00*
				Rules with MacKinlay-Pástor Single Factor Structure				
$\boxed{\hat{w}_{q,t}^{MP}}$	0.1701	0.1470	0.1480	0.2203	0.1347	0.2345	0.1700	0.1849
$\hat{w}_{p,t}^{MP}$	0.1535	0.0928	0.0705	0.0658	0.0519	0.0634	0.1001	0.0638
	0.14	*0.05*	*0.00*	*0.00*	*0.00*	*0.00*	*0.08*	*0.01*
				Rule with No-Short-Sale Constraints				
$\hat{w}_{p,t}^{NS}$	0.1905	0.1734	0.1299	0.1644	0.1700	0.2495	0.0996	0.0943
	0.03	*0.03*	*0.00*	*0.29*	*0.15*	*0.00*	*0.34*	*0.92*
	0.00	*0.00*	*0.00*	*0.06*	*0.02*	*0.00*	*0.05*	*0.36*
	0.01	*0.00*	*0.00*	*0.06*	*0.09*	*0.00*	*0.03*	*0.14*
	0.82	*0.81*	*0.21*	*0.07*	*0.86*	*0.62*	*0.09*	*0.03*

Table 2: Sharpe Ratio Comparison (Cont'd)

	Momentum $N=10$	Size-B/M $N=25$	IVOL $N=10$	OP-Inv $N=25$	NM-V (LT) $N=16$	NM-V (All) $N=46$	Industry $N=49$	Stocks $N=100$
				Other Rules from Portfolio Optimization				
$\hat{w}_{g,t}$	0.1823	0.2085	0.1088	0.1813	0.1843	0.1478	0.1227	0.0369
	0.03	*0.13*	*0.00*	*0.40*	*0.21*	*0.00*	*0.54*	*0.96*
	0.00	*0.00*	*0.00*	*0.07*	*0.03*	*0.00*	*0.05*	*0.01*
	0.01	*0.01*	*0.00*	*0.06*	*0.13*	*0.00*	*0.03*	*0.00*
	0.75	*1.00*	*0.04*	*0.04*	*0.96*	*0.04*	*0.11*	*0.00*
$\hat{w}_{g,t}^{NS}$	0.1450	0.1595	0.1639	0.1816	0.1236	0.2188	0.1717	0.1291
	0.00	*0.01*	*0.01*	*0.42*	*0.03*	*0.00*	*0.90*	*0.99*
	0.00	*0.00*	*0.01*	*0.12*	*0.00*	*0.00*	*0.57*	*0.78*
	0.00	*0.00*	*0.01*	*0.12*	*0.01*	*0.00*	*0.46*	*0.33*
	0.09	*0.71*	*0.78*	*0.09*	*0.30*	*0.32*	*0.52*	*0.03*
$\hat{w}_{KZ3,t}$	0.0640	0.1081	−0.0086	0.0361	−0.0292	−0.1277	0.0075	−0.0191
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.02*	*0.14*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	0.01	*0.17*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
				Non-Optimization Rules				
$1/N$	0.1276	0.1484	0.1224	0.1426	0.1209	0.1034	0.1527	0.1205
	0.00	*0.01*	*0.00*	*0.17*	*0.04*	*0.00*	*0.74*	*0.97*
	0.00	*0.00*	*0.00*	*0.03*	*0.00*	*0.00*	*0.38*	*0.62*
	0.00	*0.00*	*0.00*	*0.03*	*0.02*	*0.00*	*0.31*	*0.29*
	0.04	*0.52*	*0.17*	*0.01*	*0.33*	*0.00*	*0.35*	*0.05*
KO_{VT}	0.1466	0.1636	0.1619	0.1704	0.1338	0.1751	0.1902	0.1584
	0.00	*0.02*	*0.01*	*0.34*	*0.05*	*0.00*	*0.95*	*1.00*
	0.00	*0.00*	*0.01*	*0.09*	*0.00*	*0.00*	*0.76*	*0.95*
	0.00	*0.00*	*0.01*	*0.08*	*0.03*	*0.00*	*0.68*	*0.80*
	0.12	*0.75*	*0.74*	*0.05*	*0.49*	*0.07*	*0.72*	*0.16*
KO_{RT}	0.1695	0.1651	0.1438	0.1768	0.1665	0.2046	0.1534	0.1078
	0.01	*0.02*	*0.00*	*0.38*	*0.14*	*0.00*	*0.78*	*0.96*
	0.00	*0.00*	*0.00*	*0.10*	*0.02*	*0.00*	*0.35*	*0.48*
	0.00	*0.00*	*0.00*	*0.09*	*0.08*	*0.00*	*0.27*	*0.16*
	0.49	*0.72*	*0.43*	*0.08*	*0.85*	*0.23*	*0.34*	*0.02*
KO_{BT}	0.1747	0.1670	0.1274	0.1713	0.1737	0.1859	0.1585	0.1282
	0.02	*0.03*	*0.00*	*0.34*	*0.18*	*0.00*	*0.78*	*0.98*
	0.00	*0.00*	*0.00*	*0.09*	*0.03*	*0.00*	*0.43*	*0.70*
	0.01	*0.00*	*0.00*	*0.09*	*0.12*	*0.00*	*0.35*	*0.38*
	0.60	*0.74*	*0.22*	*0.07*	*0.90*	*0.13*	*0.39*	*0.06*

Table 3: Turnover Comparison

This table reports the average turnover of the portfolios studied in this paper for $h = 120$ and $\gamma = 3$, based on the eight datasets containing excess monthly returns. The four newly obtained optimal combining portfolios, i.e., $\hat{w}_{q,t}$, $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, and $\hat{w}_{q,t}^{MP}$, are highlighted with a box around. We set $\eta = 4$ for the timing strategies KO_{VT}, KO_{RT}, and KO_{BT}.

	Momentum $N=10$	Size-B/M $N=25$	IVOL $N=10$	OP-Inv $N=25$	NM-V (LT) $N=16$	NM-V (All) $N=46$	Industry $N=49$	Stocks $N=100$
w^*	0.5371	1.9138	0.6365	1.9061	2.3643	20.8989	1.6981	4.1385
Invariant Optimal Portfolio Rules								
$\boxed{\hat{w}_{q,t}}$	2.0157	3.5212	2.9934	1.9283	2.8730	34.0907	1.2583	3.8888
$\hat{w}_{p,t}$	5.0765	54.3620	7.1199	16.3314	11.2294	434.8788	91.2610	1275.5124
$\hat{w}_{u,t}$	4.3229	14.5806	6.5314	9.7989	8.6225	152.8220	40.1044	224.0102
$\hat{w}_{BS,t}$	2.3762	6.6710	3.6017	4.1109	4.2856	79.9476	5.4557	25.7444
Rules with Shrinkage Covariance Matrix Estimators								
$\boxed{\hat{w}_{q,t}^{LW2004}}$	1.0061	1.2745	1.0879	1.1136	0.9959	5.2506	0.5241	0.6093
$\hat{w}_{p,t}^{LW2004}$	2.4207	6.6031	2.6552	7.2740	3.5400	50.8673	15.2265	212.4453
$\boxed{\hat{w}_{q,t}^{LW2017}}$	1.6960	2.1229	2.3580	1.1987	2.0561	9.3306	0.4992	0.3830
$\hat{w}_{p,t}^{LW2017}$	4.2156	12.1936	5.5694	8.2761	7.5363	123.4904	23.6311	30.5442
Rules with MacKinlay-Pástor Single Factor Structure								
$\boxed{\hat{w}_{q,t}^{MP}}$	0.1460	0.1661	0.1387	0.1931	0.1711	0.1984	0.1484	0.2159
$\hat{w}_{p,t}^{MP}$	0.2321	0.2869	0.2221	0.3277	0.2636	0.2861	0.1820	0.2850
Rule with No-Short-Sale Constraints								
$\hat{w}_{p,t}^{NS}$	0.1066	0.2393	0.1808	0.1945	0.1456	0.0558	0.2092	0.2648
Other Rules from Portfolio Optimization								
$\hat{w}_{g,t}$	0.2770	0.7665	0.2640	0.5413	0.4911	1.5337	0.8227	3.7827
$\hat{w}_{g,t}^{NS}$	0.0817	0.0691	0.0088	0.0774	0.0559	0.0375	0.0733	0.1440
$\hat{w}_{KZ3,t}$	30.5852	16.5565	55.9217	8.5546	28.0002	181.5043	8.4783	274.6183
Non-Optimization Rules								
$1/N$	0.0176	0.0182	0.0172	0.0199	0.0197	0.0227	0.0341	0.0648
KO_{VT}	0.0281	0.0309	0.0188	0.0353	0.0347	0.0373	0.0481	0.0706
KO_{RT}	0.0746	0.0767	0.0886	0.1086	0.0959	0.0832	0.1369	0.1591
KO_{BT}	0.0346	0.0319	0.0383	0.0489	0.0373	0.0414	0.0710	0.1358

Table 4: CER Comparison with 20 bps Transaction Costs

This table reports the CER of the portfolios studied in this paper with $h = 120$, $\gamma = 3$, and a transaction cost of 20 bps, based on the eight datasets containing excess monthly returns. The four newly obtained optimal combining portfolios, i.e., $\hat{w}_{q,t}$, $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, and $\hat{w}_{q,t}^{MP}$, are highlighted with a box around. In the three categories to which the four new portfolios belong, one-sided tests are conducted to assess the value of using the newly derived optimal combining coefficient $\hat{c}_t$, and the p-values are reported in *italics*. For the portfolios in the remaining three categories, one-sided tests are conducted to compare them with the four newly obtained portfolios, and the corresponding p-values are reported in the four rows below the portfolio CER. We set $\eta = 4$ for the timing strategies KO_{VT}, KO_{RT}, and KO_{BT}.

	Momentum $N = 10$	Size-B/M $N = 25$	IVOL $N = 10$	OP-Inv $N = 25$	NM-V (LT) $N = 16$	NM-V (All) $N = 46$	Industry $N = 49$	Stocks $N = 100$
w^*	0.0119	0.0157	0.0094	0.0208	0.0212	0.0709	0.0171	0.0211
				Invariant Optimal Portfolio Rules				
$\boxed{\hat{w}_{q,t}}$	0.0058	0.0032	0.0006	0.0021	0.0025	−0.0324	−0.0003	−0.0147
$\hat{w}_{p,t}$	−0.0165	−5.4065	−0.0288	−0.1110	−0.0483	−17.6764	−1.4196	-627.0316
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
$\hat{w}_{u,t}$	−0.0100	−0.0595	−0.0251	−0.0566	−0.0312	−5.9473	−0.8411	−8.2292
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
$\hat{w}_{BS,t}$	0.0041	−0.0104	−0.0027	−0.0099	−0.0035	−0.4090	−0.0321	−0.2406
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
				Rules with Shrinkage Covariance Matrix Estimators				
$\boxed{\hat{w}_{q,t}^{LW2004}}$	0.0100	0.0087	0.0079	0.0061	0.0080	0.0607	0.0031	0.0008
$\hat{w}_{p,t}^{LW2004}$	0.0021	−0.0145	−0.0005	−0.0367	−0.0057	−0.3866	−0.1637	−17.9891
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
$\boxed{\hat{w}_{q,t}^{LW2017}}$	0.0078	0.0081	0.0036	0.0064	0.0053	0.0620	0.0035	0.0022
$\hat{w}_{p,t}^{LW2017}$	−0.0092	−0.0459	−0.0181	−0.0442	−0.0281	−2.7752	−0.4015	−0.4330
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
				Rules with MacKinlay-Pástor Single Factor Structure				
$\boxed{\hat{w}_{q,t}^{MP}}$	0.0044	0.0032	0.0033	0.0059	0.0027	0.0062	0.0040	0.0046
$\hat{w}_{p,t}^{MP}$	0.0034	−0.0008	−0.0014	−0.0021	−0.0022	−0.0018	0.0009	−0.0034
	0.11	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.08*	*0.00*
				Rule with No-Short-Sale Constraints				
$\hat{w}_{p,t}^{NS}$	0.0057	0.0045	0.0024	0.0041	0.0045	0.0089	−0.0001	−0.0005
	0.49	*0.67*	*0.68*	*0.73*	*0.70*	*0.98*	*0.53*	*1.00*
	0.02	*0.01*	*0.01*	*0.19*	*0.08*	*0.00*	*0.08*	*0.28*
	0.23	*0.06*	*0.37*	*0.18*	*0.41*	*0.00*	*0.06*	*0.11*
	0.87	*0.80*	*0.20*	*0.14*	*0.86*	*0.88*	*0.07*	*0.03*

Table 4: CER Comparison with 20 bps Transactions Costs (Cont'd)

	Momentum $N=10$	Size-B/M $N=25$	IVOL $N=10$	OP-Inv $N=25$	NM-V (LT) $N=16$	NM-V (All) $N=46$	Industry $N=49$	Stocks $N=100$
				Other Rules from Portfolio Optimization				
$\hat{w}_{g,t}$	0.0045	0.0044	0.0014	0.0039	0.0040	0.0004	0.0009	−0.0133
	0.35	*0.68*	*0.58*	*0.75*	*0.66*	*0.95*	*0.80*	*0.99*
	0.01	*0.00*	*0.01*	*0.10*	*0.03*	*0.00*	*0.01*	*0.00*
	0.14	*0.03*	*0.29*	*0.10*	*0.34*	*0.00*	*0.01*	*0.00*
	0.54	*0.87*	*0.02*	*0.01*	*0.87*	*0.00*	*0.02*	*0.00*
$\hat{w}^{NS}_{g,t}$	0.0033	0.0039	0.0040	0.0048	0.0024	0.0058	0.0040	0.0024
	0.23	*0.60*	*0.80*	*0.81*	*0.49*	*0.97*	*0.99*	*1.00*
	0.00	*0.00*	*0.06*	*0.27*	*0.01*	*0.00*	*0.79*	*0.96*
	0.08	*0.03*	*0.55*	*0.25*	*0.20*	*0.00*	*0.67*	*0.58*
	0.10	*0.75*	*0.82*	*0.16*	*0.39*	*0.39*	*0.49*	*0.03*
$\hat{w}_{KZ3,t}$	−0.7784	−0.1696	−3.5501	−0.1298	−1.4475	−5.3034	−0.1422	−242.7392
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
				Non-Optimization Rules				
$1/N$	0.0026	0.0036	0.0022	0.0033	0.0022	0.0013	0.0038	0.0023
	0.18	*0.55*	*0.65*	*0.65*	*0.47*	*0.95*	*0.95*	*1.00*
	0.00	*0.01*	*0.02*	*0.13*	*0.02*	*0.00*	*0.64*	*0.85*
	0.05	*0.04*	*0.36*	*0.13*	*0.20*	*0.00*	*0.56*	*0.51*
	0.07	*0.58*	*0.21*	*0.05*	*0.38*	*0.02*	*0.45*	*0.08*
KO_{VT}	0.0035	0.0043	0.0040	0.0044	0.0029	0.0046	0.0048	0.0036
	0.25	*0.64*	*0.80*	*0.77*	*0.54*	*0.96*	*1.00*	*1.00*
	0.00	*0.00*	*0.06*	*0.23*	*0.02*	*0.00*	*0.91*	*1.00*
	0.08	*0.05*	*0.55*	*0.22*	*0.24*	*0.00*	*0.84*	*0.94*
	0.16	*0.82*	*0.81*	*0.11*	*0.59*	*0.14*	*0.71*	*0.16*
KO_{RT}	0.0045	0.0044	0.0033	0.0047	0.0044	0.0059	0.0035	0.0015
	0.35	*0.65*	*0.75*	*0.80*	*0.69*	*0.97*	*0.97*	*1.00*
	0.01	*0.01*	*0.03*	*0.26*	*0.07*	*0.00*	*0.62*	*0.71*
	0.13	*0.06*	*0.47*	*0.24*	*0.40*	*0.00*	*0.51*	*0.28*
	0.54	*0.78*	*0.50*	*0.17*	*0.87*	*0.44*	*0.38*	*0.03*
KO_{BT}	0.0049	0.0046	0.0025	0.0046	0.0049	0.0052	0.0040	0.0025
	0.39	*0.67*	*0.68*	*0.78*	*0.73*	*0.97*	*0.96*	*1.00*
	0.01	*0.01*	*0.02*	*0.25*	*0.11*	*0.00*	*0.68*	*0.87*
	0.17	*0.08*	*0.39*	*0.24*	*0.46*	*0.00*	*0.60*	*0.57*
	0.69	*0.80*	*0.27*	*0.16*	*0.93*	*0.29*	*0.48*	*0.09*

Table 5: Sharpe Ratio Comparison with 20 bps Transaction Costs

This table reports the Sharpe ratio of the portfolios studied in this paper with $h = 120$, $\gamma = 3$, and a transaction cost of 20 bps, based on the eight datasets containing excess monthly returns. The four newly obtained optimal combining portfolios, i.e., $\hat{w}_{q,t}$, $\hat{w}_{q,t}^{LW2004}$, $\hat{w}_{q,t}^{LW2017}$, and $\hat{w}_{q,t}^{MP}$, are highlighted with a box around. In the three categories to which the four new portfolios belong, one-sided tests are conducted to assess the value of using the newly derived optimal combining coefficient $\hat{c}_t$, and the p-values are reported in *italics*. For the portfolios in the remaining three categories, one-sided tests are conducted to compare them with the four newly obtained portfolios, and the corresponding p-values are reported in the four rows below the portfolio Sharpe ratio. We set $\eta = 4$ for the timing strategies KO_{VT}, KO_{RT}, and KO_{BT}.

	Momentum $N=10$	Size-B/M $N=25$	IVOL $N=10$	OP-Inv $N=25$	NM-V (LT) $N=16$	NM-V (All) $N=46$	Industry $N=49$	Stocks $N=100$
w^*	0.2689	0.3070	0.2382	0.3531	0.3568	0.6713	0.3199	0.3586
				Invariant Optimal Portfolio Rules				
$\boxed{\hat{w}_{q,t}}$	0.2148	0.1707	0.1999	0.1392	0.1578	0.3848	0.0697	−0.0786
$\hat{w}_{p,t}$	0.1814	−0.0276	0.1685	0.0258	0.0841	−0.0721	−0.1628	−0.1350
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.16*
$\hat{w}_{u,t}$	0.1883	0.0751	0.1630	0.0616	0.1003	0.0380	−0.0818	−0.2101
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*
$\hat{w}_{BS,t}$	0.2109	0.1321	0.1931	0.1052	0.1391	0.2527	0.0074	−0.1836
	0.09	*0.00*	*0.01*	*0.01*	*0.03*	*0.00*	*0.02*	*0.01*
				Rules with Shrinkage Covariance Matrix Estimators				
$\boxed{\hat{w}_{q,t}^{LW2004}}$	0.2443	0.2384	0.2213	0.1925	0.2244	0.6381	0.1389	0.0765
$\hat{w}_{p,t}^{LW2004}$	0.2185	0.1575	0.1988	0.1043	0.1592	0.3511	−0.0237	−0.1266
	0.00	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
$\boxed{\hat{w}_{q,t}^{LW2017}}$	0.2292	0.2205	0.2086	0.1963	0.1826	0.6116	0.1488	0.1164
$\hat{w}_{p,t}^{LW2017}$	0.2014	0.1226	0.1831	0.1102	0.1143	0.1308	−0.0468	−0.1312
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
				Rules with MacKinlay-Pástor Single Factor Structure				
$\boxed{\hat{w}_{q,t}^{MP}}$	0.1641	0.1396	0.1413	0.2103	0.1266	0.2236	0.1623	0.1742
$\hat{w}_{p,t}^{MP}$	0.1453	0.0838	0.0628	0.0548	0.0424	0.0533	0.0926	0.0555
	0.11	*0.04*	*0.00*	*0.00*	*0.00*	*0.00*	*0.08*	*0.01*
				Rule with No-Short-Sale Constraints				
$\hat{w}_{p,t}^{NS}$	0.1867	0.1642	0.1223	0.1564	0.1648	0.2473	0.0930	0.0858
	0.20	*0.43*	*0.02*	*0.64*	*0.55*	*0.03*	*0.69*	*1.00*
	0.02	*0.01*	*0.00*	*0.20*	*0.08*	*0.00*	*0.13*	*0.59*
	0.10	*0.06*	*0.01*	*0.19*	*0.36*	*0.00*	*0.09*	*0.23*
	0.85	*0.79*	*0.19*	*0.07*	*0.88*	*0.69*	*0.09*	*0.03*

Table 5: Sharpe Ratio Comparison with 20 bps Transaction Costs (Cont'd)

	Momentum $N=10$	Size-B/M $N=25$	IVOL $N=10$	OP-Inv $N=25$	NM-V (LT) $N=16$	NM-V (All) $N=46$	Industry $N=49$	Stocks $N=100$
				Other Rules from Portfolio Optimization				
$\hat{w}_{g,t}$	0.1696	0.1702	0.0942	0.1554	0.1597	0.0681	0.0816	−0.0641
	0.10	*0.49*	*0.00*	*0.66*	*0.52*	*0.00*	*0.66*	*0.97*
	0.01	*0.01*	*0.00*	*0.13*	*0.05*	*0.00*	*0.01*	*0.00*
	0.05	*0.06*	*0.00*	*0.13*	*0.31*	*0.00*	*0.01*	*0.00*
	0.62	*0.90*	*0.02*	*0.01*	*0.88*	*0.00*	*0.02*	*0.00*
$\hat{w}^{NS}_{g,t}$	0.1414	0.1563	0.1634	0.1778	0.1211	0.2167	0.1675	0.1209
	0.03	*0.36*	*0.16*	*0.79*	*0.25*	*0.01*	*0.99*	*1.00*
	0.00	*0.01*	*0.04*	*0.37*	*0.01*	*0.00*	*0.82*	*0.96*
	0.01	*0.04*	*0.11*	*0.34*	*0.12*	*0.00*	*0.73*	*0.58*
	0.11	*0.77*	*0.86*	*0.13*	*0.40*	*0.42*	*0.56*	*0.04*
$\hat{w}_{KZ3,t}$	−0.0185	−0.0531	−0.0790	−0.0386	−0.0778	−0.2694	−0.0509	−0.0435
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.74*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.02*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
				Non-Optimization Rules				
$1/N$	0.1269	0.1477	0.1217	0.1418	0.1202	0.1026	0.1512	0.1175
	0.01	*0.29*	*0.02*	*0.52*	*0.25*	*0.00*	*0.95*	*1.00*
	0.00	*0.01*	*0.00*	*0.14*	*0.02*	*0.00*	*0.61*	*0.87*
	0.00	*0.03*	*0.01*	*0.14*	*0.13*	*0.00*	*0.52*	*0.51*
	0.07	*0.60*	*0.23*	*0.02*	*0.42*	*0.01*	*0.40*	*0.07*
KO_{VT}	0.1453	0.1622	0.1610	0.1687	0.1323	0.1732	0.1875	0.1544
	0.03	*0.42*	*0.15*	*0.72*	*0.32*	*0.00*	*1.00*	*1.00*
	0.00	*0.01*	*0.03*	*0.30*	*0.02*	*0.00*	*0.93*	*1.00*
	0.01	*0.06*	*0.10*	*0.28*	*0.17*	*0.00*	*0.87*	*0.94*
	0.17	*0.82*	*0.82*	*0.08*	*0.59*	*0.10*	*0.77*	*0.23*
KO_{RT}	0.1664	0.1622	0.1402	0.1718	0.1629	0.2008	0.1469	0.1004
	0.09	*0.42*	*0.05*	*0.75*	*0.54*	*0.01*	*0.96*	*1.00*
	0.01	*0.01*	*0.01*	*0.31*	*0.08*	*0.00*	*0.59*	*0.76*
	0.04	*0.06*	*0.03*	*0.29*	*0.35*	*0.00*	*0.48*	*0.30*
	0.55	*0.77*	*0.48*	*0.11*	*0.88*	*0.29*	*0.35*	*0.02*
KO_{BT}	0.1733	0.1658	0.1259	0.1691	0.1723	0.1840	0.1554	0.1221
	0.13	*0.45*	*0.03*	*0.72*	*0.60*	*0.00*	*0.96*	*1.00*
	0.02	*0.02*	*0.00*	*0.30*	*0.13*	*0.00*	*0.65*	*0.90*
	0.06	*0.08*	*0.02*	*0.28*	*0.42*	*0.00*	*0.56*	*0.57*
	0.70	*0.80*	*0.28*	*0.10*	*0.93*	*0.18*	*0.44*	*0.08*