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

MS 2022（3）

汇报人：周洁

2023.03.22

Raymond Kan

簡文豹

Current Appointments

Professor of Finance , Rotman School of Management, University of Toronto

Education

Ph.D. Finance, 1994, University of Chicago, Graduate School of Business

M.B.A. Finance and Statistics, 1986, University of Chicago, Graduate School of Business

B.B.A. Accounting， 1983, Chinese University of Hong Kong

research interest

Empirical Asset Pricing, Portfolio Management, Computational Statistics

- Wishart Matrix 威希特分布, forthcoming
- “Model Comparison with Sharpe Ratios” with Francisco Barillas, Cesare Robotti, and Jay Shanken, in Journal of Financial and Quantitative Analysis,2020
- “Too Good to Be True? Fallacies in Evaluating Risk Factor Models” with Nikolay Gospodinov and Cesare Robotti, in Journal of Financial Economics,2019

Xiaolu Wang

Current Appointments

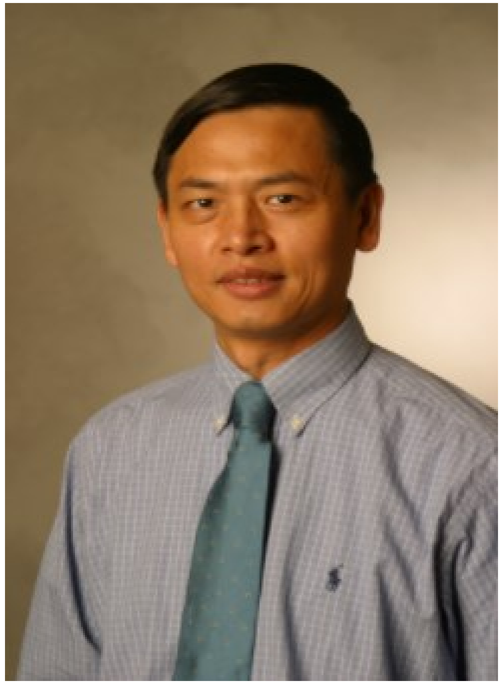
2010-至今， Associate Professor, Ivy College of Business, Iowa State University 爱荷华州立大学

Education

Ph.D. Finance, Rotman School of Management, University of Toronto, 2004-2010

- “Optimal Portfolio Choice with Unknown Benchmark Efficiency” with Raymond Kan, 2021
- “In-Sample and Out-of-Sample Sharpe Ratios of Multi-Factor Asset Pricing Models” with Raymond Kan, Xinghua Zheng, 2022

Guofu Zhou



Current Appointments

Professor of Finance, Olin Business School, Washington University

Education

Ph.D., Economics, Duke University, 1990 (08/1986–05/1990)
M.A., Mathematics, Duke University, 1987 (08/1985–05/1986)
M.S., Numerical Analysis Academia Sinica, China, 1985 (1982–1985)
B.S., Mathematics, Chengdu University of Technology, China, 1982

Research Areas

Empirical asset pricing, investor sentiment, big data, machine learning, asset allocation, portfolio optimization, anomalies, predictability, technical analysis, behavioral finance, household finance, mutual funds, hedge funds, options, Bayesian inference, and Chinese financial markets

- "Risk Momentum: A New Class of Price Patterns" with Sophia Zhengzi Li and Peixuan Yuan
- "Are Bond Returns Predictable with Real-Time Macro Data?" with Dashan Huang, Fuwei Jiang, Kunpeng Li and Guoshi Tong , Journal of Econometrics, forthcoming.
- "Winners from Winners: A Tale of Risk Factors" with Siddhartha Chib and Lingxiao Zhao Management Science, forthcoming.

- **（1）estimation risk**

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.

When estimated parameters instead of true parameters are used, the plug-in portfolio p underperforms the true optimal portfolio p∗ because of estimation errors.

- （2）**the plug-in rule**

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

In practice, μ and Σ are unknown, and they need to be estimated

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

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

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

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

- （3）**the optimal combining strategy（one of Invariant Optimal Portfolio Rules）**

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

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

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

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

- （4）Shrinkage covariance matrices estimators

压缩估计的方法来源于 *Stein*（1955）。*Stein* 估计量实际上是贝叶斯估计，即将统计量向先验信息压缩。一种简单的方法就是对估计量和先验估计量以恰当的权重取平均，而给先验估计量的权重就是压缩强度。

从统计的角度来看，在协方差矩阵估计中，样本协方差是完全基于数据的，而先验的协方差矩阵可以来源于主观判断、历史经验或者模型等。样本协方差矩阵是无偏的，但是含有大量的估计误差；而先验的协方差矩阵由于其具有较严格的假设，是有设定偏差的，但是因为待估计的参数较少，其具有较少的估计误差。因而，可以将样本协方差向一个先验的协方差矩阵压缩，以减少样本协方差的估计误差，从而在设定偏差与估计误差间达到一个最优的平衡。

从金融的角度来看，压缩估计量也可以看作两个极端的折中。样本协方差矩阵可以认为是一个N-因子模型，即每只股票都是一个因子，并且该模型中没有残差；而压缩目标是一个简单的模型。

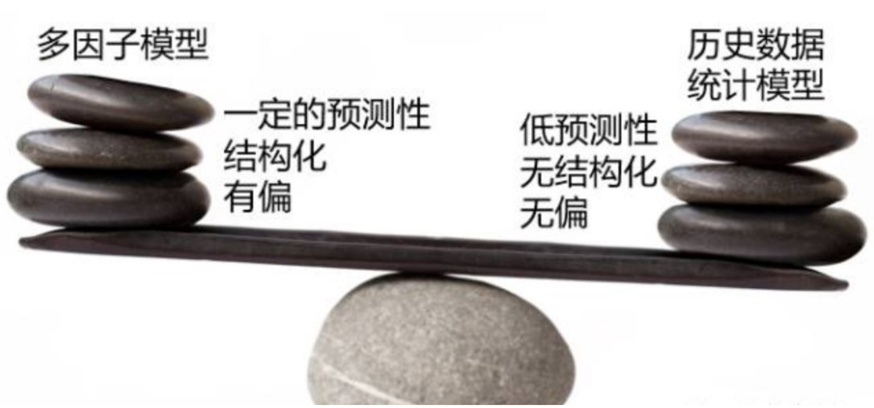
因而，协方差矩阵的压缩估计量可以表示为样本协方差矩阵与压缩目标的线性组合，即

$$\Sigma(\alpha) = \alpha F + (1-\alpha)S$$

其中，F为压缩目标，S为样本协方差矩阵，α为压缩强度。

收益率预测的贝叶斯收缩

贝叶斯收缩以多因子模型得出的收益率作为**先验**（prior），以实际收益率（历史数据）作为**新的观测值**（observation），计算出收益率均值的**后验**（posterior）作为最终预测。形象的说，该方法结合了两种方法，**以最优的比例使基于历史数据的预测向基于多因子模型的预测“收缩”。**这个最优的比例使得预测的期望误差最小。**贝叶斯收缩相当于给历史收益率数据提供了多因子模型能提供的额外有效信息**，从而得到更加有效的预测。



- （4）**Shrinkage covariance matrices 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

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

（5）Rules with MacKinlay-P'astor 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

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

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.（命题和引理中）

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

1. Introduction
2. Theoretical Results
    2.1. The Setup
    2.2. Estimation Risk （Proposition 1 Lemma 1 Lemma 2）
    2.3. Optimal Combining Coefficient （Proposition 2 Lemma 3）
3. Portfolio Rules
    3.1. Invariant Optimal Portfolio Rules
    3.2. Rules with Shrinkage Covariance Matrix Estimators
    3.3. Rules with MacKinlay-P´astor Single Factor Structure
    3.4. Rule with No-Short-Sale Constraints
    3.5. Other Rules from Portfolio Optimization
        3.5.1. Sample GMV Portfolio
        3.5.2. GMV with No-Short-Sale Constraints
        3.5.3. Normalized Kan-Zhou (2007) Three-Fund Rule
    3.6. Non-optimization Rules
        3.6.1. The 1/N Rule
        3.6.2. Volatility Timing
        3.6.3. Reward-to-Risk Timing
4. Empirical Results
    4.1. Certainty Equivalent Return
    4.2. Sharpe Ratio
    4.3. Turnover and Performance Net of Trading Costs
5. Conclusion

1. Introduction

In this paper, we focus on the case without a risk-free asset and make several contributions.

①for 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.--derive an optimal portfolio rule that minimizes the utility loss under estimation risk, and takes the form of optimally combining the sample GMV portfolio with a sample zero-investment portfolio.

②we 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 P´ astor (2000) to form new optimal portfolios.

③we 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).

2. Theoretical Results

2.1. The Setup

2.2. Estimation Risk

2.3. Optimal Combining Coefficient

2.1. The Setup

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

The investor chooses his portfolio weights w to maximize the mean-variance utility function

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

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

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

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

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

$$\sigma^2_{p^*} = \sigma^2_g + \frac{\psi^2}{\gamma^2}, \tag{5}$$

where $\sigma^2_g = 1/(1'_N \Sigma^{-1} 1_N)$ is the variance of the GMV portfolio and $\psi^2 = \mu' \Sigma^{-1} \mu - \mu^2_g / \sigma^2_g$ 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

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

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).

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 $\hat{\mu}_{g,t} = (1_N' \hat{\Sigma}_t^{-1} \hat{\mu}_t)/(1_N' \hat{\Sigma}_t^{-1} 1_N)$.

2.2. Estimation Risk

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.

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

命题1

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*

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

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*

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

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

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

引理1 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*

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

引理2 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*

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

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

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

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

2.3. Optimal Combining Coefficient

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

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.

命题2

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

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*

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

$$\sigma^2(c_t) = E[\sigma_t^2(c_t)] + E[\mu_t(c_t)^2] - E[\mu_t(c_t)]^2$$

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

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

引理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

3.1. Invariant Optimal Portfolio Rules

3.2. Rules with Shrinkage Covariance Matrix Estimators

3.3. Rules with MacKinlay-P´astor Single Factor Structure

3.4. Rule with No-Short-Sale Constraints

3.5. Other Rules from Portfolio Optimization

3.5.1. Sample GMV Portfolio

3.5.2. GMV with No-Short-Sale Constraints

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

3.6. Non-optimization Rules

3.6.1. The 1/N Rule

3.6.2. Volatility Timing

3.6.3. Reward-to-Risk Timing

3. Portfolio Rules

3.1. Invariant Optimal Portfolio Rules

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

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]

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

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

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

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 $i = 1, \cdots, N$,

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

3.2. Rules with Shrinkage Covariance Matrix Estimators

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.

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

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

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 $\hat{w}_{p,t}^{NS}$.

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

The portfolio optimally combines the risk-free asset, the sample tangency portfolio, and the sample GMV portfolio.

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.

3.6.1. The 1/N Rule

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

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

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

3.6.2. Volatility Timing

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

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

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

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

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

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.

4.1. Certainty Equivalent Return (CER)

4.2. Sharpe Ratio

4.3. Turnover and Performance Net of Trading Costs

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 (the single factor structure) continue to outperform the one using $\hat{c}_t$ alone.

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.

- 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.
- 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.