Three Essays in Asset pricing and Portfolio Choice

Botshekan, M.

2012

Chapter 4

Optimal Versus Naive Diversification in Factor Models

4.1 Introduction

Markowitz (1952) provides a solid framework for mean-variance based optimal portfolio selection. If, however, the true parameters of the return generating process are unknown and have to be estimated, the framework is subject to serious estimation error problems that limit its practical usefulness, see for example Jobson and Korkie (1980). Many studies have attempted to improve the estimation procedure and mitigate the estimation error problem, for example by Bayesian methods (see for example Jorion (1986), Pástor (2000)), shrinkage methods (see for example Ledoit and Wolf (2004), Wang (2005)), imposing a factor structure on the returns (MacKinlay and Pastor (2000)), or by optimally combining the tangency portfolio, the risk free rate and the global minimum variance portfolio (Kan and Zhou (2007)). However, DeMiguel, Garlappi, and Uppal (2009) (DGU (2009) hereafter) show that the out-of-sample performance based on such alternative methods do not really improve over the very simple, naive "$1/N$" rule, i.e., investing equally weighted across N risky assets. Tu and Zhou (2011), by contrast, show that a combination of the $1/N$ rule with one of the earlier optimization based methods proposed by Markowitz (1952), Jorion (1986), MacKinlay and Pastor (2000) or Kan and Zhou (2007) significantly outperforms with respect to the use of a single rule only.

To compare the performance of the $1/N$ rule with other theory based methods, one common approach is to generate asset returns using a factor model in a simulation setting and compare the out-of-sample performance of optimal portfolio strategies with that of the naive diversification rule. For example DGU (2009), following MacKinlay and Pastor (2000), use a one factor model for generating returns. They conclude that the estimation window needed for sample-based mean-variance strategy and its extensions to beat the $1/N$ rule is around 3000 months for a portfolio with 25 assets, and about 6000 months for a portfolio with 50 assets. Tu and Zhou (2011) also use the same simulation design for comparing the performance of their combined strategies with the $1/N$ rule. In addition, they report the results of an experiment based on a three-factor design. In contrast to the one-factor settings, the combination rules outperform naive diversification in the three-factor settings even for short estimation window of 120 months.

So far, little is known about the influence of the factor model structure underlying the return generating process on the relative performance of optimal and naive portfolio strategies. In this paper we therefore compare several portfolio strategies under a variety of data generating structures and try to characterize under what data generating structure, optimization-based portfolio strategies could outperform the naive strategy. Our focus in this paper is on the Sharpe ratio as a performance measure for different portfolio strategies. Although the Sharpe ratio has limitations when applied to non-normal data (see for example Goetzmann, Ingersoll, Spiegel, and Welch (2002)), it is still one of the most commonly used measures to compare performance of portfolio strategies in academic literature and in practice.

First, we show analytically and numerically that when the data are generated by a one-factor model, there is hardly any difference between the Sharpe ratio of the optimal tangency portfolio and that of a naive diversification strategy when there is no estimation error. As the Sharpe ratio of the optimal mean-variance strategy (without parameter uncertainty) is the highest attainable Sharpe ratio, there is very limited room for other portfolio strategies to outperform the naive diversification in an out-of-sample context. Moreover, if parameters have to be estimated, the naive diversification strategy is likely to outperform the other strategies empirically. Simulation designs to compare different portfolio strategies are therefore not very informative if based on a one-factor data generating process. We illustrate this by comparing the performance of a number of portfolio strategies proposed in the literature using data generated from a one-factor model under different values for idiosyncratic variances and mispricing.

We also drive the analytical formulas for the Sharpe ratios of the naive and mean-variance strategies when data are generated by a general factor model. We use the results to show how different parameter values affect the Sharpe ratios of the optimal and naive strategy. We characterize circumstances under which the optimal mean-variance strategy substantially outperforms the $1/N$ rule, such that some outperformance might be retained even after accounting for estimation error. We show that, however, these circumstances are hardly satisfied in practice. Consequently, also in this settings, the ability of other portfolio strategies to outperform naive diversification out-of sample is very limited.

To further investigate the effect of factor structures on the Sharpe ratios of portfolio strategies, we also consider empirical data. We use equity portfolios as well as different asset classes. When there are sufficient factors driving underlying asset returns, the difference between Sharpe ratios of the optimal and naive diversification strategies without parameter uncertainty can become substantial in empirically relevant regions. Some of this differences can be retained by a number of portfolio strategies whose parameters have to be estimated. In particular, the combination strategies proposed by Tu and Zhou (2011) (in which portfolio weights from the naive and the mean variance strategies are optimally combined), the minimum variance strategy, and the volatility timing strategy (similar to strategies proposed by Kirby and Ostdiek (2012)) outperform naive diversification in a number of empirical tests.

Considering a factor structure for asset returns has important implications for both academics and practitioners. Academics should consider that generating returns by one-factor or two-factor models in simulation studies is not informative for comparing the performance of optimal portfolio strategies with naive diversification. In a practical context, when there are more factors driving asset returns, there is more potential for optimal portfolio strategies to outperform naive diversification.

The remainder of this paper is organized as follows. In Section 4.2, we present the analytical results. Section 4.3 provides the background for the different portfolio choice methods that are used. Section 4.4 discusses the simulation results. Section 4.5 contains the empirical results. Section 4.6 concludes.

4.2 Analytical Results

In this section, we assume that the data are generated by a factor model. We derive the Sharpe ratio for the optimal tangency portfolio and for the naive diversification strategy. By comparing these two Sharpe ratios, we can assess the maximum possible increase of the mean variance Sharpe ratio over that of the naive diversification strategy. This increase holds for the case without parameter uncertainty. Parameter estimation error decreases the attainable increase in Sharpe ratio. We investigate this further in Section 4.4.

4.2.1 Mean-Variance portfolio

In the standard mean-variance (MV) model of Markowitz (1952), investors have a quadratic utility function and optimize the tradeoff between risk (measured by the variance of portfolio returns) and expected return (measured by the mean of portfolio returns). Formally, we assume that the investor selects the $N \times 1$ vector of portfolio weights w_t to maximize the utility function

$$w_t'\mu_t - \frac{\gamma}{2} w_t'\Sigma_t w_t, \tag{4.1}$$

where r_{t+1} is the vector of the risky assets' excess returns with mean $\mu_t = \mathbb{E}_t(r_{t+1})$ and covariance matrix $\Sigma_t = \mathbb{E}_t(r_{t+1}r_{t+1}') - \mu_t\mu_t'$, and with γ the investor's risk aversion coefficient.

The solution to this maximization problem is the well-known MV optimal portfolio

$$w_{mv,t} = \frac{1}{\gamma}\Sigma_t^{-1}\mu_t. \tag{4.2}$$

Any remaining wealth is invested in the risk free asset, so $w_{R_f,t} = 1 - \mathbf{1}'w_{mv,t}$, with R_f the risk free rate and $\mathbf{1} = (1, \ldots, 1)'$ an $N \times 1$ vector of ones. One important portfolio is when the investor wants to put all his wealth in the risky assets. In this case, the weights on the risky part of portfolio should add to one, namely $\mathbf{1}'w_t = 1$ and we obtain the tangency portfolio

$$w_{tan,t} = \frac{\frac{1}{\gamma}\Sigma_t^{-1}\mu_t}{|\frac{1}{\gamma}1_N'\Sigma_t^{-1}\mu_t|} = \frac{\Sigma_t^{-1}\mu_t}{|1_N'\Sigma_t^{-1}\mu_t|}. \tag{4.3}$$

Note that the Sharpe ratio of the tangency portfolio is the same as the Sharpe ratio of any linear combination of the tangency portfolio and the risk free asset.

If μ_t and Σ_t are known, the Sharpe ratio of the mean-variance strategy is given by

$$SR_{mv} = \frac{w_{mv,t}'\,\mu_t}{\sqrt{w_{mv,t}'\;\Sigma_t\;w_{mv,t}}} = \sqrt{\mu_t'\,\Sigma_t^{-1}\,\mu_t}\;. \tag{4.4}$$

The last equality follows directly from equation (4.2) or (4.3).

4.2.2 One-Factor Model

We first assume that there is one systematic (market) factor that generates the common variation in returns. The remaining variation is idiosyncratic and can be diversified in a large portfolio. The following proposition states our main result for the Sharpe ratio for this model.

Proposition 1. Consider the factor model

$$r = \beta r_m + \epsilon\,, \tag{4.5}$$

where r is an $N \times 1$ vector of excess returns, β an $N \times 1$ vector of factor loadings, r_m the market excess return with mean μ_m and variance σ_m^2, and ϵ an $N \times 1$ vector of idiosyncratic risks with zero mean and variance-covariance matrix $\sigma_\epsilon^2\, I$. If $\tilde{w}$ denotes the weights of the optimal mean-variance portfolio, then the Sharpe ratio for $\tilde{w}$ is given by

$$SR_{mv} = \frac{\mu(\tilde{w})}{\sqrt{\sigma^2(\tilde{w})}} = \frac{\mu_m}{\sqrt{\sigma_m^2 + \sigma_\epsilon^2/q}}, \tag{4.6}$$

where $q = \beta'\beta$.

For the naive diversification strategy $\bar{w} = \mathbf{1}/N$ with $\mathbf{1} = (1, \ldots, 1)' \in \mathbb{R}^{N\times 1}$ and the Sharpe ratio is given by

$$SR_e = \frac{\mu_m}{\sqrt{\sigma_m^2 + \sigma_\epsilon^2/(N\bar{\beta}^2)}}, \tag{4.7}$$

where $\bar{\beta} = \bar{w}'\beta$

Proof. See the Appendix.

As the number of assets N goes to infinity, $q = \beta'\beta$ in equation (4.6) goes to infinity and $s(\tilde{w}) \to \mu_m/\sigma_m$.

Similarly, in equation (4.7), as $N \to \infty$, the effect of idiosyncratic volatility washes out and $s(\bar{w}) \to \mu_m/\sigma_m$ as well. So if the number of assets is high and the asset returns are generated by a one factor structure, the naive diversification strategy and the mean-variance optimal strategy yield approximately the same Sharpe ratios. This already holds if the parameters are known, as long as the average β ($\bar{\beta}$) is not too small. In a one factor model, $\bar{\beta}$ is typically close to one. If we introduce parameter uncertainty, we expect the Sharpe ratio of the mean-variance strategy to deteriorate, whereas the Sharpe ratio of the $1/N$ strategy remains unchanged. As a result, we expect the $1/N$ strategy to outperform the mean-variance strategy, almost irrespective of the estimation method used. This is confirmed by our simulations in Section 4.4.

4.2.3 Multi-Factor Model

There is a large finance literature that suggests there is more than one systematic factor driving returns, see for example Fama and French (1993), Carhart (1997), Pastor and Stambaugh (2003), Campbell and Vuolteenaho (2004) and Botshekan, Kraeussl, and Lucas (2012). In this section we derive Sharpe ratios for the mean-variance optimal and the naive strategy if the data are generated by a multi-factor structure. The results are summarized in the following proposition.

Proposition 2. Consider the factor model,

$$r = \beta f + \varepsilon, \tag{4.8}$$

where r and ε are $N \times 1$ vectors, β is an $N \times K$ matrix, and f is a $K \times 1$ vector. The covariance matrices of r, f, and ε are denoted as V_r, V_f, and V_ε, respectively, and their means as μ_r, μ_f, and 0 respectively.

The Sharpe ratio for the equally weighted portfolio is

$$SR_e = \sqrt{\mu_f' \bar{B} \left(V_f^{-1} + \bar{B}\right)^{-1} V_f^{-1} \mu_f}, \tag{4.9}$$

with $\bar{\beta} = \beta' \mathbf{1}/N$ and $\bar{B} = N^2 \bar{\beta}\bar{\beta}'/(\mathbf{1}' V_\varepsilon \mathbf{1})$.

The Sharpe ratio for the mean-variance optimal portfolio is

$$SR_{mv} = \sqrt{\mu_f' B \left(V_f^{-1} + B\right)^{-1} V_f^{-1} \mu_f}, \tag{4.10}$$

where $B = \beta' V_\varepsilon^{-1} \beta$.

If without loss of generality we normalize the factors and factor loadings such that $V_f = \mathrm{I}$, the difference between the squared Sharpe ratios is

$$SR_{mv}^2 - SR_e^2 = \mu_f' (\mathrm{I} + \bar{B})^{-1} (B - \bar{B}) (\mathrm{I} + B)^{-1} \mu_f. \tag{4.11}$$

Proof. See the Appendix.

From equation (4.11), it is clear that the difference between the two Sharpe ratios only comes form the factor returns μ_f and the matrices B and $\bar{B}$. Particularly the difference between B and $\bar{B}$ plays a important role, and this in turn is closely related to the cross-sectional covariance matrix of the βs. See the Appendix for more details. This means that both the magnitude of the βs (in $\bar{B}$ and B) and their dispersion (in $B - \bar{B}$) are relevant for the difference between the Sharpe ratios.

4.3 Estimation strategies

In this section, we describe the different estimation methods and portfolio weight construction strategies to optimize mean-variance portfolio behavior. The performance of these strategies helps us to understand to what extent estimation error annihilates any portfolio gains of optimal mean-variance portfolio choice for empirical data.

4.3.1 Sample-Based Mean-Variance Portfolios

In practical situations, the mean and covariance matrix of excess returns are unknown and need to be estimated. In the sample-based mean-variance strategy, we estimate these parameters by their sample counterparts, $\hat{\mu}_t = T_w^{-1} \sum_{i=0}^{T_w-1} r_{t-i}$ and $\hat{\Sigma}_t = T_w^{-1} \sum_{i=0}^{T_w-1} (r_{t-i} - \hat{\mu}_t)(r_{t-i} - \hat{\mu}_t)'$, where T_w is the window length used for the estimation of the mean and the covariance matrix. The estimated parameters are used directly in the equation for optimal weights, $w_t \propto \Sigma_t^{-1} \mu_t$. This strategy thus ignores the potential effect of estimation risk on optimal portfolio choice.

The Sharpe ratio (and other performance metrics) for this strategy are based on a rolling window approach. Using a window of T_w observations, we estimate $\hat{\mu}$ and $\hat{\Sigma}_t$ and the optimal portfolio weights $\hat{w}_t$. The portfolio weights, in turn, are used to compute the portfolio return over period $t+1$ as $r_{\hat{m}\hat{v},t+1} = \hat{w}_t' r_{t+1}$. The estimation window is then rolled one period forward and the whole process is repeated. Based on all the portfolio returns $r_{\hat{m}\hat{v},t+1}$, we calculate the Sharpe ratio as $\hat{SR}_{\hat{m}\hat{v}} = \hat{\mu}_{\hat{m}\hat{v}} / \hat{\sigma}_{\hat{m}\hat{v}}$ where $\hat{\mu}_{\hat{m}\hat{v}}$ and $\hat{\sigma}_{\hat{m}\hat{v}}$ are the sample mean and standard deviation of the $r_{\hat{m}\hat{v},t}$ for $t = T_w + 1, \ldots, T$ where T denotes the complete sample size of the original (excess) returns r_t.

It is well-known that the Sharpe ratio of the optimal sample-based mean-variance strategy is prone to estimation error. For example, Chopra and Ziemba (1993) show that estimation error in expected returns is more costly than estimation error in the covariance matrix. Kan and Zhou (2007) further show that when the ratio of the number of assets to the length of the estimation window is small (for example, for 10 assets and a window length of 120 months), the interaction effect of the estimation error in the mean and covariance matrix can be much more sever than the individual effects of estimation error in the mean and covariance matrix added together. To disentangle these separate effects, we compute the Sharpe ratios of three additional strategies in our simulations, namely "Estimated Mean-Known Covariance", "Known Mean-Estimated Covariance" and "Known Mean-Known Covariance" strategies.

For the Sharpe ratio of the known mean-known variance strategy, denoted as $\hat{SR}_{mv}$, the approach is as follows. Using the true mean μ and covariance matrix Σ, the optimal weight $w_{mv} \propto \Sigma^{-1}\mu$ is computed and used to calculate the portfolio return $r_{mv,t+1} = w'_{mv}r_{t+1}$. The Sharpe ratio of this strategy is then given by $\hat{SR}_{mv} = \hat{\mu}_{mv}/\hat{\sigma}_{mv}$, with $\hat{\mu}_{mv}$ and $\hat{\sigma}_{mv}$ the sample mean and standard deviation of the $r_{mv,t}$, $t = T_w + 1, \ldots, T$, respectively. Similarly, the Sharpe ratio for the Estimated Mean-Known Variance strategy ($\hat{SR}_{\hat{m}v}$) and Know Mean-Estimated Variance strategy ($\hat{SR}_{m\hat{v}}$) are given by $\hat{SR}_{\hat{m}v} = \hat{\mu}_{\hat{m}v}/\hat{\sigma}_{\hat{m}v}$ and $\hat{SR}_{m\hat{v}} = \hat{\mu}_{m\hat{v}}/\hat{\sigma}_{m\hat{v}}$, respectively, where $\hat{\mu}_{\hat{m}v}$ and $\hat{\mu}_{m\hat{v}}$ and $\hat{\sigma}_{\hat{m}v}$ and $\hat{\sigma}_{m\hat{v}}$ are the sample means and variances of $r_{\hat{m}v,t+1} = w'_{\hat{m}v,t}r_{t+1}$ and $r_{m\hat{v},t+1} = w'_{m\hat{v},t}r_{t+1}$ respectively, with $w'_{\hat{m}v,t} \propto \Sigma^{-1}\hat{\mu}_t$, and $w'_{m\hat{v},t} \propto \hat{\Sigma}^{-1}\mu_t$.

Our final direct mean-variance strategy is based on objective function (4.1), with estimated mean $\hat{\mu}$ and covariance matrix $\hat{\Sigma}_t$ but with no-short-sale constraints imposed. We refer to this strategy as "constrained mean-variance strategy" and its weights as $w_{\hat{m}\hat{v},cons,t}$. This weights can be used to compute the Sharpe ratio in the same way as before.

4.3.2 Naive portfolio

In the naive (or $1/N$) strategy, we allocate a fraction of $1/N$ of current wealth to each risky asset in the portfolio. The implementation of this strategy does not require any optimization, nor does it require any data for determining the portfolio weights. As Kritzman, Page, and Turkington (2010) argue, the $1/N$ strategy has several advantages: it never shorts any asset, it avoids concentration, and at re-balancing times it sells high and buys low, thus exploiting a possible mean-reversion effect. We use two different performance metrics for this strategy.

First, we compute the Sharpe ratio of the $1/N$ or equally weighted strategy based on the true mean μ and covariance matrix Σ . The result is given by

$$SR_e = \frac{\mu_e}{\sigma_e} = \frac{w'_e\mu}{w'_e\Sigma_e w_e} = \frac{\mathbf{1}'\mu}{\sqrt{\mathbf{1}'\Sigma_u\mathbf{1}}} \tag{4.12}$$

where $w_e = \mathbf{1}/N$ and the subscript e denotes the equally weighted strategy. This Sharpe ratio can be directly computed to the optimal mean-variance from equation (4.4).

Second, we compute the Sharpe ratio of the $1/N$ strategy based on the sample returns. The result is $\hat{SR}_e = \hat{\mu}_e/\hat{\sigma}_e$, where $\hat{\mu}_e$ and $\hat{\sigma}_e$ are the mean and standard deviation of $\mathbf{1}'r_t/N$ for $t = T_w + 1, \ldots, T$. Note that $\hat{SR}_e$ can be compared directly to $\hat{SR}_{mv}$, $\hat{SR}_{\hat{m}v}$, $\hat{SR}_{m\hat{v}}$ and $\hat{SR}_{\hat{m}\hat{v}}$.

4.3.3 Combinations of the Mean-Variance and Naive portfolio

One of the advantages of the $1/N$ strategy is that it does not require any parameter estimation. This is important as estimation error may annihilate any potential gains from exploiting means, variances, and covariances of asset returns. The advantages, however, critically depend on the precision with which all of these parameters can be estimated. If the estimation can be done relatively more precisely, one could put more weight on the optimal mean-variance strategy. If estimation is hard and the sample is not very informative on the model's parameters, the converse holds and one is likely to put more emphasis on the naive strategy.

Following Tu and Zhou (2011), we therefore also report the Sharpe ratio of a combination strategy that tries to optimally combine portfolio weights from the naive and the mean variance strategies. We can write the optimal combination rule can be write as

$$w_{c,t} = \delta_t \frac{T_w - N - 2}{T_w} w_{\hat{m}\hat{v}} + (1 - \delta_t) w_e \,, \tag{4.13}$$

with $w_{\hat{m}\hat{v}}$ and w_e as defined before, and δ_t denoting the optimal combination coefficient.

Tu and Zhou (2011) estimate δ by

$$\hat{\delta}_t = \frac{\hat{\pi}_{1,t}}{\hat{\pi}_{1,t} + \hat{\pi}_{2,t}} \,, \tag{4.14}$$

where $\hat{\pi}_{1,t}$ and $\hat{\pi}_{2,t}$ are given by

$$\hat{\pi}_{1,t} = w_e' \hat{\Sigma}_t w_e - \frac{2}{\gamma} w_e' \hat{\mu}_t + \frac{1}{\gamma^2} \tilde{\theta}_t^2 \,, \tag{4.15}$$

$$\hat{\pi}_{2,t} = \frac{1}{\gamma^2}(c_1 - 1)\tilde{\theta}_t^2 + \frac{c_1}{\gamma^2} \frac{N}{T_w} \,, \tag{4.16}$$

where $c_1 = (T_w - 2)(T_w - N - 2)/(T_w - N - 1)(T_w - N - 4)$, and $\tilde{\theta}^2$ is an estimator of the squared Sharpe ratio proposed by Kan and Zhou (2007). This strategy can only be used if $T_w > N + 4$.

As it turns out later that the constrained mean-variance portfolio has several advantages over the raw mean-variance strategy, we also propose a combination of the constrained mean-variance portfolio and the naive portfolio. For this combination, we use the same combination coefficient δ_t as presented in (4.14), such that the allocation weights of this strategy are given by

$$w_{cc,t} = \delta_t \frac{T_w - N - 2}{T_w} w_{c,t} + (1 - \delta_t) w_e \,. \tag{4.17}$$

The Sharpe ratio is given by $\hat{SR}_{cc} = \hat{\mu}_{cc}/\hat{\sigma}_{cc}$, with $\hat{\mu}_{cc}$ and $\hat{\sigma}_{cc}$ the sample mean and standard deviation of $w_{cc,t-1}' r_t$ for $t = T_w + 1, \ldots, T$.

4.3.4 Bayesian Approach

Pioneered by Zellner and Chetty (1965) and Bawa, Brown, and Klein (1979), the Bayesian approach tries to incorporate estimation risk directly into the portfolio optimization using the predictive rather than the conditional (on the parameters) distribution of the returns, see also Barberis (2000) for an example in a dynamic setting. There are different implementations of this approach. In this study, we use the Bayes-Stein shrinkage portfolio. This method uses the idea of shrinkage estimation introduced by Stein (1956) and James and Stein (1961). This strategy tries to deal with estimation error in expected returns by replacing "plug-in" estimates of expected returns, $\hat{\mu}_t$, by a weighted average of $\hat{\mu}_t$ and the expected return on the global minimum variance portfolio, $\hat{\mu}_t^{min}$. The weight $\delta_{min,t}$ of $\hat{\mu}_t^{min}$ in this weighted average is calculated by

$$\delta_{min,t} = \frac{N+2}{(N+2) + T_w(\hat{\mu}_t - \hat{\mu}_t^{min})^T \, \tilde{\Sigma}_t^{-1} \, (\hat{\mu}_t - \hat{\mu}_t^{min})} \tag{4.18}$$

where $\tilde{\Sigma}_t = \dfrac{1}{T_w - N - 2} \sum_{s=t-T_w+1}^{t} (r_s - \hat{\mu}_t)(r_t - \hat{\mu}_t)$. For further details see Jorion (1986).

4.3.5 Minimum Variance

There is a general perception that optimal portfolio weights are more sensitive to estimation errors in the mean than estimation errors in the covariance matrix, see Chopra and Ziemba (1993). This motivates the use of benchmark strategies that completely ignore expected returns and only use the covariances between different assets to form optimal portfolio weights. The prime example of such a strategy is the minimum-variance strategy, which is obtained by solving

$$\min w_t' \Sigma_t w_t \quad \text{s.t} \quad \mathbf{1}' w_t = 1\,. \tag{4.19}$$

The solution to this optimization problem is given by

$$w_{minv,t} = \frac{\Sigma_t^{-1} 1_N}{1_N' \Sigma_t^{-1} 1_N}. \tag{4.20}$$

Because this strategy ignores expected returns, we expect it to be more robust than simple mean-variance strategies. This strategy can be combined with other strategies to mitigate the effect of estimation errors even further.

4.3.6 Volatility Timing strategy

Kirby and Ostdiek (2012) introduce a class of "volatility timing" strategies that ignore the expected returns and also the correlation structure of returns in forming portfolio weights. In contrast to the naive $1/N$ strategy, however, it does use the estimate of the volatility levels. We use a particular version of these strategies in which the standard deviations of asset returns are used for forming portfolio weights. As we try to minimize risk, we use the inverse of the estimated volatility level for each asset as its weight and normalize all weights to sum to one, such that the weight $w_{vi,t}$ of the volatility timing strategy for asset i at time t is given by

$$w_{vi,t} = \frac{\sigma_{i,t}^{-1}}{\Sigma_{j=1}^{N} \sigma_{j,t}^{-1}} \tag{4.21}$$

Kirby and Ostdiek (2012) use similar strategies and show that they outperform naive diversification even if transaction costs are relatively high.

4.4 Numerical Results

4.4.1 Simulation Results for One-Factor Models

To elaborate more on the analytical results provided in Section 4.2, we perform a simulation study to compare the out-of-sample performance of the different portfolio strategies introduced in Section 4.3. We consider data generating process that obey alternative factor structures. We assume that $r_t \in \mathbb{R}^{N\times 1}$ holds the excess returns in period t, with mean μ and covariance matrix Σ. Given a vector of risk factors $f_t \in \mathbb{R}^{K\times 1}$ and a matrix of risk factor sensitivities $\beta \in \mathbb{R}^{N\times K}$, we assume

$$r_t = \alpha + \beta f_t + \epsilon_t, \tag{4.22}$$

where $\epsilon_t \in \mathbb{R}^{N\times 1}$ is distributed with zero mean and diagonal covariance matrix Σ_ϵ and $\alpha \in \mathbb{R}^{N\times 1}$ is a vector of the levels of mispricing.

For the model with one factor ($K = 1$), we follow MacKinlay and Pastor (2000) and DeMiguel, Garlappi, and Uppal (2009) and use a similar set-up and similar parameter values for generating returns. We interpret the factor f_t as the (excess) market return and assume it is normally distributed with mean $\mu_f = 8\%$ and standard deviation of $\sigma_f = 16\%$ per year. The factor loadings β_i for asset i are uniformly distributed on the interval [0.5, 1.5]. For the variance-covariance matrix of the error term, we assume two specifications. First, similar to DeMiguel, Garlappi, and Uppal (2009), we assume Σ_ϵ is diagonal with square root of diagonal elements drawn from a uniform distribution with support [0.10, 0.30], so that cross-sectional average annual idiosyncratic volatility is 20%. Second, Σ_ϵ is set to be diagonal with the same diagonal elements, so the annual idiosyncratic volatility for all assets is 20%. We also allow for both zero and non-zero vector α, as the latter case is more relevant in practice. In this case, the annual mispricing α_i for asset i are evenly distributed on the interval [-2%, 2%].

To be precise, for an $N \times T$ matrix of monthly returns, our simulation steps are as follows.

1. Generate a vector $\beta \in \mathbb{R}^{N\times 1}$ where $\beta_i \sim U(0.5, 1.5)$ is the factor loading for asset i.
2. Generate a vector $\sigma_\epsilon \in \mathbb{R}^{N\times 1}$ where $\sigma_{\epsilon,i} \sim U(0.10/\sqrt{12}, 0.30/\sqrt{12})$ is the standard deviation of the idiosyncratic error term for asset i. In the case of a single standard deviation for all assets, $\sigma_{\epsilon,i} = 0.20/\sqrt{12}$ for all i.
3. Generate a vector $\alpha \in \mathbb{R}^{N\times 1}$ where $\alpha_i \sim U(-0.02/12, 0.02/12)$ is the level of mispricing for asset i. In the case of zero mispricing, $\alpha_i = 0$ for all i.
4. Generate a vector $f \in \mathbb{R}^{1\times T}$ where $f_t \sim N(0.08/12, 0.16/\sqrt{12})$ is the market excess return at time t.
5. Generate a matrix $\epsilon \in \mathbb{R}^{N\times T}$ where $\epsilon_{i,t} \sim N(0, \sigma_{\epsilon,i})$ is the error term of asset i at time t.
6. Calculate the matrix of returns using Equation (4.22).

In the above procedure, we need to simulate three $N \times 1$ vectors of uniform random variables. As these numbers are generated once and used for all T generated returns for each asset, they could influence the parameters of interest (mean and standard deviation of the portfolio strategies). For example, we assume that the average level of mispricing α_i is zero in the factor model (4.22). But when we generate one random vector α, it is highly unlikely that the average α_i is zero. This can affect the expected return of portfolio strategies. To avoid this, we used a mirror simulation method. In this method, for generating a vector of uniform random variables on the interval [0,1], we first generate $V \in \mathbb{R}^{N/2\times 1}$ uniform random variables on the interval [0,1]. The remaining $N/2$ variables are calculated by $W = 1 - V$. By this method we ensure that the average of β, α and σ_ϵ vectors are 1, 0, and $0.20/\sqrt{12}$, respectively.

For each iteration of our simulation study, we generate a matrix of $N \times 48120$ of data, which corresponds to 4010 years of monthly data to minimize Monte Carlo simulation error. All rolling window approaches from Section 4.3 are started using the first 120 months of data. Based on the estimates of means, variances and covariances over these first 10 years, the portfolio weights are constructed. Using these weights, we compute the out-of-sample portfolio return using the realized returns in month 121. Next, the window is rolled one month forward and the whole process is repeated. We continue this procedure for 48000 times. Based on these 48000 out-of-sample returns for each portfolio strategy, we calculate the Sharpe ratios. We follow DGU(2009) to compute the p-value of the difference between the Sharpe ratio of each strategy and that of $\hat{SR}_e$. The results for the one-factor set-up for the number of assets $N = 10$ and $N = 25$ are presented in Table 4.1. For each N, the results for four different settings are reported. The first two settings correspond to the case where we set $\alpha = 0$ and the other two to the case where $\alpha \neq 0$. For each case, we use single or dispersed error term variances, denoted by $\sigma^2_{\epsilon,i} \equiv \sigma^2_\epsilon$ and $\sigma^2_{\epsilon,i} \sim U$, respectively.

Table 4.1: Simulation results for the one-factor model

The table shows the Sharpe ratios for the different portfolio strategies described in section 4.3 using simulated date. Parameter estimation is based on a 120-month rolling window for estimating parameters of portfolio strategies. The results are reported for $N = 10$ and $N = 25$ assets. Data is generated by $r_t = \alpha + \beta f_t + \epsilon_t$ where r_t, α, β and $\epsilon_t \in \mathbb{R}^{N\times1}$. We assume the factor portfolio is normally distributed by mean $\mu_f = 8\%$ and standard deviation of $\sigma_f = 16\%$ per year. The factor loadings β_i for asset i are uniformly distributed on the interval [0.5, 1.5]. We assume two cases for idiosyncratic error ϵ_i for asset i. First, it is normally distributed with mean zero and annual standard deviation of 20% (or 5.77% per month) (denoted by $\sigma^2_{\epsilon,i} \equiv \sigma^2_\epsilon$). Second, it is normally distributed with mean zero and annual standard deviation that is evenly distributed on the interval [10%, 30%] (or on the [2.88%,8.66%] per month) (denoted by $\sigma^2_{\epsilon,i} \sim U$). We also assume two cases for the vector of mispricing. In the first case we set $\alpha = 0$ and the in second case, the annual α is distributed evenly on the interval [-2%,2%]. All uniform variable are generated by a mirror simulation method. In each iteration, we generate a time series of 48120 months returns for each asset, so with a rolling window of 120 month, we calculate 48000 out-of-sample returns for portfolio strategies that their Sharpe ratio are computed based on out-of-sample returns. The p-value of the null hypothesis $\hat{SR}_x < \hat{SR}_e$ is given in parentheses, where x indicates the strategy. a, b, and c denote the rejection of null hypothesis at the 10, 5, and 1 percent significance level, respectively.

	$N=10$				$N=25$			
Mispricing	$\alpha=0$		$\alpha\neq0$		$\alpha=0$		$\alpha\neq0$	
Idiosyncratic volatility	$\sigma^2_{\epsilon,i}\equiv\sigma^2_\epsilon$	$\sigma^2_{\epsilon,i}\sim U$	$\sigma^2_{\epsilon,i}\equiv\sigma^2_\epsilon$	$\sigma^2_{\epsilon,i}\sim U$	$\sigma^2_{\epsilon,i}\equiv\sigma^2_\epsilon$	$\sigma^2_{\epsilon,i}\sim U$	$\sigma^2_{\epsilon,i}\equiv\sigma^2_\epsilon$	$\sigma^2_{\epsilon,i}\sim U$
Method	1	2	3	4	1	2	3	4
SR_e	0.134	0.134	0.134	0.134	0.140	0.140	0.140	0.140
SR_{mv}	0.135	0.136	0.137	0.138	0.140	0.141	0.163	0.177
$\hat{SR}_e$	0.136	0.136	0.136	0.136	0.142	0.141	0.142	0.141
$\hat{SR}_{mv}$	0.137[a]	0.138[b]	0.139[c]	0.141[c]	0.142	0.143[b]	0.166[c]	0.180[c]
	(0.08)	(0.01)	(0.00)	(0.00)	(0.21)	(0.02)	(0.00)	(0.00)
$\hat{SR}_{m\hat{v}}$	0.132	0.133	0.135	0.136	0.124	0.124	0.144	0.155[c]
	(1.00)	(0.99)	(0.88)	(0.58)	(1.00)	(1.00)	(0.27)	(0.00)
$\hat{SR}_{\hat{m}v}$	0.001	0.002	-0.005	0.005	0.006	0.005	-0.004	0.005
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
$\hat{SR}_{\hat{m}\hat{v}}$	0.006	0.005	0.010	0.006	-0.010	0.004	-0.004	0.006
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
$\hat{SR}_{\hat{m}\hat{v},\gamma=3}$	0.059	0.060	0.061	0.063	0.035	0.034	0.047	0.055
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
$\hat{SR}_{combine}$	0.119	0.120	0.120	0.120	0.125	0.124	0.127	0.129
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
$\hat{SR}_{\hat{m}\hat{v},cons}$	0.113	0.113	0.113	0.113	0.112	0.107	0.117	0.113
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
$\hat{SR}_{cc}$	0.136	0.135	0.136	0.136	0.141	0.140	0.141	0.140
	(0.86)	(0.89)	(0.76)	(0.80)	(1.00)	(1.00)	(0.97)	(1.00)
$\hat{SR}_{bs}$	0.006	0.006	0.011	0.006	-0.010	0.004	-0.004	0.006
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
$\hat{SR}_{minv}$	0.106	0.106	0.121	0.122	0.093	0.088	0.102	0.110
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
$\hat{SR}_v$	0.136	0.136	0.137[b]	0.137[c]	0.142	0.142[b]	0.142	0.145[c]
	(1.00)	(0.99)	(0.03)	(0.00)	(0.92)	(0.01)	(0.10)	(0.00)

First we discuss the results for $N = 10$, $\alpha = 0$, and $\sigma^2_{\epsilon,i} \equiv \sigma^2_{\epsilon}$. Based on Proposition 1, we can calculate SR_{mv} and SR_e, i.e., the Sharpe ratio of the optimal mean-variance tangency portfolio and of the naive strategy. The values of both Sharpe ratios are very similar (0.135 and 0.134 respectively). This can be expected, as in the one-factor model, $q = \beta'\beta \approx N \cdot 13/12$ in Equation (4.6) is almost equal to $N\bar{\beta}^2 \approx N$ in Equation (4.7). The results for $\hat{SR}_{mv}$, where we calculate the optimal portfolio weights using the true parameters, but implement the strategy using sampled returns to calculate the portfolio return, and $\hat{SR}_e$, are also very similar (0.137 and 0.136 respectively). The p-value for the hypothesis $\hat{SR}_{mv} < \hat{SR}_e$ is 8%. The results for $N = 25$ are analogous to those for $N = 10$, except that Sharpe ratios are slightly higher (as q in Equation (4.6) and $N\bar{\beta}^2$ in Equation (4.7) are larger). So for $\alpha = 0$ and $\sigma^2_{\epsilon,i} \equiv \sigma^2_{\epsilon}$, even when there is no estimation error in the parameters, the Sharpe ratios of the optimal mean-variance strategy and that of naive diversification are almost the same. As the Sharpe ratio of the optimal mean-variance strategy when there is no estimation error is the highest Sharpe ratio we can get, we expect that no strategy that is subject to estimation error can beat naive diversification strategy's results. Results confirm this.

The same pattern emerges for the case of dispersed idiosyncratic volatility ($\alpha = 0$, and $\sigma^2_{\epsilon,i} \sim U$). As the average variance of the error terms is the same as before, the effects of higher variances for some assets cancel out against lower variances for other error terms. Adding a non zero vector for the level of mispricing ($\alpha \neq 0$) can improve the Sharpe ratio of the optimal mean-variance strategy. Mispricing has no effect on the Sharpe ratio of naive strategy, however, as the average of α_i is set to zero. In our current settings, the annual mispricing vector α is uniformly distributed on the interval [-2%,2%]. By increasing the support of the mispricing vector to for example [-4%,4%], we can further increase the difference between the Sharpe ratio of the optimal mean-variance strategy and that of naive diversification.

The Sharpe ratio of the mean-variance tangency portfolio decreases when we introduce estimation error in the variance-covariance matrix of returns ($\hat{SR}_{m\hat{v}} = 0.132$) or in the mean ($\hat{SR}_{\hat{m}v} = 0.001$) or in both ($\hat{SR}_{\hat{m}\hat{v}} = 0.006$). The cost of estimation errors in means is much higher than the cost of estimation errors in the variance-covariance matrix.

We also compare the performance of seven other portfolio strategies with naive diversification. The p-value of the null hypothesis $\hat{SR}_x < \hat{SR}_e$ is given in parentheses, where x indicates the strategy. First, using a mean-variance strategy with a fixed parameter of risk aversion, improves the performance compared to the mean-variance tangency portfolio. Here we report the results for an investor with parameter or risk aversion $\gamma = 3$. The Sharpe ratio is $\hat{SR}_{\hat{m}\hat{v},\gamma=3} = 0.059$. As Kirby and Ostdiek (2012) argue, for the mean-variance tangency portfolio, we need to normalize the optimal weights to sum to one. As Equation (4.3) shows, we therefore divide $w_{0,mv} \propto \Sigma^{-1}\mu_m$ by the denominator in equation (4.3). Due to estimation errors in the parameters, the dominator sometimes becomes very low, leading to very large weights for some assets in the portfolio. These large weights cause the mean-variance tangency portfolio to have a high volatility when implemented in an out-of-sample context and in turn leads to a poor out-of-sample performance.

Combining the portfolio weights obtained form the mean-variance strategy using $\gamma = 3$ and the naive strategy could improve the Sharpe ratio substantially ($\hat{SR}_{\hat{m}\hat{v},combine} = 0.119$). Still, this is lower than for the naive strategy. Using the mean-variance strategy with non-negativity constraints on the weights also improves the performance of the mean-variance strategy ($\hat{SR}_{\hat{m}\hat{v},cons} = 0.113$). The reason is that putting constraints on portfolio weights prevents the denominator in equation (4.3) to become very small and so prevents extreme portfolio weights. Combining the mean-variance with constrained weights strategy with the naive strategy also further improve the Sharpe ratio ($\hat{SR}_{\hat{m}\hat{v},cc}$=0.136). Still, this is not significantly higher than the Sharpe ratio of the naive strategy.

The minimum variance strategy that ignores the expected returns in forming portfolio weights also has a lower Sharpe ratio than the $1/N$ strategy. The volatility timing strategy is the only strategy that can significantly outperform naive diversification in four out of the eight settings that are reported in Table 4.1. The best performance is achieved for the case $N = 25$, $\alpha \neq 0$, and $\sigma^2_{\epsilon,i} \sim U$ ($\hat{SR}_v = 0.145$ versus $\hat{SR}_e = 0.141$). The difference between $\hat{SR}_{minv}$ and $\hat{SR}_v$ highlights the effect of estimation error in the non-diagonal elements of the variance-covariance matrix.

Overall, when data are generated by a one-factor model, even when there is no estimation error, there is hardly any difference between the Sharpe ratio of the optimal mean variance strategy and naive diversification unless we have a large value of mispricing. So for comparing the performance of portfolio strategies with naive diversification, the simulation settings that use one-factor models for generating returns are not very informative.

4.4.2 Simulation Results for Two Factor Models

In this section, we use a two-factor model to generate asset returns and estimate Sharpe ratios. For this model, we can simplify the formulas for the Sharpe ratio based on the results on Section 4.2.3. If we assume $V_\varepsilon \sim N(0, \sigma^2_\varepsilon \mathrm{I})$, we obtain

$$\bar{B} = 1/(\sigma^2_\varepsilon) \begin{bmatrix} N\bar{\beta}_1\bar{\beta}_1 & N\bar{\beta}_1\bar{\beta}_2 \\ N\bar{\beta}_2\bar{\beta}_1 & N\bar{\beta}_2\bar{\beta}_2 \end{bmatrix} \tag{4.23}$$

and

$$B = 1/(\sigma^2_\varepsilon) \begin{bmatrix} \sum \beta^2_{i1} & \sum \beta_{i1}\beta_{i2} \\ \sum \beta_{i1}\beta_{i2} & \sum \beta^2_{i2} \end{bmatrix} \tag{4.24}$$

where β_1 and β_2 are the first and second column of β with ith element β_{i1} and β_{i2}, $\bar{\beta}_1 = \mathbf{1}'\beta_1/N$, and $\bar{\beta}_2 = \mathbf{1}'\beta_2/N$. Therefore

$$C = B - \bar{B} = N/(\sigma_\varepsilon^2) \begin{bmatrix} \sigma_{\beta_1}^2 & \rho_{\beta_1,\beta_2}\sigma_{\beta_1}\sigma_{\beta_2} \\ \rho_{\beta_1,\beta_2}\sigma_{\beta_1}\sigma_{\beta_2} & \sigma_{\beta_2}^2 \end{bmatrix} \tag{4.25}$$

where $\sigma_{\beta_1}^2 = \mathrm{Var}(\beta_{i1})$, $\sigma_{\beta_2}^2 = \mathrm{Var}(\beta_{i2})$, and $\rho_{\beta_1,\beta_2} = \mathrm{Corr}(\beta_{i1}, \beta_{i2})$. We assume $\bar{\beta}_2 = a\bar{\beta}_1$ and $\sigma_{\beta_2} = b\sigma_{\beta_1}$. Based on Proposition 2, the Sharpe ratios for the two strategies and the difference between them are functions of $\bar{\beta}_1$, σ_{β_1}, $\tilde{a}$, b, ρ_{β_1,β_2}, σ_ε^2, N, and V_f. From (4.22), we see that without loss of generality we can set $V_f = \mathrm{I}$, as

$$r = \beta f + \varepsilon = \beta\,(V_f)^{1/2}(V_f)^{-1/2}\,f + \varepsilon = \tilde{\beta}\tilde{f} + \varepsilon, \tag{4.26}$$

where $\tilde{\beta} = \beta \cdot (V_f)^{1/2}$ and $\tilde{f} = (V_f)^{-1/2} \cdot f$ and $V_{\tilde{f}} = Cov(\tilde{f}) = \mathrm{I}$. Accordingly, we can calculate $\bar{\tilde{\beta}}_2$, $\bar{\tilde{\beta}}_1$, $\sigma_{\tilde{\beta}_2}$, and $\sigma_{\tilde{\beta}_1}$, similar to the previous model. The Sharpe ratios for two strategies and difference between them are functions of $\bar{\tilde{\beta}}_1$, $\sigma_{\tilde{\beta}_1}$ $\tilde{a}$, $\tilde{b}$, $\rho_{\tilde{\beta}_1,\tilde{\beta}_2}$, σ_ε^2, and N. We study the effect of each of these parameters.

To see the effect of idiosyncratic volatility, we assume that idiosyncratic volatility, σ_ε^2, is a multiple λ of systematic volatility, $\tilde{\beta}\tilde{\beta}'$. Considering
$N\sigma_\varepsilon^2 = \mathrm{trace}(V_\varepsilon) = \lambda\ \mathrm{trace}(\tilde{\beta}\tilde{\beta}') = \lambda\ \mathrm{trace}(N\tilde{C} + N\bar{\tilde{\beta}}\bar{\tilde{\beta}}') \Leftrightarrow$
$\sigma_\varepsilon^2 = \lambda[\sigma_{\tilde{\beta}_1}^2 + \sigma_{\tilde{\beta}_2}^2 + \bar{\tilde{\beta}}_1^2 + \bar{\tilde{\beta}}_2^2]$
where $\bar{\tilde{\beta}} = \tilde{\beta}'\mathbf{1}/N$ and $\tilde{C}$ is similar to C in (4.25).

To construct a benchmark and provide some reasonable values for the parameters of the model, we first estimate a general two-factor model on empirical data. Using the market and size factor from Fama and French (1993), we take individual CRSP stock returns for stocks that have at least 180 monthly observations between 196307 to 200812 and estimate β, V_f and V_ε for model (4.22). The estimates are used to calculate the parameters of interest for model (4.26), i.e, $\tilde{\beta}$ and $\mu_{\tilde{f}}$.

Table 4.2 shows the estimated parameters for model (4.26). These parameters constitute the benchmark settings for our two-factor model.

The simulation steps are as follows.

1. Set $\bar{\tilde{\beta}}_1 = 0.0464$, $\sigma_{\tilde{\beta}_1} = 0.0186$, $N = 10$, $\tilde{b} = 1$, $\rho_{\tilde{\beta}_1,\tilde{\beta}_2} = 0.5$, and $\lambda = 4.5$.

2. Select a value of $\tilde{a}$, $\tilde{a} = [-5, -4.8, ..., 0, ...4.8, 5]$.

3. For each value of $\tilde{a}$, generate 100 matrices of $\tilde{\beta}$ that consist of two columns, $\tilde{\beta}_1$ and $\tilde{\beta}_2$ with means $\bar{\tilde{\beta}}_1$ and $\bar{\tilde{\beta}}_2 = \tilde{a}\bar{\tilde{\beta}}_1$, standard deviations $\sigma_{\tilde{\beta}_1}$ and $\sigma_{\tilde{\beta}_2} = \tilde{b}\sigma_{\tilde{\beta}_1}$, and correlation $\rho_{\tilde{\beta}_1,\tilde{\beta}_2} = 0.5$.

Table 4.2: Benchmark parameters for simulation using two-factor models

The table presents the estimated parameters for model (4.26) to be used as the benchmark parameters for simulation using a two-factor settings. We use the market and size factor from Fama and French (1993) as two factors of the model and take individual CRSP stock returns for stocks that have at least 180 monthly observations between 196307 to 200812 to estimate β, V_f and V_ε for model (4.22). The estimates are used to calculate the parameters of interest for model (4.26).

Parameter	Realized value	Description
$\mu_{\tilde{f}_1}$	0.0989	Average of the first factor
$\mu_{\tilde{f}_2}$	0.0383	Average of the second factor
$\bar{\tilde{\beta}}_1$	0.0464	Average of $\tilde{\beta}_1$ in the cross section
$\sigma_{\tilde{\beta}_1}$	0.0186	Standard deviation of $\tilde{\beta}_1$ in the cross section
$\tilde{a}$	0.4538	$\bar{\tilde{\beta}}_2/\bar{\tilde{\beta}}_1$
$\tilde{b}$	1.0230	$\sigma_{\tilde{\beta}_2}$ / $\sigma_{\tilde{\beta}_1}$
$\tilde{\rho}_{\beta_1,\beta_2}$	0.5002	Correlation($\tilde{\beta}_1$,$\tilde{\beta}_2$)
λ	4.5620	Ratio of idiosyncratic variance to the systematic variance

4. For each matrix $\tilde{\beta}$, generate a 2×121 matrix of $\tilde{f}$ where $\tilde{f}_t \sim N(\mu_{\tilde{f}}, \mathrm{I})$ holds the factor returns at time t. Using the factor returns, generate a window of $N \times 121$ monthly returns, using (4.26).

5. Calculate the optimal weights for different strategies using the first 120 months of returns and then calculate the out-of-sample return for each strategy using returns in month 121.

6. Repeat steps 4 and 5 for 200 times (so for each matrix of $\tilde{\beta}$, we have 200 out-of-sample returns for each strategy) and calculate the Sharpe ratio for each strategy.

7. Average the Sharpe ratios over the 100 values of $\tilde{\beta}$ generated in step 3 for each value of $\tilde{a}$.

In the following sections, we change the parameters of interest (N, λ, etc.) one by one to investigate the effect of each on the Sharpe ratios.

Effect of correlation between betas ($\rho_{\tilde{\beta}_1,\tilde{\beta}_2}$)

We first investigate the effect of correlation between betas on the Sharpe ratios. We set $N = 10$, $\tilde{b} = 1$, and $\lambda = 4.5$, and plot graphs of the Sharpe ratio for the optimal strategy for different values of $\tilde{a}$ for $\rho_{\tilde{\beta}_1,\tilde{\beta}_2} = [-0.9, -0.5, 0, 0.5, 0.9]$. As $\rho_{\tilde{\beta}_1,\tilde{\beta}_2}$ has no effect on the Sharpe ratio of the naive strategy, we have only one curve for the naive strategy.

Figure 4.1 shows that the difference between the Sharpe ratios is largest when we have large negative values of $\tilde{a}$. This implies large negative values for $\tilde{\beta}_2$ and so large negative expected returns for some assets. In this case, the mean-variance strategy still benefits from shorting the assets with large negative returns. Imposing $1/N$, however, yields negative returns for the portfolio and so a negative Sharpe ratio. As $\tilde{a}$ increases, the Sharpe ratio of the naive strategy becomes zero at some point and then becomes positive. For the optimal strategy, as $\tilde{a}$ increases, the dominant negative effect of $\tilde{\beta}_2$ decreases and the correlation between the betas starts playing a role. As $\rho_{\tilde{\beta}_1,\tilde{\beta}_2}$ increases, the difference between expected returns of the different assets becomes more extreme, so the mean-variance strategy has more advantages with respect to naive diversification. On the other hand, a negative correlation between betas makes the return on the assets more similar, and makes the mean-variance strategy to behave more similar to the $1/N$ strategy. In the right-hand side of the graph, as $\tilde{a}$ increases, the value of $\tilde{\beta}_2$ become larger and start to dominate the expected returns. The model starts to resemble more a one-factor model again, and therefore the mean variance strategy and naive diversification yield almost the same performance.

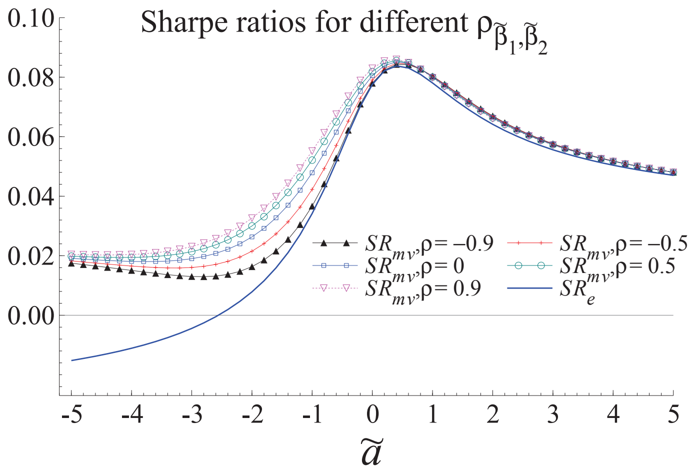

Figure 4.1: Sharpe ratios of optimal and naive strategy for different values of $\tilde{a} = \bar{\tilde{\beta}}_2/\bar{\tilde{\beta}}_1$ and $\rho_{\tilde{\beta}_1,\tilde{\beta}_2}$

Effect of number of assets

Second, we study the effect of the number of assets. We set $\rho_{\tilde{\beta}_1,\tilde{\beta}_2} = 0.50$, $\tilde{b} = 1$, $\lambda = 4.5$, and use $N = [5, 10, 50, 100]$. The left panel in Figure 4.2 shows the value of the Sharpe ratio for the optimal strategy for different values of N. The general pattern is the same as in Figure 4.1. As the number of assets increases, the Sharpe ratio of the optimal strategy also increases for all values of $\tilde{a}$ as it helps to mitigate the effect of idiosyncratic volatilities and therefore leads to lower standard deviations. The right panel in Figure 4.2 shows the same graph for the naive strategy. Increasing the number of assets improves the Sharpe ratio in the right-hand side of the graph. But in the left-hand side, because of the negative expected return for large negative values of $\tilde{a}$ the effect is reversed.

Effect of idiosyncratic volatility

Third, we study the effect of idiosyncratic volatility versus systematic volatility. We set $\rho_{\tilde{\beta}_1,\tilde{\beta}_2} = 0.50$, $\tilde{b} = 1$, and $N = 10$, and plot the Sharpe ratio for $\lambda = [0.25, 0.5, 1, 2, 4.5, 8]$ in Figure 4.3. As expected, when the ratio of idiosyncratic volatility to systematic volatility

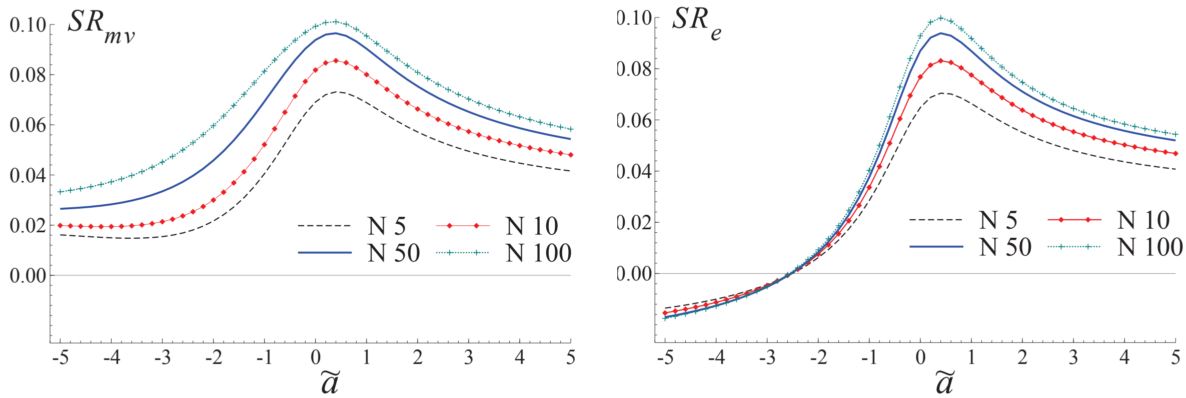

Figure 4.2: Sharpe ratios of optimal and naive strategy for different values of $\tilde{a} = \bar{\tilde{\beta}}_2/\bar{\tilde{\beta}}_1$ and number of assets N

increases, the Sharpe ratio of both the optimal and naive strategy decreases. The increased noise leads to higher standard deviations of portfolio returns and lower Sharpe ratios.

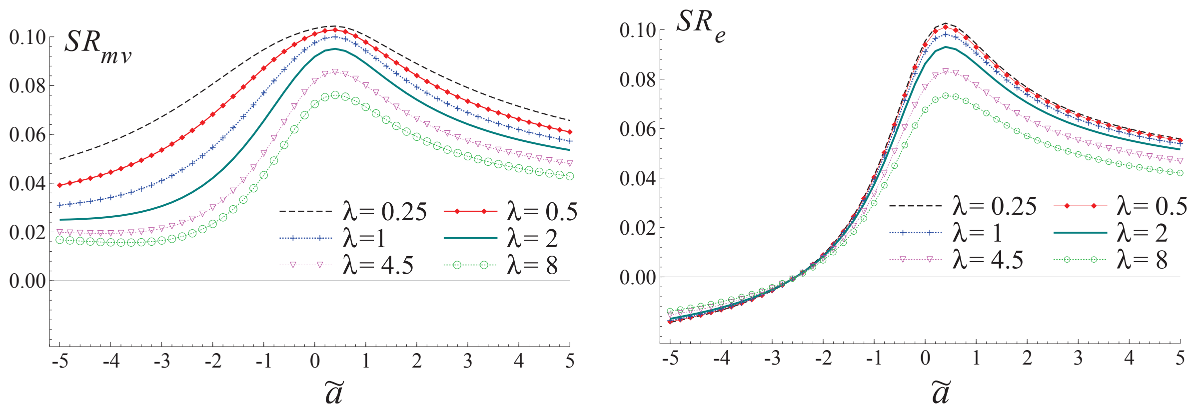

Figure 4.3: Sharpe ratios of optimal and naive strategy for different values of $\tilde{a} = \bar{\tilde{\beta}}_2/\bar{\tilde{\beta}}_1$ and λ

Effect of ratio of standard deviation of betas

Fourth, we study the effect of the ratio of standard deviations of the betas. We set $\rho_{\tilde{\beta}_1,\tilde{\beta}_2} = 0.50$, $\lambda = 4.5$, and $N = 10$, and plot the Sharpe ratios for $\tilde{b} = [0.5, 1, 2]$ in the left panel of Figure 4.4. A higher $\tilde{b}$ leads to more dispersed returns, which works in favor of the mean-variance strategy. On the other hand, $\tilde{b}$ also affects the idiosyncratic volatility, which is λ times the systematic volatility, so the Sharpe ratios of both strategies decrease when $\tilde{b}$ increases. The sum of these two effects decreases the Sharpe ratio of naive diversification at a higher speed than the optimal mean-variance strategy. This is shown in the right panel of Figure 4.4, where $SR_{mv} - SR_e$ increases with a higher value of $\tilde{b}$. As before, this difference is higher for lower values of $\tilde{a}$.

The Other Strategies

Till now, we studied the effects of different parameters on SR_{mv} and SR_e, i.e., the Sharpe ratio without parameter uncertainty. In this section, we also report the out-of-sample

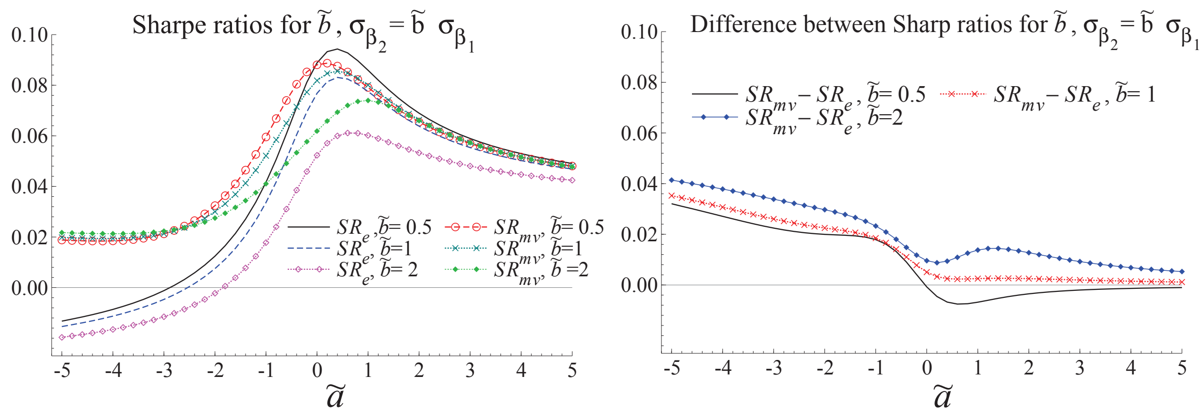

Figure 4.4: Sharpe ratios of optimal and naive strategy and the difference between them for different values of $\tilde{a} = \bar{\tilde{\beta}}_2/\bar{\tilde{\beta}}_1$ and $\tilde{b} = \sigma_{\tilde{\beta}_2}/\sigma_{\tilde{\beta}_1}$

performance of the portfolio strategies introduced in Section 4.3. We set $\rho_{\tilde{\beta}_1,\tilde{\beta}_2} = 0.5$, $\lambda = 4.5$, and $\tilde{b} = 1$, and plot the results for $N = 10$ and $N = 25$ in Figure 4.5. Except for the $\hat{SR}_{\hat{m}\hat{v}}$ and $\hat{SR}_{bs}$ ratios, which are subject to extreme weights and therefore very high levels of volatility and poor out-of-sample performance, the other strategies follow a similar pattern to that of SR_e. The reason is that except SR_{mv}, $\hat{SR}_{\hat{m}\hat{v}}$ and $\hat{SR}_{bs}$, the other strategies only allow for positive weights, such that in the case of large and negative values of $\tilde{a}$, the expected returns of these strategies become negative. Among these strategies, $\hat{SR}_v$ and $\hat{SR}_{\hat{m}\hat{v},cons}$ represent the best and the worse performance compared with the $1/N$ strategy, respectively. As before, the Sharpe ratios are higher for a larger value of N.

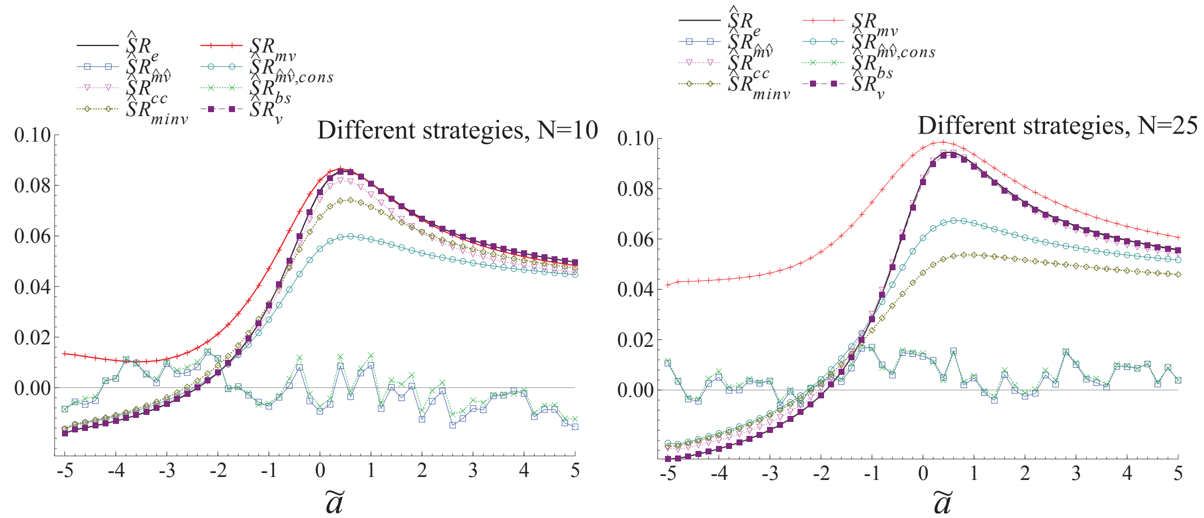

Figure 4.5: Sharpe ratios of different portfolio strategies for different values of $\tilde{a} = \bar{\tilde{\beta}}_2/\bar{\tilde{\beta}}_1$

To conclude, the ability of the optimal mean variance strategy to short assets, enable it to attain a superior performance to the naive strategy for large and negative values of $\tilde{a}$. This difference between Sharpe ratios is amplified by larger value of $\rho_{\tilde{\beta}_1,\tilde{\beta}_2}$ and $\tilde{b}$ that lead to more dispersed assets returns and works in favor of the mean-variance strategy. The larger value of N and the smaller value of λ also increase the Sharpe ratios of both strategies as the effect of idiosyncratic noise reduces. The out-of-sample performance of portfolio strategies that require parameter estimation and allow for short positions is poor. Strategies that do not allow short-sales exhibit a reasonable performance, but cannot outperform naive diversification. Therefore, if data are generated by a two-factor model, the ability of portfolio strategies to outperform naive diversification is limited, especially for parameter values found in typical empirical settings.

4.5 Empirical Results

4.5.1 Data

In Section 4.4 we observed that if the data are generated by a one-factor or two-factor model, there is hardly any room for optimization-based portfolio strategies to outperform naive diversification out-of-sample. Many studies show that financial data are generated by more than two factors, for example Fama and French (1993), Carhart (1997). Typically three, four, or more factors used for describing asset returns. Therefor, it is interesting to compare the performance of portfolio strategies using empirical data both in-sample and out-of-sample.

In our empirical tests, we use two different types of data. The first data set includes equity return indices. In particular, we consider the four sets of equity portfolios that are also used by DeMiguel, Garlappi, and Uppal (2009). The first set consists of monthly returns on three portfolios (denoted by MKT/SMB/HML): the market portfolio (MKT), a portfolio that is long on high book-to-market stocks and short on low book-to-market stocks (HML) and a portfolio that is long on small stocks and short on big stocks (SMB). The second data set includes 10 industry portfolios plus the market portfolio. The third data sets includes 21 portfolios, 20 portfolios sorted based on size and book-to-market, plus the market portfolio (denoted by FF-1-factor). The forth data set consists of 24 portfolios including 20 portfolios based on size and book-to-market, plus the market, SMB, HML and Momentum portfolios (denoted by FF-4-factors). Data sources for all data sets are Ken French's web site. We take the three-month Treasury Bill from the website of Federal Reserve Economic Data as our risk free rate.

The second type of data that we use contains different portfolios of asset classes including US and non-US equities, US and non-US bonds, hedge funds, and commodities. For this type of data, we take one month Eurodollar deposit as our risk free rate. We have three assets that are representative of equities: US equities, non-US equities and emerging market equities. Four assets are representatives of bonds: US government bonds, US high yield bonds, US credits and non-US government bonds. We have also five assets that are representative of other asset classes including real state, hedge funds, commodities, gold, and crude oil. Table 4.3 presents some background information about the data and Table 4.4 provides descriptive statistics.

Table 4.3: Description of asset classes data

The table presents some information about data of different asset classes that are used in this paper. For each asset, the source of the data, the ticker used to label it in the source, number of monthly observations, yearly average return and yearly average excess return over the risk free rate are reported.

Asset	source	ticker	number of data	Annualized return	Annualized excess return
MSCI USA equities	Datastream	MSUSAM$(RI)	485	10.53	4.21
MSCI non-U.S. equities	Datastream	MSWXUS$(RI)	485	11.28	5.47
MSCI Emerging markets equities	Datastream	MSEMKF$(MSRI)	282	15.93	11.59
JPM U.S. government bonds	Datastream	USMGUSRI	306	7.07	2.53
JPM non-U.S. government bonds	Datastream	USMGEXRI	306	8.76	4.22
U.S. credits	Datastream	LHCRPBD(IN)+100	461	8.26	1.89
U.S. high yield bonds	Datastream	LHYIELD(IN)+100	335	9.59	4.63
Hedge funds	Datastream	CSTHEDG	210	9.16	5.49
GSCI commodities index	Datastream	GSCITOT	485	11.49	4.99
U.S. real estate	Bloomberg	FNERTR Index	474	13.04	6.71
Spot gold	Datastream	GOLDBLN	485	10.78	5.00
Crude Oil	Datastream	CRUDWTC	341	9.68	4.65
One-month Eurodollar deposit	Datastream	FREDD1M	485	6.33	

Table 4.4: Description statistics for asset classes data

The table presents description statistics about data of different asset classes that are used in this paper. The average monthly excess return are reported both for the period 1994-01 to 2011-06 (that data for all assets are available) and the whole period that data for each asset is available. The correlation matrix is also reported for the period 1994-01 to 2011-06.

Asset	Equites	non-US Equities	EM Equities	Bonds	non-US Bonds	Credits	HY-Bonds	Hedge funds	Commodities	Real state	Gold	Oil
Mean (199401-201106)	0.46	0.31	0.51	0.19	0.26	0.24	0.34	0.46	0.33	0.73	0.43	1.04
Mean	0.35	0.46	0.97	0.21	0.35	0.16	0.39	0.46	0.42	0.56	0.42	0.39
Median	0.65	0.70	1.20	0.22	0.21	0.30	0.55	0.52	0.43	0.73	-0.11	0.33
StdDev	4.50	5.03	6.93	1.39	2.83	2.11	2.55	2.20	5.84	5.01	6.18	9.71
Min	-21.86	-21.38	-29.37	-4.76	-8.08	-9.17	-16.49	-8.01	-28.78	-32.25	-22.67	-32.96
Max	16.81	16.15	18.15	5.36	7.95	11.20	12.02	8.00	25.03	30.94	38.26	43.36
Skew	-0.43	-0.37	-0.69	-0.05	0.12	0.00	-0.93	-0.33	0.06	-0.70	0.98	0.45
Kurtosis	1.90	1.27	1.84	0.68	0.18	3.97	8.66	2.52	2.56	8.40	5.65	2.98
Correlations (over 199401-201106)												
non-US Equities	0.83											
EM Equities	0.73	0.81										
Bonds	-0.15	-0.17	-0.19									
non-US Bonds	0.08	0.29	0.10	0.47								
Credits	0.26	0.27	0.23	0.71	0.43							
HY-Bonds	0.61	0.63	0.62	-0.08	0.13	0.52						
Hedge funds	0.56	0.57	0.62	-0.03	-0.03	0.32	0.52					
Commodities	0.22	0.37	0.34	-0.04	0.17	0.14	0.25	0.36				
Real state	0.54	0.55	0.47	-0.04	0.18	0.30	0.61	0.32	0.19			
Gold	-0.02	0.16	0.21	0.16	0.37	0.22	0.11	0.17	0.24	0.10		
Oil	0.15	0.29	0.32	-0.10	0.10	0.07	0.21	0.27	0.88	0.10	0.20	

Table 4.5: Portfolios selected from asset classes data

The table shows the asset classes that are selected to form 7 different portfolios of asset classes. For each portfolio, N denotes the number of assets in the portfolio and T denotes the number of monthly return observations that data are available for all assets in that portfolio.

Portfolio	1	2	3	4	5	6	7
N	2	2	3	3	7	9	6
T	306	485	282	306	306	210	306
MSCI USA equities	1	1	1	1	1	1	1
MSCI non-U.S. equities	0	1	1	0	1	1	0
MSCI Emerging markets equities	0	0	1	0	0	1	0
JPM U.S. government bonds	1	0	0	1	1	1	1
JPM non-U.S. government bonds	0	0	0	0	1	1	0
U.S. credits	0	0	0	0	1	0	0
U.S. high yield bonds	0	0	0	0	0	1	0
Hedge funds	0	0	0	0	0	1	0
GSCI commodities index	0	0	0	0	1	1	1
U.S. real estate	0	0	0	1	1	1	1
Spot gold	0	0	0	0	0	0	1
Crude Oil	0	0	0	0	0	0	1

For our empirical experiment, we form 7 different portfolios of available asset classes. Table 4.5 shows the selected asset classes for each portfolio. The combination of the assets in the portfolios represents investors with different access to available asset classes. As the number of observations is different for different assets, this table also shows the number of available observations for each portfolio. Our aim here is to study whether the different combinations of asset classes that represent different factor structures influences the comparison between optimization-based portfolio strategies and the naive strategy. The main difference with equity portfolios is that there is no dominant factor (like the market factor) that drives the common variation in all assets.

4.5.2 Empirical Results for Equity portfolios

In this section, we present the results for equity portfolios. We use the same approach for computing the Sharpe ratio of different portfolio strategies as described in Section 4.3, i.e., a rolling window approach, with an estimation window of length $T_w = 120$ months. For the strategies that require the true parameters, we substitute the estimated mean and variance-covariance matrix over the whole sample, as the true parameters.

In the first empirical test, we replicate part of the empirical results of DGU (2009). We first use the same sample period as in DGU (2009), i.e., the period 1963-07 to 2004-11, and then update the data till December 2011. Table 4.6 presents the Sharpe ratios for the different portfolio strategies introduced in Section 4.3.

For each portfolio in Table 4.6, the first column is the results reported by DGU (2009). The second column contains the results we obtain for all strategies using the same sample period as in DGU(2009). The third column holds the results using data updated to December 2011. Comparing the results in column 1 and 2 for each data set, we observe that for most strategies, we get very similar Sharpe ratios to the numbers obtained by DGU(2009). Differences are larger for $\hat{SR}_{\hat{m}\hat{v}}$, $\hat{SR}_{bs}$, and $\hat{SR}_{\hat{m}\hat{v},cons}$. Part of this may be due to the data sets being downloaded at different times from Ken French's web site. Our experience shows that these data sets are continuously improving and the data for the earlier period may also be subject to change. The Sharpe ratios for the constrained mean-variance strategy as obtained by us are considerably larger than those obtained by DGU (2009) for three out of four cases. Besides the possible difference in the data, we can also attribute this to the possible use of a different solver for the quadratic programming problem of mean-variance optimization (Equation (4.1)) under constraints. Our experience show that some solvers are particularly sensitive to choosing different initial values for the optimization parameters. As a result, there is a considerable risk of ending up in a local rather than the global optimum.

Now we update the data till December 2011 and implement all portfolio strategies over the whole period. For all data sets, SR_{mv} is much higher than SR_e. This contracts with the cases in Section 4.4 where data are generated by a one-factor or two-factor model. As expected, as the number of assets increases, the difference between SR_{mv} and SR_e increases. For the FF-1-factor and FF-4-factors portfolios, the SR_{mv} is more than three times of SR_e. This suggests that there is some potential for portfolio strategies to outperform naive diversification out-of-sample. This is confirmed by the other results. For the FF-4-factors data set, $\hat{SR}_{\hat{m}\hat{v},combine} = 0.377$, $\hat{SR}_{\hat{m}\hat{v},\gamma=3} = 0.376$, $\hat{SR}_{\hat{m}\hat{v},cons} = 0.286$, $\hat{SR}_{cc} = 0.286$, and $\hat{SR}_v = 0.164$ are significantly higher than $\hat{SR}_e = 0.154$. Also for FF-1-factor $\hat{SR}_{\hat{m}\hat{v},combine} = 0.349$, $\hat{SR}_{\hat{m}\hat{v},\gamma=3} = 0.348$ $\hat{SR}_{minv} = 0.244$, and $\hat{SR}_v = 0.149$ are significantly higher than $\hat{SR}_e = 0.143$.

For the industry portfolios, $SR_{mv} = 0.228$ is almost twofold $SR_e = 0.116$. Among the out-of-sample strategies, $\hat{SR}_{minv} = 0.173$ and $\hat{SR}_v = 0.129$ outperform $\hat{SR}_e = 0.123$. For the MKT/SMB/HML portfolio, the in-sample Sharpe ratios are $SR_{mv} = 0.207$ and $SR_e = 0.178$, so there is not so much room for out-of-sample strategies to significantly outperform $\hat{SR}_e = 0.189$, certainly not after accounting for estimation error in the parameter estimates.

To conclude, adding new portfolio strategies to DGU(2009), we come to a different conclusion. For some equity portfolios, there are several portfolio strategies that outperform naive diversification out-of-sample.

Table 4.6: Sharpe ratio for equity portfolios

The table reports the Sharpe ratio for the different portfolio strategies that described in section 4.3 using four data sets of equity portfolios that also used by DGU (2009). For each portfolio, the first column is the results reported by DGU (2009) for the methods that is used both by DGU(2009) and in this paper. The second column is the results that we got using the same sample period as in DGU(2009), i.e., the period 1963-07 to 2004-11, but with data that is downloaded in 2011 from Ken French's web site. The third column is the results using data updated to December 2011. For each portfolio, N denotes the number of assets and T is the number of monthly return observation. We use a rolling approach with an estimation window length of $T_w = 120$ months for estimating parameters of each portfolio strategy. Therefore, we have $T - T_w$ out-of-sample returns to compute the Sharpe ratio for portfolio strategies that their Sharpe ratio are computed based on out-of-sample returns. The p-value of the null hypothesis $\hat{SR}_x < \hat{SR}_e$ is given in parentheses, where x indicates the strategy. a, b, and c denote the rejection of null hypothesis at the 10, 5, and 1 percent significance level, respectively.

Portfolio	MKT/SMB/HML			Industry portfolios			21 FF portfolio			24 FF portfolio		
	1	2	3	1	2	3	1	2	3	1	2	3
N	3	3	3	11	11	11	21	21	21	24	24	24
T	497	497	582	497	497	582	497	497	582	497	497	582
$T - T_w$	377	377	462	377	377	462	377	377	462	377	377	462
SR_e		0.207	0.178		0.122	0.116		0.142	0.130		0.155	0.141
SR_{mv}	0.285	0.251	0.207	0.212	0.208	0.228	0.510	0.472	0.440	0.536	0.519	0.474
$\hat{SR}_e$	0.224	0.232	0.189	0.135	0.133	0.123	0.162	0.162	0.143	0.175	0.175	0.154
$\hat{SR}_{mv}$		0.272	0.212		0.203[a]	0.227[b]		0.494[c]	0.453[c]		0.536[c]	0.478[c]
		(0.11)	(0.19)		(0.08)	(0.02)		(0.00)	(0.00)		(0.00)	(0.00)
$\hat{SR}_{m\hat{v}}$		0.245	0.192		-0.075	-0.003		0.380[c]	0.356[c]		0.162	0.107
		(0.35)	(0.45)		(1.00)	(0.98)		(0.00)	(0.00)		(0.58)	(0.77)
$\hat{SR}_{\hat{m}v}$		0.205	0.144		-0.045	-0.067		0.143	0.100		0.109	0.056
		(0.69)	(0.81)		(0.99)	(1.00)		(0.60)	(0.74)		(0.81)	(0.93)
$\hat{SR}_{\hat{m}\hat{v}}$	0.219	0.199	0.165	0.068	-0.059	-0.051	0.013	-0.045	-0.037	0.184	0.147	0.053
	(0.46)	(0.73)	(0.71)	(0.17)	(1.00)	(1.00)	(0.02)	(1.00)	(1.00)	(0.45)	(0.66)	(0.95)
$\hat{SR}_{\hat{m}\hat{v},\gamma=3}$		0.182	0.155		0.038	0.059		0.360[c]	0.348[c]		0.397[c]	0.376[c]
		(0.84)	(0.79)		(0.93)	(0.86)		(0.00)	(0.00)		(0.00)	(0.00)
$\hat{SR}_{combine}$		0.183	0.156		0.040	0.061		0.361[c]	0.349[c]		0.399[c]	0.377[c]
		(0.84)	(0.79)		(0.93)	(0.86)		(0.00)	(0.00)		(0.00)	(0.00)
$\hat{SR}_{\hat{m}\hat{v},cons}$	0.108	0.248	0.196	0.068	0.141	0.146	0.198	0.154	0.134	0.202	0.363[c]	0.286[c]
	(0.02)	(0.31)	(0.39)	(0.03)	(0.40)	(0.19)	(0.02)	(0.66)	(0.72)	(0.27)	(0.00)	(0.00)
$\hat{SR}_{cc}$		0.248	0.196		0.141	0.146		0.154	0.134		0.363[c]	0.286[c]
		(0.31)	(0.39)		(0.40)	(0.19)		(0.66)	(0.72)		(0.00)	(0.00)
$\hat{SR}_{bs}$	0.254	0.240	0.192	0.072	-0.049	-0.041	0.014	-0.038	-0.030	0.179	0.154	0.053
	(0.25)	(0.42)	(0.46)	(0.19)	(0.99)	(0.99)	(0.02)	(1.00)	(1.00)	(0.48)	(0.62)	(0.94)
$\hat{SR}_{minv}$	0.249	0.247	0.195	0.155	0.162	0.173[a]	0.278	0.281[c]	0.244[c]	-0.018	0.014	0.019
	(0.23)	(0.32)	(0.40)	(0.30)	(0.22)	(0.07)	(0.01)	(0.01)	(0.01)	(0.01)	(0.99)	(0.99)
$\hat{SR}_v$		0.231	0.185		0.138[b]	0.129[b]		0.169[c]	0.149[c]		0.189[c]	0.164[c]
		(0.53)	(0.65)		(0.04)	(0.01)		(0.00)	(0.00)		(0.00)	(0.00)

4.5.3 Empirical Results for Portfolios of Asset classes

In this section we examine the performance of our portfolio strategies using portfolios of asset classes. Table 4.7 shows the Sharpe ratios calculated for each portfolio of asset classes.

First we compare the in-sample performance of the optimal mean-variance and naive strategies for different portfolios. The difference between Sharpe ratios of the optimal and naive strategies is highest for portfolios 6 and 7, where we have a large number of different asset classes in the portfolio. It is lowest for portfolio 2, where we have only US and non-US equities in the portfolio. Obviously there is more room for out-of-sample strategies to outperform naive diversification for portfolios 6 and 7 than for portfolio 2. A principle component analysis reveals that 91, 85, 63, 64, and 59 percent of all variations in the returns of portfolios 3 to 7 can be explained by the first two principle components, respectively. Moreover, for portfolios 5 to 7, where we see the highest difference between Sharpe ratios of the mean-variance and naive strategies, we need at least 3 principle components to explain 77, 74, and 76 percent of the variations in asset returns.

We now examine the out-of-sample performance of the different portfolio strategies. For portfolio 1, 3 and 6, no strategy significantly outperforms the naive strategy. For portfolio 2, the minimum variance strategy and for portfolios 2, 4, 5, and 7, the volatility timing strategy significantly outperforms the $1/N$ strategy. A number of other strategies have higher Sharpe ratios than $\hat{SR}_e$ as well. For example, for portfolio 7, $\hat{SR}_{\hat{m}\hat{v},combine} = 0.174$, $\hat{SR}_{bs} = 0.172$, $\hat{SR}_{cc} = 0.161$ are higher than $\hat{SR}_e = 0.150$, and for portfolio 6, $\hat{SR}_{minv} = 0.177$ and $\hat{SR}_v = 0.163$ are higher than $\hat{SR}_e = 0.143$, but the differences for none of these cases are significant. This may be due to a low number of observations for some portfolios, as data for some assets are only available for only 90 out-of-sample periods, e.g., portfolio number 6.

To compensate for this, we use a simulation setting and generate simulated returns using the empirical distribution of asset return. For each portfolio, we use the empirical mean and variance-covariance of returns to generate a window of 24120 monthly returns. Next, we use these returns to calculate the Sharpe ratios for all investment strategies. Table 4.8 presents the results.

Using simulated returns, a larger number of out-of-sample strategies significantly outperforms naive diversification. For portfolios 1, 4, 5, 6, and 7, both $\hat{SR}_{minv}$ and $\hat{SR}_v$ outperform naive diversification. For portfolios 6 and 7, the three strategies $\hat{SR}_{\hat{m}\hat{v},\gamma=3}$, $\hat{SR}_{\hat{m}\hat{v},combine}$, and $\hat{SR}_{\hat{m}\hat{v},cc}$, and for portfolio 3, the $\hat{SR}_{\hat{m}\hat{v},cons}$ and $\hat{SR}_{\hat{m}\hat{v},cc}$ outperform naive diversification.

To conclude, comparing the in-sample and out-of-sample portfolio strategies for different portfolios of asset classes shows that when there are more assets in the portfolios and therefore possibly more factors driving asset returns, then there is more room for portfolio strategies to outperform naive diversification both in-sample and out-of-sample.

Table 4.7: Sharpe ratios for portfolios of different asset classes

The table shows the Sharpe ratio for the different portfolio strategies that described in section 4.3 for different portfolio of asset classes. Description of data and selected assets for each portfolio are presented in Table 4.3 and table 4.5. For each portfolio, N denotes the number of assets and T is the number of monthly return observations. We use a rolling approach with an estimation window length of $T_w = 120$ months for estimating parameters of each portfolio strategy. Therefore, we have $T - T_w$ out-of-sample returns to compute the Sharpe ratio for portfolio strategies that their Sharpe ratio are computed based on out-of-sample returns. The p-value of the null hypothesis $\hat{SR}_x < \hat{SR}_e$ is given in parentheses, where x indicates the strategy. a, b, and c denote the rejection of null hypothesis at the 10, 5, and 1 percent significance level, respectively.

Portfolio number	1	2	3	4	5	6	7
N	2	2	3	3	7	9	6
T	306	485	282	306	306	210	306
$T - T_w$	186	365	162	186	186	90	186
SR_e	0.156	0.094	0.122	0.153	0.169	0.138	0.130
SR_{mv}	0.193	0.095	0.172	0.200	0.219	0.295	0.217
$\hat{SR}_e$	0.127	0.107	0.085	0.143	0.130	0.143	0.150
$\hat{SR}_{mv}$	0.181	0.101	0.091	0.199	0.185	0.237	0.234[a]
	(0.15)	(0.80)	(0.44)	(0.15)	(0.12)	(0.14)	(0.08)
$\hat{SR}_{m\hat{v}}$	0.177	0.115	0.133	0.193	0.120	0.131	0.176
	(0.18)	(0.21)	(0.28)	(0.17)	(0.58)	(0.55)	(0.33)
$\hat{SR}_{\hat{m}v}$	0.143	-0.003	-0.139	0.129	0.044	0.062	0.187
	(0.41)	(0.96)	(0.99)	(0.59)	(0.89)	(0.78)	(0.32)
$\hat{SR}_{\hat{m}\hat{v}}$	0.145	-0.038	0.038	0.125	-0.003	0.026	0.156
	(0.39)	(0.98)	(0.68)	(0.63)	(0.98)	(0.87)	(0.47)
$\hat{SR}_{\hat{m}\hat{v},\gamma=3}$	0.122	0.041	-0.047	0.130	0.015	0.063	0.158
	(0.53)	(0.93)	(0.94)	(0.58)	(0.96)	(0.78)	(0.46)
$\hat{SR}_{combine}$	0.090	0.056	-0.001	0.113	0.040	0.109	0.174
	(0.76)	(0.93)	(0.90)	(0.74)	(0.97)	(0.66)	(0.35)
$\hat{SR}_{\hat{m}\hat{v},cons}$	0.101	0.101	0.056	0.094	0.066	0.001	0.127
	(0.69)	(0.64)	(0.84)	(0.86)	(0.89)	(0.99)	(0.64)
$\hat{SR}_{cc}$	0.089	0.102	0.084	0.114	0.121	0.095	0.161
	(0.88)	(0.73)	(0.54)	(0.86)	(0.65)	(0.98)	(0.37)
$\hat{SR}_{bs}$	0.154	-0.027	0.028	0.151	0.061	0.072	0.172
	(0.34)	(0.99)	(0.71)	(0.45)	(0.85)	(0.77)	(0.39)
$\hat{SR}_{minv}$	0.147	0.120[b]	0.022	0.161	0.164	0.177	0.150
	(0.41)	(0.03)	(0.98)	(0.41)	(0.33)	(0.37)	(0.50)
$\hat{SR}_v$	0.176	0.110[b]	0.079	0.182[a]	0.165[a]	0.163	0.186[a]
	(0.12)	(0.02)	(0.90)	(0.06)	(0.09)	(0.13)	(0.07)

Table 4.8: Sharpe ratio for simulated returns using different asset classes

The table reports the Sharpe ratio for the different portfolio strategies described in section 4.3, using simulated returns generated by empirical distribution of different portfolios of asset returns. Description of data and selected assets for each portfolio are presented in Table 4.3 and table 4.5. N denotes the number of assets in each portfolio. We generate a $T = 24120$ monthly returns for each portfolio, so with a rolling window of 120 months, we calculate 24000 out-of-sample returns for the portfolio strategies that use a rolling window for estimating parameters of the model. The p-value of the null hypothesis $\hat{SR}_x < \hat{SR}_e$ is given in parentheses, where x indicates the strategy. a, b, and c denote the rejection of null hypothesis at the 10, 5, and 1 percent significance level, respectively.

Portfolio number	1	2	3	4	5	6	7
N	2	2	3	3	7	9	6
T	24120	24120	24120	24120	24120	24120	24120
$T - T_w$	24000	24000	24000	24000	24000	24000	24000
SR_e	0.156	0.094	0.122	0.153	0.169	0.138	0.130
SR_{mv}	0.193	0.095	0.172	0.200	0.219	0.295	0.217
$\hat{SR}_e$	0.158	0.101	0.126	0.149	0.167	0.133	0.124
$\hat{SR}_{mv}$	0.194$^{\mathbf{c}}$	0.102	0.170$^{\mathbf{c}}$	0.199$^{\mathbf{c}}$	0.221$^{\mathbf{c}}$	0.300$^{\mathbf{c}}$	0.208$^{\mathbf{c}}$
	(0.00)	(0.21)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\hat{SR}_{m\hat{v}}$	0.194$^{\mathbf{c}}$	0.101	0.166$^{\mathbf{c}}$	0.198$^{\mathbf{c}}$	0.218$^{\mathbf{c}}$	0.288$^{\mathbf{c}}$	0.202$^{\mathbf{c}}$
	(0.00)	(0.73)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\hat{SR}_{\hat{m}v}$	0.006	0.002	0.011	0.008	0.009	0.006	0.006
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
$\hat{SR}_{\hat{m}\hat{v}}$	0.010	-0.006	0.012	0.017	0.004	0.017	0.005
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
$\hat{SR}_{\hat{m}\hat{v},\gamma=3}$	0.157	0.062	0.123	0.151	0.142	0.218$^{\mathbf{c}}$	0.141$^{\mathbf{b}}$
	(0.55)	(1.00)	(0.66)	(0.34)	(1.00)	(0.00)	(0.01)
$\hat{SR}_{combine}$	0.150	0.084	0.130	0.151	0.150	0.208$^{\mathbf{c}}$	0.145$^{\mathbf{c}}$
	(0.95)	(1.00)	(0.17)	(0.38)	(1.00)	(0.00)	(0.00)
$\hat{SR}_{\hat{m}\hat{v},cons}$	0.124	0.093	0.132$^{\mathbf{b}}$	0.115	0.120	0.124	0.112
	(1.00)	(1.00)	(0.01)	(1.00)	(1.00)	(0.97)	(0.99)
$\hat{SR}_{cc}$	0.147	0.099	0.130$^{\mathbf{c}}$	0.140	0.158	0.138$^{\mathbf{b}}$	0.130$^{\mathbf{b}}$
	(1.00)	(0.94)	(0.00)	(1.00)	(1.00)	(0.04)	(0.01)
$\hat{SR}_{bs}$	0.017	-0.005	0.014	0.039	0.022	0.030	0.009
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
$\hat{SR}_{minv}$	0.182$^{\mathbf{c}}$	0.098	0.096	0.185$^{\mathbf{c}}$	0.191$^{\mathbf{c}}$	0.231$^{\mathbf{c}}$	0.184$^{\mathbf{c}}$
	(0.00)	(1.00)	(1.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\hat{SR}_v$	0.193$^{\mathbf{c}}$	0.101	0.122	0.186$^{\mathbf{c}}$	0.201$^{\mathbf{c}}$	0.171$^{\mathbf{c}}$	0.169$^{\mathbf{c}}$
	(0.00)	(0.99)	(1.00)	(0.00)	(0.00)	(0.00)	(0.00)

4.5.4 Alternative methods for estimating means and (co)variances

There is some further discussions in the literature about possibilities to enhance the performance of portfolio optimization strategies by different ways of estimating means and variances. Kritzman, Page, and Turkington (2010) (KPT (2010) hereafter) try to improve performance using three different methods for estimating means and covariances. The first method is to estimate the means and covariances recursively, using all data available at every point in time. Panel A of Table 4.9 shows the empirical results for this method, starting from an initial 120-month estimation window.

Comparing the results in Table 4.9 with Table 4.7, estimating the parameters recursively improves the out-of-sample performance portfolio strategies in most of the cases (the main exceptions are $\hat{SR}_{\hat{m}\hat{v},cons}$ and $\hat{SR}_{minv}$). Improvement, however, is not substantial and only $\hat{SR}_{\hat{m}\hat{v},cc}$ for portfolio 5 add to the cases that significantly outperforms the $1/N$ strategy. On the other hand, the difference for $\hat{SR}_{minv}$ and $\hat{SR}_{v}$ for portfolio 2 become insignificant.

The other method proposed by KPT (2010) is to use a constant risk premium (estimated form a long history of available data) instead of estimating risk premia using a 120-month rolling window. The covariance matrix is again estimated recursively. Although this method depends on having a long history of asset returns, it seems to be feasible for many assets in our data set. We used the average excess return over the longest available period for each asset as our estimate of the constant risk premium for that asset. Panel B of Table 4.9 shows the empirical results for this method. Comparing the results with the results in table 4.7, performance of the portfolio strategies that involve the mean improve considerably. Among these strategies, $\hat{SR}_{\hat{m}\hat{v}}$ for portfolio 3, $\hat{SR}_{combine}$ for portfolio 1 and 5, and $\hat{SR}_{\hat{m}\hat{v},cons}$ and $\hat{SR}_{cc}$ for portfolio 1, add to the cases that are significantly higher than $\hat{SR}_{e}$.

Third method uses a constant risk premium for the means and a rolling window for estimating the covariance matrix. The results are presented in panel C of Table 4.9. Compared to panel B, the evidence is mixed. Sharpe ratios of some strategies increase, while those for some others decrease.

Although the feasibility of the last two methods is controversial, they may provide some way forward for optimization based strategies to outperform naive diversification. The results is, however, conditional on the financial analyst being able to provide a reasonable estimate of risk premia for each asset.

Table 4.9: Sharpe ratio for different portfolio of asset classes: alternative estimation methods

The table reports the Sharpe ratios for the different portfolio strategies sing some alternative methods proposed by KPT (2010) for estimating the mean and variance-covariance of asset returns. We use the same data sets that used in Table 4.7. In panel A, we estimate $\hat{\mu}$ and $\hat{\Sigma}$ recursively, starting from an initial 120-month estimation window. In panel B, we use a constant risk premium $\bar{\mu}$ and a recursive window for estimating $\hat{\Sigma}$. In panel C, we use a constant risk premium $\bar{\mu}$ and a 120-month rolling window to estimate $\hat{\Sigma}$. The p-value of the null hypothesis $\hat{SR}_x < \hat{SR}_e$ is given in parentheses, where x indicates the strategy. a, b, and c denote the rejection of null hypothesis at the 10, 5, and 1 percent significance level, respectively.

Portfolio	1	2	3	4	5	6	7	1	2	3	4	5	6	7
N	2	2	3	3	7	9	6	2	2	3	3	7	9	6
T	306	485	282	306	306	210	306	306	485	282	306	306	210	306
$T - T_w$	186	365	162	186	186	90	186	186	365	162	186	186	90	186
$\hat{SR}_e$	0.127	0.107	0.085	0.143	0.130	0.143	0.150	0.127	0.107	0.085	0.143	0.130	0.143	0.150
	Panel A: recursive window to estimate μ and Σ							Panel B: constant risk premium $\bar{\mu}$ and recursive window to estimate Σ						
$\hat{SR}_{\hat{m}\hat{v}}$	0.149	0.003	0.079	0.137	0.053	0.000	0.082	0.173	0.105	0.134[b]	0.181	0.165	0.233	0.184
	(0.31)	(0.95)	(0.56)	(0.54)	(0.95)	(0.94)	(0.85)	(0.27)	(0.61)	(0.01)	(0.25)	(0.31)	(0.11)	(0.27)
$\hat{SR}_{\hat{m}\hat{v},\gamma=3}$	0.144	0.049	0.024	0.133	0.054	0.005	0.085	0.172	0.103	0.099	0.177	0.161	0.206	0.168
	(0.35)	(0.96)	(0.87)	(0.57)	(0.95)	(0.93)	(0.86)	(0.27)	(0.76)	(0.29)	(0.28)	(0.34)	(0.20)	(0.38)
$\hat{SR}_{combine}$	0.119	0.097	0.063	0.123	0.102	0.081	0.138	0.172[a]	0.107	0.078	0.165	0.156[a]	0.143	0.151
	(0.64)	(0.80)	(0.79)	(0.87)	(0.80)	(0.90)	(0.63)	(0.07)	(0.63)	(0.76)	(0.11)	(0.09)	(0.49)	(0.48)
$\hat{SR}_{\hat{m}\hat{v},cons}$	0.082	0.091	0.070	0.080	0.057	0.038	0.066	0.167[a]	0.107	0.102	0.128	0.132	0.170	0.143
	(0.95)	(0.82)	(0.76)	(0.94)	(0.97)	(0.93)	(0.93)	(0.10)	(0.44)	(0.12)	(0.65)	(0.48)	(0.24)	(0.55)
$\hat{SR}_{cc}$	0.117	0.105	0.085	0.141	0.147[b]	0.124	0.156	0.138[b]	0.107	0.080	0.144	0.131	0.143	0.154
	(0.82)	(0.65)	(0.50)	(0.58)	(0.02)	(0.85)	(0.37)	(0.01)	(0.65)	(0.74)	(0.44)	(0.41)	(0.49)	(0.27)
$\hat{SR}_{bs}$	0.162	0.025	0.051	0.167	0.124	0.114	0.156	0.156	0.111	0.068	0.177	0.183	0.212	0.188
	(0.31)	(0.95)	(0.89)	(0.37)	(0.54)	(0.63)	(0.46)	(0.37)	(0.18)	(0.90)	(0.34)	(0.22)	(0.22)	(0.29)
$\hat{SR}_{minv}$	0.155	0.111	0.036	0.168	0.165	0.198	0.186	0.155	0.111	0.036	0.168	0.165	0.198	0.186
	(0.38)	(0.18)	(0.97)	(0.39)	(0.31)	(0.29)	(0.31)	(0.38)	(0.18)	(0.97)	(0.39)	(0.31)	(0.29)	(0.31)
$\hat{SR}_v$	0.179	0.108	0.078	0.180[a]	0.165[a]	0.163	0.184[a]	0.179	0.108	0.078	0.180[a]	0.165[a]	0.163	0.184[a]
	(0.11)	(0.14)	(0.91)	(0.06)	(0.07)	(0.12)	(0.08)	(0.11)	(0.14)	(0.91)	(0.06)	(0.07)	(0.12)	(0.08)
Panel C: constant risk premium $\bar{\mu}$ and rolling window to estimate Σ														
$\hat{SR}_{\hat{m}\hat{v}}$	0.164	0.115	0.156	0.183	0.176	0.223	0.152							
	(0.30)	(0.21)	(0.13)	(0.25)	(0.26)	(0.17)	(0.49)							
$\hat{SR}_{\hat{m}\hat{v},\gamma=3}$	0.158	0.111	0.108	0.171	0.168	0.231	0.124							
	(0.33)	(0.38)	(0.29)	(0.32)	(0.31)	(0.15)	(0.66)							
$\hat{SR}_{combine}$	0.167	0.101	0.077	0.173	0.172	0.219[b]	0.137							
	(0.19)	(1.00)	(0.79)	(0.17)	(0.24)	(0.05)	(0.70)							
$\hat{SR}_{\hat{m}\hat{v},cons}$	0.148	0.118[a]	0.101	0.134	0.127	0.163	0.134							
	(0.17)	(0.05)	(0.16)	(0.58)	(0.53)	(0.31)	(0.61)							
$\hat{SR}_{cc}$	0.141[b]	0.104	0.089	0.146	0.129	0.140	0.154							
	(0.02)	(0.97)	(0.31)	(0.38)	(0.52)	(0.65)	(0.31)							
$\hat{SR}_{bs}$	0.148	0.120[b]	0.103	0.169	0.180	0.193	0.150							
	(0.40)	(0.02)	(0.30)	(0.37)	(0.25)	(0.30)	(0.50)							
$\hat{SR}_{minv}$	0.147	0.120[b]	0.022	0.161	0.164	0.177	0.150							
	(0.41)	(0.03)	(0.98)	(0.41)	(0.33)	(0.37)	(0.50)							
$\hat{SR}_v$	0.176	0.110[b]	0.079	0.182[a]	0.165[a]	0.163	0.186[a]							
	(0.12)	(0.02)	(0.90)	(0.06)	(0.09)	(0.13)	(0.07)							

4.6 Conclusions

We compared several portfolio strategies proposed in the literature under a variety of data generating structures. The focus of our comparison was to characterize under what data generating structure, means and covariances can be exploited to outperform a naive investment strategy, known as the $1/N$ rule. We showed analytically that when the data are generated by a one-factor model, there is hardly any difference between the Sharpe ratio of the optimal tangency portfolio and that of the $1/N$ strategy. Accounting for estimation error, therefore, we can *a priori* expect no other portfolio strategy to significantly outperform the $1/N$ strategy out-of-sample, regardless of the estimation method used. A simulation approach to compare different estimators and portfolio strategies is therefore not very informative if based on a one-factor data generating process. We also derived the analytical formulas for the Sharpe ratios of naive and mean-variance strategies when data are generated by a general factor model. As an illustration, we used a simulation setting. For data sets that are generated by a two-factor model, we showed how different parameters of data generating process affect the Sharpe ratios of two strategies. We showed that there are circumstances under which in-sample optimal mean-variance strategy can significantly outperform the $1/N$ strategy, but these conditions are hardly satisfied in practice. Consequently, also in two-factor models, the ability of optimization-based portfolio strategies to outperform naive diversification out-of sample is very limited.

We also used different types of empirical data and compared the in-sample and out-of-sample performance of different portfolio strategies proposed in the literature. We used equity portfolios and different asset classes. When there are sufficient factors driving underlying asset returns in the portfolio, there is more room for portfolio strategies to outperform naive diversification both in sample and out-of-sample. In particular, the combined strategy proposed by Tu and Zhou (2011), the minimum variance strategy, and the volatility timing strategy outperform naive diversification in a number of empirical tests.

4.A Appendix: Proofs

4.A.1 Proof of Proposition 1

From the one-factor structure of the model, we obtain

$$Var[r] = \beta Var[r_m]\beta' + Var(\epsilon) = \sigma_m^2 \beta\beta' + \sigma_\epsilon^2 \mathrm{I}, \tag{A1}$$

$$E[r] = \beta E[r_m] = \beta\mu_m. \tag{A2}$$

The optimal mean-variance weights are proportional to

$$\tilde{w}_0 \propto (Var[r])^{-1}\beta\mu_m. \tag{A3}$$

Using the well-known result

$$(A + buv')^{-1} = A^{-1} + \frac{b}{1 + bv'A^{-1}u} A^{-1}uv'A^{-1}, \tag{A4}$$

for $A \in \mathbb{R}^{N \times N}$, $u, v \in \mathbb{R}^{N \times 1}$, and $b \in \mathbb{R}$. Using $A = \sigma_\epsilon^2 \mathrm{I}$, $u = v = \beta$, $b = \sigma_m^2$, and $q = \beta'\beta$, we obtain

$$\tilde{w}_0 \propto (\sigma_\epsilon^{-2} - \frac{\sigma_\epsilon^{-2} q \sigma_m^2}{\sigma_\epsilon^2 + q\sigma_m^2})\beta\mu_m \propto \beta \tag{A5}$$

With these optimal weights, the Sharpe ratio becomes

$$\mu(\tilde{w}) = \frac{\tilde{w}_0'\beta\mu_m}{\tilde{w}_0'\mathbf{1}} = \frac{q\mu_m}{\tilde{w}_0'\mathbf{1}} \tag{A6}$$

$$\sigma^2(\tilde{w}) = var(\tilde{w}'r_t) = \frac{\beta'(\sigma_m^2\beta\beta' + \sigma_\epsilon^2 \mathrm{I})\beta}{(\tilde{w}_0'\mathbf{1})^2} = \frac{(q^2\sigma_m^2 + q\sigma_\epsilon^2)}{(\tilde{w}_0'\mathbf{1})^2}, \tag{A7}$$

$$SR_{mv} = \frac{\mu(\tilde{w})}{\sqrt{\sigma^2(\tilde{w})}} = \frac{\mu_m}{\sqrt{\sigma_m^2 + \sigma_\epsilon^2/q}}. \tag{A8}$$

For the naive diversification strategy $\bar{w} = 1/N$, we have

$$\mu(\bar{w}) = \bar{w}'\beta\mu_m = \bar{\beta}\mu_m, \tag{A9}$$

$$\sigma^2(\bar{w}) = var(\bar{w}'r_t) = \bar{w}'(\sigma_m^2\beta\beta' + \sigma_\epsilon^2 \mathrm{I})\bar{w} = \bar{\beta}^2\sigma_m^2 + \frac{\sigma_\epsilon^2}{N}, \tag{A10}$$

$$SR_e = \frac{\mu_m}{\sqrt{\sigma_m^2 + \sigma_\epsilon^2/(N\bar{\beta}^2)}}, \tag{A11}$$

with $\bar{\beta} = \bar{w}'\beta$.

4.A.2 Proof of Proposition 2

The optimal mean-variance weights are given by

$$\tilde{w} \propto (V_\varepsilon + \beta V_f \beta')^{-1} \beta\mu_f, \tag{A12}$$

and the naive diversification strategy's weight by $\bar{w} = \mathbf{1}/N$. The Sharpe ratio for the equally weighted portfolio is

$$SR_e = \frac{\bar{\beta}'\mu_f}{\sqrt{\bar{\beta}'V_f\bar{\beta} + \mathbf{1}'V_\varepsilon\mathbf{1}/N^2}} \tag{A13}$$

$$= \sqrt{\mu_f'\bar{\beta}\left(\bar{\beta}'V_f\bar{\beta} + \frac{\mathbf{1}'V_\varepsilon\mathbf{1}}{N^2}\right)^{-1}\bar{\beta}'\mu_f} \tag{A14}$$

$$= \sqrt{\mu_f'\bar{B}\left(V_f^{-1} + \bar{B}\right)^{-1}V_f^{-1}\mu_f}, \tag{A15}$$

with $\bar{\beta} = \beta'\mathbf{1}/N$ and $\bar{B} = N^2\bar{\beta}\bar{\beta}'/(\mathbf{1}'V_\varepsilon\mathbf{1})$, while that for the mean-variance optimal portfolio is

$$\begin{aligned}
SR_{mv} &= \sqrt{\mu_f'\beta'\left(V_\varepsilon + \beta V_f\beta'\right)^{-1}\beta\mu_f} && \text{(A16)}\\
&= \sqrt{\mu_f'\beta'\left(V_\varepsilon^{-1} - V_\varepsilon^{-1}\beta\left(V_f^{-1} + \beta'V_\varepsilon^{-1}\beta\right)^{-1}\beta'V_\varepsilon^{-1}\right)\beta\mu_f} && \text{(A17)}\\
&= \sqrt{\mu_f'B\mu_f + \mu_f'B\left(V_f^{-1} + B\right)^{-1}B\mu_f} && \text{(A18)}\\
&= \sqrt{\mu_f'B\left(V_f^{-1} + B\right)^{-1}V_f^{-1}\mu_f} && \text{(A19)}\\
&= \sqrt{\mu_f'B\left(\mathrm{I} + V_fB\right)^{-1}\mu_f} && \text{(A20)}\\
&= \sqrt{\mu_f'\left(B^{-1} + V_f\right)^{-1}\mu_f}, && \text{(A21)}
\end{aligned}$$

where $B = \beta'V_\varepsilon^{-1}\beta$, and where the last equality only holds if B is invertible. Otherwise, (A19) is takes as the final results.

As we can transform the factors f and factor loadings β to $\tilde{f} = V_f^{-1/2}f$ and $\tilde{\beta} = \beta V_f^{1/2}$ without changing the factor model itself, we can without loss of generality set $V_f = \mathrm{I}$. Based on Equations (A19) and (A15), the squared Sharpe Ratio for the mean-variance optimal strategy and the naive diversification strategy are given by

$$SR_{mv}^2 = \mu_f'B(\mathrm{I} + B)^{-1}\mu_f, \tag{A22}$$

and

$$SR_e^2 = \mu_f'\bar{B}(\mathrm{I} + \bar{B})^{-1}\mu_f, \tag{A23}$$

respectively. Define

$$C = B - \bar{B} \quad \Leftrightarrow \quad B = C + \bar{B}. \tag{A24}$$

Subtracting (A23) from (A22), we get

$$\begin{aligned}
SR_{mv}^2 - SR_e^2 &= \mu_f'\left(B(\mathrm{I} + B)^{-1} - \bar{B}(\mathrm{I} + \bar{B})^{-1}\right)\mu_f && \text{(A25)}\\
&= \mu_f'\left(B - \bar{B}(\mathrm{I} + \bar{B})^{-1}(\mathrm{I} + B)\right)(\mathrm{I} + B)^{-1}\mu_f && \text{(A26)}\\
&= \mu_f'\left(\bar{B} + C - \bar{B}(\mathrm{I} + \bar{B})^{-1}(\mathrm{I} + \bar{B} + C)\right)(\mathrm{I} + B)^{-1}\mu_f && \text{(A27)}\\
&= \mu_f'\left(C - \bar{B}(\mathrm{I} + \bar{B})^{-1}C\right)(\mathrm{I} + B)^{-1}\mu_f && \text{(A28)}\\
&= \mu_f'\left(\mathrm{I} - \bar{B}(\mathrm{I} + \bar{B})^{-1}\right)C(\mathrm{I} + B)^{-1}\mu_f && \text{(A29)}\\
&= \mu_f'(\mathrm{I} + \bar{B})^{-1}(B - \bar{B})(\mathrm{I} + B)^{-1}\mu_f. && \text{(A30)}
\end{aligned}$$

Note that if $V_\varepsilon = \sigma_\varepsilon^2\mathrm{I}$, $C = N \cdot \mathrm{Cov}_N(\beta)/\sigma_\varepsilon^2$, i.e., N/σ_ε^2 times the sample covariance matrix of the βs.