Optimal Asset Allocation: How Many Miles are we From the Promised Gains?
(Preliminary and incomplete. Please, do not circulate.)

Pedro Engel Caio Almeida
April 9, 2017

Abstract
Garlappi Uppal tested different models in a mean variance portfolio problem and found out that with a large dataset, none were consistently better than the 1/N rule in terms of Sharpe ratio, certainty-equivalent return, or turnover. We evaluate here the performance of different portfolio strategies suggested by the literature with an up-to-date dataset and try to achieve a better portfolio strategy that consider not only out of sample performance but take into account the in sample results and estimation risk as possible references for a good portfolio criteria.
1 Introduction

In this paper we seek to answer the following questions: Does portfolio optimization models give good results for portfolio managers when compared to "naive" strategies? Is it important to incorporate uncertainty about return distribution when building a portfolio strategy? These are only a few but important questions that we are going to investigate here.

(DeMiguel, Garlappi, and Uppal 2009) evaluate the out-of-sample performance of optimal portfolio strategies and considered that none was consistently better than the 1/N rule in terms of Sharpe ratio, certainty equivalent return, or turnover. They conclude: "This suggests that there are still many miles to go before the gains promised by optimal portfolio choice can actually be realized out of sample". We evaluate their claim in light of recent data from industry portfolio returns and show that although no single portfolio strategy beat the naive strategy, the combination of the naive strategy with the minimum variance portfolio seems promising.

When evaluating an optimal strategy, there are some basic principles we must assume. What is the individual objective function? What set of information is available at the investment date? What is known and what is unknown and must be estimated from the data. What are the individuals believes about the unknown? All of that must be carefully treated when carrying on an investment analysis. More than that it is important to know how robust the model is with respect to the data. For instance if a model take very different results with only a little more or less of information it cannot be considered robust for a practical purpose. We are going to see that individual believes take a great deal in turning the portfolio model robust in the sense that it takes a great amount of data to change the perception of the investor which in turns bring some desirable robustness to the portfolio problem.

The problem of portfolio allocation in economics is an old and recurrent topic. The first appealing approach was due to (Markowitz 1952). Markowitz described a preference ordering where individuals were eager for more return but dislike volatility. Since more returns reflects more consumption and volatility reflects consumption uncertainty, these should be the main concerns for any allocation problem. The partial ordering then implies that for any 2 portfolios, the one that has more expected return given a fixed level of bearable risk (variance) should be preferred, and given a fixed level of return, the portfolio with less variance should be the one preferred. One problem that comes along with this line of reasoning is that when an individual solves for the optimal portfolio he does not know the true mean or covariance of the portfolio assets. (Black and Litterman 1992) points out that the resulting optimal portfolio is very sensitive to mean estimation. They propose a more robust estimation for the mean that allows to incorporate investors views. (Glasserman and Xu 2014) incorporate individuals uncertainty about parameters estimate into the portfolio problem.
(Garlappi, Uppal, and Wang 2007) Develop a model of Portfolio Selection with parameter and model uncertainty. Others like Jessica and Missaka consider the case if individuals should or should not time the market. The idea for market timing comes from the fact that when more information is available, better predictability about the return distribution can be made. In this line, (Amihud and Hurvich 2004), Ang, Andrew and Geert Bekaert, 2007 and (Avramov 2004) show how some market information can be useful to predict stock return.

The article is divided as follows: Section 2 introduces the basic mean variance portfolio model, its known caveats, and extensions that seek to achieve reduced estimation risk; sections 3 presents a more detailed construction of the Bayesian models with particular description of Jorion Bayes-stein portfolio strategy and Pastor Stambaugh model of investors with asset pricing beliefs. Section 4 briefly describes the models to be tested; section 5 the empirical exercise and the performance criteria; section 6 presents the results; and for the last section we conclude.

2 The Basic Model and Its Extensions

We start analyzing the standard portfolio problem to assess its weaknesses and evaluate its ramifications. We consider here the case where the individuals maximize their expected utility of the terminal wealth

$$E[U(W_T)]$$

where the terminal wealth is given by

$$W_T = W_0(1 + r_p)$$

Letting $r_p = (1 - \sum_{i=1}^{n} a_i) r_f + \sum_{i=1}^{n} a_i r_i$, or equivalently $r_p(a) = r_f + \sum_{i=1}^{n} a_i r_i^e$ where $r_i^e = r_i - r_f$. It is easy to see that we can describe the preferences in terms of assets excess return with respect to the risk free rate. In particular, (Kandel and Stambaugh 1996), Stambaugh(1999) and Barberis (2000), consider the case where the investor’s preferences over terminal wealth are described by constant relative-risk aversion power utility function of the form

$$v(W) = \frac{W^\gamma}{\gamma}$$

Following Jurczenko and Maillet (2006), the individual problem is then

$$\max_a E[U(W_0(1 + r_p(a)))]$$

which yields as first order condition

$$E[U'(W)r_i^e] = 0$$
By doing a second Taylor expansion of the marginal utility over the expected wealth $\bar{W} = W_0(1 + \mu)$, where $\mu = E[r_p]$, we get

$$U'(W) \approx U'(\bar{W}) + U''(W)W_0(r_p - \mu) + \frac{1}{2}U'''(W)W_0^2(r_p - \mu)^2$$

which yields the approximate first order condition

$$E[r_i] + \frac{U''(W)}{U'(W)}W_0E[r_i(r_p - \mu)] + \frac{1}{2}\frac{U'''(W)}{U'(W)}W_0^2E[r_i(r_p - \mu)^2] = 0$$

The term $-U''(.)W/U'(.)$ is known as the Arrow Pratt relative risk aversion coefficient and the term $(1/2)U'''(.)W^2/U'(.)$ is known as the prudence coefficient. Note that for the CRRA case, where $U(W) = W^{1-\gamma}/(1-\gamma)$, we have

$$-U''(.)W/U'(.) = \gamma$$

and

$$(1/2)U'''(.)W^2/U'(.) = (1/2)\gamma(\gamma + 1).$$

So, if we consider just the first order approximation, we get

$$E[r_i] - \gamma\sum_{i=1}^n a_i^*E[(r_i - \mu_e)(r_j - \mu_j)] = 0$$

that can be conveniently written in vector form

$$\tilde{\mu} - \gamma\tilde{\Sigma}a^* = 0$$

where $\tilde{\mu}$ is the vector of expected excess returns of the risk assets and $\tilde{\Sigma}$ is its covariance matrix.

By observing equation (3), we can see that this is the FOC of the quadratic problem

$$\max_a \{a'\tilde{\mu} - \frac{\gamma}{2}a'\tilde{\Sigma}a\}$$

and its solution is given by $a^* = \gamma^{-1}\tilde{\Sigma}^{-1}\tilde{\mu}$.

Even in this simple framework that considers only the first 2 moments of the excess return distribution it is important to note that they are unknown parameters for the portfolio manager. Since the parameters are unknown, they must be estimated from the data available at the investment date. The most simple way to estimate these parameters and make this formula useful for a practical portfolio decision problem is to compute the sample mean and the sample covariance matrix of excess returns $\hat{\mu}$ and $\hat{\Sigma}$.

In summary, the feasible problem of an investor is

$$\max_a a'\hat{\mu} - \frac{\gamma}{2}a'\hat{\Sigma}a$$
where $\hat{\mu}$ and $\hat{\Sigma}$ must be estimated from the data. The simple choice is to use the sample mean and the sample covariance of excess returns given by

$$
\hat{\mu}_i = \frac{1}{T} \sum_{t=1}^{T} r_{it} 
$$

(6)

$$
\hat{\Sigma}_{ij} = \frac{1}{T-1} \sum_{t=1}^{T} (r_{it} - \hat{\mu}_i)(r_{jt} - \hat{\mu}_j)
$$

(7)

One problem with this approach pointed out by Black and Litterman is that the portfolio choice is very sensitive to the mean estimate. That is, it translates into very different portfolios depending on the sample used. Another problem is that if we are concerned about estimation risk as is going to be defined in the next section, the sample mean and sample variance should not be the best estimators to be used.

2.1 Problems of the Basic Model

In a mean-variance framework the objective function of an investor is given by the utility function

$$
U(\bar{r}) = E[u(a'\bar{r})] = E[a'\bar{r} - \frac{\gamma}{2} a'\bar{r} a] = a'\hat{\mu} - \frac{\gamma}{2} a'\hat{\Sigma} a
$$

where $\bar{r} = r_p - \mu$ is the portfolio excess return and $\bar{r}$ is the random vector of assets excess returns with respect to the risk free rate.

If the true parameters $\theta = (\hat{\mu}, \hat{\Sigma})$ were known, then the choice $a^* = a(\theta) = \gamma^{-1}\Sigma^{-1}\hat{\mu}$ would be optimal by construction, and

$$
F(\theta, \theta) = a(\theta)'\hat{\mu} - \frac{\gamma}{2} a(\theta)'\hat{\Sigma} a(\theta) = F_{\text{max}}
$$

Since the parameters are unknown, $\theta$ must be estimated from the data, $\hat{\theta}(z) = (\hat{\mu}(z), \hat{\Sigma}(z))$, which implies a decision rule $\hat{a} = a(\hat{\theta}) = \gamma^{-1}\hat{\Sigma}^{-1}\hat{\mu}$ that necessarily leads to a lower utility level, that is,

$$
F(\theta, \hat{\theta}(z)) = a(\hat{\theta})'\hat{\mu} - \frac{\gamma}{2} a(\hat{\theta})'\hat{\Sigma} a(\hat{\theta}) \leq F_{\text{max}}
$$

This loss of utility due to parameter uncertainty is known as estimation risk. An estimation loss function is a non-negative function $L(\theta, \hat{\theta})$ that achieves its minimum at $\hat{\theta} = \theta$. If we define

$$
L(\theta, \hat{\theta}) = F(\theta, \hat{\theta}) - F(\theta, \hat{\theta})
$$

Since the optimal rule for the problem is by construction $a(\theta)$, its clear that $L(\theta, \hat{\theta}) \geq 0$ for $\theta \neq \hat{\theta}$ and $L(\theta, \hat{\theta}) = 0$ if $\theta = \hat{\theta}$. 
The risk function for an estimator \( \hat{\theta}(z) \) is defined as

\[
R(\theta, \hat{\theta}) = \int L(\theta, \hat{\theta}(z)) f(z|\theta) d\theta
\]

where \( f(z|\theta) \) is the likelihood function of the sample \( z \). It is clear that for a given estimator \( \hat{\theta} \), the risk function is a function of the parameter \( \theta \in \Theta \).

An estimator \( \hat{\theta} \) is said to be inadmissible if there exist another estimator \( \hat{\theta} \) with at least equal and sometimes lower risk for any possible value of the true unknown parameter \( \theta \). That is, \( \hat{\theta} \) is inadmissible if there exists some \( \hat{\theta} \) with

\[
R(\theta, \hat{\theta}) \leq R(\theta, \hat{\theta}) \quad \forall \theta \in \Theta
\]

With the inequality strict for some \( \theta \).

If we treat \( \theta \) as random, with an a priori density \( p(\theta) \), then the average loss from use of an estimator \( \hat{\theta} \) is

\[
r(p, \hat{\theta}) = E_p[R(\theta, \hat{\theta})] = \int \left( \int L(\theta, \hat{\theta}(z)) f(z|\theta) dz \right) p(\theta) d\theta
\]

\[
= \int \left( \int L(\theta, \hat{\theta}(z)) g(\theta|z) d\theta \right) f(z) dz
\]

where

\[
f(z) = \int f(z|\theta) p(\theta) d\theta
\]

and

\[
g(\theta|z) = \frac{f(z|\theta)p(\theta)}{f(z)}
\]

that is, the marginal density of \( z \) and the a posteriori density of \( \theta \) given \( z \).

Given a priori density \( p \), the estimator \( \hat{\theta} \) that minimizes \( r(p, \hat{\theta}) \) is the Bayes estimator, and the resulting minimum is the Bayes risk. The estimator that minimizes \( r(p, \hat{\theta}) \) is one that for each \( z \) minimizes the expression in parenthesis, that is, the expectation of \( L(\theta, \hat{\theta}(z)) \) with respect to the a posteriori distribution.

If, for instance, \( \theta \) and \( \hat{\theta}(z) \) for a given \( z \) are vectors and \( L(\theta, \hat{\theta}(z)) = (\theta - \hat{\theta})' Q (\theta - \hat{\theta}) \), where \( Q \) is positive definite, then

\[
E_{\theta|z}[L(\theta, \hat{\theta}(z))] = E_{\theta|z}[(\theta - E[\theta|z])' Q (\theta - E[\theta|z])] + (E[\theta|z] - \hat{\theta}(z))' Q (E[\theta|z] - \hat{\theta}(z))
\]

which is minimized at \( \hat{\theta}(z) = E[\theta|z] \), the mean of the a posteriori distribution.
Going back to our particular case we have \( a^* = a(\theta) = \gamma^{-1} \tilde{\Sigma}^{-1} \tilde{\mu} \), which yields a resulting utility of

\[
F(\theta, \theta) = a(\theta)' \tilde{\mu} - \frac{\gamma}{2} a(\theta)' \tilde{\Sigma} a(\theta) = \frac{1}{2\gamma} \tilde{\mu} \tilde{\Sigma}^{-1} \tilde{\mu} = \frac{S^2}{2\gamma}
\]

where \( s^2 \) is the squared Sharpe ratio of the ex ante tangency portfolio of the risky assets. We have seen that given \( \theta \), the risk function is given by

\[
R(\theta, \hat{\theta}(z)) = E_{\tilde{z}|\theta}[L(\theta, \hat{\theta}(z))] = F(\theta, \theta) - E_{\tilde{z}|\theta}[F(\theta, \hat{\theta}(z))]
\]

Considering \( z = \{\tilde{r}_1, ..., \tilde{r}_T\} \) is the matrix of \( T \) observations of vector excess returns and \( \hat{\theta} = (\hat{\mu}, \hat{\Sigma}) \) is the sample mean and sample covariance computed as in (6) and (7). Following Kan and Zoh, it is interesting to consider the case when \( \tilde{r} \) are independent and normally distributed random vectors each of which has common mean and variance \( \tilde{\mu} \) and \( \tilde{\Sigma} \). Under these assumptions, it is well known that \( \hat{\mu} \) and \( \hat{\Sigma} \) are independent of each other and have the following exact distributions,

\[
\hat{\mu} \sim N(\tilde{\mu}, \tilde{\Sigma}),
\]

\[
T \hat{\Sigma} \sim W(T - 1, \Sigma),
\]

where \( W(T - 1, \Sigma) \) denotes a Wishart distribution with \( T - 1 \) degrees of freedom and covariance matrix \( \Sigma \). Since \( E[\Sigma] = T \Sigma^{-1} / (T - N - 2) \), we have

\[
E[\hat{\mu}] = E[\gamma^{-1} \tilde{\Sigma}^{-1} \hat{\mu}] = \frac{T}{T - N - 2} a(\theta)
\]

When \( \Sigma \) is known, \( \hat{\mu} = \Sigma^{-1} \hat{\mu} / \gamma \), and since \( T \hat{\mu}' \Sigma^{-1} \hat{\mu} \sim \chi^2(T \mu' \Sigma \mu) \), we have

\[
E_{\tilde{z}|\theta}[F(\theta, \hat{\theta}(z))] = E[\hat{\mu}]' \mu - \frac{\gamma}{2} E[\hat{\mu}' \Sigma \hat{\mu}]
\]

\[
= \frac{\mu' \Sigma \mu}{\gamma} - \frac{1}{2\gamma} E[\hat{\mu}' \Sigma^{-1} \hat{\mu}]
\]

\[
= \frac{\mu' \Sigma \mu}{\gamma} - \frac{1}{2\gamma} \left( \frac{N + T \mu' \Sigma \mu}{T} \right)
\]

\[
= \frac{S^2}{2\gamma} - \frac{N}{2\gamma T}
\]

As a result, the risk function from using \( \hat{\theta} \) instead of \( \theta \) is

\[
R(\theta, \hat{\theta}(z)) = F(\theta, \theta) - E_{\tilde{z}|\theta}[F(\theta, \hat{\theta}(z))] = \frac{N}{2\gamma T}
\]

The result is intuitive. As large is the sample size \( T \), more is learned about the true parameter \( \theta \) and the estimation loss is reduced. In the limiting case where
$T \to \infty$ the estimator $\hat{\theta} \to \theta$ and the loss from estimation risk goes to zero. On the other hand, as big is the size $N$ of the vector of unknown parameters, more uncertainty is present, since more parameters must be estimated, bigger is the loss. Finally, the bigger $\gamma$, more risk averse is the investor, so he invest less in the risky assets and the loss due to uncertainty is smaller.

To consider the more general case where $\mu$ and $\Sigma$ are unknown, we must use 2 known results about the Wishart distribution. Let $W = \Sigma^{-1/2} \Sigma^{-1} \sim W_N(T - 1, I_N)/T$. The inverse moments of $W$ are

\[
E[W^{-1}] = \left( \frac{T}{T - N - 2} \right) I_N
\]

\[
E[W^{-2}] = \left[ \frac{T^2(T - 2)}{(T - N - 1)(T - N - 2)(T - N - 4)} \right] I_N
\]

With these results and the fact that $\hat{\mu}$ and $\hat{\Sigma}$ are independent, we can easily compute the expected out-of-sample performance

\[
E_{z|\theta}[F(\theta, \hat{\theta}(z))] = \frac{1}{\gamma} E[\hat{\mu}' \hat{\Sigma}^{-1} \mu] - \frac{1}{2\gamma} E[\hat{\mu}' \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \hat{\mu}]
\]

\[
= \frac{1}{\gamma} E[\hat{\mu}' \Sigma^{-1/2} W^{-1} \Sigma^{-1/2} \mu] - \frac{1}{2\gamma} E[\hat{\mu}' \Sigma^{-1/2} W^{-2} \Sigma^{-1/2} \hat{\mu}]
\]

Now, it is easy to see that

\[
E[\hat{\mu}' \Sigma^{-1/2} W^{-1} \Sigma^{-1/2} \mu] = \left( \frac{T}{T - N - 2} \right) S^2
\]

and with a little extra work, we get

\[
E[\hat{\mu}' \Sigma^{-1/2} W^{-2} \Sigma^{-1/2} \hat{\mu}] = E[tr(\Sigma^{-1/2} \hat{\mu} \hat{\Sigma}^{-1/2} \mu - \Sigma^{-1/2} \hat{\mu} \hat{\Sigma}^{-1/2} \hat{\mu})]
\]

\[
= tr(E[\Sigma^{-1/2} \hat{\mu} \hat{\Sigma}^{-1/2} W^{-2}])
\]

\[
= \left[ \frac{T^2(T - 2)}{(T - N - 1)(T - N - 2)(T - N - 4)} \right] tr(E[\Sigma^{-1/2} \hat{\mu} \hat{\Sigma}^{-1/2} W^{-2}])
\]

\[
= \left[ \frac{T^2(T - 2)}{(T - N - 1)(T - N - 2)(T - N - 4)} \right] E[\hat{\mu}' \Sigma^{-1} \hat{\mu}]
\]

\[
= \left[ \frac{T^2(T - 2)}{(T - N - 1)(T - N - 2)(T - N - 4)} \right] \left( \frac{N + T \mu' \Sigma^{-1} \mu}{T} \right)
\]

(11)

combining (10) with (11), we have

\[
E_{z|\theta}[F(\theta, \hat{\theta}(z))] = k_1 \frac{S^2}{2\gamma} - \frac{NT(T - 2)}{2\gamma(T - N - 1)(T - N - 2)(T - N - 4)}
\]

(12)

where

\[
k_1 = \left( \frac{T}{T - N - 2} \right) \left[ 2 - \frac{T(T - 2)}{(T - N - 1)(T - N - 4)} \right]
\]
Hence, the out-of-sample performance of the MLE for $\theta$ when $\mu$ and $\Sigma$ are unknown is

$$R(\theta, \hat{\theta}(z)) = F(\theta, \theta) - E_{z}[F(\theta, \hat{\theta}(z))]$$

$$= (1 - k_1) \frac{S^2}{2\gamma} + \frac{NT(T - 2)}{2\gamma(T - N - 1)(T - N - 2)(T - N - 4)}$$

This is a closed form solution that relates the expected loss with respect to $N, T, \gamma$, and $S^2$. The intuition remains, as $N$, or $S^2$ increases, the loss increases. On the other hand, as $T$ or $\gamma$ increases, the loss decreases.

Next we are going to show that the sample covariance mean and Covariance is inadmissible in the sense that there exist another estimator that provides a smaller loss, regardless of the values of the true parameters.

### 2.2 The Bayesian Solution

The Bayesian approach assumes that the investor cares about the expected utility under the predictive distribution $p(r_{t+1}|z)$, which is determined by the historical data and the prior. Brown (1976), Klein and Bawa (1976), and (Stambaugh 1997) show under the diffuse prior on $\mu$ and $\Sigma$,

$$p(\mu, \Sigma) \propto |\Sigma|^{-\frac{N+1}{2}}$$

The optimal portfolio weights computed by the predictive moments in this case is

$$\tilde{a} = \frac{1}{\gamma} \left( \frac{T - N - 2}{T + 1} \right) \tilde{\Sigma}^{-1} \tilde{\mu} = \left( \frac{T - N - 2}{T + 1} \right) \tilde{a}$$

The question is if this approach actually improves out of sample performance in terms of reduced loss.

$$E_{z}[F(\theta, \tilde{\theta}(z))] = \left( \frac{T - N - 2}{T + 1} \right)^{\frac{1}{2}} E[\mu^T \tilde{\Sigma}^{-1} \mu]$$

and using (10) and (11) we get

$$E_{z}[F(\theta, \tilde{\theta}(z))] = k_2 \frac{S^2}{2\gamma} - \frac{NT(T - 2)(T - N - 2)}{2\gamma(T + 1)^2(T - N - 1)(T - N - 4)}$$

where $T > N + 4$ and

$$k_2 = \left( \frac{T}{T + 1} \right) \left[ 2 - \frac{T(T - 2)(T - N - 2)}{(T + 1)(T - N - 1)(T - N - 4)} \right]$$
The resulting Bayesian risk in this case is

\[
R(\theta, \tilde{\theta}(z)) = F(\theta, \theta) - E_z[F(\theta, \tilde{\theta}(z))]
\]

\[
= (1 - k_2) \frac{S^2}{2\gamma} + \frac{NT(T - 2)(T - N - 2)}{2\gamma(T + 1)^2(T - N - 1)(T - N - 4)} \tag{15}
\]

Now we can compare the risk of the sample mean and covariance \( \hat{\theta} \) with the risk of the Bayes estimator \( \tilde{\theta} \). We have,

\[
R(\theta, \hat{\theta}(z)) - R(\theta, \tilde{\theta}(z)) = (k_2 - k_1) \frac{S^2}{2\gamma} + \frac{NT(T - 2)(2T - N - 1)(N + 3)}{2\gamma(T + 1)^2(T - N - 1)(T - N - 2)(T - N - 4)}
\]

It is easy to see that

\[
k_2 - k_1 = \frac{T^2(T - 2)(2T - N - 1)(N + 3)}{(T + 1)^2(T - N - 1)(T - N - 2)(T - N - 4)} - 2 \frac{(T + 1)(T - N - 1)(T - N - 4)(N + 3)}{(T + 1)^2(T - N - 1)(T - N - 2)(T - N - 4)} > 0
\]

because \( T^2 > 2(T + 1) \) for any \( T > 2 \). Thus the Bayesian portfolio rule always strictly outperforms the sample mean and sample covariance by yielding lower expected loss regardless of the value of true parameters \( \theta \). That means that \( \hat{\theta} \) is inadmissible as defined before.

### 2.3 Kan-Zhou Optimal Two-Fund Rule

(Kan and Zhou 2007) consider the class of Two-funded portfolio rules that have weights

\[
a(\tilde{\theta}) = \frac{c}{\gamma} \tilde{\Sigma}^{-1} \tilde{\mu} \tag{16}
\]

where \( c \) is a constant scalar. Clearly, the previous rules are particular cases of this class. In the first case we have \( c_1 = 1 \) and in the second case \( c_2 = (T - N - 2)/(T + 1) \). This rule can be viewed as a plug-in estimator that estimates \( \Sigma \) by using \( \tilde{\Sigma}_* = \tilde{\Sigma}/c \).

It is easy to compute the expected out of sample performance of this class of portfolio rules. Let \( \hat{\theta}_* = (\hat{\mu}, \hat{\Sigma}_*) \). Using equations (10) and (11), we have

\[
E_z[F(\theta, \hat{\theta}_*(z))] = \frac{cS^2}{\gamma} \left( \frac{T}{T - N - 2} \right) - \frac{c^2}{2\gamma} \left( S^2 + \frac{N}{T} \right) \left[ \frac{T^2(T - 2)}{(T - N - 1)(T - N - 2)(T - N - 4)} \right]
\]

differentiating with respect to \( c \), the optimal \( c \) is

\[
c^* = \left[ \frac{(T - N - 1)(T - N - 4)}{T(T - 2)} \right] \left( \frac{S^2}{S^2 + \frac{N}{T}} \right) \tag{17}
\]
Although $c^*$ is optimal, it is not a feasible strategy since $\theta$ is unknown in practice. Nevertheless, the second factor is close to 1 for high values of $S^2$ and or high values of $T$. It suggests taking the suboptimal rule

$$\hat{a}_* = \frac{c_3}{\gamma} \hat{\Sigma}^{-1} \hat{\mu},$$

where

$$c_3 = \frac{(T - N - 1)(T - N - 4)}{T(T - 2)} \tag{18}$$

This rule is parameter independent and suggests investing $\hat{a}_*$ in the risky assets and $1'_{N\gamma} \hat{a}_*$ in the risk free asset. It is not difficult to show that this strategy dominates over the Bayesian strategy. In fact, since $f(c) = E_{z|\theta}[F(\theta, \hat{\theta}_*(z))]$ is a quadratic function of $c$, the expected out of sample performance is a decreasing function of $c$ for $c \geq c^*$. So we have the result by checking that $c_2 > c_3 > c^*$. To see this, just note that

$$c_2 = \frac{T - N - 2}{T + 1} > \frac{T - N - 4}{T} > \left( \frac{T - N - 4}{T} \right) \left( \frac{T - N - 1}{T - 2} \right) = c_3 > \left[ \frac{(T - N - 1)(T - N - 4)}{T(T - 2)} \right] \left( \frac{S^2}{S^2 + \frac{N}{T}} \right) = c^*$$

2.4 The Three-Fund Rule and Mixed Portfolio Strategies

Kan and Zhou propose a mixed portfolio strategy to reduce estimation risk. The idea is that since if two portfolio have estimation errors that are not perfect correlated, estimation risk can be diversified. As the alternative risk portfolio they propose using the global minimum variance portfolio since it depend only on $\hat{\Sigma}$ but not on $\hat{\mu}$. The idea is to consider the portfolio rule:

$$\hat{a} = \hat{a}(c, d) = \frac{1}{\gamma} (c\hat{\Sigma}^{-1} \hat{\mu} + d\hat{\Sigma}^{-1}1_N) \tag{19}$$

and find $c$ and $d$ that minimizes estimation risk. We show in the appendix that this strategy suggests choosing

$$c^* = c_3 \left( \frac{\psi^2}{\psi^2 + \frac{N}{T}} \right),$$

$$d^* = c_3 \left( \frac{N}{\psi^2 + \frac{N}{T}} \right)$$

where

$$\psi^2 = \mu' \Sigma^{-1} \mu - \left( \frac{\mu' \Sigma^{-1} 1_N}{1' \Sigma^{-1} 1_N} \right)^2 = (\mu - \mu_g 1_N)' \Sigma^{-1} (\mu - \mu_g 1_N)$$
\[ \mu_g = \frac{1}{N} \Sigma^{-1} \mu \]  

is the expected excess return of the ex-ante global minimum-variance portfolio. The problem with this strategy is that the values of \( c \) and \( d \) in this case depend on population parameters that are unknown. So we are going to analyze the a posteriori naive strategy of \( c = d = 1/2 \).

### 3 The Bayesian Framework

#### 3.1 The Bayesian Prior-Posterior Approach to Model Uncertainty

Some important concepts of bayesian statistics are the definitions of the prior, posterior and the predictive distributions. The idea is that the investor may consider a number of possible return distributions. An investor who ignores the uncertainty in the model parameters uses the distribution of future returns conditional on both past data and fixed parameter values \( \theta \), \( f(r_{t+T}|\theta, z) \), where \( z = (z_1, ..., z_t)' \). In contrast, the investor who takes parameter uncertainty into account samples from the predictive distribution, conditional only on past data and not on the parameters, \( f(r_{t+T}|z) \). To be more precise, suppose that the initial prior pdf for a parameter vector \( \theta \) is \( p(\theta) \) and the investor observes a set of data \( z \) with pdf \( p(z|\theta) \). Let \( p(z, \theta) \) the joint probability density function for a random observation vector \( z \) and a parameter vector \( \theta \), also considered random. Since

\[ p(z, \theta) = p(z|\theta)p(\theta) \]  

\[ = p(\theta|z)p(z) \]  

we have that

\[ p(\theta|z) \propto p(\theta)p(z|\theta) \]  

The term \( p(\theta) \) is the prior distribution, \( p(\theta|z) \) is the posterior pdf for the parameter vector \( \theta \), and \( p(z|\theta) \), viewed as a function of \( \theta \), is the likelihood function. As pointed out by (Zellner 1996), the joint posterior p.d.f. \( p(\theta|z) \), has all the prior and sample information incorporated in it.

An important prior to be considered in some cases is the prior of ignorance, also known, as an uninformative prior in Bayesian statistics. Jeffreys’ prescription for representing ignorance about a value of \( \theta \), which can assume values from \(-\infty \) to \(+\infty \), is to take

\[ p(\theta)d\theta \propto d\theta \]  

When considering parameters, like the standard deviation \( \sigma \), which by their nature, can assume values from 0 to \(+\infty \), Jeffreys suggests taking its logarithm
uniform; that is, considering $\theta = \log \sigma$, where $\theta$ takes values from $-\infty$ to $+\infty$. Since $d\theta = d\sigma/\sigma$, we have

$$p(\sigma)d\sigma \propto \frac{d\sigma}{\sigma}, \quad 0 < \sigma < +\infty$$

(24)

A more detail explanation about Jeffreys’rule will be given in the following section. Now, suppose that, given our sample information $z$, we are interest in making inferences about other observations that are still unobserved. Let $\tilde{z}$ represent a vector of yet unobserved observations. We have

$$p(\tilde{z}, \theta | z) = p(\tilde{z} | \theta, z)p(\theta | z)$$

(25)

Note that on the right we have the conditional pdf for $\tilde{z}$, given $\theta$ and $z$, whereas $p(\theta | z)$ is the posterior pdf for $\theta$. To obtain the predictive pdf, $p(\tilde{z} | z)$, we merely integrate the above equation with respect to $\theta$; that is

$$p(\tilde{z} | z) = \int p(\tilde{z}, \theta | z)d\theta$$

(26)

$$= \int p(\tilde{z} | \theta, z)p(\theta | z)d\theta$$

(27)

To contextualize this framework, we analyze here how much a portfolio manager decides to invest in each of the available risky assets. Let $U(R)$ be the utility function, where $R$ is the return distribution from the investment, and $g(R | \theta)$ the conditional density of asset returns given the set of parameters $\theta$. If $\theta$ is known, the conditional expected utility of the investor is

$$E[U(R | \theta)] = \int U(R)g(R | \theta)dR$$

(28)

In practice, however, $\theta$ is unknown and needs to be estimated from data. In the presence of parameter uncertainty, we need to infer the posterior density, $p(\theta | z)$, from the data, where $z = (r_1, ..., r_T)$ is the vector of past returns. The expected utility is then given by

$$E[U(R | z)] = E[E[U(R | \theta) | z]] = \int \int U(R)g(R | \theta)p(\theta | z)d\theta dR$$

(29)

Let $p(\theta)$ be the unconditional prior about the unknown parameter. Then the posterior density given $z$ is

$$p(\theta | z) = \frac{g(z | \theta)p(\theta)}{p(z)} = \frac{g(z | \theta)p(\theta)}{\int g(z | \theta)p(\theta)d\theta}$$

(30)

If the returns are i.i.d., then

$$p(\theta | z) = \frac{\prod_{t=1}^{T} g(r_t | \theta)p(\theta)}{\int g(z | \theta)p(\theta)d\theta}$$

(31)
and the predictive density, given z, is

\[ g(R|z) = \int g(R|\theta)p(\theta|z)d\theta = \int g(R|\theta) \left( \prod_{t=1}^{T} g(r_t|\theta)p(\theta) \right) d\theta \]  \hspace{1cm} (32)

Using the predictive density, the expected utility of the investor is given by

\[ E[U(R|z)] = \int U(R)g(R|z)dR \]  \hspace{1cm} (33)

For instance, if we consider a quadratic utility in the investments returns, that is, if \( U(R) = R - \frac{\gamma}{2} R^2 \), then using the predictive density, the expected utility is

\[ E[U(R|z)] = E[R|z] - \frac{\gamma}{2} E[R^2|z] \]  \hspace{1cm} (34)

That makes clear that in a Bayesian framework with unknown return distribution a straightforward generalization of Markowitz original problem urges us to compute the first and second predictive moments of the assets returns which are, respectively

\[ E[R|z] = \int Rg(R|z)dR, \]
\[ E[R^2|z] = \int R^2g(R|z)dR \]

We follow (Barberis 2000) and analyze 2 distinct portfolio problems: a static buy-and-hold problem and a dynamic problem with optimal rebalancing. We define regular intervals for rebalancing the portfolio. We treat here uncertainty about the parameters, known as estimation risk, using a Bayesian approach. To understand the Bayesian approach it is useful to think that we can solve the portfolio in 2 different ways. The simple way is to construct the distribution of future returns conditional on fixed parameter estimates. The second is to consider the parameters as functions of the Data and integrate the parameters over the posterior distribution. This allows for the construction of the predictive distribution for future returns, conditional only on observed data. Comparing the results obtained from considering the parameters fixed with the one obtained with the predictive distribution is the way considered in the literature to evaluate the effect of parameter uncertainty. Barberis analyzes the separate effects of model uncertainty and predictability by considering uncertainty in an i.i.d. case first. Secondly he analyzes the effect of buy and hold by comparing it with an optimal rebalancing problem. It is important to point out here that our analyses differ from Barberis in the sense that he considers only the problem where the investor must decide how much to invest in the risk free and how much to invest in the risky asset.
3.2 Jeffreys’ Rule for Prior Ignorance

Since the prior distribution of $\theta$, $p(\theta)$ plays an important role in Bayesian models, we are going to give here a brief discussion based on (Box and Tiao 2011).

Let $y = (y_1, ..., y_n)$ be a random sample from a distribution $p(y|\theta)$. With certain regularity conditions, for sufficiently large $n$, the likelihood function of $\theta$ is approximately Normal, and remains Normal under mild one-to-one transformations of $\theta$. In such case, the log-likelihood is quadratic, so that

$$L(\theta|y) = L(\hat{\theta}|y) - \frac{n}{2}(\theta - \hat{\theta})^2 \left( -\frac{1}{n} \frac{\partial^2 L}{\partial \theta^2}(\hat{\theta}) \right)$$

Where $\hat{\theta}$ is the Maximum Likelihood Estimator of $\theta$. In general, the quantity

$$J(\hat{\theta}) = -\frac{1}{n} \frac{\partial^2 L}{\partial \theta^2}(\hat{\theta})$$

is a function of all the data $y$. But it is interesting to note that for a large $n$, $\hat{\theta}$ converges to $\theta_0$ and the average of the above function converges to the true mean of the function

$$E \left[ -\frac{\partial^2 L}{\partial \theta^2}(\theta_0) \right] = -\int \frac{\partial^2 L}{\partial \theta^2}(\theta_0)p(y|\theta_0)dy$$

Then we have $J(\hat{\theta})$ is approximately $\mathcal{F}(\hat{\theta})$, where $\mathcal{F}(\hat{\theta})$ is the function

$$\mathcal{F}(\hat{\theta}) = -E \left[ \frac{\partial^2 L}{\partial \theta^2}(\theta) \right] = E \left[ \frac{\partial L}{\partial \theta}(\theta) \right]^2$$

Consequently, we use $\mathcal{F}(\hat{\theta})$, that depends on $\hat{\theta}$ only, to approximate $J(\hat{\theta})$. Now suppose $\phi(\theta)$ is a one to one transformation. Then,

$$J(\hat{\theta}) = \left( -\frac{1}{n} \frac{\partial^2 L}{\partial \phi^2}(\hat{\phi}) \right) = \left( -\frac{1}{n} \frac{\partial^2 L}{\partial \theta^2}(\hat{\theta}) \right) \left( \frac{d\theta}{d\phi} \right)^2 = J(\hat{\theta}) \left( \frac{d\theta}{d\phi} \right)^2$$

or by using the approximation,

$$\mathcal{F}(\hat{\phi}) = \mathcal{F}(\hat{\theta}) \left( \frac{d\theta}{d\phi} \right)^2$$

It follows that if $\phi(\hat{\theta})$ is chosen such that

$$\left| \frac{d\theta}{d\phi} \right| \propto \mathcal{F}^{-1/2}(\hat{\theta})$$

Then $J(\hat{\theta})$ will be a constant independent of $\hat{\phi}$, and the likelihood will be approximately data translated in terms of $\phi$. That is, we find that the metric $\phi(\theta)$ for which a locally uniform prior is approximately noninformative is such that

$$\frac{d\phi}{d\theta} \propto \mathcal{F}^{1/2}(\theta)$$
Equivalently, the noninformative prior for $\theta$ should be chosen so that, locally,

$$p(\theta) \propto F^{1/2}(\theta)$$

Now, for the multiple parameters case, if the distribution of $y$, depending on $k$ parameters $\theta$, obeys certain regularity conditions, then, for sufficiently large samples, the likelihood function for $\theta$ and for mild transformations of $\theta$ approaches a Multivariate Normal distribution. The log likelihood is thus quadratic,

$$L(\theta|y) = \log l(\theta|y) = L(\hat{\theta}|y) - \frac{n}{2} (\theta - \hat{\theta})' D_\hat{\theta} (\theta - \hat{\theta}),$$

where $\hat{\theta}$ is the vector of Maximum Likelihood Estimates of $\theta$ and $-nD_\theta$ is a $k \times K$ matrix of second derivatives of the parameters evaluated at $\hat{\theta}$,

$$D_\theta = \left\{ -\frac{1}{n} \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} (\hat{\theta}) \right\}, \quad i, j = 1, ..., k$$

which, for a large $n$, can be approximated by $F_n(\hat{\theta})$ which is a function of $\hat{\theta}$ only, where

$$F_n(\theta) = E \left[ \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} (\theta) \right]$$

where the expectation is taken with respect to the data distribution $p(y|\theta)$. That is, $F_n(\theta)$ is the information matrix associated with the sample $y$.

Now, the idea is to find a transformation $\phi$ which ensures that the content of the approximate likelihood region of $\phi$,

$$(\phi - \hat{\phi})' F_n(\hat{\phi})(\phi - \hat{\phi}) < \text{const.}$$

remains constant for different $\hat{\phi}$. This is equivalent to asking for a transformation for which the $|F_n(\phi)|$ is independent of $\phi$. Since

$$F_n(\phi) = A F_n(\theta) A'$$

where $A$ is the $k \times k$ matrix of partial derivatives of $\theta$ with respect to $\phi$. Thus,

$$|F_n(\phi)| = |A|^2 |F_n(\theta)|$$

Then, the above requirement is fulfilled if $\phi$ is such that

$$|A| \propto |F_n(\theta)|^{-1/2}$$

The corresponding noninformative prior in $\theta$ is then

$$p(\theta) = p(\phi)|A|$$

that is,

$$p(\theta) \propto |F_n(\theta)|^{1/2}$$

16
So, we can conclude the Jeffreys’ rule for multiparameter problems is to take the prior distribution for a set of parameters to be proportional to the square root of the determinant of the information matrix.

To illustrate the argument, we are going to show that the determinant of the information matrix for $\Sigma$ and for $\Sigma^{-1}$ in the multivariate regression model are respectively proportional to $|\Sigma|^{-(m+1)}$ and $|\Sigma|^{m+1}$. To see this first we have that the m-dimensional Normal distribution has density

$$p(y|\mu, \Sigma) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp \left[ -\frac{1}{2} tr \Sigma^{-1} (y - \mu)(y - \mu)' \right]$$  \hspace{1cm} (35)$$

and its corresponding log likelihood is given by

$$L(\mu, \Sigma) = -\frac{m}{2} \log 2\pi + \frac{1}{2} \log |\Sigma|^{-1} - \frac{1}{2} tr \Sigma^{-1} (y - \mu)(y - \mu)'$$  \hspace{1cm} (36)$$

differentiating with respect to $\sigma_{ij}$ we get

$$\frac{\partial L}{\partial \sigma_{ij}} = \frac{1}{2} \frac{1}{|\Sigma|^{-1}} \frac{\partial |\Sigma|^{-1}}{\partial \sigma_{ij}} - (y_i - \mu_i)(y_j - \mu_j)$$  \hspace{1cm} (37)$$

Since $\frac{\partial |\Sigma|^{-1}}{\partial \sigma_{ij}}$ is just the cofactor of $\sigma_{ij}$ (see Appendix), the first term on the right hand side is $(1/2)\sigma_{ij}$. Thus, the second derivatives are

$$\frac{\partial^2 L}{\partial \sigma_{ij} \partial \sigma_{kl}} = \frac{1}{2} \frac{\partial \sigma_{ij}}{\partial \sigma_{kl}}$$  \hspace{1cm} (38)$$

so the determinant of the information matrix is proportional to

$$|F(\Sigma^{-1})| = \left| \mathbb{E} \left[ \frac{\partial^2 L}{\partial \sigma_{ij} \partial \sigma_{kl}} \right] \right| \propto \left| \frac{\partial \Sigma}{\partial \Sigma^{-1}} \right|$$  \hspace{1cm} (39)$$

since we know that (see appendix)

$$\left| \frac{\partial \Sigma}{\partial \Sigma^{-1}} \right| = |\Sigma|^{m+1}$$  \hspace{1cm} (40)$$

we have the second result. To get the first, we only need to remind that

$$|F(\Sigma)| = |F(\Sigma^{-1})| \left| \frac{\partial \Sigma}{\partial \Sigma^{-1}} \right|^{-2}$$  \hspace{1cm} (41)$$

We conclude that the uninformative prior for a covariance matrix is

$$p(\Sigma) \propto |F(\Sigma)|^{1/2} \propto |\Sigma|^{-(m+1)/2}$$  \hspace{1cm} (42)$$
3.3 A Simple Bayesian Portfolio Model

In a simple framework, the problem facing a Bayesian investor is to estimate the $N$-dimensional vector of means $\mu$ from i.i.d. population $y_t \sim N(\mu, \Sigma)$, $t = 1, \ldots, T$. The key result in (Jorion 1986) can be summarized as follows. Assume the following 3 conditions: (i) Investors have an informative prior on $\mu$ of the form

$$p(\mu|\hat{\mu}, \nu_\mu) \propto \exp\left[ -\frac{1}{2}(\mu - \bar{\mu}1_N)^T(\nu_\mu\Sigma^{-1})(\mu - \bar{\mu}1_N) \right]$$

with $\bar{\mu}$ being the grand mean and $\nu_\mu$ giving an indication of prior precision; (ii) Investors have diffuse prior on the grand mean $\bar{\mu}$; (iii) The density $p(\nu_\mu|\mu, \bar{\mu}, \Sigma)$ is a Gamma function with mean at $(N + 2)/d$ where $d$ is defined as

$$d = (\mu - \bar{\mu}1_N)^T\Sigma^{-1}(\mu - \bar{\mu}1_N)$$

and is replaced by its sample estimate

$$(\hat{\mu} - \mu_{MIN}1_N)^T\Sigma^{-1}(\hat{\mu} - \mu_{MIN}1_N)$$

where

$$\mu_{MIN} = \frac{\bar{\mu}^T\Sigma^{-1}1_N}{1_N\Sigma^{-1}1_N}$$

Then, the predictive density for the return distribution $g(r|y, \Sigma, \nu_\mu)$, conditional on $\Sigma$ and the precision $\nu_\mu$ is a multivariate normal with predictive Bayes-Stein mean, $\mu_{BS}$, equal to

$$\mu_{BS} = (1 - \phi_{BS})\hat{\mu} + \phi_{BS}\mu_{MIN}1_N$$

where $\hat{\mu}$ is the sample mean, $\mu_{MIN}$ is the minimum variance portfolio,

$$\phi_{BS} = \frac{\nu_\mu}{T + \nu_\mu} = \frac{N + 2}{(N + 2) + T(\hat{\mu} - \mu_{MIN}1_N)^T\Sigma^{-1}(\hat{\mu} - \mu_{MIN}1_N)}$$

and covariance matrix

$$Var(r) = \Sigma \left( 1 + \frac{1}{T + \nu_\mu} \right) + \frac{\nu_\mu}{T(T + 1 + \nu_\mu)} 1_N1_N^T$$

The term $\phi_{BS}$ is known as the shrinkage coefficient, since it shrink the sample mean towards the mean of the minimum variance portfolio. It is easy to see that the case of zero precision, $\nu_\mu = 0$, corresponds to the Bayes diffuse prior case considered in Bawa, Brown, and Klein (1979) and Zellner and Chetty in which the sample mean is the predictive mean but the covariance matrix is inflated by the factor $(1 + 1/T)$. Finally, for $\nu_\mu \rightarrow \infty$ the predictive mean is the mean of the minimum variance portfolio and the covariance matrix is given by $\Sigma + (1/T)1_N1_N^T(1/1_N\Sigma^{-1}1_N)$.
Substituting the Bayes Stein estimator $\mu_{BS}$, it is easy to see that the optimal portfolio weights can be written as

$$\omega_{BS}(\nu_\mu) = (1 - \phi_{BS}(\nu_\mu))\omega_{MV} + \phi_{BS}(\nu_\mu)\omega_{MIN}$$  \hspace{1cm} (45)$$

where the global minimum variance portfolio is given by

$$\omega_{MIN} = \frac{1}{1_N} \Sigma^{-1} 1_N$$  \hspace{1cm} (46)$$

and the mean variance portfolio is given by

$$\omega_{MV} = \frac{1}{\gamma} \Sigma^{-1} (\hat{\mu} - \hat{\mu}_{0} 1_N)$$

In practice, $\Sigma$ is unknown, and is usually replaced by

$$\hat{\Sigma}_{BS} = \frac{T - 1}{T - N - 2} \hat{\Sigma}$$  \hspace{1cm} (47)$$

where $\hat{\Sigma}$ is the usual unbiased sample covariance matrix.

### 3.4 Incorporating Pricing Model Restrictions

(Pastor and Stambaugh 2000) consider a framework where individuals can invest in assets that represent benchmark positions and non-benchmark assets. It suggests partition the return distribution $r_t = (r_{1,t}, r_{2,t})$ where $r_{2,t}$ contains the payoffs on $k$ benchmark positions and $r_{1,t}$ the payoff of the remaining assets.

They consider then a multivariate regression,

$$r_{1,t} = \alpha + Br_{2,t} + u_t$$  \hspace{1cm} (48)$$

where $B$ is a $(n - 3) \times 3$ matrix where each of its rows is a vector of betas of the first $(n - 3)$ non-benchmark positions, $(\beta_{1,i}, \beta_{2,i}, \beta_{3,i})$, and

$$u_t \sim N(0, \Sigma)$$  \hspace{1cm} (49)$$

This regression equation implies moments restrictions that the parameters must obey. In particular,

$$\alpha = E[r_{1,t}] - BE[r_{2,t}]$$  \hspace{1cm} (50)$$

and

$$\Sigma = V_{11} - BV_{22}B'$$  \hspace{1cm} (51)$$

The approach they used impose restrictions on $\alpha$ but no restrictions on $B, \Sigma, E_2$ and $V_{22}$. In Bayesian statistics that means to use uninformative prior distributions to this set of parameters. The prior distribution for $\Sigma$ is specified as inverted Wishart,

$$\Sigma^{-1} \sim W(H^{-1}, v)$$  \hspace{1cm} (52)$$
with degrees of freedom \( v = 15 \). From the properties of the inverted Whishart distribution, the prior expectation of \( \Sigma \) equals \( H/(v - m - 1) I_m \), so that \( E[\Sigma] = s^2 I_m \). The value of \( s^2 \) is set equal to the average of the diagonal elements of the sample estimate of \( \Sigma \). The joint prior distribution for the remaining parameters \((B, E_2, V_{22})\) is assumed to be diffuse and independent of \( \alpha \) and \( \Sigma \). The density function for these parameters are

\[
p(\Sigma) \propto |\Sigma|^{-(v+m+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} H \Sigma^{-1} \right\}
\]

\[
p(B) \propto 1
\]

\[
p(E_2) \propto 1
\]

\[
p(V_{22}) \propto |V_{22}|^{-k+1}
\]

Following their approach, we consider as factor based the CAPM, in which \( k = 1 \), and the 3 factor FF model, in which \( k = 3 \). The pricing model impose \( \alpha = 0 \). To allow for mispricing uncertainty, they consider a prior distribution for \( \alpha \) specified as a normal distribution,

\[
\alpha|\Sigma \sim N(0,\sigma^2_\alpha(\frac{1}{s^2} \Sigma))
\] (53)

so, its density function is represented by

\[
p(\alpha|\Sigma) \propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} \alpha' \left( \frac{\sigma^2_\alpha}{s^2} \Sigma \right)^{-1} \alpha \right\}
\]

Where \( \sigma^2_\alpha \) reflects the investor’s prior degree of mispricing uncertainty. When \( \sigma_\alpha = 0 \), the investor believes dogmatically in the model. When \( \sigma_\alpha = \infty \), the investor regards the model as useless.

Letting \( \theta = (\alpha, B, \Sigma, E_2, V_{22}) \), the joint prior distribution of the parameters can be easily obtained by the factor

\[
p(\theta) = p(\alpha|\Sigma)p(\Sigma)p(B)p(E_2)p(V_{22})
\]

To understand how these benchmark models alter our original mean variance problem, it is useful to consider the moments partitioned as follows

\[
E = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}, \quad V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}
\] (54)

The vector of betas for each of the non-benchmark positions is a row of the matrix

\[
B = V_{12}V_{22}^{-1}
\]

Also, the pricing restriction implies that \( \alpha = 0 \) or equivalently

\[
E_1 = BE_2
\] (55)
Substituting in the portfolio optimization formula, we have

\[
w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \frac{1}{A} \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}^{-1} \begin{bmatrix} V_{12}V_{22}^{-1}E_2 \\ E_2 \end{bmatrix} = \frac{1}{A} \begin{bmatrix} 0 \\ V_{22}^{-1}E_2 \end{bmatrix}
\]

(56)

Since from a partitioned inverse result due to Duncan (1944) we have

\[
\begin{bmatrix} V_{11} \\ V_{21} \\ V_{22} \end{bmatrix}^{-1} = \begin{bmatrix} (V_{11} - V_{12}V_{22}^{-1}V_{21})^{-1} - (V_{11} - V_{12}V_{22}^{-1}V_{21})^{-1}V_{12}V_{22}^{-1} \\ -(V_{22} - V_{21}V_{11}^{-1}V_{12})^{-1}V_{21}V_{11}^{-1} \\ (V_{22} - V_{21}V_{11}^{-1}V_{12})^{-1} \end{bmatrix}
\]

That is, if the investor fully believes in the pricing model, \( \sigma_\alpha = 0 \), the optimal portfolio involves only the benchmark positions. Then, these would be the exact moments he should estimate from the data. Writing in terms of the predictive distribution, the optimal portfolio in this case is

\[
w^* = (1/A)V_{22}^{-1}E_2^*
\]

On the other hand, if the investor consider the pricing model useless, \( \sigma_\alpha = \infty \). It would be sound to consider the case in between. In the next subsection we are going to show how to replace \( E \) and \( V \) by moments of the Bayesian predictive distributions corresponding to varying degrees of prior confidence in the pricing model.

### 3.4.1 Finding the Predictive Moments of the model

To find the predictive moments, we follow Pastor and Stambaugh and define \( Y = (r_{1,1}, ..., r_{1,T})', \ X = (r_{2,1}, ..., r_{2,T})', \) and \( Z = (1_T, X) \), where \( 1_T \) denotes a \( T \) vector of ones. Also define the \((k + 1) \times m \) matrix \( A = (\alpha, B)' \), and let \( a = vec(A) \). For the \( T \) observations \( t = 1, ..., T \), the regression model can be written as

\[
Y = ZA + U, \quad vec(U) \sim N(0, \Sigma \otimes I_T),
\]

(57)

Where \( U = (u_1, ..., u_T)' \). The matrix \( R = (Y, X) \) contains the entire sample. Define the statistics \( \hat{A} = (Z'Z)^{-1}Z'Y, \hat{a} = vec(A), \hat{\Sigma} = (Y - Z\hat{A})(Y - Z\hat{A})/T, \ E_2 = X'1_T/T, \) and \( \hat{V}_{22} = (X - 1_T\hat{E}_2)(X - 1_T\hat{E}_2)/T. \)

We factor the likelihood function as

\[
p(R|\theta) = p(Y, X|\theta) = p(Y|\theta, X)p(X|\theta),
\]

(58)

where

\[
p(Y|\theta, X) \propto |\Sigma|^{-T/2} \exp \left\{ -\frac{1}{2} tr(Y - ZA)'(Y - ZA)\Sigma^{-1} \right\}
\]

\[
\propto |\Sigma|^{-T/2} \exp \left\{ -\frac{T}{2} tr\hat{\Sigma}\Sigma^{-1} - \frac{1}{2} tr(A - \hat{A})'Z'Z(A - \hat{A})\Sigma^{-1} \right\}
\]

\[
\propto |\Sigma|^{-T/2} \exp \left\{ -\frac{1}{2} trT\hat{\Sigma}\Sigma^{-1} - \frac{1}{2} (a - \hat{a})'(\Sigma^{-1} \otimes Z'Z)(a - \hat{a}) \right\}
\]

21
where the second line follows from the fact that \((Y - XB)'(Y - XB) = (Y - X\hat{B})'(Y - X\hat{B}) + (B - X\hat{B})'X'X(B - X\hat{B})\), where \(\hat{B} = (X'X)^{-1}X'Y\) and the third line follows from the facts \(trA'B = vec(A)'vec(B)\) and \(vec(ABC) = (C' \otimes A)vec(B)\). And

\[
p(X|\theta) \propto |V_{22}|^{-T/2} \exp \left\{ -\frac{1}{2} tr(X - 1_T E_2')'(X - 1_T E_2')V_{22}^{-1} \right\}
\]
\[
\propto |V_{22}|^{-T/2} \exp \left\{ -\frac{T}{2} tr\hat{V}_{22}V_{22}^{-1} - \frac{T}{2} tr(E_2 - \hat{E}_2)(E_2 - \hat{E}_2)'V_{22}^{-1} \right\}
\]

As we have seen, the joint prior distribution of all parameters is

\[
p(\theta) = p(\alpha|\Sigma)p(\Sigma)p(B)p(E_2)p(V_{22})
\]

where

\[
p(\alpha|\Sigma) \propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} \alpha' \left( \frac{\sigma^2}{s^2} \Sigma \right)^{-1} \alpha \right\}
\]
\[
p(\Sigma) \propto |\Sigma|^{-s+\frac{k+1}{2}} \exp \left\{ -\frac{1}{2} trH\Sigma^{-1} \right\}
\]
\[
p(B) \propto 1
\]
\[
p(E_2) \propto 1
\]
\[
p(V_{22}) \propto |V_{22}|^{-\frac{k+1}{2}}
\]

The priors of \(B, E_2,\) and \(V_{22}\) are diffuse. The prior of \(\Sigma\) is inverted Wishart with a small number of degrees of freedom which is uninformative. The prior on \(\alpha\) given \(\Sigma\) is normal and centered at the pricing restriction.

Note that

\[
\alpha'(\frac{\sigma^2}{s^2} \Sigma)^{-1} \alpha = \alpha' (\Sigma^{-1} \otimes D) \alpha
\]

where \(D\) is a \((k + 1) \times (k + 1)\) matrix whose \((1, 1)\) element is \(\frac{\sigma^2}{s^2}\) and all other elements are zero. So, we can rewrite the density of \(\alpha|\Sigma\) as

\[
p(\alpha|\Sigma) \propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} \alpha' (\Sigma^{-1} \otimes D) \alpha \right\}
\]

Combining the prior with the likelihood function we get the posterior distribution of \(\theta\):

\[
p(\theta|R) \propto p(R|\theta)p(\theta)
\]

An interesting fact is that the posterior distribution can be factored into two parts, one that involves the regression parameters \((a, \Sigma)\) and another that involves the benchmark moments \((E_2, V_{22})\). To see this, note that

\[
p(\theta|R) \propto p(R|\theta)p(\theta) \propto p(\theta)p(Y|\theta,X)p(X|\theta)p(\alpha|\Sigma)p(\Sigma)p(B)p(E_2)p(V_{22})
\]
\[
\propto p(Y|\theta,X)p(\alpha|\Sigma)p(\Sigma)p(B)p(E_2)p(V_{22})p(X|E_2,V_{22})\]
\[
\propto p(a,\Sigma|R)p(E_2, V_{22}|X)
\]
The joint posterior of the regression parameters is just proportional to the product \( p(Y|\theta, X)p(\alpha|\Sigma)p(\Sigma)p(B) \), that is
\[
p(a, \Sigma|R) \propto \left| \Sigma \right|^{-\frac{k+1}{2}} \exp \left\{ -\frac{1}{2} \left[ a'(\Sigma^{-1} \otimes D)a + (a - \hat{a})'(\Sigma^{-1} \otimes Z'Z)(a - \hat{a}) \right] \right\} \times \left| \Sigma \right|^{-\frac{2T+k+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(H + T\hat{\Sigma})\Sigma^{-1} \right\}.
\]

Letting \( \tilde{a} = (I_m \otimes F^{-1}Z'Z)\hat{a}, F = D + Z'Z, Q = Z'(I_T - ZF^{-1}Z')Z \) we have
\[
p(a, \Sigma|R) \propto \left| \Sigma \right|^{-\frac{k+1}{2}} \exp \left\{ -\frac{1}{2} [(a - \tilde{a})'(\Sigma^{-1} \otimes F)(a - \tilde{a})] \right\} \times \left| \Sigma \right|^{-\frac{2T+k+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(H + T\hat{\Sigma} + \hat{A}'Q\hat{A})\Sigma^{-1} \right\},
\]

Now, from \( p(a, \Sigma|R) = p(a|\Sigma, R)p(\Sigma|R) \), it is clear that \( \Sigma|R \) have a IW\((H + T\hat{\Sigma} + \hat{A}'Q\hat{A}), T + v - k, m)\). It follows that
\[
\begin{align*}
a|\Sigma, R &\sim N(\tilde{a}, \Sigma \otimes F^{-1}) \\
\Sigma^{-1}|R &\sim W((H + T\tilde{\Sigma} + \hat{A}'Q\hat{A})^{-1}, T + v - k, m)
\end{align*}
\]

Therefore,
\[
\begin{align*}
\hat{a} &= E(a|R) = (I_m \otimes F^{-1}Z'Z)\hat{a} \\
\hat{\Sigma} &= E(\Sigma|R) = \frac{1}{T + v - k - m - 1}(H + T\hat{\Sigma} + \hat{A}'Q\hat{A}) \\
\text{Var}(a|R) &= \hat{\Sigma} \otimes F^{-1}.
\end{align*}
\]

Where the third equality follows from the Law of Total Conditional Variance which states that \( \text{Var}[a|R] = E[\text{Var}(a|\Sigma, R)|R] + \text{Var}[E(a|\Sigma, R)|R] \).

Now, the joint posterior of the benchmark moments is just proportional to
\[
p(E_2, V_{22})p(X|E_2, V_{22})
\]

, that is
\[
p(E_2, V_{22}|X) \propto V_{22}^{-\frac{T+k+1}{2}} \exp \left\{ -\frac{T}{2} \text{tr}(\hat{V}_{22}V_{22}^{-1}) - \frac{T}{2} \text{tr}(E_2 - \hat{E}_2)(E_2 - \hat{E}_2)'V_{22}^{-1} \right\}
\]

It follows that
\[
\begin{align*}
E_2|V_{22}, R &\sim N(\hat{E}_2, \frac{1}{T}V_{22}) \\
V_{22}^{-1}|R &\sim W((T\hat{V}_{22})^{-1}, T - 1, k).
\end{align*}
\]
Therefore,
\[
\tilde{E}_2 = E(E_2|R) = \bar{E}_2
\]
\[
\tilde{V}_{22} = E(V_{22}|R) = \frac{T}{T-k-2} \tilde{V}_{22}
\]
\[
Var(E_2|R) = E[Var(E_2|V_{22}, R)|R] = \frac{1}{T} E[V_{22}|R]
\]
\[
= \frac{1}{T-k-2} \tilde{V}_{22}.
\]

Where the third equality, again, follows the Law of Total Conditional Variance.

Now we only need to note that the predictive mean obeys the relation,
\[
E^* = \tilde{E} = E(r_{T+1}|R) = E(E_2|\theta, R) = \bar{E}.
\]

Since \(B\) and \(E_2\) are independent in the posterior, the mean of the predictive distribution is
\[
E^* = \tilde{E} = E\left(\frac{\alpha + BE_2}{E_2}\right) = \left(\frac{\tilde{\alpha} + \tilde{B}\tilde{E}_2}{\tilde{E}_2}\right)
\]
(64)
where \(\tilde{\alpha}\) and \(\tilde{B}\) are obtained from
\[
\tilde{a} = (I_m \otimes F^{-1}Z'Z)\hat{a}
\]
where \(\tilde{a} = vec((\tilde{\alpha} \quad \tilde{B})').\) To compute \(V_{22}^*\) we only need to note that
\[
V^* = Var(r_{T+1}|R) = E(V|R) + Var(E|R) = \tilde{V} + Var(E|R),
\]
We can partition the predictive covariance matrix as
\[
V^* = \begin{bmatrix}
V_{11}^* & V_{12}^*
\end{bmatrix}
\]
Applying this rule to the lower right submatrix gives
\[
V_{22}^* = \tilde{V}_{22} + Var(E_2|R),
\]
And applying it to the off diagonal submatrices gives
\[
V_{12}^* = V_{21}^* = E(BV_{22}|R) + Cov(\alpha + BE_2, E_2'|R)
\]
\[
= \tilde{B}\tilde{V}_{22} + \tilde{B}Var(E_2|R)
\]
A little bit of extra work is needed for the \(V_{11}^*\) elements. We use the decomposition
\[
Cov(y_i, T+1, y_j, T+1|R) = E(Cov(y_i, T+1, y_j, T+1|a, R)|R)
\]
\[
+ Cov(E(y_i, T+1|a, R), E(y_j, T+1|a, R)|R)
\]
24
where
\[ E(Cov(y_{i,T+1}, y_{j,T+1}|a, R)|R) = \tilde{b}_i^* V_{22}^* \tilde{b}_j^* + Tr[V_{22}^* Cov(\tilde{b}_i, \tilde{b}_j^*)] + \tilde{\sigma}_{ij} \]

and
\[ Cov(E(y_{i,T+1}|a, R), E(y_{j,T+1}|a, R)|R) = [1 \ E_2^i] Cov(a_i, a_j'|R) [1 \ E_2^j]' \]

Here, \( Cov(b_i, b_j'|R) \) and \( Cov(a_i, a_j'|R) \) are submatrices of \( Var(a|R) \).

It is interesting to note that if the investor believes dogmatically in the model \( \sigma_\alpha = 0 \), then the predictive mean is the sample mean and the covariance is inflated to
\[ V_{22}^* = \left( \frac{T+1}{T-k-2} \right) \tilde{V}_{22} \approx \left( 1 + \frac{1}{T} \right) \tilde{V}_{22} \]

which is the Bayes Stein estimator with zero precision. The prescription to find the others submatrix from the predictive second moment is given in the appendix.

4 Brief Description of Asset Allocation Strategies Considered

In this section we summarize the asset allocation models considered. The main difference between this models is how to estimate the unknown population parameters \( \mu \) and \( \Sigma \) from the data.

4.1 Naive Risk Portfolio

The naive risk portfolio is defined as the portfolio that invest equal weights, \( 1/N/N \), in the risk assets. This strategy is considered naive since it disregard the data and any sort of optimization problem. Note that in this case no amount is invested in the risk free asset. We later consider a naive strategy that takes the risk free asset under consideration.

4.2 Sample-based mean-variance portfolio

The sample mean-variance portfolio follows the classic ”plug-in” approach, that is, it substitutes the population parameters \( \mu \) and \( \Sigma \) by its sample counterparts in the optimization rule. This strategy ignores the possibility of estimation errors. In this case we have \( a^{mv} = (1/\gamma) \tilde{\Sigma}^{-1} \tilde{\mu} \).

4.3 Bayesian diffuse-prior portfolio

Barry (1974), Klein and Bawa (1976), and Brown (1979) show that if the prior for \( \Theta = (\mu, \Sigma) \) is diffuse, and the conditional likelihood is normal, then the predictive distribution is a student-t with mean \( \tilde{\mu} \) and variance \( \tilde{\Sigma}(1 + 1/M) \).
where $M$ is the estimation window. Since in our study $M = 120$ months, the difference from the sample mean-variance portfolio is negligible and the results of this strategy are not reported.

4.4 Jorion-Bayes-Stein portfolio

As described before, this strategy consider correcting estimation error by setting

\[
\hat{\mu}^{JBS} = (1 - \hat{\phi})\hat{\mu} + \hat{\phi}\hat{\mu}_{min}
\]

\[
\hat{\phi} = \frac{N + 2}{(N + 2) + T(\hat{\mu} - \hat{\mu}_{min})\hat{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_{min})}
\]

where

\[
\hat{\Sigma}^{JBS} = \frac{T - 1}{T - N - 2} \hat{\Sigma}
\]

and $\mu_{MIN} = a'_{MIN}\hat{\mu}$ is the average excess return on the sample global minimum variance portfolio. This sample estimator is called a shrink estimator because it shrink the sample mean towards the mean of the minimum variance portfolio.

4.5 Pastor-Stambaugh Bayesian portfolio

As we described here, Pastor Stambaugh adress the arbitrariness of the choice of a shrinkage target $\hat{\mu}$, and of the shrinkage factor, $\phi$, by using the investor’s belief about the validity of an asset pricing model. As for the asset pricing model we consider the Capital Asset Pricing Model (CAPM) and the (Fama and French 1993) three-factor model.

4.6 Global Minimum-Variance portfolio

As its name suggests, the minimum variance portfolio strategy is the one that minimizes the variance of returns. To implement this strategy we only need to estimate the covariance matrix of asset returns. This strategy suggests $a_{MIN} = \Sigma^{-1}_N 1_N\Sigma^{-1}_N$. Note that this strategy does not take the risk free asset under consideration.

4.7 Kan and Zhou two-fund portfolio

The Kan and Zhou optimal two fund rule suggest choosing

\[
a(\hat{\theta}) = \frac{c}{\gamma} \hat{\Sigma}^{-1} \hat{\mu}
\]

where the value of $c$ is the one that minimizes expected loss. Since the exact value of $c$ is not feasible, we set $c$ to the approximate value that does not depend of the unknown population values. That is, we set

\[
c = \frac{(T - N - 1)(T - N - 4)}{T(T - 2)}
\]
We favor here the Kan-Zhou two portfolio rule with respect to the three fund rule because of its simplicity and for the fact that the three fund rule is not feasible, so the sample analog add estimation error that turns its optimality questionable.

4.8 Mixture of equally weighted and minimum-variance portfolios

Garllapi Uppal consider the case where investors combine the equally weighted with the minimum-variance portfolio. Their motivation is that because expected returns are more difficult to estimate than covariances, one way to ignore the estimates of mean returns but not the estimates of covariances is to consider this strategy. This portfolio strategy is

\[ \hat{a}^{cw-min} = c \frac{1}{N} 1_N + d \Sigma^{-1} 1_N \]  

(69)

where \( c \) and \( d \) are chosen to maximize expected out of sample performance.

4.9 Naive Mixture of Portfolios

We consider a new portfolio strategy that has not been studied in the existing literature. This is strategy is a naive combination of two portfolio rules. Our motivation is that although mean and variance are difficult to estimate, both may be useful to reduce estimation uncertainty, that is, different believes may result in different estimation risks that maybe diversified. Since we do not want to favor any of the chosen models, and since usually, the optimal weights that maximize out of sample performance are not feasible, we set equal weights to both portfolio rules.

The natural candidates are combinations of the sample based optimal mean variance portfolio with the global minimum variance portfolio, that is

\[ \hat{a}^{nmv-min} = \frac{1}{2} \left( \frac{\hat{\Sigma}^{-1} \hat{\mu}}{\gamma} + \frac{\hat{\Sigma}^{-1} 1_N}{1_N \Sigma^{-1} 1_N} \right) \]

and the naive risk portfolio with the global minimum variance portfolio,

\[ \hat{a}^{new-min} = \frac{1}{2} \left( \frac{1}{N} 1_N + \frac{\hat{\Sigma}^{-1} 1_N}{1_N \Sigma^{-1} 1_N} \right) \]

We also consider the Jorion-Bayes-Stein optimal portfolio with the global minimum variance portfolio,

\[ \hat{a}^{njbs-min} = \frac{1}{2} \left( \frac{\hat{\Sigma}^{-1} \mu^{JBS}}{\gamma} + \frac{\hat{\Sigma}^{-1} 1_N}{1_N \Sigma^{-1} 1_N} \right) \]
5 Data Description and Methodology for Evaluating Performance

We consider a data set of 12 industry portfolios as the risk assets and 1 risk-free asset that here is the 1 month T-bill. We use 120 months of data retrieved from french.com for the industry portfolios and the T-bill retrieved from fed.com.

We follow here the approach used in the literature that relies on a "rolling sample". This approach considers an estimation window of size M. In each month t, starting from t=M+1, we use the data in the previous M months to estimate the parameters needed to implement a particular strategy. The outcome of this rolling-window approach is a series of T-M monthly out of sample returns generated by each of the portfolio strategies listed.

Given the time series of monthly out of sample returns generated by each strategy and in each dataset, we compute 3 quantities. The out of sample Sharpe ratio of strategy k:

\[ \hat{SR}_k = \frac{\hat{\mu}_k}{\hat{\sigma}_k} \]  

(70)

The certainty equivalent return, defined as the risk free rate that an investor is willing to accept rather than adopting a particular risky portfolio strategy:

\[ \hat{CEQ}_k = \hat{\mu}_k - \gamma \frac{1}{2} \hat{\sigma}^2_k \]  

(71)

In which \( \hat{\mu}_k \) and \( \hat{\sigma}^2_k \) are the mean and variance of out of sample portfolio returns for strategy k, and \( \gamma \) is the risk aversion. For the third evaluation method we compute the return-loss with respect to the 1/N strategy. The return-loss is defined as the additional return needed for strategy k to perform as well as the 1/N strategy in terms of the Sharpe ratio. Let \( \mu_{ew} \) and \( \sigma_{ew} \) be the monthly out of sample mean and volatility of the returns from 1/N strategy. The return-loss from strategy k is

\[ return - loss_k = \frac{\mu_{ew}}{\sigma_{ew}} \sigma_k - \mu_k \]  

(72)

It is important to point out that to compute the welfare loss of uncertainty, we must compare what should be the individual utility when the investor is certain about the pricing model. That is, if \( \sigma_\alpha = 0 \) in the Bayesian framework. We can interpret the results in the following manner. How much money an investor would be willing to give up in order to be sure about the true pricing model.

6 Results

6.1 Results from Empirical Datasets

The first strategy considered is the one the individual invest in the naive risk portfolio. In this case the individual disregard the risk free asset and invest all in the risk assets with equal amounts in each asset.
The second strategy considered is the one the individual invests in the global minimum variance portfolio. In this case the investor disregard the risk free asset and invest all the amount in the portfolio that minimizes the variance of the risk portfolio return. Performance, read, utility, is computed for \( \gamma = 10 \).

The third strategy we consider is the mean variance portfolio with the sample mean and sample variance as estimates for the populational moments. The table below reports the sharpe ratio in sample and out of sample of the resulting portfolio for each value of \( t \) ranging from 61 to 119 for the selected value of \( \gamma = 10 \).

The fourth strategy we consider is the Jorion-Bayes Stein estimate for the mean. The idea is to shrink the mean toward the grand mean that here is the minimum variance mean. With this strategy, a smaller estimation risk is achieved.

The fifth strategy is Kan-Zhou two fund portfolio. This strategy attains an even lower estimation risk than Jorion suggestion.

The next strategy we are now going to consider is Pastor and Stambaugh framework that consider individual beliefs in the asset pricing model. The beliefs can be summarized in the value of the parameter \( \sigma_\alpha \). If \( \sigma_\alpha = 0 \) we say that the individual truly believes in the asset pricing model, if \( \sigma_\alpha = \infty \) we say that the individual consider the asset pricing model as useless. We follow Garlappi-Uppal and consider \( \sigma_\alpha = 1% \). The benchmark asset considered here is based in the CAPM model where we consider the SP500 as the benchmark.

As a matter of comparison, we are going to seek combinations of each individual strategy with the minimum variance portfolio. Since the minimum variance portfolio only depends on sample estimation of the covariance matrix, but not on the individual assets mean, it is a good portfolio to be considered for combined strategies.

### 6.2 Table of results

**In Sample Sharpe ratios for empirical data**

<table>
<thead>
<tr>
<th>Strategy/t</th>
<th>61</th>
<th>71</th>
<th>81</th>
<th>91</th>
<th>101</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/N</td>
<td>0.0681</td>
<td>0.0881</td>
<td>0.1358</td>
<td>0.1560</td>
<td>0.1659</td>
<td>0.1485</td>
</tr>
<tr>
<td>min</td>
<td>0.3058</td>
<td>0.3899</td>
<td>0.3445</td>
<td>0.3214</td>
<td>0.3500</td>
<td>0.3334</td>
</tr>
<tr>
<td>mv</td>
<td>0.5939</td>
<td>0.5404</td>
<td>0.4963</td>
<td>0.4423</td>
<td>0.4648</td>
<td>0.4494</td>
</tr>
<tr>
<td>JBS</td>
<td>0.5599</td>
<td>0.5075</td>
<td>0.4584</td>
<td>0.4007</td>
<td>0.4333</td>
<td>0.4187</td>
</tr>
<tr>
<td>KZ</td>
<td>0.5939</td>
<td>0.5404</td>
<td>0.4963</td>
<td>0.4423</td>
<td>0.4648</td>
<td>0.4494</td>
</tr>
<tr>
<td>PS</td>
<td>0.6129</td>
<td>0.5715</td>
<td>0.5417</td>
<td>0.5064</td>
<td>0.5164</td>
<td>0.4904</td>
</tr>
<tr>
<td>mv − min</td>
<td>0.5633</td>
<td>0.5212</td>
<td>0.4724</td>
<td>0.4196</td>
<td>0.4452</td>
<td>0.4288</td>
</tr>
<tr>
<td>ew − min</td>
<td>0.1643</td>
<td>0.2117</td>
<td>0.2356</td>
<td>0.2416</td>
<td>0.2604</td>
<td>0.2417</td>
</tr>
<tr>
<td>JBS − min</td>
<td>0.4710</td>
<td>0.4559</td>
<td>0.4055</td>
<td>0.3585</td>
<td>0.3930</td>
<td>0.3782</td>
</tr>
</tbody>
</table>
### Out of Sample Sharpe ratios for empirical data

<table>
<thead>
<tr>
<th>Strategy/t</th>
<th>61</th>
<th>71</th>
<th>81</th>
<th>91</th>
<th>101</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/N</td>
<td>0.3720</td>
<td>0.3724</td>
<td>0.2746</td>
<td>0.2123</td>
<td>0.1580</td>
<td>0.6741</td>
</tr>
<tr>
<td>min</td>
<td>0.2887</td>
<td>0.2114</td>
<td>0.2304</td>
<td>0.2355</td>
<td>0.0469</td>
<td>−0.0814</td>
</tr>
<tr>
<td>mv</td>
<td>0.0727</td>
<td>0.1135</td>
<td>0.0570</td>
<td>0.0708</td>
<td>−0.1017</td>
<td>−0.2672</td>
</tr>
<tr>
<td>JBS</td>
<td>0.1558</td>
<td>0.1586</td>
<td>0.1362</td>
<td>0.1575</td>
<td>−0.0409</td>
<td>−0.2072</td>
</tr>
<tr>
<td>KZ</td>
<td>0.0727</td>
<td>0.1135</td>
<td>0.0570</td>
<td>0.0708</td>
<td>−0.1017</td>
<td>−0.2672</td>
</tr>
<tr>
<td>PS</td>
<td>0.1434</td>
<td>0.1675</td>
<td>0.1138</td>
<td>0.1048</td>
<td>−0.0371</td>
<td>−0.1843</td>
</tr>
<tr>
<td>mv − min</td>
<td>0.1519</td>
<td>0.1490</td>
<td>0.1203</td>
<td>0.1349</td>
<td>−0.0548</td>
<td>−0.2200</td>
</tr>
<tr>
<td>ew − min</td>
<td>0.3832</td>
<td>0.3333</td>
<td>0.2881</td>
<td>0.2593</td>
<td>0.1203</td>
<td>0.3184</td>
</tr>
<tr>
<td>JBS − min</td>
<td>0.2181</td>
<td>0.1812</td>
<td>0.1788</td>
<td>0.1942</td>
<td>−0.0041</td>
<td>−0.1624</td>
</tr>
</tbody>
</table>

### In Sample Certainty Equivalent for empirical data

<table>
<thead>
<tr>
<th>Strategy/t</th>
<th>61</th>
<th>71</th>
<th>81</th>
<th>91</th>
<th>101</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/N</td>
<td>−0.0123</td>
<td>−0.0097</td>
<td>−0.0062</td>
<td>−0.0044</td>
<td>−0.0033</td>
<td>−0.0041</td>
</tr>
<tr>
<td>min</td>
<td>0.0046</td>
<td>0.0067</td>
<td>0.0057</td>
<td>0.0051</td>
<td>0.0059</td>
<td>0.0054</td>
</tr>
<tr>
<td>mv</td>
<td>0.0176</td>
<td>0.0146</td>
<td>0.0123</td>
<td>0.0098</td>
<td>0.0108</td>
<td>0.0101</td>
</tr>
<tr>
<td>JBS</td>
<td>0.0139</td>
<td>0.0115</td>
<td>0.0098</td>
<td>0.0077</td>
<td>0.0088</td>
<td>0.0083</td>
</tr>
<tr>
<td>KZ</td>
<td>0.0147</td>
<td>0.0128</td>
<td>0.0111</td>
<td>0.0090</td>
<td>0.0101</td>
<td>0.0096</td>
</tr>
<tr>
<td>PS</td>
<td>0.0184</td>
<td>0.0156</td>
<td>0.0145</td>
<td>0.0128</td>
<td>0.0132</td>
<td>0.0120</td>
</tr>
<tr>
<td>mv − min</td>
<td>0.0141</td>
<td>0.0124</td>
<td>0.0105</td>
<td>0.0084</td>
<td>0.0094</td>
<td>0.0088</td>
</tr>
<tr>
<td>ew − min</td>
<td>−0.0007</td>
<td>0.0013</td>
<td>0.0021</td>
<td>0.0024</td>
<td>0.0031</td>
<td>0.0024</td>
</tr>
<tr>
<td>JBS − min</td>
<td>0.0096</td>
<td>0.0090</td>
<td>0.0077</td>
<td>0.0062</td>
<td>0.0073</td>
<td>0.0068</td>
</tr>
</tbody>
</table>

### Out of Sample Certainty Equivalent for empirical data

<table>
<thead>
<tr>
<th>Strategy/t</th>
<th>61</th>
<th>71</th>
<th>81</th>
<th>91</th>
<th>101</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/N</td>
<td>0.0066</td>
<td>0.0067</td>
<td>0.0037</td>
<td>0.0016</td>
<td>−0.0006</td>
<td>0.0109</td>
</tr>
<tr>
<td>min</td>
<td>0.0041</td>
<td>0.0014</td>
<td>0.0024</td>
<td>0.0026</td>
<td>−0.0034</td>
<td>−0.0063</td>
</tr>
<tr>
<td>mv</td>
<td>−0.0187</td>
<td>−0.0149</td>
<td>−0.0139</td>
<td>−0.0107</td>
<td>−0.0251</td>
<td>−0.0416</td>
</tr>
<tr>
<td>JBS</td>
<td>−0.0032</td>
<td>−0.0033</td>
<td>−0.0024</td>
<td>−0.0009</td>
<td>−0.0102</td>
<td>−0.0187</td>
</tr>
<tr>
<td>KZ</td>
<td>−0.0054</td>
<td>−0.0045</td>
<td>−0.0059</td>
<td>−0.0048</td>
<td>−0.0151</td>
<td>−0.0277</td>
</tr>
<tr>
<td>PS</td>
<td>−0.0056</td>
<td>−0.0036</td>
<td>−0.0076</td>
<td>−0.0100</td>
<td>−0.0210</td>
<td>−0.0373</td>
</tr>
<tr>
<td>mv − min</td>
<td>−0.0036</td>
<td>−0.0048</td>
<td>−0.0038</td>
<td>−0.0025</td>
<td>−0.0122</td>
<td>−0.0215</td>
</tr>
<tr>
<td>ew − min</td>
<td>0.0066</td>
<td>0.0054</td>
<td>0.0041</td>
<td>0.0034</td>
<td>−0.0007</td>
<td>0.0039</td>
</tr>
<tr>
<td>JBS − min</td>
<td>0.0012</td>
<td>−0.0007</td>
<td>0.0003</td>
<td>0.0010</td>
<td>−0.0065</td>
<td>−0.0121</td>
</tr>
</tbody>
</table>

6.3 Results from Simulated Data

TO BE INCLUDED
7 Conclusion

By analyzing the monthly returns from the 12 industry portfolio for the 10 years period from January 2016 to January 2017 we were able to form portfolio strategies based on expected loss minimization and predictive moments from Bayesian beliefs. When comparing this models with the naive portfolio, that invests equal amount in each of the risk assets, none were able to outperform it. This confirms Garlappi-Uppal findings that we are still miles away when looking at only single portfolio strategies. Besides this fact, we were able to form a simple mixture of the naive portfolio with the minimum variance portfolio that improves in sample performance whether maintaining similar out of sample performance. This suggests an investigation of possible optimal strategies that combine the risk free asset with this mixture of naive risk asset and minimum variance portfolio.
A Appendix

A.1 The derivative of the determinant as a cofactor

Let $B = b_{ij}$ be a $p \times p$ matrix. Then

$$\frac{\partial |B|}{\partial b_{ij}} = B_{ij}$$

To see this we only need to expand $|B|$ by the elements of the $i$th row

$$|B| = \sum_{h=1}^{p} b_{ih} B_{ih}$$

since $B_{ih}$ does not contain $b_{ij}$, the result follows.

A.2 The Jacobian of the Inverse

Let $A = G^{-1}$ be a PDS matrix. The Jacobian of the transformation can be shown to be $|A|^{-(n+1)}$. To see this, write $AG = I$. Then

$$\frac{\partial A}{\partial \theta} G + A \frac{\partial G}{\partial \theta} = 0$$

or

$$\frac{\partial G}{\partial \theta} = -G \frac{\partial A}{\partial \theta} G$$

If $\theta = a_{ij}$, we have $\partial g_{\alpha \beta}/\partial a_{ij} = -g_{\alpha \beta} g_{\beta j}$, for $\beta \leq \alpha$ and $j \leq i$, since $A$ and $G$ are symmetric and the transformation from elements of $G$ to those of $A$ involves just $m(m+1)/2$ distinct elements of $G$. On forming the Jacobian matrix and taking its determinant, we have $|G|^{m+1} = |A|^{-(n+1)}$.

A.3 Proof of the First Moment of the Inverted Wishart

We show here that if $A$ has a $W(\Sigma, n, p)$ distribution, then

$$E[A^{-1}] = (1/(n-p-1))\Sigma^{-1}$$

To see this, let $C$ be such that $\Sigma = CC'$. Then we can write

$$A = CBC'$$

where $B$ has a $W(I, n)$ distribution and

$$E[A^{-1}] = C'^{-1} E[B^{-1}] C^{-1}$$

By symmetry we can write

$$E[B^{-1}] = k_1 I + k_2 ee'$$
Now note that for every orthogonal matrix $Q$, $QBQ'$ has a $W(I, n)$ distribution. This is true, since if $B = ZZ'$ where $Z$ is $N(0, I)$, then $QBQ' = QZZ'Q'$ where $QZ$ is also $N(0, I)$. This imply that $E[(QBQ')^{-1}] = QE[B^{-1}]Q' = E[B^{-1}]$ for every orthogonal matrix $Q$. For that, we must have $k_2 = 0$ and

$$E[B^{-1}] = k_1 I$$

Now we just need to note that the diagonal terms of $B^{-1}$ are $IA^2(n - p + 1)$. Since $E[IA^2(n - p + 1)] = (n - p - 1)^{-1}$, we have

$$E[B^{-1}] = (n - p - 1)^{-1} I$$

We conclude that

$$E[A^{-1}] = C'^{-1}E[B^{-1}]C^{-1} = (n - p - 1)^{-1}(CC')^{-1} = (n - p - 1)^{-1} \Sigma^{-1}$$
References


