Mean Variance Optimization and Beyond: Improve Optimal Portfolio Construction with Bayesian Regularization

Abstract

Mean variance optimization algorithm seeks to form portfolios with the maximum trade off between expected return and risk. Investors, however, do not know the true value of expected return and risk of the investable universe, but use the historical sample to estimate the expected moments. In the presence of estimation errors, the algorithm has the fundamental limitation of maximizing the estimation error by overweighing assets with higher returns and lower risks and underweighing assets with lower returns and higher risks. To illustrate the effect of estimation error on portfolio weights, we compare two portfolio decision rules based on: maximum likelihood estimator without accounting for estimation error, and Bayesian MAP estimator with estimation error modeled by Bayesian priors. It is shown that portfolio decision rule with Bayesian regularization reduces the loss of portfolio value from that of the true optimal portfolio based on true values of expected return and risk. The reduction, however, comes with a price of efficiency. In the Bayesian approach, Sharpe ratio of the tangency portfolio shrinks by a factor of $\frac{T-N-2}{T+1}$ as compared to that yielded by the maximum likelihood approach.

1 Introduction

Mean variance optimization (MVO), along with Capital Asset Pricing Model (CAPM) and Black Sholes formula for option pricing, are considered to be the three major workhorses of modern finance theory. MVO was first proposed by Markowitz (1955), who applied statistical decision theory to guide portfolio construction and subsequently won Nobel Prize in economics in 1990. The objective of MVO is to form a diversified portfolio of assets so as to maximize expected returns for any risk level or minimize risk for a given level of return.

While theoretical model focus on ex ante values, practical implementation of MVO calls for ex post results. In the MVO setting, an investor is given a return history of N assets and faced with the problem of choosing assets among the investable universe to achieve the maximum expected return and risk tradeoff. Using the historical data, the investor estimates the expected return and risk, and plugs the sample estimates into the MVO optimizer and solves for the weights, i.e. proportions of investable wealth for each asset. The resulting set of portfolio weights describe optimal solutions, and the set of optimal portfolios for all possible levels of risk or return forms the mean variance frontier.

The optimization output, i.e. portfolio weights can be viewed as “plug-in” estimator of the true weights since the inputs are *computed* using historic data but *justified* on the basis of predicted relationships. Since investors do not know the true values of the expected moments, and their sample estimates are inevitably subject to estimation error. In the presence of estimation error, MVO is often criticized as “estimation-error maximizer”. To understand the criticism, a simple thought experiment shall suffice. MVO significantly overweighs those assets that have large estimated returns, small variances, negative correlations with other assets, and underweighs those with small estimated returns, large variances and positive correlations. These assets are, of course, most likely to suffer from large estimation errors. A practical and inevitable consequence of the error maximization procedure is that, any estimates of the statistical properties of the optimized portfolio may be significantly biased if the statistical properties are the object of optimization. If the optimal portfolio is solved by minimizing the portfolio risk, the variance of the output portfolio may be a significant underestimate of the true level of risk.

Estimation error significantly undermines robustness of the optimization output from MVO procedure. It is necessary, therefore, to look for more robust estimators for the expected return and risk. This is a well-posed regularization problem in machine learning. A natural candidate is the Bayesian MAP (maximum a posterior) estimator wherein Bayesian priors are introduced to model the uncertainty in parameters. We show that under the distributional assumption of multivariate normality, the MAP estimator for the expected return and variance can be solved from the Euler equations in close form. When there is no diffusion or uncertainty in the prior, MAP estimator is reduced to maximum likelihood estimator. Maximum likelihood estimator is derived under the distributional assumption of multivariate normality and it is informationally efficient absent of estimation errors. Bayesian MAP estimator is motivated to account for estimation error in the unknown parameters. To demonstrate the difference between the competing estimators and the resulting portfolio choice rules, we calculate the optimal portfolios using monthly data on 4 indices. It is found that portfolio rule based on ML estimation is more optimal as measured by the maximum Sharpe ratio of the efficiency frontier (Sharpe ratio of the tangency portfolio), but suffers bigger loss in portfolio value from that of the true optimal portfolio.

Our paper contributes to the literature by using Bayesian regularization to address one of the fundamental limitations of the MVO approach for portfolio construction: error maximization. We derive close-form solutions for the conditional moments by maximizing the posterior and solving the Euler equations. Portfolio rules based on different estimation methods are compared with a risk function coherent with the investment objective.

The remainder of the paper is organized as follows. Section 2 sets up the mean variance optimization problem to solve for optimal portfolio weights. Section 3 proposes two methods to estimate the inputs: ML estimator and Bayesian MAP estimator. Section 4 describes the data and presents the empirical results. Section 5 concludes.

2 Mean Variance Optimization

Let $\Phi_T = \{R_{i,t}\}_{i=1,...,N,t=1,...T}$ denote the sample of return history on N risky assets. Consider an investor with a one-period investment horizon who, after observing this sample, must make an investment decision at the end of period T. It is assumed that the investor finds the historical evidence useful and assesses the characteristic of potential investment rule based on conditional distribution $P(R_{T+1} \mid \Phi_T)$. Suppose the investor's objective is to form a portfolio from one risk free asset and N risky asset to maximize the risk and return trade off, based on Markowitz's modern portfolio choice theory.

Let W denote the N-vector of weighs or percentage of investable wealth in risky assets, then the return on the investor's overall portfolio is given by

$$R_{p,T+1} = R_f + W'R_{T+1} \tag{1}$$

We can write the mean variance optimization problem as follows

$$\underset{W}{Max} W'E - \frac{\lambda}{2} W'VW \tag{2}$$

where $E = E\{R_{p,T+1} \mid \Phi_T\}$ is the next-period return conditioning on history and $V = Cov\{R_{p,T+1}, R_{p,T+1}' \mid \Phi_T\}$ is the next-period conditional covariance. λ is a scalar parameter measuring the level of risk aversion of the investor.

Since the objective function is quadratic, the necessary and sufficient condition for optimality is to solve the optimal weights from the first order conditions. The solution to (2) can be easily verified to be

$$W^* = \frac{1}{\lambda} AB$$

$$where \quad A = \frac{V^{-1}E}{1'V^{-1}E} \tag{3}$$

$$and \quad B = \frac{A'E}{A'VA}$$

where A gives the weights on N risky assets to form the portfolio with maximum Sharpe ratio, i.e. the tangency portfolio, and B gives the Sharpe ratio of the tangency portfolio, which measures how much return is required by the investor for per unit of risk.

3 Estimating Conditional Mean and Covariance and Plug-in Estimator of Portfolio Choice

3. 1 ML Estimator and Bayesian MAP Estimator

The objective of mean-variance optimization can be justified with the assumption that the returns follow a multivariate normal distribution. Intuitively, the first and second moments are sufficient statistics to summarize the information of normal distribution. In the framework of multivariate normal distributions, we can estimate the value of E and V consistently with ML. This is equivalent to having an infinitely long history of returns.

For simplicity, we assume that the T observations of $R_{i,t}$ are independent realizations from a multivariate normal distribution with unconditional mean E_0 and unconditional covariance V_0. Then independence implies that $E = E_0$, and $V = V_0$.

Given the sample $\Phi_T = \{R_{i,t}\}_{i=1,...,N,t=1,...T}$, the likelihood function can be written as

$$p(R_{11},...,R_{1,T},...R_{N,1},...,R_{N,T}) = \prod_{t=1}^{T} (\frac{1}{(2\pi)^N} |V|^{-.5} \exp\{-\frac{1}{2}(R_{.,t} - E)'V^{-1}(R_{.,t} - E)\}) \quad (4)$$

and the ML estimator of E and V are given by

$$\hat{E}_{ML} = \frac{1}{T} R_{,t} '1 \quad (5)$$

and $\hat{V}_{ML} = \frac{1}{T}(R_{,t} - 1\hat{E}_{ML} ')'(R_{,t} - 1\hat{E}_{ML} ')$ (6)

Plugging in (5) and (6) into (3), we can solve for the optimal weights for the tangency portfolio.

However the investor does not know the true values of E and V , the maximum likelihood estimates based on finite sample deviate from the true values, giving rise to estimation error. Consequently, the MVO procedure in fact maximizes the sample estimates present with estimation errors instead of the true moments in the population. Moreover, in the presence of estimation error, the conditional distribution $p(R_{t+1} | \Phi_t)$ is generally not multivariate normal.

To mitigate the effect of estimation error on portfolio choice, we use Bayesian estimator wherein the uncertainty in the parameters E and V is modeled by the prior density $P(E,V)$. We use the following standard prior to represent investors' belief about the parameters of multivariate normal distribution.

$$P(E,V) \propto |V|^{-\frac{N+1}{2}} \quad (7)$$

The posterior density is given by

$$P(E,V|\Phi_T) \propto P(E,V)P(\Phi_T|E,V) \quad (8)$$

where $P(\Phi_T|E,V)$ is the likelihood function given by (4).

Maximizing the posterior, we can derive the Bayesian estimator from solving the Euler equation

$$\hat{E}_{Bay} = \hat{E}_{ML} = \frac{1}{T} R_{,t} '1 \quad (9)$$

and $\hat{V}_{Bay} = \frac{T+1}{T-N-2}\hat{V}_{ML}$ (10)

Since $\frac{T+1}{T-N-2} \geq 1$, the Bayesian estimate of the covariance matrix dominates the ML estimate in the positive semi-definite sense. Hence we obtain the shrinkage estimate on e optimal weights

$$W^*_{Bay} = \frac{T-N-2}{T+1} W^*_{ML} \quad (11)$$

3.2 Certainty Equivalent Loss as a Measure for Generalization Error

To gauge the out-of-sample performance of the portfolios constructed with different decision rules, we need to construct a risk function for different portfolio rules. We propose “certainty equivalent loss” as a form of generalization error, which measures the loss in portfolio value due to estimation error.

Let θ denote the parameter space $\{E,V\}$, W be a portfolio decision rule calculated from (3) based on true moments, and $W_p(\Phi)$ be the portfolio decision rule p associated with a given sample Φ. The certainty equivalent loss is given by $L(\theta,W,W_p(\Phi)) = C - C_p$

where $C_p = E[R_{p,T+1} \mid \Phi_T] - \frac{\lambda}{2} VAR[R_{p,T+1} \mid \Phi_T]$ (12)

C_p is called certainty equivalent is because portfolio p achieves the same value for the objective function as does a portfolio providing a riskless return of C_p. Similarly C is the value of the objective function associated with true value of the moments. By viewing each method’s portfolio weights as a function of the sample, we define the portfolio rule’s risk function $r(\theta) = E\{L(\theta,W,W_q)\}$ with expectation taken over the distribution of Φ for a given set of θ.

$L(\theta,W,W_p(\Phi)) = C - C_p$ can be interpreted as a specific form of generalization error since it captures the deviation of portfolio rule from the true optimal rule. The risk function associated with different portfolio choice rules can be computed as follows: (1) generate sets of random samples which follow multivariate normal distribution with known values of the mean and covariance; (2) for each sample, calculate the certainty equivalent loss $C - C_q$; (3) take the average of the certainty equivalent loss across all the samples.

4. Empirical Results

In this section, we derive the portfolio decision rule using a universe of five index assets: risk free asset as given by the 3-month Treasury bill rate, and four risky assets consisting of Russell 3000 (RAY), MSCI EAFE (MXEA), Goldman Sachs’ Commodity Index (SPGSCI), and SP500 (SPX) with their Bloomberg ticker given in the bracket. The data contains monthly returns on the indices from Jan 1980 to Oct 2009, and they are obtained from Yahoo finance and Bloomberg. The empirical results are summarized in Table 1 and Graph 1. Graph 1 illustrates how we can locate the tangency portfolio, which is given by the line passing the riskfree asset and tangent to the efficiency frontier. The tangency portfolio gives the allocation between the risk free asset and the frontier portfolio on the mean variance frontier and has the maximum Sharpe ratio within the budge constraint. To computer the risk function, 5000 random samples of the same length are generated based on standard multivariate normal distribution. We repeat the calculation of the portfolio weights based on the ML and Bayesian MAP for each sample, and average the certainty equivalent loss over 5000 random samples. The certainty equivalent loss is presented in basis points as shown in Table 1. Bayesian estimator shrinks the portfolio weights and Sharpe ratio by the factor of $\frac{T-N-2}{T+1}$ at the price of reducing estimation error, evidenced in the reduced certainty equivalent loss. Hence MVO algorithm regularized with Bayesian prior makes the portfolio weights more robust to estimation error, but reduces the efficiency as shown in the drop in Sharpe ratio of the tangency portfolio.

Table 1: Summary Statistics, Parameter Estimates and Performance Measures

Bayesian Estimator

	Mean	Std	Skew	Kurtosis	Tangent Portfolio	Sharpe Ratio	Certainty Equivalent Loss
RAY Index	0.076	1.901	-1.094	4.033	13.9%	0.603	24.36
MXEA Index	0.079	2.127	-0.615	1.432	48.6%		
SPGSCI Index	0.031	2.315	-0.420	3.782	9.8%		
SPX Index	0.065	1.752	-0.680	2.574	26.3%		

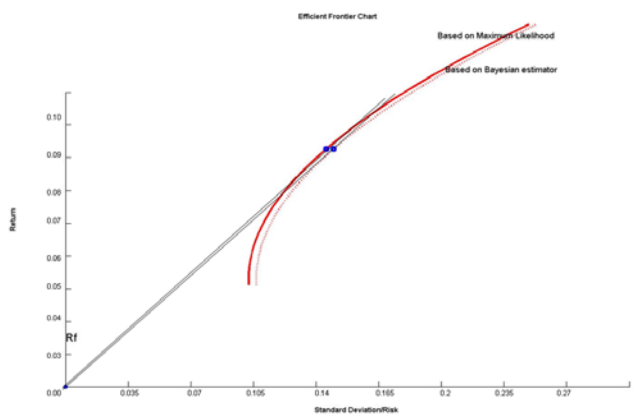
**Data Source: Bloomberg, Yahoo Finance*

Maximum Likelihood

	Mean	Std	Skew	Kurtosis	Tangent Portfolio	Sharpe Ratio	Certainty Equivalent Loss
RAY Index	0.076	1.901	-1.094	4.033	14.1%	0.612	42.15
MXEA Index	0.079	2.127	-0.615	1.432	49.3%		
SPGSCI Index	0.031	2.315	-0.420	3.782	9.9%		
SPX Index	0.065	1.752	-0.680	2.574	26.7%		

**Data Source: Bloomberg, Yahoo Finance*

Graph 1: Mean Variance Frontier and Tangency Portfolio



5. Conclusion

Mean variance optimization algorithm seeks to form portfolios with the maximum trade off between expected return and risk. As the expected returns and risk are estimated, the algorithm has the fundamental limitation of maximizing the estimation error by overweighing assets with higher returns and lower risks and underweighing assets with lower returns and higher risks. To illustrate the effect of estimation error on portfolio choice, we use two estimation methods: maximum likelihood and Bayesian MAP. It is shown that portfolio decision rule with Bayesian regularization reduces the deviation of portfolio value from that based on true moments. The reduction comes with a price of efficiency. Sharpe ratio of the tangency portfolio based on Bayesian regularization shrinks by a factor of $\frac{T-N-2}{T+1}$.