Bayesian Optimization

MFM Practitioner Module: Risk & Asset Allocation

John Dodson

February 18, 2015
Outline

Motivation

Estimation Risk

Bias & Efficiency

Von Neumann-Morgenstern

Robust Optimization

Michaud Re-sampling

Black-Litterman
No introduction to portfolio optimization would be complete without acknowledging the significant contribution of the Markowitz mean-variance efficient frontier concept which lies at the heart of the two-step procedure.

Nonetheless, the mean-variance analysis has two major shortcomings which subsequent researchers and practitioners have tried to address:

- It is not robust, and
- there is no clear accommodation for expert views.

N.B.: These were not issues for Markowitz’ original application, because he argued that all investors were attempting to solve the same problem with all available information, and therefore a broad capitalization-weighted index should serve as a proxy for $\alpha_{SR}$ for all\(^1\).

\(^1\)Meucci argues that in fact $\alpha_{MV}$ is the implicit benchmark.
Refinements

Meucci discusses four approaches to these shortcomings.

Coherent Allocations
It may be valuable to re-cast the optimization problem in an explicitly Bayesian setting.

- Von Neumann-Morgenstern utility maximization
- Robust optimization

In the former case we use subjective probabilities; in the latter case we introduce a risk functional.

Heuristic Techniques
These are becoming more common in practice.

- Michaud re-sampling
- Black-Litterman subjective characterization

These address each of the mean-variance shortcomings in turn and can be applied separately or together within the two-step procedure.
Recall that we denote by \( \theta \) the parameters that we have chosen to characterize the market vector random variable \( M \sim \cdot(\theta) \). Since the investor satisfaction \( S_\theta(\alpha) \) and the feasible set \( C_\theta \) associated with an portfolio \( \alpha \in C_\theta \) are estimable, the optimal portfolio

\[
\alpha^* = \arg \max_{\alpha \in C_\theta} S_\theta(\alpha)
\]

also depends in an estimable (non-random) way on \( \theta \).

**Estimation Risk**

- But we can only estimate \( \theta \); and
- we may have doubts about the characterization of \( M \).

These are two varieties of estimation risk. How can we express our “risk aversion” to estimation risk in our allocation decision?
Estimation Risk

This is not a theoretical or academic issue. The procedures we have outlined so far effectively seek out and try to exploit fine-grained details in the markets in their attempt to optimize satisfaction.

Robust Decisions
We would be ill-advised to incur transaction costs and undue risks if these details are artificial. We need our portfolio allocation procedure to be robust. In analogy to the concept of estimator loss that was introduced in chapter 4, Meucci introduces here the concept of opportunity cost,  

$$OC_\theta(\alpha) = S_\theta(\alpha^*) - S_\theta(\alpha) \geq 0$$

where $\alpha$ plays the role of the estimator and $\theta$ plays the role of the random sample.
Bias & Efficiency

In contrast to the situation in classical estimation theory with quadratic loss, we do not have a direct decomposition of expected loss into contributions from bias and inefficiency.

Decision Strategies
But Meucci shows us two examples of extreme allocation decision strategies that attempt to illustrate this trade-off.

- prior allocation strategy
  - high bias
  - low inefficiency

- sample-based allocation strategy
  - low bias
  - high inefficiency
We have already discussed the limitations of expected utility as an index of satisfaction; in particular objective support and treatment of diversification. One advantage that utility has is it readily accommodates estimation risk. As before, denote the unknown parameters of the market vector by \( \theta \). As long as the constraints do not depend upon on these parameters, we can write

\[
\arg\max_{\alpha \in C} CE_\theta(\alpha) = \arg\max_{\alpha \in C} E \left( u \left( \Psi_\alpha \right) \right) = \arg\max_{\alpha \in C} \iint u(\alpha' m) f_{M|\theta}(m) f_{\theta|Y^{(N)}}(\theta) \ d\theta \ dm
\]

where \( f_{\theta|Y^{(N)}}(\cdot) \) is posterior parameter density, based on the prior density \( f_\theta(\cdot) \) and the data \( y^{(N)} = (x_1, \ldots, x_N)' \).
Utility Maximization

Two-Step Procedure
As long as the utility $u(\cdot)$ is increasing, this index is consistent with weak stochastic dominance, so the integral needs to be evaluated only along the efficient frontier based upon the predictive characterization with

$$ E M = \int \int m f_{M|\theta}(m) f_{\theta|Y(N)}(\theta) \, d\theta \, dm $$

$$ \text{cov} M = \int \int mm' f_{M|\theta}(m) f_{\theta|Y(N)}(\theta) \, d\theta \, dm - E M E M' $$

N.B.: Often the two-step procedure can be complicated by constraints that depend upon the parameters of the market vector. In this case, the approach will probably fail because we may not be able to pull the maximization out of the outer integral.
Robust Optimization

Robust optimization is based on the concept of opportunity cost.

\[ OC_\theta(\alpha) = -S_\theta(\alpha) + \max_{\alpha^* \in C_\theta} S_\theta(\alpha^*) \geq 0 \]

In classical optimization, \( \theta \) is known with certainty and we can choose the allocation with zero opportunity cost. When we only have an estimate for \( \theta \), we can obtain a more robust result using a minimax criterion.

Minimax Criterion
Say we can be confident that the unknown true value of the market vector parameters lies in some range \( \theta \in \Theta \). If the allocation constraints depend upon \( \theta \), we can define the smallest possible feasible set for this range, \( C_\Theta \), and choose the optimal allocation as

\[ \alpha_\Theta = \arg \min_{\alpha \in C_\Theta} \max_{\theta \in \Theta} OC_\theta(\alpha) \]
Robust Optimization

There is a trade-off in selecting the range for the market vector parameters, $\theta \in \Theta$.

- large enough so that you can be confident that it includes the true value
- small enough so that the actual opportunity cost incurred from sub-optimality is not too large

The author proposes the minimum-measure ellipsoid at some level of confidence based on the posterior characterization of the market vector parameters.

Aversion to Market & Estimation Risk

The problem for the mean-variance efficient frontier can be reduced to second-order cone programming.

$$\arg \max_{\alpha \in C} \alpha' \hat{\mu} - \gamma_m \sqrt{\alpha' \hat{\Sigma} \alpha} - \gamma_e \sqrt{\alpha' T \alpha}$$

for $M|\mu \sim \mathcal{N}(\mu, \hat{\Sigma})$ and $\mu \sim \mathcal{N}(\hat{\mu}, T)$. 
Michaud Re-sampling

Michaud re-sampling\(^2\) is neither motivated nor justified by probability or decision theory (nor is it re-sampling in the statistical sense). It is a heuristic intended to directly address the instability of optimal portfolios.

Simulations

Once the parameters for the market vector have been estimated from \(y^{(N)}\), instead of using \(\hat{\theta}\) to maximize \(S_\theta(\alpha)\),

1. use them to generate a large number of random samples of the invariants each of length \(N\), \(\{y^{(N,1)}, \ldots, y^{(N,Q)}\}\)

2. use the new samples to re-estimate the parameters, \(\{\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(Q)}\}\)

3. use the re-estimated parameters to re-solve the optimization problem, \(\{\alpha^{(1)}, \ldots, \alpha^{(Q)}\}\)

report back the mean allocation, \(\alpha_{RS} = \frac{1}{Q} \sum_{q=1}^{Q} \alpha^{(q)}\)

\(^2\)The text acknowledges relevant intellectual property rights.
Michaud Re-sampling

Presumably an argument based upon the Law of Large Numbers can be made to show that this procedure converges. But the author points out several considerations:

▶ If the constraints are not linear in the allocations, there is nothing to prevent the procedure from converging to an infeasible portfolio.
▶ We know nothing about the relationship between the resampled allocation and the true optimal allocation.

This technique is comparable to other heuristics for stabilizing the optimal portfolio such as the inclusion of transactions penalties or exogenous diversification rules.

Pseudo-Data Interpretation

My own speculation is that one could develop a more coherent version of this technique with the pseudo-data interpretation of the conjugate prior, where the pseudo-data vector of length $NQ$ is drawn from a prior based on $\hat{\theta}$. 
The Black-Litterman procedure can be interpreted as a technique for specifying a prior on the market vector parameters based on manager advantage. It can be generalized, but was originally presented and is most easily understood in terms of linear views on a normal model for the market invariants.

**Objective Market Model**

We assume that the market vector is distributed according to $M \sim \mathcal{N}(\mu, \Sigma)$ where $\mu$ and $\Sigma$ are given (perhaps by $\Sigma = \bar{\Sigma}$ and $\mu = \frac{1}{\zeta_0} \bar{\Sigma} \alpha_0$ in equilibrium).

**Manager Expertise**

The manager’s expertise may be limited to a subset of the market, defined by a **picks matrix** $P$ that defines the linear combinations of factors for which the manager is prepared to provide a view.
Manager View
The manager expresses his view, $v$, on the picks and a scalar confidence, $0 \leq c \leq 1$

$$V \mid PM \sim \mathcal{N} \left( v, \left( \frac{1}{c} - 1 \right) P \Sigma P' \right)$$

Black-Litterman Subjective Measure
Conditioning the objective characterization with the view leads to the subjective characterization

$$M \mid V \sim \mathcal{N} \left( \mu_{BL}, \Sigma_{BL} \right) \quad \text{with}$$

$$\mu_{BL} = \mu + c \Sigma P' \left( P \Sigma P' \right)^{-1} (v - P \mu)$$

$$\Sigma_{BL} = \Sigma - c \Sigma P' \left( P \Sigma P' \right)^{-1} P \Sigma$$
Black-Litterman

We can then proceed with classical equivalent optimization using the subjective characterization of the market vector.

Consistency of The Subjective Measure

It may be difficult to quantify the manager’s confidence, $c$.

- To that end, it is useful to measure how consistent the manager’s subjective characterization is with the objective characterization.

Focusing on the subjective location, the author notes that Mahalanobis distance (squared) for a normal is chi-squared. He defines the consistency between $\mu_{BL}$ and $\mu$ as

$$C = 1 - F_{\chi^2_N} ((\mu_{BL} - \mu)' \Sigma^{-1} (\mu_{BL} - \mu))$$

where $\chi^2_N \sim G \left( \frac{N}{2}, \frac{1}{2} \right)$.

A low level of consistency reflects bold views.
The Black-Litterman approach is reminiscent of updating a prior to obtain a posterior, but it is not strictly Bayesian.

**Bayesian Interpretation**
For one, we take the market vector to be normal with mean and covariance parameters, but we never specify a prior characterization for those parameters; rather, we condition the parameters directly on the view. This is an application of Bayes’ Theorem, but not an application of Bayesian estimation theory.

**Opinion Pooling**
In subsequent publications\(^3\), Meucci introduces a more general version of this model, which has the immediate practical appeal of allowing one to extend the concept of a view into non-normal characterizations of the market.

\(^3\)E.g. RISK Magazine, 2/06.