Incorporating Black-Litterman Views in Portfolio Construction when Stock Returns are a Mixture of Normals

Burak Kocuk and Gérard Cornuéjols

July 18, 2017

Abstract

In this paper, we consider the basic problem of portfolio construction in financial engineering, and analyze how market-based and analytical approaches can be combined to obtain efficient portfolios. As a first step in our analysis, we model the asset returns as a random variable distributed according to a mixture of normal random variables. We then discuss how to construct portfolios that minimize the Conditional Value-at-Risk (CVaR) under this probabilistic model. Since the CVaR measure does not have a closed form expression in this case, we propose an algorithm which can numerically compute this measure and solve the resulting convex program. We also construct a second-order cone representable approximation of the CVaR under the mixture model. Furthermore, we incorporate the market equilibrium information into this procedure through the well-known Black-Litterman approach via an inverse optimization framework by utilizing the proposed approximation. Our computational experiments on a real dataset show that this approach with an emphasis on the market equilibrium typically yields less risky portfolios than a purely market-based portfolio while producing similar returns on average.

1 Introduction

Portfolio construction is one of the most fundamental problems in financial engineering: Given n risky assets, historical information about their returns and market capitalization of these assets, construct a portfolio that will produce the maximum expected return at a minimum risk. Maximizing expected return and minimizing risk are often conflicting objectives, and hence a compromise should be made by investors based on their risk aversion.

We consider two paradigms for solving this problem: an "analytical" approach and one that is "market-based". In the analytical approach, key parameters such as the vector of mean asset returns, denoted by μ, and covariance between the asset returns, denoted by Σ, are estimated from historical data. Then, a combination of the expected portfolio return and a risk measure is optimized by solving a problem of the form:

$$\max_{x \in \mathcal{X}} \ \mu^T x - \delta \mathrm{R}(\mathrm{x}). \tag{1}$$

Here, $R(x)$ is the risk of portfolio x under a chosen risk measure, δ is a positive, predetermined risk aversion factor and $\mathcal{X}$ is the set of feasible portfolios. The earliest example of this approach is Markowitz's seminal work [25] in 1952, in which a mean-variance (MV) optimization problem is proposed with $n = 3$ assets, $R(x) = x^T \Sigma x$ and $\mathcal{X} = \Delta_n := \{x \in \mathbb{R}^n : \ \sum_{j=1}^n x_j = 1, \ x \geq 0\}$.

In the market-based approach, one merely invests in a "market portfolio" proportional to the current market capitalization of the assets. The logic behind this market-based approach is the Efficient-Market Hypothesis [17], which loosely states that the price of an asset captures all the information about that asset.

There are advantages and disadvantages to each approach. The major advantage of the analytical approach is that if the parameter predictions are accurate, then it can yield provably optimal portfolios. Unfortunately, this almost never happens in practice. In particular, the estimation of the mean return vector μ is error-prone and even a small perturbation in the parameter estimation can yield completely different portfolios due to what Michaud calls the "error-maximization property" [27]. Robust optimization techniques are proposed to circumvent the difficulties caused by parameter estimation [18, 16, 38, 9, 13, 41]. Another issue with the analytical approach is the determination of the risk aversion parameter δ. Choosing smaller values of δ puts more emphasis on the expected return and since the estimates of μ are generally inaccurate, it may lead to poor portfolios in practice. One may alleviate this issue by simply choosing δ infinitely large, thus reducing the generic portfolio optimization (1) to a "risk minimizing" portfolio problem (see, for instance [13]).

Another issue regarding the analytical approach is to specify an adequate risk measure. Although variance (or standard deviation) is a typical risk measure, others are also used, for instance, Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR), which can better quantify the downside risk. One difficulty with these measures is that a distributional assumption should be made, which requires modeling stock returns as random variables. A first choice is a multivariate normal distribution, which allows easy-to-solve risk minimization problems for both VaR and CVaR measures. However, the normal distribution typically does not provide a good fit to the stock return data due to heavy tails. Other probabilistic models have been proposed including stable distribution [23], t-distribution [6], and mixture of multivariate normals [11, 8, 10] among others. As the probabilistic model becomes more complex, it might be difficult to solve optimization problems involving VaR or CVaR terms.

The main advantage of the market-based approach is its simplicity since it does not require any parameter estimation, optimization or distribution fitting procedure. One can, for instance, simply track Standard & Poor's 500 index, which is arguably a representative proxy for the United Stated stock market. The disadvantage of the market-based approach is its inflexibility. For instance, if an investor believes that a certain stock will outperform another, this approach does not allow to incorporate this view.

Several studies try to combine the two approaches. For instance, in the early 1990s, Black and Litterman [4, 5] proposed a way of combining the market portfolio and investor views into the classical MV approach. In practice, portfolios obtained using the Black-Litterman (BL) methodology tend to be more robust to data perturbations. But there are also some issues with the BL methodology: For instance, the derivation of the estimates is not very intuitive and a number of papers [19, 31, 15, 26] attempt to clarify it from different perspectives. Moreover, the BL derivation is based on strong assumptions, one of them being the normality of the random return vector. Furthermore, parameters have to be determined exogenously to incorporate the confidence in the investor views.

Recently, the BL model has been generalized using an interpretation as an inverse optimization problem [3]. Different from the classical derivation of the BL model, mean and covariance of the returns are determined as the solution of a conic program. This approach is flexible enough to eliminate some of the shortcomings of the BL methodology such as its inability to include views on variance, necessity to exogenously choose several parameters etc. Other recent extensions of the BL methodology include [21], in which inverse optimization incorporates views on the variance, and [35] in which views are created using Verbal Decision Analysis.

In this paper, we address several issues raised above and extend recent work. We start our analysis by focusing on the vector of stock returns modeled as a random variable. Using Standard & Poor's (S&P) 500 dataset, we show that the returns are not normally distributed via statistical tests, and we propose an enhanced probabilistic model, namely a mixture of normal random variables. Then, we discuss how to construct risk minimizing portfolios under different probabilistic models (normal and mixture of normals) and risk measures (standard deviation and CVaR). We also propose a BL-type approach, which incorporates market-based information into CVaR minimizing portfolios.

Buckley, Saunders and Seco [8] model the stock returns as a mixture of two normal random variables, considering several objective functions for the mean and variance of each "regime". The main difference in our work is that we optimize portfolios directly with respect to the CVaR of a mixture distribution. Since this measure does not have a closed form expression, we first propose an algorithm to compute its value given a portfolio vector, and use this algorithm as a subroutine in the actual portfolio construction problem.

Our key contributions in this paper are summarized below:

1. We model the vector of stock returns as a random variable coming from a mixture of multivariate normals. Our main contribution in this part is to propose two different approaches for estimating the resulting parameters.

2. We analyze how to construct portfolios that minimize CVaR under the mixture model. Although CVaR minimization under the normal distribution is straightforward, resulting in a second-order cone program, the case with the mixture is less obvious since the CVaR of a mixture distribution does not have a closed form expression. We analyze how it can be numerically computed to

machine precision via a root finding algorithm. We also propose a closed form, second-order cone representable approximation of CVaR in this case.

3. We extend the work on the BL approach via inverse optimization to CVaR minimization under both normal and mixture distribution models. In the latter case, we propose a sophisticated approach, which combines our previous two contributions. In particular, our model is governed by the market equilibrium equation, and the parameters of the mixture distribution are treated as investor views.

4. We present computational results applied to the S&P 500 dataset. Empirically, we observe that, as expected, market-based portfolios typically have higher reward and higher risk than risk minimizing portfolios. However, we show on the same dataset that a certain combination of the market-based and risk minimizing portfolios obtained through a BL-type approach may yield portfolios with similar rewards and smaller risk.

The rest of the paper is organized as follows: In Section 2, we provide a statistical analysis of the S&P 500 dataset and model the stock returns as a mixture of normal random variables. In Section 3, we present our portfolio optimization problem with different risk measures. In Section 4, we propose a new approach to combine CVaR minimizing portfolios with market information to obtain BL-type portfolios. In Section 5, we present our computational experiments on the S&P 500 dataset, and compare market-based, risk minimizing and combination portfolios. Finally, Section 6 concludes our paper with further remarks.

2 Statistical Analysis of Stock Returns

As opposed to the standard Markowitz approach, which does not require any distribution information on the asset returns to construct the portfolios, the VaR and CVaR measures need explicit forms of the distributions. It is not uncommon in the finance literature to model the stock returns as normally distributed random variables. However, stock returns are rarely normally distributed, and typically have heavier tails. Therefore, it is crucial to use a different probabilistic model to capture the heavy tail effect, especially the left tail, which is closely related to the risk of a portfolio when the VaR and CVaR measures are used.

In this section, by using a real dataset, specifically the stocks in Standard & Poor's (S&P) 500 index over a 30-year time span, we provide statistical evidence that the stock returns does not come from a normal distribution. Then, we propose an alternative probabilistic model, namely, we model the stock returns as a mixture of normal random variables and explain how the parameters of the mixture distribution can be estimated. Finally, we again statistically show that the proposed models based on mixture distributions are in fact better fits to the S&P 500 dataset.

2.1 Data Collection and Normality Tests

In order to test the hypothesis that the stock returns are normally distributed, we first collected historical stock returns and market capitalization information from the Wharton Research Database Services (WRDS). Since working with tens of thousands of different stocks is not appropriate for this study, we focused on stocks in the S&P 500 index. Following [3], we further simplified our analysis by focusing on 11 sectors according to the Global Industrial Classification Standard (GICS). Using WRDS, we collected the return and market capitalization information for all the stocks that have been in the S&P 500 between January 1987-December 2016, spanning a 30-year period. We then computed the return of a sector for each month as a weighted average of the returns of the companies in S&P 500 in that particular time period, where the weights are taken as the market capitalization of each stock in that sector. This procedure gave us 360 sector return vectors of size 11, denoted by R^t, $t = 1, \dots, 360$. We also recorded the percentage market capitalization of each sector j in month t, denoted by M_j^t, $j = 1, \dots, 11$, $t = 1, \dots, 360$.

We formally test the hypothesis that the sector returns are normally distributed using several statistical tests. We use an `R` package called `MVN` [22] to test the multivariate normality of the sector return vectors. We also test whether the returns of individual sectors are normally distributed. According to our extensive tests, which can be seen in Appendix A.1, we conclude that there is significant evidence that the neither the sector return vector nor the returns of individual sectors are normally distributed.

2.2 Modeling Returns as Mixture of Normal Random Variables

The fact that the vector of sector returns is unlikely to come from a multivariate normal distribution motivates us to search for an alternate probabilistic model that better explains the randomness of the stock returns. We will try to construct such a model using mixtures of (multivariate) normal random variables.

This choice for our model can be explained from two perspectives. First, we note that the stock returns have typically heavier left tails, which can be considered as the most critical part since it directly relates to the investment risk. This was previously observed by many researchers, including the J.P. Morgan Asset Management group [34]. In this paper, we try to capture this effect by introducing a mixture of random variables. An intuitive explanation for this phenomenon is offered as follows: Under "regular" conditions, the market, in fact, behaves following an approximate normal distribution. However, every once in awhile, a "shock" happens and shifts the mean of the return distribution to the left with possibly higher variance. This can explain the relatively heavier left tails of the empirical distribution. Second, as we will demonstrate below, introducing this more sophisticated probabilistic model greatly improves the fit to the data. However, this better fit comes with a cost of more complicated procedures for data fitting and for portfolio optimization. The data fitting issue will be addressed in the remainder of this section. As for the portfolio optimization procedures under these more complicated distributions, they will be discussed in Section 3.3.

Formally, let us assume that the random return is distributed as a mixture of two multivariate normal random variables, that is, with some probability ρ_i, returns are normally distributed with mean μ^i and covariance matrix Σ^i, for $i = 1, 2$. In other words, we have

$$r_M = \begin{cases} r_{M,1} & \text{w.p. } \rho_1 \\ r_{M,2} & \text{w.p. } \rho_2 \end{cases} \text{ where } r_{M,i} \sim \mathrm{N}(\mu^i, \Sigma^i),\ \ i = 1, 2. \tag{2}$$

Note that if ρ_i, μ^i, Σ^i, $i = 1, 2$, are given, we can compute the expectation and covariance of r_M as follows:

$$\mathrm{E}[r_M] = \rho_1\mu^1 + \rho_2\mu^2 \text{ and } \mathrm{Cov}(r_M) = \rho_1\Sigma^1 + \rho_2\Sigma^2 + \rho_1\rho_2(\mu^1 - \mu^2)(\mu^1 - \mu^2)^T. \tag{3}$$

In the next two subsections, we propose two models to estimate these parameters. We also provide statistical evidence that both of the proposed models improve the fits to the S&P 500 dataset in Appendix A.2 through Likelihood Ratio Test.

2.2.1 Market-wise Model

A straightforward approach to estimate the parameters of the mixture distribution is to use maximum-likelihood estimation, which can be achieved by the Expectation-Maximization (EM) Algorithm [14]. The EM algorithm can also be seen as a clustering method in the sense that it partitions a given dataset into a predetermined number of subsets. For our purposes, we can call the EM algorithm with 2 clusters and obtain the estimates of the parameters of the mixture distribution. We note that in this methodology, all the parameters of the market are estimated simultaneously, therefore, we call this model "market-wise".

Although the implementation of the EM algorithm is quite straightforward, there are a couple of points to be clarified. For instance, the likelihood function of a mixture distribution is not concave. Therefore, the EM algorithm may give some estimates that are only locally optimal. One can initialize the algorithm with random starting points to try to overcome this issue. Another important point, which might be more subtle, is the meaningfulness of the estimates: Suppose that $\hat{\mu}^2$ and $\hat{\mu}^1$ represent the mean return vectors for regular (good) and non-regular (bad) market conditions respectively. We would expect that $\hat{\mu}^2 > \hat{\mu}^1$. Although the algorithm itself does not guarantee such an outcome, we do actually obtain $\hat{\mu}^2 > \hat{\mu}^1$ and $\hat{\rho}_2 > \hat{\rho}_1$, which is in accordance with our intuition (see Table 1).

2.2.2 Sector-wise Model

As pointed out above, the Market-wise Model has some shortcomings in the estimation of parameters of the mixture model. Now, we present a more explicit model, which we think can alleviate some of these disadvantages. Our "Sector-wise Model" is constructed as follows: For each sector j, we partition the sample space into "Good" and "Bad" months in a way that β fraction of the time periods with the worst returns are put in the set $\mathcal{B}_j$ and the rest is stored in the set $\mathcal{G}_j$. Here, β is

Sector	Normal μ	Normal σ	Mixture ($\rho_1 = 0.19$, $\rho_2 = 0.81$) μ^1	μ^2	σ^1	σ^2
Energy	1.1769	6.2823	-0.0686	1.4687	8.5162	5.5799
Consumer Discretionary	1.5112	5.3195	0.4788	1.7532	8.2673	4.3019
Consumer Staples	1.3905	4.1470	0.6265	1.5696	6.7426	3.2166
Real Estate	1.1514	7.2451	0.3782	1.3327	11.9825	5.5492
Industrials	1.2887	5.1505	0.1638	1.5523	8.1720	4.0790
Financials	1.3322	6.2810	0.7178	1.4762	10.0766	4.9656
Telecommunications Services	1.0318	5.4780	0.3681	1.1873	8.7565	4.3408
Information Technology	1.7264	7.1032	1.3907	1.8051	11.8056	5.4234
Materials	1.3898	5.6955	0.0673	1.6998	8.4971	4.7437
Health Care	1.4164	4.6432	1.3203	1.4389	6.9005	3.9212
Utilities	1.0140	4.2801	0.6794	1.0924	6.1977	3.6764

Table 1: Estimates for normal and mixture of normal fits under a Market-wise Model with 360-month data (all figures are in percentage). Covariances between the sectors are not reported for brevity. Here, σ^i is a vector consisting of the standard deviation of sectors, $i = 1, 2$.

a predetermined parameter, say 0.01 or 0.05.

The above partitioning allows us to distinguish good and bad months for each sector individually. Hence, this model is sector-wise. However, we need to determine which months are good/bad for the market overall. In order to come up with a partition with respect to the market, we define the good months as the intersection of good months for each sector, that is, $\mathcal{G} = \cap_j \mathcal{G}_j$, and the bad months as the rest, that is, $\mathcal{B} = \{1, \ldots, T\} \setminus \mathcal{G}$. In words, a month is bad for the market if it is bad for at least one sector, and good for the market if it is good for all the sectors. Finally, when we obtain this partition of good/bad months, we compute a multivariate normal fit for these two clusters, and set $\rho_1 = |\mathcal{B}|/T$ and $\rho_2 = |\mathcal{G}|/T$.

Let us discuss the advantages and disadvantages of the Sector-wise Model. First of all, the estimation of the parameters is quite straightforward and does not suffer from local optimality. Also, the difference in the mean return vectors μ^1 and μ^2 is more significant (see Table 2). On the other hand, a potential disadvantage of this model is its dependency on the choice of the parameter β.

Sector	$\beta = 0.05$ ($\rho_1 = 0.21$, $\rho_2 = 0.79$) μ^1	μ^2	σ^1	σ^2	$\beta = 0.01$ ($\rho_1 = 0.05$, $\rho_2 = 0.95$) μ^1	μ^2	σ^1	σ^2
Energy	-4.1096	2.5447	6.5642	5.4354	-6.6418	1.5884	9.4547	5.8049
Consumer Discretionary	-3.5689	2.8257	6.2170	4.1605	-8.6638	2.0468	6.3325	4.6946
Consumer Staples	-2.3382	2.3553	4.6186	3.4140	-5.7599	1.7669	5.8934	3.6787
Real Estate	-5.0460	2.7550	8.0399	6.0837	-11.1181	1.7972	9.2416	6.5300
Industrials	-4.2306	2.7168	5.6924	3.8936	-9.5334	1.8583	6.8897	4.3651
Financials	-4.7850	2.9150	6.9737	5.0008	-10.1492	1.9365	8.9733	5.4949
Telecommunications Services	-3.3425	2.1636	6.4104	4.5849	-5.9919	1.4014	6.4749	5.1732
Information Technology	-3.5388	3.0887	9.5042	5.5998	-9.6614	2.3257	9.0944	6.4645
Materials	-4.3457	2.8739	5.7548	4.6504	-8.2799	1.8988	7.7008	5.0989
Health Care	-2.4096	2.4063	5.3480	3.8790	-4.4674	1.7261	6.4402	4.3250
Utilities	-1.7655	1.7331	5.4648	3.5919	-2.3301	1.1900	6.4592	4.0725

Table 2: Estimates mixture of normal fits under Sector-wise Model with 360-month data.

3 Risk Minimizing Portfolio Construction

In this section, we discuss how to obtain risk minimizing portfolios under different probabilistic models and risk measures. We start with standard deviation minimization in Section 3.1, which does not require any distribution information on the returns. Then, we consider CVaR minimization when the sector returns are modeled as multivariate normal distributions in Section 3.2. These two optimization problems are well-known in the financial engineering literature. Our main contribution is presented in Section 3.3, in which we show how CVaR can be minimized when the sector returns are modeled as a mixture of normals.

3.1 Standard Deviation Minimization

Assuming that the covariance matrix Σ is estimated from historical data, we can solve the following problem to minimize the standard deviation of the portfolio return:

$$\min\left\{\sqrt{x^T\Sigma x}:\ x\in\Delta_n\right\}. \tag{4}$$

We note that problem (4) can be solved efficiently either as a quadratic program (after squaring the objective function) or as a second-order cone program (SOCP) in a lifted space.

3.2 CVaR Minimization under Normal Distribution

Let us assume that the vector of sector returns, denoted by r_N, is modeled to come from a multivariate normal distribution with mean parameter μ and covariance matrix Σ, estimated from historical data. Then, we can obtain a portfolio minimizing CVaR by solving

$$\min\left\{\mathrm{CVaR}_\alpha(r_N^Tx):\ x\in\Delta_n\right\}, \tag{5}$$

where the α-level CVaR of the return of a portfolio x is computed as

$$\mathrm{CVaR}_\alpha(r_N^Tx) = -\mu^Tx + \frac{\phi(\Phi^{-1}(\alpha))}{\alpha}\sqrt{x^T\Sigma x}.$$

Here, ϕ and Φ are respectively the probability density function (pdf) and the cumulative distribution function (cdf) of the standard normal distribution.

We again note that problem (5) can be formulated as an SOCP in a lifted space.

3.3 CVaR Minimization under Mixture Distribution

Now, let us assume that the vector of sector returns, denote by r_M, is modeled to come from a mixture of normal distributions with parameters ρ_i, μ^i, Σ^i, $i=1,2$, obtained from the historical data using one of the techniques proposed in Section 2.2. In this case, we would like to obtain a

CVaR minimizing portfolio by solving the following optimization problem:

$$\min \left\{ \mathrm{CVaR}_\alpha(r_M^T x) : \ x \in \Delta_n \right\}. \tag{6}$$

Since CVaR is a convex function [29], the optimization problem (6) is again a convex program. However, the major difficulty with this problem is that CVaR of a mixture distribution does not have a closed form expression. In the sequel, we will analyze how it can be computed to machine precision. For notational purposes, let us define $\nu_i := {\mu^i}^T x$, $\sigma_i^2 := x^T \Sigma^i x$, $i = 1, 2$ and $V := \mathrm{VaR}_\alpha(r_M^T x)$.

3.3.1 Computing CVaR under Mixture Distribution

We first note that $\mathrm{CVaR}_\alpha(r_M^T x)$ can be computed analytically if V is at hand (also derived in [7]) through

$$\begin{aligned} \mathrm{CVaR}_\alpha(r_M^T x) &= -\frac{1}{\alpha} \int_{-\infty}^{-V} y \sum_{i=1}^{2} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y-\nu_i)^2}{2\sigma_i^2}} dy \\ &= -\frac{1}{\alpha} \sum_{i=1}^{2} \int_{-\infty}^{-V} y \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y-\nu_i)^2}{2\sigma_i^2}} dy \\ &= \frac{1}{\alpha} \sum_{i=1}^{2} \rho_i [\sigma_i^2 \phi(\nu_i, \sigma_i^2, -V) - \nu_i \Phi(\nu_i, \sigma_i^2, -V)]. \end{aligned} \tag{7}$$

Here, $\phi(\nu, \sigma^2, y)$ and $\Phi(\nu, \sigma^2, y)$ are respectively the pdf and cdf of the normal distribution with mean ν and variance σ^2 evaluated at the point y.

However, $V = \mathrm{VaR}_\alpha(r_M^T x)$ does not have a closed form expression either. Therefore, the problem reduces to computing V. To this end, we will adopt the following procedure: First, by the definition of VaR, we have

$$\alpha = \mathrm{P}(r_M^T x < -V) = \sum_{i=1}^{2} \rho_i \mathrm{P}(r_{M,i}^T x < -V).$$

Let us define a function $f : \mathbb{R} \to \mathbb{R}$ as

$$f(v) := \sum_{i=1}^{2} \rho_i \mathrm{P}(r_{M,i}^T x < -v) - \alpha. \tag{8}$$

Clearly, the root of the function f is $V = \mathrm{VaR}_\alpha(r_M^T x)$. Hence, computing the VaR of the mixture distribution reduces to a single-variable root finding problem, which can be solved by available algorithms to machine precision. Many of these algorithms, such as bisection search, require an interval which contains the root. Below, we provide under and over-approximations of the function $\mathrm{VaR}_\alpha(r^T x)$, which can be used to initialize such algorithms.

Proposition 1. *Let $\alpha \le \rho_i$, $i = 1, 2$. Then,* $\max_{i=1,2}\{VaR_{\alpha/\rho_i}(r_{M,i}^T x)\} \le VaR_\alpha(r_M^T x)$.

Proof. For any $k \in \{1, 2\}$, we have

$$\begin{aligned}
\mathrm{P}\left(r^T x \le -\mathrm{VaR}_{\alpha/\rho_k}(r_{M,k}^T x)\right) =& \sum_{j=1}^{2} \rho_j \mathrm{P}\left(r_j^T x \le -\mathrm{VaR}_{\alpha/\rho_k}(r_k^T x)\right) \\
\ge& \rho_k \mathrm{P}\left(r_k^T x \le -\mathrm{VaR}_{\alpha/\rho_k}(r_{M,k}^T x)\right) \\
=& \alpha.
\end{aligned}$$

This implies that $\mathrm{VaR}_{\alpha/\rho_i}(r_{M,i}^T x) \le \mathrm{VaR}_\alpha(r_M^T x)$, hence the result follows. □

Let $\mathrm{ri}(\mathcal{S})$ denote the relative interior of the set $\mathcal{S}$.

Proposition 2. *Let $\alpha\theta_i \le \rho_i$, $i = 1, 2$ and $\theta \in ri(\Delta_2)$. Then, $VaR_\alpha(r_M^T x) \le \max_{i=1,2}\{VaR_{\alpha\theta_i/\rho_i}(r_{M,i}^T x)\}$.*

Proof. The result follows since we have

$$\begin{aligned}
\mathrm{P}\left(r_M^T x \le -\max_{i=1,2}\{\mathrm{VaR}_{\alpha\theta_i/\rho_i}(r_{M,i}^T x)\}\right) =& \sum_{j=1}^{2} \rho_j \mathrm{P}\left(r_j^T x \le -\max_{i=1,2}\{\mathrm{VaR}_{\alpha\theta_i/\rho_i}(r_{M,i}^T x)\}\right) \\
\le& \sum_{j=1}^{2} \rho_j \mathrm{P}\left(r_j^T x \le -\mathrm{VaR}_{\alpha\theta_j/\rho_j}(r_{M,j}^T x)\right) \\
=& \sum_{j=1}^{2} \rho_j \alpha\theta_j/\rho_j \\
=& \alpha.
\end{aligned}$$

□

Note that $\mathrm{VaR}_\alpha(r_M^T x)$ can be sandwiched between two expressions containing $\mathrm{VaR}_{\tilde{\alpha}}(r_{M,i}^T x)$ by utilizing Propositions 1 and 2. The advantage of this approach is that since $\mathrm{VaR}_{\tilde{\alpha}}(r_{M,i}^T x)$ can be computed explicitly as

$$\mathrm{VaR}_{\tilde{\alpha}}(r_{M,i}^T x) = -\nu_i^T x - \Phi^{-1}(\tilde{\alpha})\sigma_i, \tag{9}$$

we can obtain analytical expressions for lower and upper bounds of an interval which contains V. The above procedure is summarized in Algorithm 1.

Algorithm 1 Computation of $\mathrm{CVaR}_\alpha(r_M^T x)$.

Choose $\theta \in \mathrm{ri}(\Delta_2)$ s.t. $\alpha\theta_i \le \rho_i$, $i = 1, 2$.
Compute $L := \max_{i=1,2}\{\mathrm{VaR}_{\alpha/\rho_i}(r_{M,i}^T x)\}$ and $U := \max_{i=1,2}\{\mathrm{VaR}_{\alpha\theta_i/\rho_i}(r_{M,i}^T x)\}$ by using (9).
Compute V as the root of the function f defined in (8) on the interval $[L, U]$ using the bisection method.
Compute $\mathrm{CVaR}_\alpha(r_M^T x)$ by equation (7).

Now, let us go back to the CVaR minimization problem (6). Although we do not have a closed form expression for $\mathrm{CVaR}_\alpha(r_M^T x)$, we can still query its value given a portfolio vector x by using

Algorithm 1 as an oracle. For many optimization algorithms, first and second order derivatives are also needed. In our case, they can be approximated through a finite difference method since we have the ability to carry out the function evaluations. Hence, it is possible to numerically solve optimization problems involving CVaR for a mixture random variable.

3.3.2 Approximating CVaR under Mixture Distribution

In the previous section, we showed how $\mathrm{CVaR}_\alpha(r_M^T x)$ can be computed to machine precision via a root finding algorithm. Now, we propose an over-approximation of $\mathrm{CVaR}_\alpha(r_M^T x)$, which can be computed explicitly. Moreover, this approximation is second-order cone representable.

Proposition 3. $CVaR_\alpha(r_M^T x) \le \sum_{i=1}^{2} CVaR_{\alpha/\rho_i}(r_{M,i}^T x)$.

Proof. For notational purposes, let us define $V_i := \mathrm{VaR}_\alpha(r_{M,i}^T x)$. Note that $V \ge V_i$ for all $i = 1, 2$ due to Proposition 1. Then, we have

$$\begin{aligned}
\mathrm{CVaR}_\alpha(r_M^T x) &= -\frac{1}{\alpha}\int_{-\infty}^{-V} y \sum_{i=1}^{2} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y-\nu_i)^2}{2\sigma_i^2}}\, dy \\
&= -\frac{1}{\alpha}\sum_{i=1}^{2}\int_{-\infty}^{-V} y \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y-\nu_i)^2}{2\sigma_i^2}}\, dy \\
&\le -\frac{1}{\alpha}\sum_{i=1}^{2}\int_{-\infty}^{-V_i} y \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y-\nu_i)^2}{2\sigma_i^2}}\, dy \\
&= \frac{1}{\alpha}\sum_{i=1}^{2} \rho_i [\sigma_i^2 \underbrace{\phi(\nu_i, \sigma_i^2, -V_i)}_{\frac{1}{\sigma_i}\phi(\Phi^{-1}(\alpha/\rho_i))} - \nu_i \underbrace{\Phi(\nu_i, \sigma_i^2, -V_i)}_{\alpha/\rho_i}] \\
&= \sum_{i=1}^{2}\left[-\nu_i + \sigma_i \frac{\phi\left(\Phi^{-1}(\alpha/\rho_i)\right)}{\alpha/\rho_i}\right] \\
&= \sum_{i=1}^{2} \mathrm{CVaR}_{\alpha/\rho_i}(r_{M,i}^T x).
\end{aligned}$$

□

4 Incorporating Market Information into Risk Minimizing Portfolios

In the previous section, we presented ways to obtain risk minimizing portfolios under different probabilistic models and risk measures. In this section, our objective is to combine market information with these risk minimizing portfolios, which will be achieved through a BL-type approach. After reviewing the basic construction of the BL approach through Theil's mixed estimation procedure, we also summarize a modern interpretation via inverse optimization in Section 4.1. Then, we extend the inverse optimization model from MV optimization to CVaR minimization under both normal and mixture distributions in Section 4.2.

4.1 Review of the Black-Litterman Model

The classical BL model is driven by two factors: market equilibrium and investor views. Let us first discuss how market equilibrium is obtained through "reverse optimization". Consider the MV optimization problem with $\mathcal{X} = \mathbb{R}^n$:

$$\max_{x \in \mathbb{R}^n} \; \mu^T x - \delta x^T \Sigma x. \tag{11}$$

Assuming that Σ is estimated from the historical data as $\hat{\Sigma}$, δ is predetermined and the percentage market capitalization vector is given as x^{m}, we can compute the "implied returns" which induce this market as

$$\Pi := 2\delta\hat{\Sigma}x^{\mathrm{m}}, \tag{12}$$

by writing down the optimality conditions for the unconstrained optimization problem (11). This derivation is the basis of the capital asset pricing model (CAPM). In the absence of any other views, an investor should invest proportional to x^{m}.

The second force that drives the portfolio away from the market equilibrium is the investor views. The BL model incorporates these views in the portfolio allocation via a certain "mixing" procedure. The basic construction is as follows: Suppose that the random return vector is distributed normally with mean μ and covariance matrix $\hat{\Sigma}$, that is, $r \sim \mathrm{N}(\mu, \hat{\Sigma})$. However, according to this model, μ itself is random with

$$\mu \sim \mathrm{N}(\Pi, \tau\hat{\Sigma}), \tag{13}$$

where τ is chosen exogenously, and a smaller value implies strong confidence in the market equilibrium. The investor views are expressed with linear equations of the form $P\mu = q$ with certain confidence level quantified by the covariance matrix Ω. To be more precise, the BL model assumes that

$$P\mu \sim \mathrm{N}(q, \Omega), \tag{14}$$

where Ω is a diagonal matrix (so that views are independent) with positive diagonal entries. Again, smaller values indicate stronger confidence in the views.

At this point, we remind the reader the solution of the generalized least squares problem, which is a key to obtain the BL estimates:

Proposition 4. *[1] Consider the generalized least squares problem $Ax = b + \epsilon$, where $\epsilon \sim N(0, \Omega)$ with $\Omega \succ 0$. Then,*

$$\hat{x} := (A^T\Omega^{-1}A)^{-1}A^T\Omega^{-1}b,$$

minimizes the $||\cdot||_\Omega$ norm of the error, where $||v||_\Omega := \sqrt{v^T\Omega^{-1}v}$.

In order to obtain an estimate for μ, one can use Theil's Mixed Estimation [37, 36] procedure, which is just a corollary of Proposition 4. In particular, we rewrite equations (13)-(14) as

$$\begin{bmatrix} I \\ P \end{bmatrix} \mu = \begin{bmatrix} \Pi \\ q \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}, \quad \text{where} \quad \epsilon_1 \sim \mathrm{N}(0, \tau\hat{\Sigma}), \ \epsilon_2 \sim \mathrm{N}(0, \Omega). \tag{15}$$

The BL estimate for the mean return vector μ is obtained by

$$\mu^{\mathrm{BL}} = ((\tau\hat{\Sigma})^{-1} + P^T\Omega^{-1}P)^{-1}((\tau\hat{\Sigma})^{-1}\Pi + P^T\Omega^{-1}q). \tag{16}$$

Alternative derivations for μ^{BL} and Σ^{BL} can be seen in [19, 31] based on a Bayesian interpretation and in [3] via inverse optimization. Here, we will outline the latter approach since it has motivated the extension presented in the next section. Recall that the optimality condition of the MV problem (11) can be written as $\mu = 2\delta\Sigma x^*$. In the inverse optimization problem, given a solution x^*, we search for the parameters that make this solution optimal. Assuming that $x^* = x^{\mathrm{m}}$, δ is given and $\Sigma \succeq 0$, the solution of the inverse problem satisfies the following linear equation in μ and Σ:

$$\mu = 2\delta\Sigma x^{\mathrm{m}}. \tag{17}$$

This is the market equilibrium in the BL model. Similarly, investor views are expressed as linear equations in μ as

$$P\mu = q. \tag{18}$$

In this approach, the aim is to solve the system (17)-(18) and $\Sigma \succeq 0$ simultaneously to obtain estimates for μ and Σ in a certain least squares sense. In particular, one solves the following norm minimization problem:

$$\min_{\mu, \Sigma \succeq 0} \left\| \begin{matrix} \mu - 2\delta\Sigma x^{\mathrm{m}} \\ P\mu - q \end{matrix} \right\| \tag{19}$$

The flexibility of this approach allows us to use different norms, which can lead to tractable conic programs. For instance, if the underlying norm is $||v||_{\bar{\Omega}}$ where $\bar{\Omega} = \begin{bmatrix} \tau\hat{\Sigma} & 0 \\ 0 & \Omega \end{bmatrix}$ and $\Sigma = \hat{\Sigma}$, then the optimal solution of the problem (19) is precisely μ^{BL} in (16). Hence, the BL estimate is just a special case of the inverse optimization approach. Although this derivation allows one to specify views on the variance estimate Σ through, for instance, factor models, we will use historical variance estimates directly not to over-complicate our analysis.

4.2 Extension of the Black-Litterman Model to CVaR Minimization Problems

In this section, we extend the BL Model from MV optimization to CVaR minimization problems under both normal and mixture distributions. We use the inverse optimization framework [3] outlined in the previous section, and also present an equivalent interpretation in terms of Theil's Mixed Estimation principle.

4.2.1 Normal Distribution

Let us consider the CVaR minimization problem under a normal distribution defined in (5), which can be explicitly stated as

$$\min\big\{ -\mu^T x + z\sqrt{x^T\Sigma x} : \ e^T x = 1, x \geq 0\big\}, \tag{20}$$

where e is the vector of ones and $z := \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{\alpha}$. Let us associate dual variables λ and γ to the equality and inequality constraints in problem (20). Noting that the problem is a convex program, first-order necessary and sufficient conditions are:

$$-\mu + z\frac{\Sigma x}{\sqrt{x^T\Sigma x}} - \lambda e - \gamma = 0,\ e^T x = 1,\ x \geq 0,\quad \gamma_j x_j = 0,\ \gamma \geq 0. \tag{21}$$

Let us now consider the inverse problem, where we will treat a market portfolio x^{m} as an optimal primal solution to problem (20), and we will search for the parameters μ, Σ and dual variables λ, γ that satisfy the optimality conditions in (21). At this point, we will assume that the covariance matrix Σ is estimated by the sample covariance matrix $\hat{\Sigma}$. There are two main reasons for this assumption: First, it is generally accepted in the financial engineering literature that the estimation of the covariance matrix is reasonably accurate as opposed to the mean estimation. Second, this assumption allows us to have an inverse problem that is linear in the remaining unknowns. A further simplification can be made by taking into account the fact that each sector is invested a positive amount in the market portfolio since the percentage market capitalization is positive for each sector, i.e., $x^{\mathrm{m}} > 0$. Hence, we have that $\gamma = 0$. Therefore, the inverse problem is given below, where μ and λ are the only remaining unknowns:

$$\mu + \lambda e = \tilde{\mu}_N := z\frac{\hat{\Sigma}x^{\mathrm{m}}}{\sqrt{x^{\mathrm{m}T}\hat{\Sigma}x^{\mathrm{m}}}}. \tag{22}$$

We will call (22) the market equilibrium equation under normal distribution.

In our approach, the investor views are expressed simply as

$$\mu = \hat{\mu}, \tag{23}$$

where $\hat{\mu}$ is the sample average estimated from the data. We then solve the following optimization problem to obtain adjusted μ estimates

$$\min_{\mu,\lambda} \left\| \begin{matrix} \mu + \lambda e - \tilde{\mu}_N \\ \mu - \hat{\mu} \end{matrix} \right\|_{\bar{\Omega}}, \tag{24}$$

where $\bar{\Omega} = \begin{bmatrix} \tau\hat{\Sigma} & 0 \\ 0 & \hat{\Sigma} \end{bmatrix}$, with $\tau > 0$. Note that as τ goes to zero, we put more emphasis on the market equilibrium and investor views become less important. In the other extreme, as τ gets larger, then the investor views become dominant and we recover the sample estimate. We would like to point out that our model allows more general investor views than the one used above. For instance, similar to the original BL model, linear equations in terms of the mean return vector given as $P\mu = q$ can be incorporated into our model as well.

Note that problem (24) has a closed form solution. Below, we also provide an equivalent interpretation in terms of Theil's Mixed Estimation principle.

Proposition 5. *The solution of* (24) *is equivalent to the solution of the following mixed estimation problem:*

$$\begin{bmatrix} I & e \\ I & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \lambda \end{bmatrix} = \begin{bmatrix} \tilde{\mu}_N \\ \hat{\mu} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}, \quad \textit{where} \;\; \epsilon_1 \sim N(0, \tau\hat{\Sigma}), \;\; \epsilon_2 \sim N(0, \hat{\Sigma}).$$

4.2.2 Mixture Distribution

Let us now consider the CVaR minimization problem under the mixture distribution defined in (6). Recall that $\mathrm{CVaR}_\alpha(r_M^T x)$ does not have a closed form expression, hence, we cannot directly apply the inverse optimization idea. However, due to Proposition 3, we have an explicit expression that over-approximates $\mathrm{CVaR}_\alpha(r_M^T x)$. To be more precise, let us consider the following convex program

$$\min \left\{ \sum_{i=1}^{2} \left(-{\mu^i}^T x + z_i \sqrt{x^T \Sigma^i x} \right) : e^T x = 1, x \geq 0 \right\}, \tag{25}$$

where $z_i = \frac{\phi\left(\Phi^{-1}(\alpha/\rho_i)\right)}{\alpha/\rho_i}$. Let us again associate dual variables λ and γ to the equality and inequality constraints, and write down the first-order necessary and sufficient conditions as:

$$\sum_{i=1}^{2} \left(-\mu^i + z_i \frac{\Sigma^i x}{\sqrt{x^T \Sigma^i x}} \right) - \lambda e - \gamma = 0, \;\; e^T x = 1, \;\; x \geq 0, \quad \gamma_j x_j = 0, \;\; \gamma \geq 0 \tag{26}$$

Using the inverse optimization framework, we will again treat a market portfolio x^{m} as an optimal primal solution to problem (25), and we will search for the parameters μ^i, Σ^i, $i = 1, 2$, and dual variables λ, γ that satisfy the optimality conditions (26). We again make the assumption that the covariance matrices Σ^i's are estimated from the historical data, using one of the techniques developed in Section 2.2. By the same reasoning as above, we have that $\gamma = 0$. Finally, the inverse problem is given below, where μ^i, $i = 1, 2$, and λ are the remaining unknowns:

$$\sum_{i=1}^{2} \mu^i + \lambda e = \tilde{\mu}_M := \sum_{i=1}^{2} z_i \frac{\hat{\Sigma}^i x^{\mathrm{m}}}{\sqrt{{x^{\mathrm{m}}}^T \hat{\Sigma}^i x^{\mathrm{m}}}}. \tag{27}$$

We will call (27) the market equilibrium equation under a mixture distribution.

Similarly, the investor views are expressed as

$$\mu^i = \hat{\mu}^i, \;\; i = 1, 2, \tag{28}$$

where $\hat{\mu}^i$'s are estimated from the historical data. We then solve the following optimization problem to obtain adjusted μ^i estimates

$$\min_{\mu,\lambda} \left\| \begin{matrix} \sum_{i=1}^{2} \mu^i + \lambda e - \tilde{\mu}_M \\ \mu^1 - \hat{\mu}^1 \\ \mu^2 - \hat{\mu}^2 \end{matrix} \right\|_{\bar{\Omega}}, \tag{29}$$

where $\bar{\Omega} = \begin{bmatrix} \tau\hat{\Sigma} & 0 & 0 \\ 0 & \hat{\Sigma}^1 & 0 \\ 0 & 0 & \hat{\Sigma}^2 \end{bmatrix}$, with $\tau > 0$.

We point out a difference between the two models under normal and mixture distributions. In the former case, as τ approaches zero, we recover the market portfolio. However, in the latter, we do not necessarily recover the market portfolio due to added flexibility in the mixture model. On the other hand, as τ gets larger, the investor views become more important and we recover the estimates μ^i, as before.

We provide an equivalent interpretation of the problem (29) below.

Proposition 6. *Solution of* (29) *is equivalent to the solution of the following mixed estimation problem:*

$$\begin{bmatrix} I & I & e \\ I & 0 & 0 \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} \mu^1 \\ \mu^2 \\ \lambda \end{bmatrix} = \begin{bmatrix} \tilde{\mu}_M \\ \hat{\mu}^1 \\ \hat{\mu}^2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}, \quad \textit{where} \;\; \epsilon_1 \sim N(0, \tau\hat{\Sigma}), \;\; \epsilon_2 \sim N(0, \hat{\Sigma}^1), \;\; \epsilon_3 \sim N(0, \hat{\Sigma}^2).$$

5 Computational Results

In this section, we present the results of our computational experiments on the S&P 500 dataset. We explain the experimental setting in Section 5.1, including a proper definition of two types of market-based portfolios, our criterion for preferable portfolios and the rolling horizon based back-testing. In Section 5.2, we provide a comparison of these two market portfolios with five other risk minimizing portfolios obtained through the analytical approaches outlined in Section 3. In Section 5.3, we apply the BL-approach proposed in Section 4.2 for CVaR minimization and show that the resulting portfolios may dominate the market-based approach in terms of both expected return and risk measures. In Section 5.4, we propose a simple portfolio shrinkage idea, which provides non-dominated portfolios. Finally, in Section 5.5, we conduct extensive simulations to replicate the results on a synthetic dataset.

5.1 Experimental Setting

Let us start by explaining the details of our experimental setting. We use the 360-month S&P dataset mentioned in Section 2.1. Seven portfolios, two market-based and five risk minimizing, are constructed for each month $t = 181, \ldots, 360$, using only the historical return and market capitalization data in the interval $[t-H, t-1]$, where H is a rolling horizon window. This procedure gives us 180 portfolios for each strategy. For a fixed strategy S, these portfolios, denoted by x_S^t, are evaluated using the return information at R^t, by simply computing $p_S^t := {R^t}^T x_S^t$. The performance of a portfolio construction strategy is evaluated with respect to average reward, standard deviation and 1% CVaR measures by using the dataset $\{p_S^t : \ t = 181, \ldots, 360\}$. We also report two risk-adjusted performance measures: the ratio of average reward to standard deviation (also known as the Sharpe ratio) and 1% CVaR.

We now discuss how these seven portfolios are obtained. We construct two market-based portfolios for $t = 181, \ldots, 360$ using the market capitalization information M^t:

- Last Market Portfolio (LstM): We simply choose $x_{\text{LstM}}^t = M^{t-1}$.
- Average Market Portfolio (AvgM): We compute $x_{\text{AvgM}}^t = \frac{1}{H} \sum_{t'=t-H}^{t-1} M^{t'}$.

We construct five risk minimizing portfolios for $t = 181, \ldots, 360$ using the returns R^t:

- Standard Deviation Minimizing Portfolio (StDev): We obtain x_{StDev}^t by solving (4), where Σ is estimated from the data.
- CVaR Minimizing Portfolio under Normal Distribution (CVaR): We obtain x_{CVaR}^t by solving (5) with $\alpha = 0.01$, where μ and Σ are estimated from the data.
- CVaR Minimizing Portfolio under Mixture Distribution with Market-wise Model (EM): We obtain x_{EM}^t by solving (6) with $\alpha = 0.01$, where ρ_i, μ^i and Σ^i are estimated from the data using the EM algorithm.
- CVaR Minimizing Portfolio under Mixture Distribution with Sector-wise Model with $\beta = 0.05$ (GB(0.05)): We obtain $x_{\text{GB}(0.05)}^t$ by solving (6), where ρ_i, μ^i and Σ^i are estimated from the data using the Good-Bad partitioning with $\beta = 0.05$.
- CVaR Minimizing Portfolio under Mixture Distribution with Sector-wise Model with $\beta = 0.01$ (GB(0.01)): Same as GB(0.05) except β is chosen as 0.01.

We use MOSEK [28] to solve problems (4) and (5), and IPOPT [40] to solve problem (6).

5.2 Market-Based vs. Risk Minimizing Portfolios

We first compare the performance of two market-based and five risk minimizing portfolios with respect to different H values in Figure 1. We have several observations: First and foremost, market-based approaches result in higher reward, higher risk portfolios compared to risk minimizing portfolios. An interesting observation is that AvgM performs better with increasing values of H in terms of both average return and risk measures, and that it has a higher average return than LstM with typically slightly larger risk (except for CVaR when $H = 180$). This suggests that historical data can be useful even in market-based portfolio construction. We also observe that most of the risk minimizing portfolios, especially StDev, CVaR and EM, have better risk-adjusted performance compared to the market-based ones.

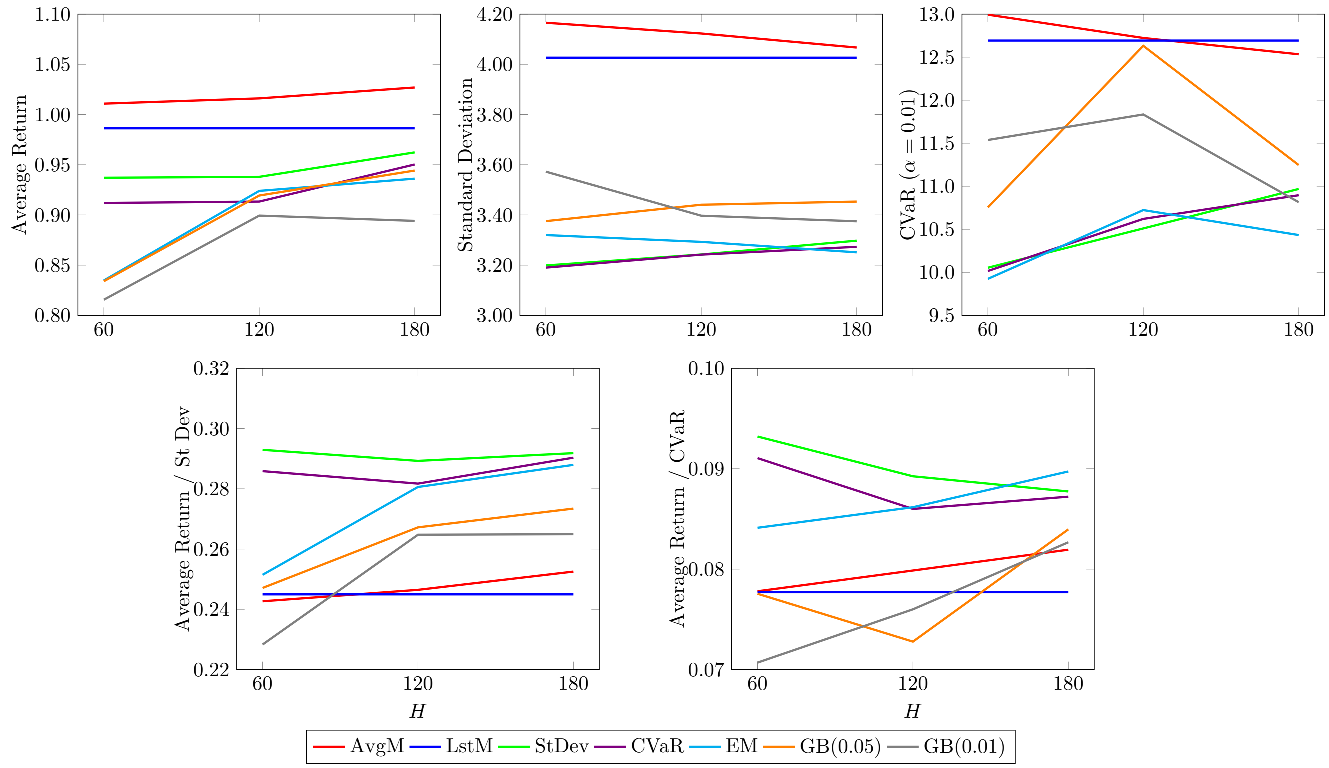

Figure 1: Performance comparison of market-based vs. risk minimizing portfolios for different H values. Vertical axes are in percentage.

If we focus on the analytical approaches, we observe a general trend that larger H values give rise to portfolios with a significantly higher average return, especially for EM, GB(0.01) and GB(0.05) while the performances of StDev and CVaR are less affected. The associated risks for the StDev, CVaR and EM methods are more or less stable while the risk performance of GB(0.01) and GB(0.05) methods are very much dependent on the rolling horizon window H, especially with the CVaR measure. Finally, we note that the risk-adjusted performance measures typically improve with larger H values under the mixture models EM, GB(0.01) and GB(0.05).

5.3 Combining Risk Minimizing Portfolios with Market Portfolios via the Black-Litterman Approach

In the previous subsection, we had two main observations: (i) The risk minimizing approaches, in fact, do generate lower risk portfolios than market-based approaches, but also with a lower reward. (ii) The performance of all portfolios is better in terms of the average return when H is large. In this subsection, we will focus on $H = 180$, and apply the BL approach developed in Section 4.2, which may result in portfolios with higher average return, lower risk than the market-based portfolios.

Note that our previous results with LstM, AvgM and StDev are not affected. They will serve as benchmarks in this subsection. We modify the other four portfolio construction methods as follows:

- CVaR: Solve problem (24) with respect to a given market capitalization x^{m} to update the mean estimate μ.
- EM, GB(0.05), GB(0.01): Solve problem (29) with respect to a given market capitalization x^{m} to update the mean estimates μ^i, $i = 1, 2$.

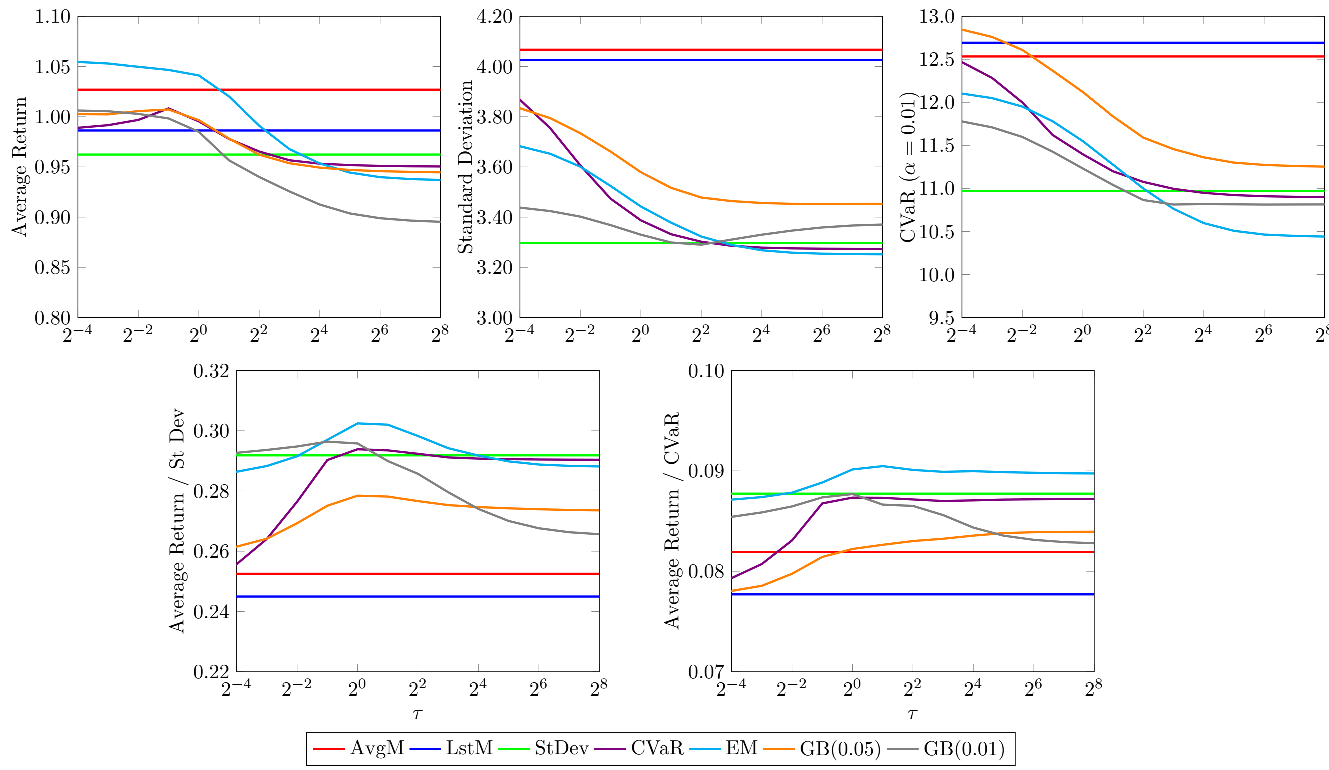

Figure 2: Performance comparison of market-based vs. risk minimizing portfolios after BL modification ($H = 180$, $x^{\mathrm{m}} = x_{\mathrm{LstM}}$).

In Figure 2, we choose the last market portfolio as x^{m} for each t. Although not shown in the figure, we verify that CVaR matches LstM when $\tau = 0$. Interestingly, CVaR outperforms LstM for $\tau \leq 1$ in both average return and risk measures. Arguably, the best performing method is EM, which has higher average return and lower risk than LstM and AvgM when $\tau \leq 4$ and $\tau \leq 1$, respectively. GB(0.05) and GB(0.01) do not perform as well as the EM approach. The GB(0.05) method typically performs better in average return and worse in risk as compared to the GB(0.01) method. We also note the success of risk-minimizing portfolios in terms of both of the risk-adjusted performance measures, especially the EM approach. A final observation is that when τ is large enough, we can recover the risk minimizing portfolios as expected.

In Figure 3, we choose the average market portfolio as x^{m} for each t. We again verify that CVaR matches LstM when $\tau = 0$. This time CVaR does not perform better than LstM in terms of average return, but its performance is very close for $\tau \leq 0.5$, with a sharp decrease in both risk measures. EM is again the winner as it performs better than both market-based portfolios in similar ranges mentioned above with even better average returns. It is also interesting to note that the EM approach outperforms the other portfolios in terms of the risk-adjusted measures for a large range of τ values considered. Although GB(0.05) and GB(0.01) are not very competitive against EM, they still perform well enough to produce non-dominated portfolios with respect to LstM and AvgM.

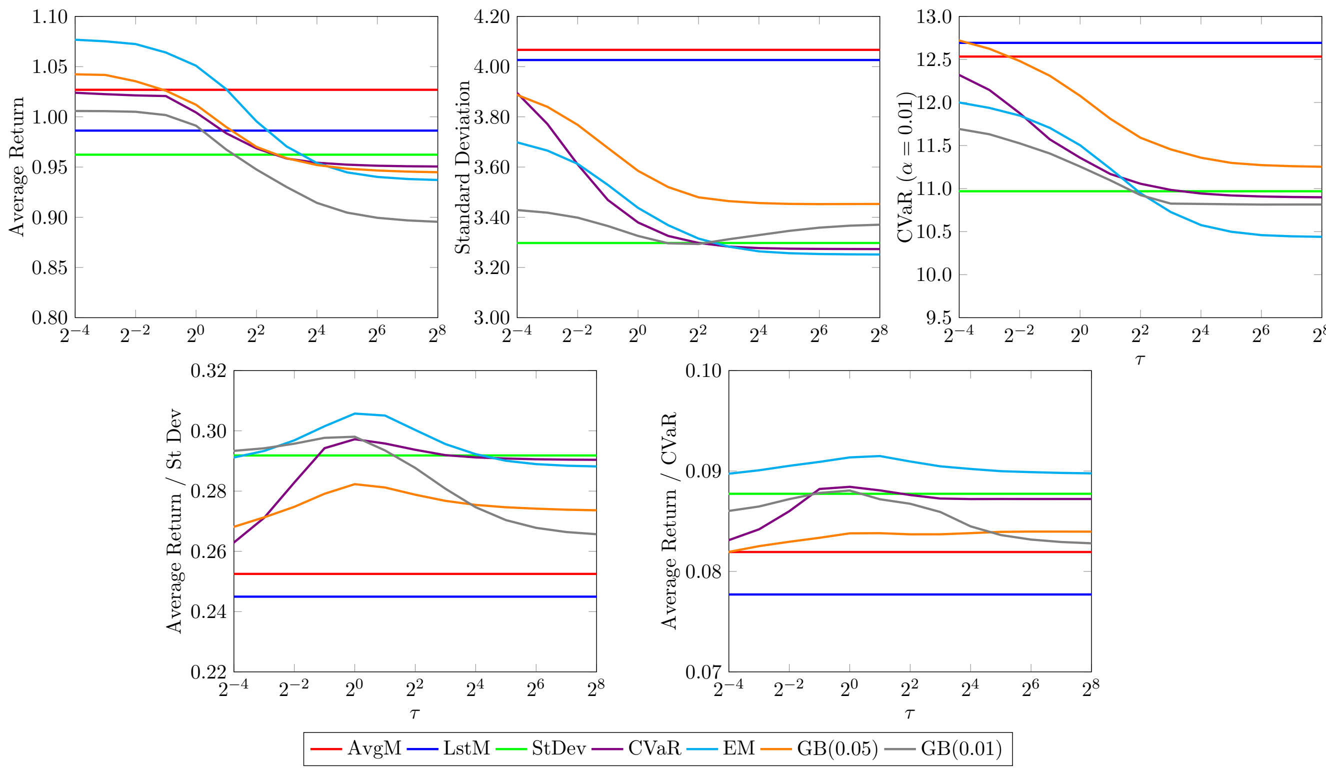

Figure 3: Performance comparison of market-based vs. risk minimizing portfolios after BL modification ($H = 180$, $x^{\mathrm{m}} = x_{\mathrm{AvgM}}$).

Our general conclusion regarding the experiments in this subsection is that one can benefit from the BL modification the most if the market portfolio is chosen according to the average percentage market capitalization and the value of τ is chosen small enough, for instance, close to 1. This typically results in portfolios with similar average return and lower risk compared to market portfolios. Such a strategy may even lead to portfolios that dominate the market-based portfolios when the EM method is used. Moreover, the risk-minimizing portfolio constructions methods, especially the EM approach, typically yield portfolios which are better in terms of the risk-adjusted measures.

Finally, we would like to discuss the choice of the rolling horizon window H in our results. When we repeated the above experiments with $H = 60$ and $H = 120$, we observed that CVaR and EM consistently give portfolios with similar average returns and lower risk compared to the market portfolios for small τ. However, the performances of GB(0.05) and GB(0.01) are quite sensitive to H. When H is chosen as 60, they both perform worse than EM in terms of both average return and risk criterion. On the other hand, when H is selected as 120, GB(0.05) provides high-return, high-risk portfolios whereas GB(0.01) gives low-return, low-risk portfolios compared to EM. Hence, our experience suggests that the market-wise model EM performs typically better than the sector-wise models GB(0.05) and GB(0.01). We again observe that the risk-minimizing portfolios have better risk-adjusted performance compared to the market-based portfolios. However, it is interesting to note that the methods based on the mixture distribution typically have worse risk-adjusted performance than the ones based on the normal distribution when H is selected as 60. This is somewhat expected since the mixture models have a larger number of parameters to be estimated, hence, they may require a reasonably large dataset to have more accurate estimators than the normal models.

5.4 Combining Risk Minimizing Portfolios with Market Portfolios via a Portfolio Shrinkage Approach

In this subsection, we propose a simple idea to combine market-based and risk minimizing portfolios by simply taking their convex combination. We will call this approach "portfolio shrinkage" (PS) as we can shrink a given portfolio towards a market portfolio by decreasing its weight in the convex combination. To be more precise, suppose that x^* is a risk minimizing portfolio obtained by solving one of the optimization problems (4), (5), (6), and we have a market portfolio x^{m} at hand. For $\kappa \in [0,1]$, we will obtain another portfolio x^*_κ by

$$x^*_\kappa := (1-\kappa)x^* + \kappa x^{\mathrm{m}}.$$

As opposed to the BL approach, which updates the input of the portfolio optimization problem, the PS approach works with the output of the optimization problem directly, and combines the market information with a risk minimizing portfolio more explicitly. The advantage of the PS approach compared the BL approach is its simplicity as the latter requires an elaborate analysis and additional optimization problems to be solved. On the other hand, by construction, the PS approach cannot produce portfolios that dominate the market portfolio it is based on whereas this is possible for the BL approach.

Interesting conclusions can be drawn when the average market portfolio is chosen as x^{m} in the numerical experiments, shown in Figure 4. As we shrink towards the better performing market portfolio, we can significantly increase the average return of PS portfolios. For instance, StDev dominates LstM when $\kappa \in [0.4, 0.9]$ while CVaR, EM and GB(0.0.5) all dominate LstM when $\kappa \in [0.6, 0.9]$ in terms of both average return and risk measures. By construction, we also obtain portfolios that neither dominate nor are dominated by AvgM.

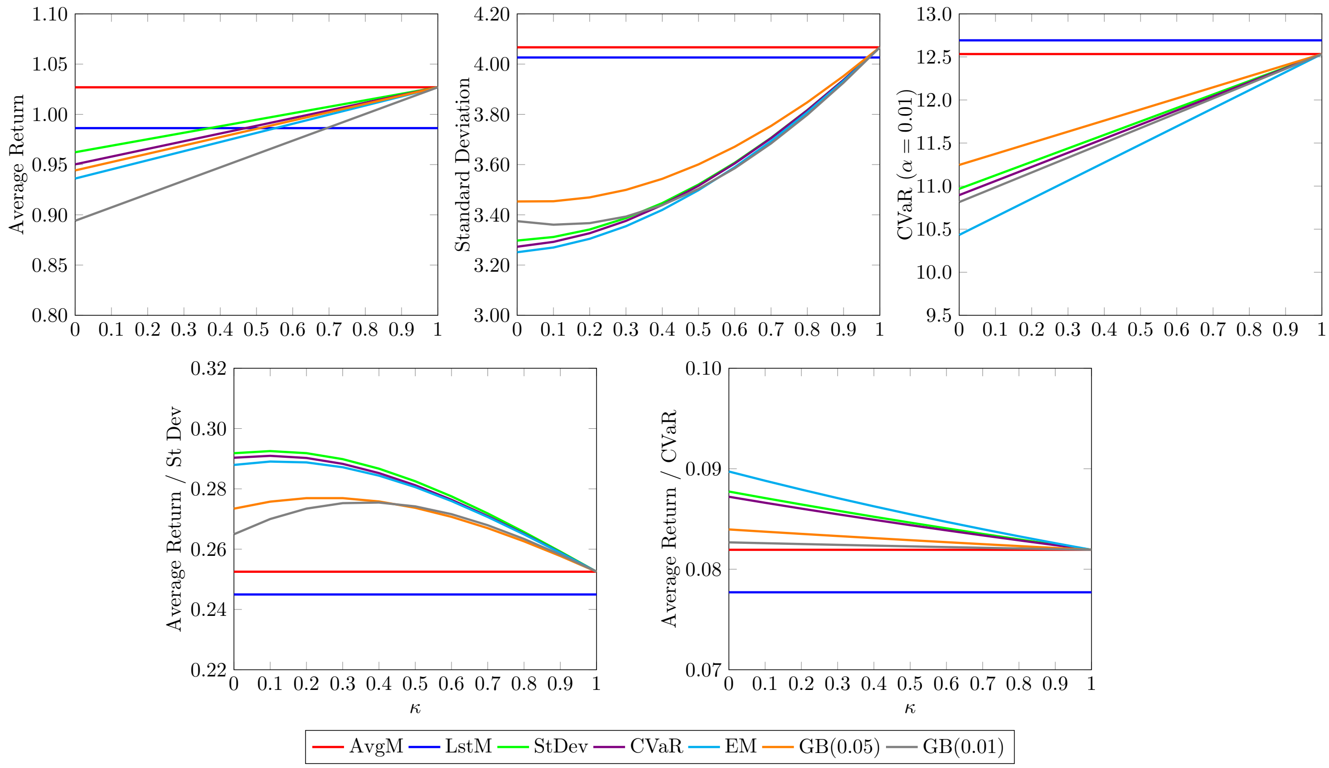

Figure 4: Performance comparison of market-based vs. risk minimizing portfolios after PS ($H = 180$, $x^{\mathrm{m}} = x_{\mathrm{AvgM}}$).

As a final remark, we would like to point out the fundamental similarity between the BL and PS approaches. Although constructed entirely differently, both methods work the best when the average percentage market capitalization is used as the market portfolio, and we put more emphasis on the market information (small τ for BL and large κ for PS). Hence, we can conclude that investing according to the market-based approaches is a good starting point for portfolio construction, but it can be enhanced by incorporating analytical approaches, which often leads to less risky and sometimes more profitable portfolios.

5.5 Replicating the Results on a Synthetic Dataset

In the previous subsections, our experiments were based on a single stream of data, namely the monthly S&P 500 dataset over a 30-year period. Now, we will try to replicate these results using a synthetic dataset. In order to carry out a reasonable simulation, we assume that the *true* distribution for the return vector is a mixture of two normal random variables with the parameters given in Table 1.

5.5.1 True Distribution is Known

Under the assumption that the true distribution is known to the investors, we can obtain an *ideal* CVaR minimizing portfolio by solving problem (6), or an *approximately ideal* portfolio under a normal distribution with parameters computed using (3) by solving problem (5). We compare the performance of these portfolios (EM and CVaR) with an equally weighted market portfolio, that is, $x^{\mathrm{m}} = \frac{1}{n}e$ in Table 3 and six other portfolios obtained through the BL approach. By construction, the EM portfolio has the smallest CVaR value but its expected return is also much smaller than that of the market portfolio. However, when combined with the market information through the BL approach, the EM portfolios retain a similar expected return while having a significantly reduced risk compared to the market portfolio. For example, the expected return of the EM($\tau = 1/4$) portfolio is only 0.66% smaller in relative terms than that of the equally weighted market portfolio whereas the standard deviation and CVaR measures are improved by about 9.1% and 11.6%, respectively.

We observe that the CVaR approach, which is obtained via a normal approximation, also produces a lower-reward, lower-risk portfolio compared to the market portfolio. Interestingly, its average reward is slightly better than the EM portfolio with higher risk in terms of the CVaR measure. We also see that the performance of the CVaR portfolios with the BL modification is very similar compared to their EM counterparts, especially when $\tau \geq 1/4$, suggesting that the BL approach may not be very sensitive to the exact distribution.

	Avg	St Dev	1% CVaR	Avg/St Dev	Avg/1% CVaR
Market	1.31	4.17	13.01	0.31	0.10
EM($\tau = 1/16$)	1.31	3.86	11.79	0.34	0.11
EM($\tau = 1/4$)	1.30	3.79	11.50	0.34	0.11
EM($\tau = 1$)	1.29	3.61	10.67	0.36	0.12
EM	1.20	3.41	9.38	0.35	0.13
CVaR($\tau = 1/16$)	1.31	4.04	12.50	0.32	0.10
CVaR($\tau = 1/4$)	1.31	3.79	11.52	0.34	0.11
CVaR($\tau = 1$)	1.29	3.52	10.33	0.37	0.12
CVaR	1.24	3.39	9.62	0.37	0.13

Table 3: Performance comparison of market-based, risk minimizing and BL-type portfolios under the true distribution. The average return (Avg), standard deviation (St Dev), 1% CVaR and two risk-adjusted performance measures are reported.

Because we assume here that the true distribution of the returns is known, it is interesting to observe what happens when we add an expected return constraint of the form $\hat{\mu}^T x \geq \mu_0$ to the CVaR minimization problems. In Table 4, we report the performance of the optimal portfolios under this additional constraint with two different values of μ_0. These results demonstrate that one can obtain portfolios with higher expected returns and smaller risk than the market portfolio under the assumption that the true distribution is known. We once again observe the similarity of the performances of the EM and CVaR portfolios.

5.5.2 True Distribution is not Known

So far in this subsection, we assumed that the true distribution of the returns is known. Of course this is not the case in reality. We now relax this assumption and design the following experiment: Suppose that we are given 181 random return vectors drawn from the mixture distribution. We

		Avg	St Dev	1% CVaR
EM	$\mu_0 = \hat{\mu}^T x^{\mathrm{m}}$	1.31	3.49	9.67
	$\mu_0 = 1.05\hat{\mu}^T x^{\mathrm{m}}$	1.38	3.65	10.22
CVaR	$\mu_0 = \hat{\mu}^T x^{\mathrm{m}}$	1.31	3.46	9.86
	$\mu_0 = 1.05\hat{\mu}^T x^{\mathrm{m}}$	1.38	3.62	10.37

Table 4: Performance of risk minimizing portfolios under the true distribution with an expected return constraint.

use the first 180 of these vectors to estimate the parameters of the true distribution under the market-wise and normal models. Based on these estimated parameters we obtain the EM and CVaR portfolios and their BL versions with varying τ values using the optimization algorithms described earlier. Finally, we evaluate these portfolios together with the equally weighted market portfolio using the last return vector. We repeat this experiment 10000 times and report the statistics in Table 5. We again note that the BL version of the EM approach yields portfolios with practically the same average return as the market portfolio for all the values of τ considered with a significantly reduced risk. For example, the average return of the EM($\tau = 1$) portfolio is only 0.40% lower than that of the market portfolio in relative terms whereas the standard deviation and 1% CVaR measures are improved by about 12.4% and 17.4%, respectively. We note that the reduction in the CVaR measure is even more dramatic when smaller values of α are considered. We also point out that the statistics of the EM and CVaR portfolios with the BL modification are almost indistinguishable for $\tau \geq 1/4$ while the performance of the EM($\tau = 1/16$) portfolio is slightly better than that of the CVaR($\tau = 1/16$) portfolio.

	Avg	St Dev	1% CVaR	0.1% CVaR	0.05% CVaR	Avg/St Dev	Avg/1% CVaR
Market	1.31	4.08	12.46	17.59	18.74	0.32	0.10
EM($\tau = 1/16$)	1.31	3.80	11.27	15.79	16.73	0.35	0.12
EM($\tau = 1/4$)	1.31	3.72	10.94	15.34	16.10	0.35	0.12
EM($\tau = 1$)	1.30	3.57	10.29	14.36	15.10	0.36	0.13
EM	1.24	3.49	9.75	13.90	14.78	0.35	0.13
CVaR($\tau = 1/16$)	1.31	3.95	11.91	16.90	17.88	0.33	0.11
CVaR($\tau = 1/4$)	1.31	3.72	10.94	15.53	16.25	0.35	0.12
CVaR($\tau = 1$)	1.29	3.51	10.03	14.16	14.88	0.37	0.13
CVaR	1.25	3.43	9.58	13.69	14.52	0.37	0.13

Table 5: Performance comparison of market-based, risk minimizing and BL-type portfolios under the synthetic data with 10000 replications.

As a final test, we repeat the above experiment with an expected return constraint and report the results in Table 6. When compared to the case without the expected return constraint (Table 5), the EM portfolio has higher average return and higher risk. Also, its performance is similar to the market portfolio in terms of the average return and it has a smaller risk. Nevertheless, a BL modified EM portfolio without the expected return constraint, such as EM($\tau = 1/4$), would still be preferable.

On the other hand, the CVaR portfolio with an expected return constraint produces counterintuitive results: Its average return is smaller compared to the case without the expected return constraint and the risk is higher. This observation suggests that when the true parameters are not known and an additional expected return constraint is enforced, the CVaR approach, which assumes a normal distribution for the return vector, may produce poor results and its behavior can be fundamentally different than that of the EM approach, which assumes a mixture distribution.

	Avg	St Dev	1% CVaR
EM	1.31	3.83	11.12
CVaR	1.22	3.52	9.98

Table 6: Performance of risk minimizing portfolios under the synthetic data with 10000 replications with an expected return constraint $\hat{\mu}^T x \geq \hat{\mu}^T x^{\mathrm{m}}$.

6 Conclusion

In this paper, we addressed the basic question of portfolio construction, combining analytical and market-based approaches. As an important component of the analytical approaches, we proposed to model stock returns as the mixture of two normals. We then explored the possibility of constructing risk minimizing portfolios under different probabilistic models (including the mixture distribution) and risk measures (including Conditional Value-at-Risk). We also proposed a Black-Litterman type approach using an inverse optimization framework to incorporate market information into a portfolio that minimizes the Conditional Value-at-Risk under the mixture distribution. Our computational experiments showed that market portfolios typically yield high-reward, high-risk portfolios, and that risk can be reduced by combining analytical and market-based approaches while achieving similar average returns.

Acknowledgments

We wish to thank Dr. Rob Stubbs for his insightful comments on an earlier version of this paper. This work was supported in part by NSF grant CMMI1560828 and ONR grant N00014-12-10032.

A Detailed Statistical Tests

A.1 Normality Tests

We first tested the multivariate normality of the sector return vectors using the 360-month dataset. The `MVN` package provides three statistical tests, namely, Mardia [24], Henze-Zirkler [20] and Royston [30], to perform this test. The p-values for these tests are extremely small, ranging from 0 to 10^{-37}. Therefore, we can confidently reject the hypothesis that the sector return vector follows a multivariate normal distribution.

One may question the validity of the above test since the market dynamics may have changed over a time span of 30 years. Hence, we also test the null hypothesis that the returns corresponding to certain subintervals may come from a multivariate normal distribution. To do this, we first fix a "horizon" H, and consider the interval $[t, t+H-1]$, $t = 1, \ldots, 361-H$, and repeat the statistical tests for each of these intervals. We report the percentage of subintervals with p-values less than 1% in Table 7. We observe that the null hypothesis can be rejected for the majority of the subintervals, especially when H is 120 or 180.

Test	$H = 60$	$H = 120$	$H = 180$
Mardia (Skewness)	49.50	100.00	100.00
Mardia (Kurtosis)	64.78	97.93	100.00
Henze-Zirkler	28.57	89.21	100.00
Royston	53.49	95.85	100.00

Table 7: Percentage of intervals with p-values less than 1% for multivariate normality tests with respect to different H values.

Although the sector return vector does not follow a multivariate normal distribution, it is still possible that individual sectors may come from normal distributions. In Table 8, we reported the p-values for four normality tests, namely Shapiro-Wilk (SW) [33], Anderson-Darling (AD) [2], Shapiro-Francia (SF) [32], Cramer-von Mises (CVM) [12, 39], provided by the `MVN` package. We observe that the SW, AD and SF Tests have small p-values for all the sectors, indicating that the sector returns are not normally distributed. However, there is some evidence that the returns of the Energy and Information Technology sectors may follow a normal distribution according to the CVM Test.

Sector	SW	AD	SF	CVM
Energy	0.0005	0.0317	0.0002	0.0761
Consumer Discretionary	0.0000	0.0002	0.0000	0.0006
Consumer Staples	0.0000	0.0000	0.0000	0.0004
Real Estate	0.0000	0.0002	0.0000	0.0005
Industrials	0.0000	0.0000	0.0000	0.0000
Financials	0.0000	0.0000	0.0000	0.0000
Telecommunications Services	0.0000	0.0014	0.0000	0.0073
Information Technology	0.0074	0.0088	0.0040	0.0134
Materials	0.0000	0.0024	0.0000	0.0079
Health Care	0.0173	0.0044	0.0093	0.0048
Utilities	0.0046	0.0006	0.0037	0.0005

Table 8: p-values of normality tests for individual sectors with 360-month data.

A.2 Goodness-of-Fit Tests for the Mixture Models

In Section A.1, we demonstrated through statistical tests that the sector returns do not follow a multivariate normal distribution. Now, we conduct Likelihood Ratio Test (LRT) to show that the models proposed in Section 2.2 are actually better fits to the S&P 500 dataset than the normal distribution. In LRT, the null hypothesis is that the S&P 500 dataset follows a normal distribution while the alternative hypothesis is that this dataset comes from one of the mixture distributions from the previous section. We report the corresponding p-values in Table 9, which shows that the mixture models are in fact better fits to the S&P 500 dataset.

Mixture Model	p-value
Market-wise	0
Sector-wise ($\beta = 0.05$)	2.67×10^{-10}
Sector-wise ($\beta = 0.01$)	5.56×10^{-6}

Table 9: p-values for LRT for different mixture distributions with 360-month data.

Finally, we also apply the RLT test to certain subintervals. Again, for a fixed length H, we consider the interval $[t, t + H - 1]$, where $t = 1, \ldots, 361 - H$, and repeated the statistical tests for each of these intervals. We report the percentage of subintervals with p-values less than 1% in Table 10. We observe that especially for $H = 120$ and $H = 180$, we can confidently say that mixture distributions are better fits to our dataset than the normal fit.

Mixture Model	$H = 60$	$H = 120$	$H = 180$
Market-wise	65.55	100.00	100.00
Sector-wise ($\beta = 0.05$)	48.95	96.23	100.00
Sector-wise ($\beta = 0.01$)	64.80	100.00	100.00

Table 10: Percentage of intervals with p-values less than 1% for LRT.

References

[1] AC Aitken. IV. – On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48, 1936.

[2] TW Anderson and DA Darling. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23(2):193–212, 1952.

[3] D Bertsimas, V Gupta, and IC Paschalidis. Inverse optimization: A new perspective on the Black-Litterman model. *Operations Research*, 60(6):1389–1403, 2012.

[4] F Black and RB Litterman. Asset allocation: Combining investor views with market equilibrium. *Journal of Fixed Income*, 1(2):7–18, 1991.

[5] F Black and RB Litterman. Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43, 1992.

[6] RC Blattberg and NJ Gonedes. A comparison of the stable and student distributions as statistical models for stock prices. *The Journal of Business*, 47(2):244–280, 1974.

[7] SA Broda and MS Paolella. *Expected shortfall for distributions in finance*, pages 57–99. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[8] I Buckley, D Saunders, and L Seco. Portfolio optimization when asset returns have the Gaussian mixture distribution. *European Journal of Operational Research*, 185(3):1434–1461, 2008.

[9] S Ceria and RA Stubbs. Incorporating estimation errors into portfolio selection: Robust portfolio construction. In *Asset Management*, pages 270–294. Springer, 2006.

[10] R Chen and L Yu. A novel nonlinear value-at-risk method for modeling risk of option portfolio with multivariate mixture of normal distributions. *Economic Modelling*, 35:796–804, 2013.

[11] E. Chin, A. S. Weigend, and H. Zimmermann. Computing portfolio risk using gaussian mixtures and independent component analysis. In *Proceedings of the IEEE/IAFE Conference on Computational Intelligence for Financial Engineering*, pages 74–117, 1999.

[12] H Cramér. On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal*, 1928(1):13–74, 1928.

[13] V DeMiguel and FJ Nogales. Portfolio selection with robust estimation. *Operations Research*, 57(3):560–577, 2009.

[14] AP Dempster, NM Laird, and DB Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 39(1):1–38, 1977.

[15] W Drobetz. How to avoid the pitfalls in portfolio optimization? Putting the Black-Litterman approach at work. *Financial Markets and Portfolio Management*, 15(1):59–75, 2001.

[16] L El Ghaoui, M Oks, and F Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.

[17] EF Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970.

[18] D Goldfarb and G Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.

[19] G He and RB Litterman. The intuition behind Black–Litterman model portfolios. 1999.

[20] N Henze and B Zirkler. A class of invariant consistent tests for multivariate normality. *Communications in Statistics-Theory and Methods*, 19(10):3595–3617, 1990.

[21] X Jia and J Gao. Extensions of Black–Litterman portfolio optimization model with downside risk measure. In *Control and Decision Conference (CCDC), 2016 Chinese*, pages 1114–1119. IEEE, 2016.

[22] S Korkmaz, D Goksuluk, and G Zararsiz. MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2):151–162, 2014.

[23] B Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.

[24] KV Mardia. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, pages 519–530, 1970.

[25] H Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[26] A Meucci. Black–Litterman approach. *Encyclopedia of Quantitative Finance*, 2010.

[27] RO Michaud. The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1):31–42, 1989.

[28] MOSEK. *MOSEK Modeling Manual*. MOSEK ApS, 2013.

[29] GC Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic constrained optimization*, pages 272–281. Springer, 2000.

[30] P Royston. Approximating the Shapiro-Wilk W-test for non-normality. *Statistics and Computing*, 2(3):117–119, 1992.

[31] S Satchell and A Scowcroft. A demystification of the Black–Litterman model: Managing quantitative and traditional portfolio construction. *Journal of Asset Management*, 1(2):138–150, 2000.

[32] SS Shapiro and RS Francia. An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67(337):215–216, 1972.

[33] SS Shapiro and MB Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, 1965.

[34] AZ Sheikh and H Qiao. Non-normality of market returns: A framework for asset allocation decision making. *Journal of Alternative Investments*, 12(3):8–35, 2010.

[35] T Silva, PR Pinheiro, and M Poggi. A more human-like portfolio optimization approach. *European Journal of Operational Research*, 256(1):252–260, 2017.

[36] H Theil. On the use of incomplete prior information in regression analysis. *Journal of the American Statistical Association*, 58(302):401–414, 1963.

[37] H Theil and AS Goldberger. On pure and mixed statistical estimation in economics. *International Economic Review*, 2(1):65–78, 1961.

[38] RH Tütüncü and M Koenig. Robust asset allocation. *Annals of Operations Research*, 132(1-4):157–187, 2004.

[39] R Von Mises. Statistik und wahrheit. *Julius Springer*, 1928.

[40] A Wächter and LT Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.

[41] S Zhu and M Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, 57(5):1155–1168, 2009.