Portfolio optimization

Zhenyuan Liu (teamed with Josiah Davis, Matthew Rainey)

05/08/2017

Contents

Nomenclature and Definition

book value the net asset value of a company, calculated as total assets minus intangible assets (patents, goodwill) and liabilities

book-to-market ratio the ratio of a company's book value to its market value (market equity)

compounded monthly return the geometric mean of monthly return in a given year

CVaR$_\alpha$ conditional value-at-risk at level α, the expected value of loss that is higher than VaR_α, i.e. $E(l \mid l \geq VaR_\alpha)$, where l is the loss of a portfolio, which is a random variable. In other words, expected value of loss in the worst $1-\alpha$ scenarios

equally distributed portfolio a portfolio such that the weight of any equity is $\frac{1}{N}$

market equity total dollar market value of all of a company's outstanding shares, also called the market value or market capacity

portfolio period a period that includes the training years and the test year of a portfolio for a given test year

Sharpe ratio return minus the risk-free return, divided by the standard deviation of return

utility function a non-decreasing function that has diminishing marginal gains

value equally distributed portfolio a portfolio such that the weight of an equity is the ratio of the company's market value to the market value of all companies

VaR$_\alpha$ value-at-risk at level α, a value γ such that $P(l \geq \gamma) = 1 - \alpha$, where l is the loss of a portfolio, which is a random variable

1 Introduction

In the current project, we tried to build an "optimal" portfolio. What for a portfolio can be called "optimal"? One with low out-of-sample variance, and high return.
To build "optimal" portfolios, we used the adjusted monthly return of various equities as well as multiple "risk-free" government bonds, and the cross-sectional characteristics/factors such as market equity, book-to-market ratio etc.
In our project, we first explored several portfolio optimization methods that only use historic monthly return data in solving some forms of optimization problems. It's noted in[1] that many such methods fail to outperform the equally distributed portfolio consistently, due to the large estimation error associated with these methods. Our results verified these conclusions.
We also explored two factor-based methods. These methods not only use historic monthly return data, but also historic cross-sectional characteristics/factors. One such method models the returns of equities directly, similar to the Fama-French three-factor model[2]. The other models the portfolio weights directly by adjusting the weights of a value equally distributed portfolio using 3 cross-sectional characteristics: the market equity (me), the book-to-market ratio (btm), and the compounded historic return. These cross-sectional characteristics/factors are from the fundamentals data revealed by companies' quarterly/annual earning reports. We used annual fundamentals data instead of quarterly data following the method proposed in[3].

2 Data description

The historic monthly return data used for this project can be easily obtained by using R's *quantmod* package. When provided with the equity symbol, start data, and end date, the *quantmod* package can retrieve the historic data of an equity from several sources and store it as a time series object in *xts* format. The raw data contains the daily open, close, high, low, adjusted prices, and volume. we used *quantmod*'s *monthlyReturn* function to pre-process the data and extracted the adjusted monthly return for equities.
The historic cross-sectional characteristics/factors data of companies can be obtained using the *Wharton Research Data Services* (WRDS)[4]. Hundreds of factors of companies are available from WRDS, among which we only used the equity price, the shares outstanding, and the book value per share. Together with the monthly return data, we were able to calculate the 3 cross-sectional characteristics: the market equity, the book-to-value ratio, and the compounded monthly return.
We selected 556 equities with full records (monthly return, equity price, shares outstanding, and book value per share) from 1986 to 2015 from thousands of equities from NYSE and NASDAQ . This filter makes solving our optimization problems much more easily, so that we can concentrate on studying the performance of various portfolio optimization methods rather than spending much

time dealing with missing data, new equities, equities unlisted from NYSE or NASDAQ etc. Another way to filter data is to select all equities with full records for a given portfolio period (5 training years and 1 test year). For example, select all the equities with full records from 1986 to 1991 in building a portfolio for 1991. We did some test cases and found that these two different filter methods have essentially no difference on the results. Therefore, we used the same 556 equities for all portfolio periods (test years 1991 to 2015, training years 1986 to 2014). The data we used have a size of a few MB. So we didn't have to deal with over-sized data in our project.

3 Optimization methods

3.1 Mean-variance optimization with LASSO

Classical Markowitz portfolio optimization methods seek to minimize portfolio variance for a given portfolio expected return or maximize the portfolio expected return for a given portfolio variance. There are several different ways to formulate the optimization problem under the Markowitz portfolio theory. One such formulation is as follows:

$$\begin{aligned} \min_x \quad & x^T \Sigma x + \tau ||x||_1 \\ s.t. \quad & \hat{r}^T x \geq \mu \\ & x^T \mathbf{1} = 1 \\ & x \in \mathcal{X} \end{aligned} \tag{1}$$

Define r as the monthly return vector, which is a random variable, then $\hat{r}$ is the expected value of the r. x is the weights vector, with all $x_i's$ sum up to one. A negative x_i indicates a short position in asset i. $\Sigma = E(r-\hat{r})(r-\hat{r})^T$, is the covariance matrix of the return vector r. $\mathcal{X}$ is a convex set of all feasible $x's$, which is formed by constraints like budget, transaction costs, no-short restrictions, sector bounds, and diversification etc[5]. τ is a tuning parameter that controls the sparsity of the portfolio. μ is a preset benchmark target for the expected return. This problem has been studied in[6].
Theoretically, if $\hat{r}$ and Σ are known, then the portfolios built by solving problem (1) are indeed optimal (more precisely, sub-optimal because of the regularization). However, $\hat{r}$ and Σ are by no means known to anyone. As a result, $\hat{r}$ and Σ have to be estimated using the sample mean and the sample covariance matrix of r. We adopted the same methodology in training and testing our portfolios as in[6] . Each portfolio period has six years, the monthly return data of the first 5 years (60 months) were used to estimate the $\hat{r}$ and Σ in problem (1), and the monthly return data of the last year were used to evaluate the out-of-sample performance of the portfolio. 47 different τ spanning a wide range were used as tuning parameters to control sparsity. The general form of constraint $x \in \mathcal{X}$ was not used so far. The benchmark μ was chosen as the return of an equally

distributed portfolio.
For each period, different τ led to different optimal solutions. Model selection became important in the next step of building our portfolios. We restricted our portfolios to be long only, meaning that $x_i's$ are non-negative. It's proved in[6] that when τ is sufficiently large, $x_i's$ would be non-negative. Therefore, there's no need to add the constraint $x \geq 0$ in problem (1). In general, we first selected from such $\tau's$ that $x^* \geq 0$, where $*$ indicates the optimal solution of x. Next, we chose the portfolio with the lowest in-sample variance

3.2 Minimum-variance optimization with LASSO

Due to the notoriously large estimation error in using the sample mean to estimate $\hat{r}$, researchers[1][7] argued that it's a better strategy to drop the constraint $\hat{r}^T x \geq \mu$ in problem (1), leading to the so-called minimum-variance optimization problem:

$$\begin{aligned} \min_x \quad & x^T \Sigma x + \tau ||x||_1 \\ s.t. \quad & x^T \mathbf{1} = 1 \\ & x \in \mathcal{X} \end{aligned} \tag{2}$$

To solve problem (2), we adopted the same training, testing and model selection rules as in section 3.1.

3.3 CVaR optimization with LASSO

Value-at-risk (VaR_α) and conditional value-at-risk ($CVaR_\alpha$) are broadly used in risk management. Let $l = -r^T x$ be the loss of a portfolio, which is a random variable. $VaR_\alpha(x)$, the value-at-risk of a portfolio x at probability level $\alpha \in (0, 1)$ is defined as:

$$VaR_\alpha(x) = F^{-1}(\alpha)$$

where F is the cumulative distribution function of l. In other words, $VaR_\alpha(x)$ is a value γ such that $P(l \geq \gamma) = 1 - \alpha$. This means, the probability that the loss of a portfolio is larger than or equal to $VaR_\alpha(x)$ is equal to $1 - \alpha$. For example, $VaR_{0.95}(x) = 0.01$ means the probability that the monthly loss of a portfolio is larger than or equal to 0.01 is 0.05. Clearly, a small number for $VaR_\alpha(x)$ is highly desired.
The conditional value-at-risk ($CVaR_\alpha$) is the expected value of loss that is higher than VaR_α:

$$CVaR_\alpha = E(l \; | l \geq VaR_\alpha)$$

In other words, $CVaR_\alpha$ is the expected value of loss in the worst $1-\alpha$ scenarios. Rockafellar[8] proved that low $CVaR_\alpha$ always implies low VaR_α and showed $CVaR_\alpha$ is a better tool than VaR_α in risk management. Therefore, we explored the following portfolio optimization problem using $CVaR_\alpha$:

$$\begin{aligned} \min_x \quad & CVaR_\alpha(x) + \tau ||x||_1 \\ s.t. \quad & \hat{r}^T x \geq \mu \\ & x^T \mathbf{1} = 1 \\ & x \in \mathcal{X} \end{aligned} \tag{3}$$

The only difference between the problem above and problem (1) is that we replaced $x^T \Sigma x$, which is the variance of a portfolio, with $CVaR_\alpha(x)$, which is the conditional value-at-risk of that portfolio. The intuition here is that, we want to minimize the expected value of loss in the worst $1-\alpha$ scenarios (plus a regularization term) by generating scenarios using training data (explained in next paragraph). To solve problem (3), we need to reformulate it as a convex optimization problem. Rockafellar[8] demonstrated with rigorous proof that the constraint $CVaR_\alpha(x) \leq \gamma$ can be reformulated with several linear constraints, at the cost of introducing new variables. Problem (3) is then equivalent to the following problem:

$$\begin{aligned} \min_{x,\ y_0,\ y_j' s,\ \gamma} \quad & \gamma + \tau ||x||_1 \\ s.t. \quad & y_0 + \frac{1}{1-\alpha} \sum_{j=1}^{J} p_j y_j \leq \gamma \\ & -(r_j)^T x - y_0 \leq y_j, \quad j = 1, ..., J \\ & 0 \leq y_j, \quad j = 1, ..., J \\ & \hat{r}^T x \geq \mu \\ & x^T \mathbf{1} = 1 \\ & x \in \mathcal{X} \end{aligned} \tag{4}$$

where y_o and $y_j' s$ are auxiliary variables (they don't have physical meanings here, it's very common to introduce new variables in solving convex optimization problem). γ is the value of $CVaR_\alpha$, α is the probability level, and τ is a tuning parameter for sparsity. x, y_o, $y_j' s$, and γ are all decision variables, i.e. problem (4) minimizes the objective function over these variables.
p_j is the probability of the jth **scenario**, and r_j is the return vector of the jth **scenario**. Scenarios come into play as the data-driven nature of problem (4). To solve problem (4), we need some kind of scenarios generating mechanisms by using the historic monthly return data. There are two different ways to do this. The first one is to use bootstrapping, i.e. we can simulate scenarios by sampling with replacement from historic monthly return data. For example, we used as training data the monthly return data of 5 years preceding the test year. We drew 1000 bootstrapping samples with replacement from these 60 monthly return data. For each of these scenarios, p_j simply equals $\frac{1}{1000}$. The other way is to assume Gaussian distribution for the return vector r. The mean and covariance of r can be estimated by the sample mean and sample covariance matrix.

After that, we can generate scenarios using the estimated Gaussian distribution. By definition, $CVaR_\alpha$ is the expected value of loss in the worst $1-\alpha$ scenarios. So we used the above mentioned scenarios generating mechanisms to simulate enough scenarios to capture the expected loss in the worst cases[8].
We adopted the similar methodology in training and testing our portfolios as in section 3.1. Each portfolio period has six years, the monthly return data of the first 5 years (60 months) were used to generate J (1000) scenarios, and the monthly return data of the last year were used to evaluate the out-of-sample performance of the portfolio. 47 different τ spanning a wide range were used as tuning parameters to control sparsity. The general form of constraint $x \in \mathcal{X}$ was not used so far. The benchmark μ was chosen as the return of an equally distributed portfolio.
For each test year, different τ led to different optimal solutions. We used a very similar model selection rule as in section 3.1. In general, we first selected from such $\tau' s$ that $x^* \geq 0$, where $*$ indicates the optimal solution of x. Next, we chose the portfolio with the **lowest conditional value-at-risk** for the training period. Note that $CVaR$ can't be calculated directly, it's obtained immediately after we solve problem (4), more exactly, $CVaR = \gamma^*$, where γ^* is γ at optimum.

3.4 Cross-sectional 3-factor method

Another broad class of portfolio optimization methods are factor-based methods. These are portfolio optimization methods that not only use historic monthly return data, but also historic cross-sectional characteristics/factors to build portfolios. In the famous Fama-French three-factor paper[2], Fama and French used three factors, namely an overall market factor, a factor related to the firm size, and the book-to-market ratio to model the return of equities.
A variant of the Fama-French three factor model[3] models the weights of a portfolio directly using 3 cross-sectional characteristics: the market equity (me), the book-to-market ratio (btm), and the compounded monthly return (r_comp). The market equity is total dollar market value of all of a company's outstanding shares. The book-to-market ratio is the ratio of a company's book value to its market value, where the book value is the net asset value of a company. The compounded monthly return is the geometric mean of monthly returns in a given year. The portfolio weights are calculated as follows:

$$x_{i,T+1} = y_{i,T} + \frac{1}{N_T}(\theta_1^{T+1} me_{i,T} + \theta_2^{T+1} btm_{i,T} + \theta_3^{T+1} r_{comp,i,T})$$

where $x_{i,T+1}$ is the weight of the ith equity for time $T+1$, the test year. $y_{i,T}$ is the weight of the ith equity of a **value equally distributed portfolio** at time T, the year preceding the test year. A **value equally distributed portfolio** is a portfolio such that the weight of an equity is the ratio of the company's market value to the market value of all companies. N_T is the number of equities at time T. $me_{i,T}$, $btm_{i,T}$, and $r_{comp,i,T}$ are the market equity, book-to-market ratio, and compounded return of the ith equity at time T, respectively. θ_1^{T+1},

θ_2^{T+1}, and θ_3^{T+1} are the three coefficients that are the same for all equities, note that they are also the same for the entire duration from time 0 to time $T+1$, i.e. both the training period and test period. These three coefficient can be calculated by maximizing some form of utility as follows:

$$\begin{gathered}\max_{\theta} \quad \sum_{t=0}^{T-1} u(r_{p,t+1}) = \sum_{t=0}^{T-1} u(\sum_{i=1}^{N_t} x_{i,t+1} r_{i,t+1}) = \\ \sum_{t=0}^{T-1} u(\sum_{i=1}^{N_t} (y_{i,t} + \frac{1}{N_t}(\theta_1^{T+1} me_{i,t} + \theta_2^{T+1} btm_{i,t} + \theta_3^{T+1} r_{comp,i,t})) r_{i,t+1})\end{gathered} \tag{5}$$

In the unconstrained optimization problem above, T is the last year for training, and 0 is the first year for training. In our project, we used the monthly return data of the 5 years preceding the test year for training. For example, if we want to build a portfolio for the year 1991, which is $T+1$, then $t=0$ is the year 1986, and T is the year 1990. N_t is the total number of equities at time t, $y_{i,t}$ is the weight of the ith equity of a value equally distributed portfolio at time t. $me_{i,t}$, $btm_{i,t}$, and $r_{comp,i,t}$ are the market equity, book-to-market ratio, and compounded return of the ith equity at time t, respectively. Note that $me_{i,t}$ and $btm_{i,t}$ were transformed and standardized in solving problem (5)[3]. $u()$ is the constant relative risk aversion (CRRA) utility function that we want to maximize. Empirically this utility function guarantees the existence and uniqueness of an optimal solution to problem (5). The CRRA utility function with parameter θ is shown as follows:

$$u(c) = \begin{cases} \frac{1}{1-\theta} c^{1-\theta} & if\ \theta > 0, \theta \neq 1 \\ ln(c) & if\ \theta = 1 \end{cases}$$

Following[3], we used the CRRA utility function with parameter 5. Solving problem (5) is extremely fast because it has only 3 decision variables for each test year! In comparison, problem (1) and problem (2) have 556 decision variables, whereas problem (4) has more than 2500 decision variables, if we draw 1000 scenarios.

3.5 Other portfolio optimization methods

We also explored several other optimization methods. For some of these methods, we did several test runs to evaluate their performance and efficacy. For others, we concluded their performance and efficacy from reading papers. These methods either have very bad out-of-sample performance (huge variance and losses) or can't deliver better performance than the mean-variance optimization method. To focus on more promising methods, we decided not to delve into these methods.

3.5.1 Sharpe Ratio portfolio optimization

Sharpe Ratio is defined as the return minus the risk-free return, divided by the standard deviation of return. A Sharpe ratio portfolio optimization problem writes:

$$\begin{aligned} \max_{x} \quad & \frac{\hat{r}^T x - r_f}{\sqrt{x^T \Sigma x}} \\ s.t. \quad & \hat{r}^T x \geq r_f \\ & x^T \mathbf{1} = 1 \\ & x \in \mathcal{X} \end{aligned} \tag{6}$$

where $r_f \geq 0$ is the return of a risk-free asset. The problem in this form is not convex. However, we can add a slack variable to convert it to an equivalent convex problem, more exactly a SOCP (second order cone programming) problem[5]. Details are omitted here, because this method is of no practical use. This problem seeks to maximize return and minimize variance at the same time, which is desirable. But the performance is terrible, as shown by our test runs. The underlying reason is, any optimization problem that seeks to maximize $\hat{r}^T x$ in some sense has terrible out-of-sample performance, due to the huge estimation error of using the sample mean to estimate $\hat{r}$.

3.5.2 Another mean-variance optimization with LASSO

Another form of mean-variance optimization is as follows:

$$\begin{aligned} \min_{x} \quad & x^T \Sigma x - \theta \hat{r}^T x + \tau ||x||_1 \\ s.t. \quad & \hat{r}^T x \geq \mu \\ & x^T \mathbf{1} = 1 \\ & x \in \mathcal{X} \end{aligned} \tag{7}$$

where $\theta \geq 0$ is a tuning parameter. If θ is very small, then this problem converges to the mean-variance optimization in problem (1). For a θ large enough, the out-of-sample performance is terrible, because this problem seeks to maximize $\hat{r}^T x$. Once again, any optimization problem that seeks to maximize $\hat{r}^T x$ in some sense has terrible out-of-sample performance, due to the huge estimation error of using the sample mean to estimate $\hat{r}$.

3.5.3 Robust portfolio optimization

In theory, robust portfolio optimization takes into account the uncertainly in the constraints and performs better in worst cases[5].
A robust version of the mean-variance problem (1) writes:

$$
\begin{aligned}
\min_{x,\ \Sigma,\ \hat{r}} \quad & x^T \Sigma x + \tau ||x||_1 \\
s.t. \quad & \hat{r}^T x \geq \mu \\
& x^T \mathbf{1} = 1 \\
& x \in \mathcal{X} \\
& \Sigma_{min} \preceq \Sigma \preceq \Sigma_{max} \\
& \Sigma \succeq 0 \\
& r_{min} \leq \hat{r} \leq r_{max}
\end{aligned}
\tag{8}
$$

In this robust version of problem (1), Σ_{min} and Σ_{max} are lower and upper bounds for Σ. Similarly, r_{min} and r_{max} are lower and upper bounds for $\hat{r}$. These bounds can be estimated from historic monthly return data. The major difference is that Σ and $\hat{r}$ are decision variables in problem (8). In our project, we have 556 equities in total, then Σ alone has 154,846 unique elements, causing a large number of decision variables in problem (8). Therefore, it's very slow to solve problem (8). In addition, we have very limited historic monthly data to estimate Σ_{min}, Σ_{max}, r_{min} , and r_{max}. As a result, solving problem (8) is not a promising strategy.
This is verified in [9]. The out-of-sample performance of using robust optimization is not better than its non-robust counterpart in terms of Sharpe Ratio[9].

3.6 Numerical solvers

The optimization problems defined above are convex problems or can be converted to a convex problem easily. Convex problems are guaranteed to have (a) global optimal solution (if feasible) and can be solved with various state-of-the-art solvers efficiently. In the current project, we used CVXPY[10] to solve our optimization problems.

4 Results

4.1 Mean-variance optimization with LASSO

In this section, we discuss the results of solving problem (1) in section 3.1. For each test year from 1991 to 2015, we used the monthly return data of the preceding 5 years as training data to build a portfolio. In other words, we built 25 different portfolios for each year from 1991 to 2015.
The out-of-sample performance of the portfolios built with the mean-variance methods and that of equally distributed portfolios are listed in Table 1. For each period (either 5 years or 25 years), we calculated the mean of monthly return, denoted as m, the standard deviation of monthly return, denoted as σ, and the Sharpe ratio, denoted as S, and the number of assets, denoted as n.

Period	**Optimal**				**Equal Weight**			
	m	σ	S	n	m	σ	S	n
01/91-12/15	.012	.029	.420	36	.014	.043	.324	556
01/91-12/95	.012	.019	.632	32	.020	.028	.711	556
01/96-12/00	.015	.031	.490	44	.015	.039	.388	556
01/01-12/05	.016	.031	.531	41	.015	.041	.366	556
01/06-12/10	.005	.036	.151	35	.010	.062	.165	556
01/11-12/15	.013	.027	.491	28	.010	.038	.257	556

Table 1: Best mean-variance portfolios vs. equally distributed portfolios. m is the mean of monthly return, σ is the standard deviation of monthly return, S is the Sharpe ratio, n is the number of assets, for the given period, all results are out-of-sample

First note that, the average Sharpe ratio S for 25 test years of mean-variance portfolios is 0.420, much higher than that of equal weight portfolios, which is 0.324. The reason is that mean-variance portfolios have much lower σ in all periods, leading to higher Sharpe ratio. The much lower σ results from the fact that problem (1) attempts to minimize the in-sample variance (plus some regularization term). Although there's huge estimation error in using sample mean and sample covariance matrix, the out-of-sample variance of mean-variance portfolios is much smaller than equal weight portfolios. The reason behind this is, portfolios with low variance in the past tend to have low variance in the future as well.

However, we can't conclude from the higher Sharpe ratio that portfolios built by the mean-variance methods can beat equally distributed portfolios consistently. Because the out-of-sample return of mean-variance portfolios is lower than equal weight portfolios. For mean-variance portfolios, an average monthly return of 0.012 may seem only slightly smaller than 0.014 of equal weight portfolios, but the yield over 25 years is much less. The time evolution of cumulative value of portfolios with an initial value of 1 dollar in 1990 is plotted in figure 1 (transaction costs and dividends are not considered). Clearly, the equally distributed portfolios can achieve a higher cumulative value in almost every year.

This may seem frustrating, yet the equally distributed portfolios have been recognized unanimously as a very tough strategy[1][6]. DeMiguel[1] showed that the estimation window needed for the mean-variance methods to outperform the equally distributed portfolio is around 3000 months for a portfolio with 25 assets and about 6000 months for a portfolio with 50 assets! Moreover, training data and test data were drawn from the same distribution in his simulation. This is hardly true for an equity's past and future return. Efforts attempted to reduce the estimation error in mean-variance methods by using shrinkage or factor analysis were also shown to be not effective[1][11].

Although our "best" mean-variance portfolios are not superior to the equally distributed portfolios in earning money, they're indeed sparse! As shown in

Table 1, the average number of assets is only 36 out of 556 equities in total. In real life, a sparse portfolio is highly desired, meaning lower management cost.

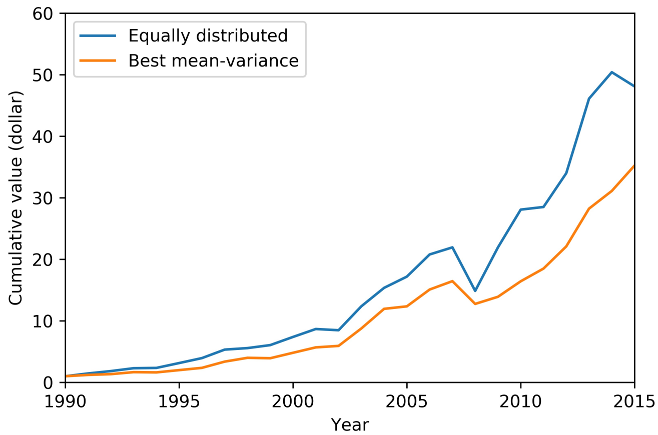

Figure 1: Cumulative value of portfolios from 1991 - 2015

4.2 Minimum-variance optimization with LASSO

In this section, we discuss the results of solving problem (2) in section 3.2. As explained before, the only difference between problem (1) and problem (2) is that problem (2) doesn't have the constraint $\hat{r}^T x \geq \mu$.

The out-of-sample performance of the minimum-variance portfolios and that of mean-variance portfolios are listed in Table 2. For each period (either 5 years or 25 years), we calculated the mean of monthly return, denoted as m, the standard deviation of monthly return, denoted as σ, and the Sharpe ratio, denoted as S. As shown in Table 2 and Figure 2, the out-of-sample performance of minimum-variance and mean-variance portfolios are quite close to each other. In Figure 2, the two cumulative value curves almost overlap with each other from 1991 to 2007, differences start to occur only after 2008, when rare events happened in the crisis. It was shown[1][7] that the out-of-sample performance of minimum-variance methods is even slightly better than mean-variance methods. The underlying reason is that the constraint $\hat{r}^T x \geq \mu$ is of little practical use or even misleading, due to the huge error in estimating $\hat{r}$ using sample mean. Sample mean of r simply doesn't contain much information about future r. Thankfully, minimizing the in-sample variance does lead to portfolios with low out-of-sample variance. Furthermore, portfolios with small variance do yield good return. As a result, both mean-variance and minimum-variance methods' performance is acceptable.

Period	Minimum-variance				Mean-variance			
	m	σ	S	n	m	σ	S	n
01/91-12/15	.012	.030	.401	36	.012	.029	.420	36
01/91-12/95	.013	.020	.644	30	.012	.019	.632	32
01/96-12/00	.014	.032	.441	38	.015	.031	.490	44
01/01-12/05	.016	.031	.528	42	.016	.031	.531	41
01/06-12/10	.004	.038	.116	38	.005	.036	.151	35
01/11-12/15	.012	.026	.470	32	.013	.027	.491	28

Table 2: Best minimum-variance portfolios vs. best mean-variance. m is the mean of monthly return, σ is the standard deviation of monthly return, S is the Sharpe ratio, n is the number of assets, for the given period, all results are out-of-sample

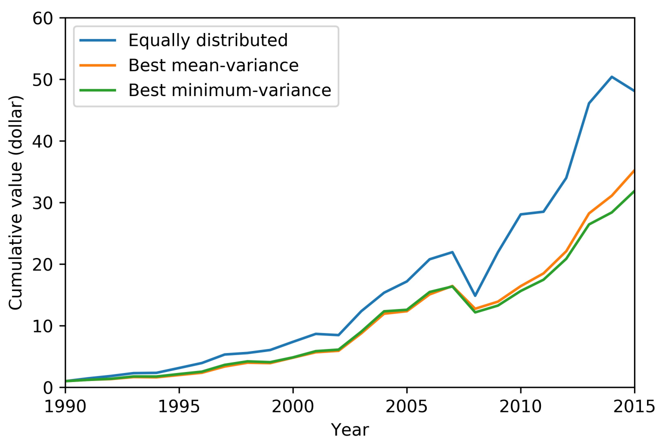

Figure 2: Cumulative value of portfolios from 1991 - 2015

4.3 CVaR optimization with LASSO

In this section, we discuss the results of solving problem (4) in section 3.3. α was chosen as 0.99, meaning that we minimized the expected loss in the worst 1% cases in-sample. The out-of-sample performance of the CVaR portfolios and that of equally distributed portfolios are listed in Table 3. For each period (either 5 years or 25 years), we calculated the mean of monthly return, denoted as m, the standard deviation of monthly return, denoted as σ, and the Sharpe ratio, denoted as S.
Similar to mean-variance portfolios, CVaR portfolios have lower variance and lower average return than equal weight portfolios out-of-sample. The perfor-

mance of CVaR portfolios are slightly worse than mean-variance portfolios, as shown in Figure 3.

Clearly, in problem (4) CVaR plays a similar role as variance does in problem (1). Moreover, small CVaR implies small variance, and vice versa. In section 4.2, we argued that portfolios with low variance in the past tend to have low variance in the future as well. This rule also applies to CVaR. Portfolios with low CVaR in the past tend to have low CVaR in the future as well. In solving problem (4), we minimized in-sample CVaR. Therefore, the out-of-sample CVaR should be low as well. However,it's hard to quantify the out-of-sample CVaR, because we only have 12 monthly return data for the test period. To capture the expected loss in the worst 0.01 scenarios, we need to generate many, say 1000 scenarios. Using 12 monthly return data to generate 1000 samples can hardly capture the worst 0.01 cases. Note that we can't even calculate the in-sample CVaR directly, we only get to know it after we solve problem (4). For example, in building a portfolio for 1991, we used the monthly return data from 1986 to 1990 to solve problem (4). After solving problem (4), we knew that $\gamma^* = 0.026$, which is the decision variable γ at optimal. Then we conclude that $CVaR_{0.99} = \gamma^* = 0.026$, which means, in the worst 1% cases, the expected monthly loss would be 0.026 in-sample.

Generally, low in-sample CVaR implies low out-of-sample CVaR, low out-of-sample CVaR implies low out-of-sample variance. As a result, the CVaR portfolios are actually low variance portfolios just like mean-variance portfolios. It's postulated that different α values may lead to different variance of portfolios out-of-sample. Once again, CVaR portfolios can't yield higher return than equal weight portfolios, because they used 60 monthly return data in sampling scenarios, whereas these data can't provide accurate information about the true distribution of r.

Period	**Optimal**			**Equal Weight**		
	m	σ	S	m	σ	S
01/91-12/15	.012	.038	.318	.014	.043	.324
01/91-12/95	.016	.028	.546	.020	.028	.711
01/96-12/00	.013	.044	.289	.015	.039	.388
01/01-12/05	.014	.036	.378	.015	.041	.366
01/06-12/10	.005	.046	.102	.010	.062	.165
01/11-12/15	.014	.032	.425	.010	.038	.257

Table 3: Best CVaR method portfolios vs. equally distributed portfolios. m is the mean of monthly return, σ is the standard deviation of monthly return, S is the Sharpe ratio, n is the number of assets, for the given period, all results are out-of-sample

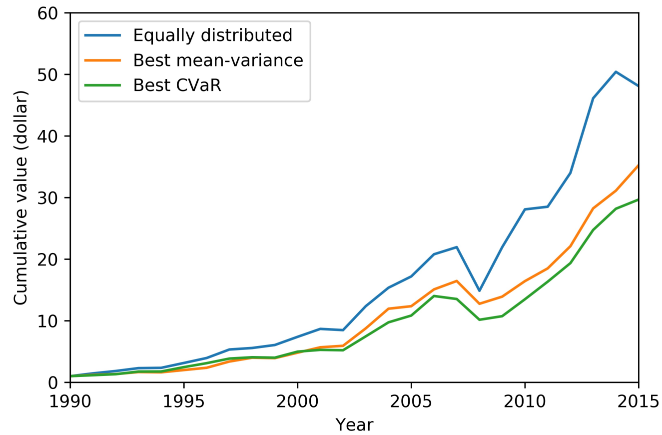

Figure 3: Cumulative value of portfolios from 1991 - 2015

4.4 Cross-sectional 3-factor model

In this section, we discuss the results of solving problem (5) in section 3.4. This method belongs to another class of methods called factor-based methods, which also use historic cross-sectional characteristics/factors to build portfolios. The cross-sectional 3-factor model tries to adjust the weights of a value equally distributed portfolio, which has already a very good performance. This adjustment incorporates extra information about companies, i.e. cross-sectional factors like market equity, and book-to-market ratio etc. This adjustment was shown to be very effective in building portfolios from 1964 to 2002[3].

The out-of-sample performance of the portfolios built with the cross-sectional 3-factor model and that of equally distributed portfolios are listed in Table 4. For each period (either 5 years or 25 years), we calculated the mean of monthly return, denoted as m, the standard deviation of monthly return, denoted as σ, and the Sharpe ratio, denoted as S.

As shown in Table 4, the 25-year average Sharpe ratio of cross-sectional 3-factor model portfolios is slightly higher than that of equally distributed portfolios. Unlike mean-variance or minimum-variance portfolios, the higher Sharpe ratio is not a result of lower σ, which is the denominator of the Sharpe ratio. Rather, the average variance of cross-sectional 3-factor model portfolios is slightly higher than equal weight portfolios. As discussed in section 3.5, if we maximize the in-sample return directly in any form, the out-of-sample performance of portfolios would be terrible, having absurdly huge variance and losses. We didn't have this issue with cross-sectional 3-factor model portfolios, because problem (5) doesn't maximize the in-sample return directly, rather it maximizes the CRRA utility, which takes as input the in-sample return. The CRRA utility function takes into account the relative risk aversion. As a result, the variance is kept

low, though slightly higher than that of equal weight portfolios.
The most exciting about cross-sectional 3-factor model portfolios is that they do have higher average return than equal weight portfolios, namely 0.017 vs. 0.014. As shown in Figure 4, the yield of cross-sectional 3-factor model portfolios is much higher than equal weight portfolios over 25 years. One drawback of cross-sectional 3-factor model portfolios is that they have very large turnover, i.e. portfolio weights change drastically from year to year. However, their superior performance is more than enough to offset this drawback. In plotting Figure 4, we manually subtract the yearly gain by 1% to account for the transaction costs due to large turnover for each year from 1991 to 2015. The cumulative value of cross-sectional 3-factor model portfolios is still much higher in all years.
Given the superior performance of cross-sectional 3-factor models, is it true that cross-sectional 3-factor model portfolios will consistently beat the equal weight portfolios in terms of return in the future? Unfortunately, most likely no. As shown in Table 4, for the period 01/91 to 12/95, average return of cross-sectional 3-factor model portfolios is much higher than that of equal weight portfolios, namely 0.029 vs. 0.020. However, for the period 01/11 to 12/15, they are the same. There's a clear trend that the superiority of cross-sectional 3-factor model portfolios decayed vs. time. The reason behind this is that, the three factors we used may be called "anomalies", which means factors that are able to explain return of companies to a large extent, when people first used them. However, they were not any more "anomalies", as more and more people became aware of them and used them. Keep in mind, Fama and French proposed the three-factor model for the first time in 1993[2]. After that, people did intensive and extensive research in factor-based models. Most common factors have become gradually not anomalous at all since then. To beat the equal weight portfolios, which outperformed more than 90% hedge funds in 2016, many more factors and more advanced machine learning methods are needed to build portfolios.

Period	**Optimal**			**Equal Weight**		
	m	σ	S	m	σ	S
01/91-12/15	.017	.047	.369	.014	.043	.324
01/91-12/95	.029	.040	.721	.020	.028	.711
01/96-12/00	.017	.045	.372	.015	.039	.388
01/01-12/05	.019	.045	.418	.015	.041	.366
01/06-12/10	.012	.061	.193	.010	.062	.165
01/11-12/15	.010	.036	.275	.010	.038	.257

Table 4: Cross-sectional 3-factor model portfolios vs. equally distributed portfolios. m is the mean of monthly return, σ is the standard deviation of monthly return, S is the Sharpe ratio, n is the number of assets, for the given period, all results are out-of-sample

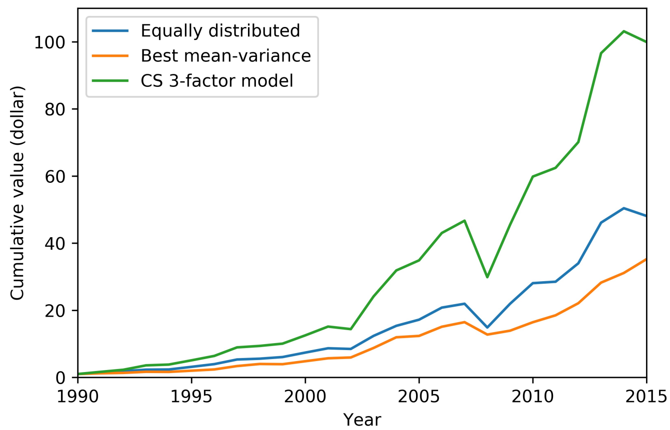

Figure 4: Cumulative value of portfolios from 1991 - 2015

5 Conclusions

5.1 Mean-variance and minimum-variance methods

We first explored mean-variance method and minimum-variance method by solving problem (1) and problem (2). These two methods have very similar performance, because they all try to minimize the in-sample portfolio variance (plus a regularization term). Both used historic monthly return to estimate the mean and covariance matrix of the return vector. The only difference between them is that minimum-variance method doesn't use the constraint $\hat{r}^T x \geq \mu$. Adding this constraint is of little practical use, due to the huge error in estimating $\hat{r}$ using sample mean. Past monthly return simply doesn't predict future return well. Fortunately, minimizing the in-sample variance does lead to portfolios with low out-of-sample variance. Meanwhile, equities with small variance do yield good return. As a result, both mean-variance and minimum-variance methods' performance is acceptable. Researchers tried to reduce the estimation error by methods like shrinkage method and factor analysis[7][11], but the performance was not improved. We also tried to add bonds to our portfolios. Although bonds are deemed as risk-free and have very low risk, portfolios with bonds actually have lower cumulative value than mean-variance portfolios, because they have much lower return in good years.

In[6], the authors claimed the portfolios built by solving problem (1) is consistently better than the naive equally distributed portfolios. This is because they used a relatively small number of assets in building their portfolios (25 and 100, respectively). In addition, they actually used portfolios grouped by market equity and book-to-market ratio, rather than equities to build portfolios. Therefore, the sample mean has less estimation error in their cases. Besides, they only compared Sharpe ratio and didn't compare the cumulative value of portfolios over time. After all, the most important standard in evaluating portfolios is how much yield it can achieve.

5.2 CVaR method

We also explored CVaR method by solving problem (4). The performance of CVaR portfolios is slightly worse than mean-variance portfolios. The CVaR method requires resampling from historic monthly return data. It minimizes the in-sample CVaR instead of in-sample variance of portfolios. There are two rather intuitively true facts: small CVaR implies small variance, and vice versa; portfolios with low in-sample CVaR/variance in the past tend to have low CVaR/variance in the future. Because of these facts, the CVaR portfolios have low out-of-sample CVaR/variance. As a result, the CVaR portfolios are actually low variance portfolios just like mean-variance portfolios. Thus, CVaR portfolios have acceptable performance. Once again, CVaR portfolios can't yield higher return than equal weight portfolios, because they only used finite historic monthly return, though with 1000 resamplings.

5.3 Cross-sectional 3-factor model

The cross-sectional 3-factor model tries to adjust the weights of a value equally distributed portfolio, which already has very good performance. This adjustment incorporates extra information about companies, i.e. cross-sectional factors like market equity, and book-to-market ratio etc. The variance of cross-sectional 3-factor model portfolios is only slightly higher than equal weight portfolios, because the CRRA utility function used takes into account the relative risk aversion. Cross-sectional 3-factor model portfolios can yield a much higher return than equal weight portfolios, even after compensating for the large turnover they have. The superiority of cross-sectional 3-factor model portfolios would decay vs. time. The factors we used became less and less effective, as more and more people became aware of them and used them. To build portfolios with high performance, we need to find new "anomalies". This has become increasingly difficult, as everyone on the market is playing "optimally". Advanced techniques using feature engineering, machine learning etc would be necessary to find new "anomalies".

5.4 Final remarks

In general, there's a huge gap between theories and real-life applications. Methods with rigorous theories like $CVaR$ is not better than the simpler and much faster mean-variance methods. Methods with less theories like the cross-sectional 3-factor models have the best performance. We should always be skeptical of theories and know their limitations.

References

[1] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953, 2009.

[2] Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

[3] Michael W. Brandt, Pedro Santa-Clara, and Rossen Valkanov. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, 22(9):3411, 2009.

[4] https://wrds-web.wharton.upenn.edu/wrds/.

[5] G.C. Calafiore and L. El Ghaoui. *Optimization Models*. Control systems and optimization series. Cambridge University Press, October 2014.

[6] Joshua Brodie, Ingrid Daubechies, Christine De Mol, and Domenico Giannone. Sparse and stable markowitz portfolios. (6474), 2007.

[7] Tze Leung Lai, Haipeng Xing, and Zehao Chen. Mean–variance portfolio optimization when means and covariances are unknown. *The Annals of Applied Statistics*, 5(2A):798–823, 2011.

[8] R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2000.

[9] Andre Santos. The out-of-sample performance of robust portfolio optimization. *Brazilian Review of Finance*, 8(2):141–166, 2010.

[10] Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

[11] Michael W. Brandt. Portfolio choice problems. In *Handbook of Financial Econometrics, forthcoming*. North-Holland, 2004.