Sparse and Stable Optimal Portfolios

Presented as Part of MATH8550

Daniel Mckenzie

December 6, 2017

What is a portfolio?

- Suppose there are N stocks available to an investor.
- Let $p_i(t)$ be the price of the i-th stock at time t (in, say, USD).
- A Portfolio is a vector of weights $\mathbf{w} = (w_1, \ldots, w_N)^T$, where w_i is the amount of capital to be invested in asset i.
- If w is total available capital, then require that $\sum_{i=1}^{N} w_i = \mathbf{w} \cdot \mathbf{1} = w$
- **Assumption 1:** An investor can make a purchase of any size in any stock (eg. can buy \$57.6543 worth of AAPL shares).

The return on a portfolio

Definition 1

1. **simple return:** $x_i := \frac{p_i(t)-p_i(t-1)}{p_i(t-1)}$.
2. **continuous return:** $x_i := \ln(p_i(t)/p_i(t-1))$.

We will only use simple return, and refer to it just as 'return'.

Definition 2

1. *The* **return** *on a portfolio* $\mathbf{w}$ *at time* t *is* $r(t) := \mathbf{x} \cdot \mathbf{w} = \sum_{i=1}^{N} w_i x_i(t)$
2. *The* **value** *of a portfolio* $\mathbf{w}$ *at time* t *is* $w(1 + r(t))$
   (recall $w = \sum_{i=1}^{N} w_i$*)*

Henceforth will normalize and set $w = 1$.

Optimal Portfolio Theory

1. Markowitz, 1952: '. . . the investor should consider expected return a desirable thing and variance of return an undesirable thing'.[1].

[1] Markowitz 1952.

Optimal Portfolio Theory

1. Markowitz, 1952: '...the investor should consider expected return a desirable thing and variance of return an undesirable thing'.[1].
2. **Assumption 2:** Fixed time window. Create portfolio at $t = 0$. Cash out at $t = T$. Cannot change portfolio for $0 < t < T$.

[1] Markowitz 1952.

Optimal Portfolio Theory

1. Markowitz, 1952: '...the investor should consider expected return a desirable thing and variance of return an undesirable thing'.[1].
2. **Assumption 2:** Fixed time window. Create portfolio at $t = 0$. Cash out at $t = T$. Cannot change portfolio for $0 < t < T$.
3. Expected return:

$$\mathbb{E}[r(T)] = \mathbb{E}\left[\sum_{i=1}^{N} w_i x_i(T)\right] = \sum_{i=1}^{N} w_i \mathbb{E}\left[x_i(T)\right] = \sum_{i=1}^{N} w_i \mu_i = \mathbf{w}^\top \boldsymbol{\mu}$$

where $\mu_i := \mathbb{E}[x_i(T)]$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$.

[1] Markowitz 1952.

Optimal Portfolio Theory 2

Variance of return *(Let $\bar{r} := \mathbb{E}[r(T)] = \mathbf{w}^T \boldsymbol{\mu}$)*:

$$
\begin{aligned}
\mathbb{E}\left[(r(T) - \bar{r})^2\right] &= \mathbb{E}\left[\left(\mathbf{w}^\top \mathbf{x}(T) - \mathbf{w}^\top \boldsymbol{\mu}\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathbf{w}^\top (\mathbf{x}(T) - \boldsymbol{\mu})\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathbf{w}^T (\mathbf{x}(T) - \boldsymbol{\mu})\right)\left(\mathbf{w}^T (\mathbf{x} - \boldsymbol{\mu})\right)\right] \\
&= \mathbb{E}\left[\mathbf{w}^T (\mathbf{x}(T) - \boldsymbol{\mu})(\mathbf{x}(T) - \boldsymbol{\mu})^T \mathbf{w}\right] \\
&= \mathbf{w}^T \mathbb{E}\left[(\mathbf{x}(T) - \boldsymbol{\mu})(\mathbf{x}(T) - \boldsymbol{\mu})^T\right] \mathbf{w} \\
&=: \mathbf{w}^T S \mathbf{w}
\end{aligned}
$$

S is called the *Covariance Matrix*

Optimal Portfolio Theory 3

Definition 3

$S := \mathbb{E}\left[(\mathbf{x}(T) - \boldsymbol{\mu})(\mathbf{x}(T) - \boldsymbol{\mu})^T\right]$ *is the covariance matrix.*
$S_{ij} = \mathbb{E}\left[(x_i(T) - \mu_i)(x_j(T) - \mu_j)\right]$

Optimal Portfolio Theory 3

Definition 3

$S := \mathbb{E}\left[(\mathbf{x}(T)-\boldsymbol{\mu})(\mathbf{x}(T)-\boldsymbol{\mu})^T\right]$ *is the covariance matrix.*
$S_{ij} = \mathbb{E}\left[(x_i(T)-\mu_i)(x_j(T)-\mu_j)\right]$

Recall we want to maximize return while minimizing risk. (At least) three problem formulations:

1. $\mathbf{w}_{opt} := \operatorname{argmin}\left\{\mathbf{w}^T S\mathbf{w} : \ \mathbf{w}\in\mathbb{R}^N, \mathbf{w}^\top\mathbf{1} = 1 \text{ and } \mathbf{w}^\top\boldsymbol{\mu} = r^*\right\}$ OR,
2. $\mathbf{w}_{opt} := \operatorname{argmax}\left\{\mathbf{w}^\top\boldsymbol{\mu} : \mathbf{w}\in\mathbb{R}^N, \mathbf{w}^\top\mathbf{1} = 1 \text{ and } \mathbf{w}^T S\mathbf{w} = v^*\right\}$ OR,
3. $\mathbf{w}_{opt} := \operatorname{argmin}\left\{\mathbf{w}^T S\mathbf{w} - \lambda\mathbf{w}^\top\boldsymbol{\mu} : \ \mathbf{w}\in\mathbb{R}^N \text{ and } \mathbf{w}^\top\mathbf{1} = 1\right\}$ for some $\lambda > 0$

Making This an Honest Optimization Problem

Will focus on problem formulation 1:

$$\mathbf{w}_{opt} := \operatorname{argmin} \left\{ \mathbf{w}^T S \mathbf{w} : \ \mathbf{w} \in \mathbb{R}^N, \mathbf{w}^\top \mathbf{1} = 1 \text{ and } \mathbf{w}^\top \boldsymbol{\mu} = r^* \right\}$$

[2] And see Brodie et al. 2009.

Making This an Honest Optimization Problem

Will focus on problem formulation 1:

$$\mathbf{w}_{opt} := \operatorname{argmin}\left\{\mathbf{w}^T S \mathbf{w} : \ \mathbf{w} \in \mathbb{R}^N, \mathbf{w}^\top \mathbf{1} = 1 \text{ and } \mathbf{w}^\top \boldsymbol{\mu} = r^*\right\}$$

skipping some steps[2], can show that the above is equivalent to:

$$\mathbf{w}_{opt} := \operatorname{argmin}\left\{\mathbb{E}\left[\left(\mathbf{x}(T)^\top \mathbf{w} - r^*\right)^2\right] : \mathbf{w} \in \mathbb{R}^N, \mathbf{w}^\top \mathbf{1} = 1 \text{ and } \mathbf{w}^\top \boldsymbol{\mu} = r^*\right\}$$

[2] And see Brodie et al. 2009.

Making This an Honest Optimization Problem

Will focus on problem formulation 1:

$$\mathbf{w}_{opt} := \operatorname{argmin} \left\{ \mathbf{w}^T S \mathbf{w} : \ \mathbf{w} \in \mathbb{R}^N, \mathbf{w}^\top \mathbf{1} = 1 \text{ and } \mathbf{w}^\top \boldsymbol{\mu} = r^* \right\}$$

skipping some steps[2], can show that the above is equivalent to:

$$\mathbf{w}_{opt} := \operatorname{argmin} \left\{ \mathbb{E}\left[\left(\mathbf{x}(T)^\top \mathbf{w} - r^* \right)^2 \right] : \mathbf{w} \in \mathbb{R}^N, \mathbf{w}^\top \mathbf{1} = 1 \text{ and } \mathbf{w}^\top \boldsymbol{\mu} = r^* \right\}$$

Assumption 3: Returns are stationary. Thus can use historic data $\{x_i(-\tilde{T}), \ldots, x_i(-1)\}$ to approximate $\mu_i = \mathbb{E}[x_i(T)]$:

$$\mu_i \approx \hat{\mu}_i := \tilde{T}^{-1} \sum_{t=-\tilde{T}}^{1} x_i(t)$$

$$\mathbb{E}\left[\left(\mathbf{x}(T)^\top \mathbf{w} - r^* \right)^2 \right] \approx \tilde{T}^{-1} \sum_{t=-\tilde{T}}^{-1} \left(\mathbf{x}(t)^\top \mathbf{w} - r^* \right)^2 = \tilde{T}^{-1} \| X\mathbf{w} - r^* \mathbf{1}_{\mathbb{R}^{\tilde{T}}} \|_2^2$$

[2] And see Brodie et al. 2009.

Making This an Honest Optimization Problem 2

Assumption 4: We don't allow short positions, thus $w_i \geq 0$ for all i
Thus problem becomes:

$$\mathbf{w}_{opt} = \text{argmin}\{\|X\mathbf{w} - r^*\mathbf{1}\|_2^2 : \ \mathbf{w} \in \mathbb{R}^N_{\geq 0}, \ \mathbf{w}^\top \mathbf{1} = 1 \text{ and } \mathbf{w}^\top \hat{\boldsymbol{\mu}} = r^*\} \quad (1)$$

Where:

$$X = \begin{bmatrix} x_1(-\tilde{T}) & \cdots & x_N(-\tilde{T}) \\ \vdots & \ddots & \vdots \\ x_1(-1) & \cdots & x_N(-1) \end{bmatrix}$$

Problem: Columns of X are highly correlated. Hence X is (approximately) rank deficient, so this minimization problem is unstable.

Solving this problem

1. Ignore all data. Choose equally weighted portfolio $\mathbf{w}_{ew} = \frac{1}{N}\mathbf{1}$.
2. Solve (1) using constrained least squares anyway, *eg* using `CVX`[3].
3. Add an ℓ_1 term (See Brodie et al. 2009) while allowing short positions (*i.e.* dropping assumption that all $w_i \geq 0$). Turns problem into a LASSO problem:

$$\operatorname{argmin}\{\|X\mathbf{w} - r^*\mathbf{1}\|_2^2 + \lambda\|\mathbf{w}\|_1 : \ \mathbf{w}^\top\mathbf{1} = 1 \text{ and } \mathbf{w}^\top\hat{\boldsymbol{\mu}} = r^*\}$$

4. In DeMiguel, Garlappi, and Uppal 2007 many (10+) other approaches are tested. None consistently outperform $\mathbf{w}_{ew}$.
5. `Sparse-Convex` approach. Find a k-sparse portfolio $\mathbf{w}_k$ that replicates performance of the (very non-sparse) equally weighted portfolio $\mathbf{w}_{ew}$. Amounts to solving:

$$\operatorname{argmin}\{\|X\mathbf{w} - X\mathbf{w}_{ew}\|_2^2 : \mathbf{w} \in \Delta \cap \Sigma_k\}$$

Where Σ_k is k-sparse vectors and $\Delta = \{\mathbf{w} \in \mathbb{R}^N_{\geq 0}, \ \mathbf{w}^\top\mathbf{1} = 1\}$.

[3] CVX Research 2012.

Sparse-Convex approach continued

Theorem 1

Let Σ_k be k-sparse vectors and $\Delta = \{\mathbf{w} \in \mathbb{R}^N_{\geq 0},\ \mathbf{w}^\top \mathbf{1} = 1\}$. In (Kyrillidis et al. 2013) an exact expression for $\mathcal{P}_{\Delta \cap \Sigma_k}$ is given.

Conclusion: Can exactly solve the `Sparse-Convex` problem via Projected Gradient descent.

Algorithm 1 `Sparse-Convex`

Input Historic data $X \in \mathbb{R}^{\tilde{T} \times N}$
Initialize $\mathbf{w}^{(0)} = (1, 0, \ldots, 0)^\top \in \mathbb{R}^N$
while True **do**
  Determine step size $\alpha^{(n)}$
  $\mathbf{w}^{(n+1)} = \mathcal{P}_{\Delta \cap \Sigma_k}\left(\mathbf{w}^{(n)} - \alpha^{(n)} X^\top X(\mathbf{w}^{(n)} - \mathbf{w}_{ew})\right)$
  if Stopping Criterion Met **then**
    Return
Output Sparse portfolio $\mathbf{w}^*$

Experiment 1: Comparing Stability - Set up

1. Use a common data set, the Fama-French 48 industry data set (FF48).
2. Every US stock is classified into one of 48 industries ('Agriculture', 'Tobacco', 'Chemicals' etc.).
3. Monthly returns for each industry are calculated from July 1926 to October 2017 [4].
4. Chose a five year period between July 1971 and June 2016 at random as our training data X
5. Considered two additional data sets $X_i = X + \epsilon E_i$, for $i = 1, 2$ where $E_{ij} \sim_{i.i.d} \mathcal{N}(0, 1)$ and $\epsilon = 0.005$.
6. Ran `CVX` to solve problem (1) and used `Sparse-Convex` to solve problem (2) for X, X_1 and X_2

[4] available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french`

Experiment 1: Comparing Stability - Results

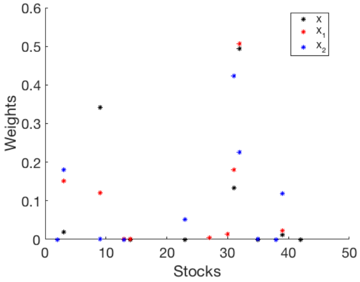

Figure: Constrained Least Squares Approach.

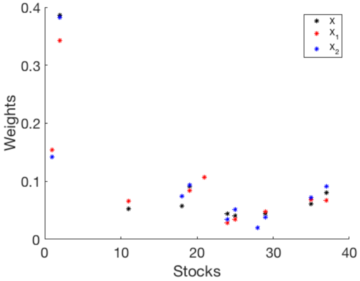

Figure: `Sparse-Convex` approach.

Experiment 2: Comparing Returns - Set Up

1. We use the same setup as (Brodie et al. 2009).
2. Used same FF48 data set.
3. Starting from July 1976, used previous five years of data as training data for four different methods: `Equally Weighted`, Constrained Least Squares (using `CVX`), `Sparse-Convex` with $k = 5$ and `Sparse-Convex` with $k = 10$.
4. Calculated returns on this portfolio over year July 1976 - June 1977.
5. 'Re-adjust' portfolios: Use previous five years of data (from June 1977 backwards) to build 4 new portfolios using the same methods.
6. Repeat up until June 2016.

Experiment 2: Comparing Returns - Results 1

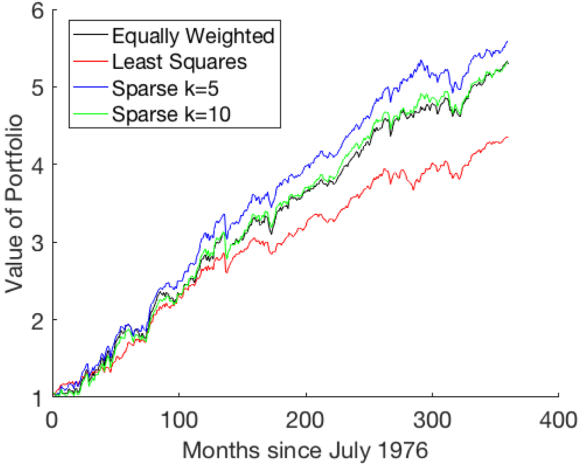

Figure: Portfolio value vs. time, for four different methods of selecting an optimal portfolio (using FF48 data set)

Experiment 2: Comparing Returns - Results 2

Method	μ	σ	S
Equally weighted	1.19%	4.50%	26.52
`CVX`	0.93%	3.48%	26.64
`Sparse Convex`, $k = 5$	1.27%	4.66%	27.25
`Sparse Convex`, $k = 10$	1.20%	4.56%	26.28

Table: Comparing mean monthly return (μ), standard deviation of monthly return (σ) and Sharpe ratio ($S := \mu/\sigma$) for different methods

Code

Code available at
https://github.com/DanielMckenzie/SparseOptimalPortfolios.
Use at own risk!

References

Joshua Brodie et al. "Sparse and stable Markowitz portfolios". In: *Proceedings of the National Academy of Sciences* 106.30 (2009), pp. 12267–12272.

Inc CVX Research. *CVX: Matlab Software for Disciplined Convex Programming, version 2.0*. `http://cvxr.com/cvx`. 2012.

Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" In: *The review of Financial studies* 22.5 (2007), pp. 1915–1953.

Anastasios Kyrillidis et al. "Sparse projections onto the simplex". In: *Proceedings of The 30th International Conference on Machine Learning*. 2013, pp. 235–243.

Harry Markowitz. "Portfolio selection". In: *The journal of finance* 7.1 (1952), pp. 77–91.