

Optimal Sparse Markowitz portfolios

Daniel McKenzie *

May 23, 2017

Abstract

We investigate the construction of optimal (in the sense of Markowitz) portfolios with the additional assumption that the portfolios are sparse. That is, that there are few active positions compared to the number of tradeable assets. Moreover, we impose a ‘no short positions’ constraint, in order to make our results of more use to the individual, non-institutional investor. Using historic data, we construct sparse portfolios which accurately duplicate the performance of an equally weighted portfolio, which is known to be a highly effective portfolio selection strategy.

1 Introduction

Simply put, a portfolio is a collection of investments in some subset of a given set of assets. These investments can be either long (i.e. ownership of some portion of the asset) or short (borrowing of some portion of the asset). Given a fixed set of assets to invest in, a natural question is, ‘what is the optimal portfolio made out of them’?. Clearly the answer to this question depends on the individual investor, but in 1952 Markowitz ([Mar52]) provided the key insight that was to become the basis of Modern Portfolio Theory (MPT). He realized that central to an investors choice of portfolio is the balancing of risk and reward. Using this, he provided the following rule of thumb to guide investors in creating their portfolios: ‘Given several portfolios with the same expected return, the optimal one is the one with smallest standard deviation of return.’ Given N tradeable assets, with estimated returns μ_1, \dots, μ_N and correlation matrix Σ , this suggests that one fix a desired return μ^* and solve the optimization problem:

$$\mathbf{w}_{opt} = \arg \min_{\mathbf{w} \in \mathbb{R}_+^N} \mathbf{w}^T \Sigma \mathbf{w} \quad \text{subject to } \boldsymbol{\mu}^T \mathbf{w} = \mu^* \text{ and } \sum_{i=1}^N w_i = 1 \quad (1.1)$$

eq:LSMarkowitz

where $\boldsymbol{\mu} = [\mu_1, \dots, \mu_N]^T$. Unfortunately, problem (1.1) is notoriously difficult to solve in practice (see DeMiguel, Garlappi et al), essentially because it is unstable. That is, small discrepancies in the estimation of Σ can create arbitrarily large differences in \mathbf{w}_{opt} . Clearly this is computationally problematic, as in any real-world scenario Σ will be computed from empirical data, and thus liable to contain errors. In light of this, it is tempting to opt for the ‘naïve’ portfolio, \mathbf{w}_n , which allocates $1/N$ -th of the available capital to each available asset.

This approach, while generally successful ([?]) is not always ideal for small investors once transaction costs are taken into account. Indeed, if an investor is to be charged a brokerage fee of K dollars per transaction, just to set up \mathbf{w}_n will cost KN dollars, an amount likely comparable to the

*mckenzie@math.uga.edu. Department of Mathematics, University of Georgia, Athens, GA 30602.

total available capital of the investor! In this short report, we discuss how one can use ideas from signal processing to create a portfolio with only a few active positions (a ‘sparse’ portfolio) whose performance matches that of \mathbf{w}_n .

2 Related work

To the best of the author’s knowledge, the first groups to connect Optimal Portfolio Theory with Compressed Sensing were Brodie *et al.* ([BDDM⁺09]) and independently DeMiguel *et al.* ([DGNU09]), who (independently) proposed modifying (1.1) by adding an ℓ_1 -norm penalty function:

$$\mathbf{w}_{opt} = \arg \min_{\mathbf{w} \in \mathbb{R}_+^N} (\mathbf{w}^T \Sigma \mathbf{w} + \lambda \|\mathbf{w}\|_1) \quad (2.1) \quad \text{eq:LASSO}$$

$$\text{subject to } \boldsymbol{\mu}^T \mathbf{w} = \mu^* \text{ and } \sum_{i=1}^N w_i = 1 \quad (2.2) \quad \text{eq:Constraint}$$

Problems of the form (2.1), without the constraints (2.2), are known as LASSO (Least Absolute Shrinkage and Selection Operator) problems, and were introduced in the Statistics literature in 1996 by Robert Tibshirani ([Tib96]). Importantly, as noticed by Tibshirani, solutions to the LASSO tend to have few non-zero entries (they are *sparse*). It can also be shown that the LASSO problem is stable. Brodie *et al.* showed, both theoretically and experimentally, that the solution to problem (2.1) - (2.2) is a sparse, (Markowitz-) optimal portfolio, which in many cases outperforms the naive portfolio \mathbf{w}_n .

On the other hand, Compressed Sensing, a sub-field of mathematical Signal Processing, has developed many other problem formulations (as well as algorithms) akin to (2.1), which are designed to find sparse solutions to linear problems like $A\mathbf{x} = \mathbf{y}$ or $\arg \min \|A\mathbf{x} - \mathbf{y}\|_2$. However, problems where additional constraints need to be imposed (such as (2.2)) are understudied in the literature. Fortunately, a recent thesis of Kyrillidis ([Kyr14], see also [KBCK13]) studies the problem:

$$\mathbf{x}_{opt} = \arg \min f(\mathbf{x}) \quad (2.3)$$

$$\text{subject to: } \|\mathbf{x}\|_0 \leq s \quad \sum_{i=1}^N x_i = \lambda \text{ and } x_i \geq 0 \text{ for all } i \quad (2.4)$$

where $f : \mathbb{R}^N \rightarrow \mathbb{R}$ is convex and differentiable and $\|\mathbf{x}\|_0 := \#\{i : x_i \neq 0\}$. He also develops a Projected Gradient Descent (PGD) algorithm capable of efficiently approximating \mathbf{x}_{opt} . He then adapts this to the portfolio optimization problem, studying:

$$\mathbf{w}_{opt} = \arg \min \mathbf{w}^T \Sigma \mathbf{w} - \tau \boldsymbol{\mu}^T \mathbf{w} \quad (2.5) \quad \text{eq:Kyr1}$$

$$\text{subject to: } \|\mathbf{w}\|_0 \leq s \quad \sum_{i=1}^N w_i = 1 \text{ and } w_i \geq 0 \text{ for all } i \quad (2.6) \quad \text{eq:KyrConstra}$$

where τ is a tuneable parameter balancing risk and reward. Furthermore, he demonstrates experimentally (using the Fama-French 48 data set) that the portfolios obtained for $s = 4$ and $s = 10$ (i.e. 4 or 10 active positions) are surprisingly competitive, frequently outperforming the naive $1/n$ strategy and a more sophisticated quadratic programming strategy.

A third related line of research is that of *sparse index tracking*. We shall think of an index as a special kind of financial data time series. This could be the hourly/ daily/ weekly/ etc. returns of an idealized portfolio (such as the SP500 or the DJIA), commodity prices, or something more exotic such as investor sentiment ([BDDM⁺09]). In any case, given such a time series \mathbf{y} , as well as a $T \times N$ matrix of returns R for all available assets (over the same time interval as \mathbf{y} , the goal is to create a portfolio \mathbf{w}_y which tracks the performance of the index \mathbf{y} as well as possible. Mathematically, this is a simple least squares problem:

$$\mathbf{w}_y = \arg \min_{\mathbf{w}} \|\mathbf{y} - R\mathbf{w}\|_2 \quad (2.7) \quad \text{eq:Tracker}$$

For reasons similar to those outlined in the Introduction, it is often useful to add a sparsity constraint (and we shall also add a ‘no-shorts’ constraint) to problem (2.7), resulting in the (no-shorts) sparse index tracking problem:

$$\mathbf{w}_y = \arg \min_{\mathbf{w}} \|\mathbf{y} - R\mathbf{w}\|_2 \text{ subject to: } \|\mathbf{w}\|_0 \leq s \text{ and } w_i \geq 0 \quad (2.8) \quad \text{eq:SparseTrac}$$

This problem is studied thoroughly in [ZXL15], although they use a different algorithm to Kyrillidis’ PGD. We shall consider the index tracking problem with index given by the returns of the equally-weighted portfolio \mathbf{w}_n . That is: $\mathbf{y} = R\mathbf{w}_n$.

3 Problem Formulation

Before formulating the problem we wish to solve, let us introduce some notation. For each asset $i = 1, \dots, N$, let $r_{i,t}$ denote the return on asset i over the t -th time interval (week/ month/ day etc.) where t runs from 1 to T . Collect these into a vector $\mathbf{r}_t = [r_{1,t}, \dots, r_{N,t}]^\top$. Let R denote the $T \times N$ matrix with \mathbf{r}_t^\top as its t -th row. Note that given a portfolio \mathbf{w} , the t -th entry of $R\mathbf{w} \in \mathbb{R}^T$ gives the return of this portfolio over the t -th time interval. Hence, we consider the problem:

$$\mathbf{w}_s = \arg \min_{\mathbf{w}} \|R\mathbf{w}_n - R\mathbf{w}\|_2 \text{ subject to: } \|\mathbf{w}\|_0 \leq s \text{ and } w_i \geq 0 \quad (3.1) \quad \text{eq:SparseTrac}$$

We solve this problem using the PGD algorithm, which in this case amounts to a slight variation of the well-known compressed sensing algorithm Iterative Hard Thresholding (IHT). Code is available at danielmckenzie.github.io.

4 Numerical Experiment

We conducted two numerical experiments; one using the Fama-French FF48 data set ¹, consisting of 48 industry sector portfolios (such as ‘Medical Equipment’, ‘Chemicals’ or ‘Entertainment’.) and one using data for the 500 stocks making up the SP500, obtained from Yahoo! finance. Unfortunately as Yahoo! appears to no longer be supporting their Finance API, the second data set is no longer easily available. For the sake of reproducibility, we thus include only the results of the first experiment. We follow the experimental methodology of [BDDM⁺09]. For each year between 1976 and 2006, we constructed 10 portfolios with sparsities $s = 2, 4, 6, \dots, 20$ using the previous five year’s returns as training data. We then simulated holding these portfolios for one year, and calculated their final returns. Thus, we end up with a data set of returns of 30 portfolios for each sparsity level, each held for a year. As a control group, we also calculate the return on \mathbf{w}_n for each year from 1976 to 2006.

¹available at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

The percentage mean monthly returns (μ), percentage monthly risk (i.e. standard deviation, σ) and Sharpe Ratio $S = \mu/\sigma$ are tabulated below:

Sparsity	μ	σ	S
10	1.30	5.04	0.26
12	1.27	4.96	0.26
14	1.26	4.90	0.26
16	1.28	4.84	0.26
18	1.25	4.80	0.26
20	1.25	4.67	0.27
22	1.24	4.67	0.26
24	1.21	4.63	0.26
26	1.21	4.59	0.26
28	1.23	4.58	0.27
30	1.20	4.59	0.26
32	1.22	4.60	0.26
34	1.21	4.59	0.26
36	1.22	4.58	0.27
38	1.19	4.60	0.26
40	1.20	4.56	0.26
42	1.20	4.56	0.26
44	1.20	4.54	0.26
46	1.19	4.52	0.26
48	1.19	4.50	0.27
\mathbf{w}_n	1.19	4.50	0.27

Table 1: The percentage monthly returns, risk and Sharpe ratio for sparse portfolios and the evenly weighted portfolio \mathbf{w}_n .

While the mean annual return and risk is presented graphically in figure 1. Recall that the goal of this exercise is to duplicate the performance of \mathbf{w}_n using a sparse portfolio. Unsurprisingly the ‘sparse’ portfolios with 40+ active positions mimic the performance of \mathbf{w}_n very closely (in fact the portfolio with 48 active positions frequently coincides with \mathbf{w}_n). What is pleasantly surprising is that there are several portfolios with fewer than 24 active positions (i.e. less than 50% sparse) whose performance still tracks \mathbf{w}_N closely. Equally unexpected is that several of these truly sparse portfolios have a higher annual Sharpe ratio than \mathbf{w}_n , as seen in figure 1 (a higher Sharpe ratio is more desirable).

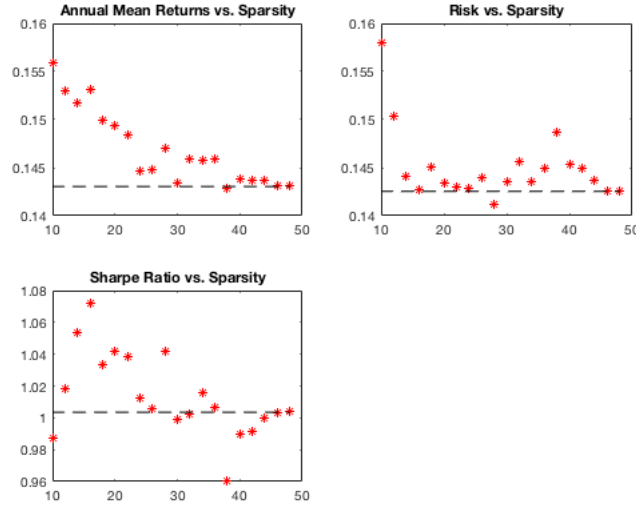


Figure 1: Annual performance of Sparse portfolios, indicated by red stars. The performance of \mathbf{w}_n is indicated by a dashed black line

5 Concluding Remarks

The results of section §4 corroborate the claim made in [BDDM⁺09] and [ZXL15] that sparse portfolios perform favourable when compared against other portfolios, at least given a fairly small ‘market’ of available securities. It would be interesting to test the behaviour of sparse portfolios found using our approach on larger data sets, as in [ZXL15]. We remark that the experimental setup (constructing an entirely new portfolio each year) is not meant to mimic investor behaviour - an investor is likely to construct a portfolio and then make minor alterations to it each year. This can also be modelled using the Compressed Sensing framework, by considering sparse portfolio adjustments, as in [BDDM⁺09]. We leave this consideration for future work.

References

- [BDDM⁺09] Joshua Brodie, Ingrid Daubechies, Christine De Mol, Domenico Giannone, and Ignace Loris. Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences*, 106(30):12267–12272, 2009.
- [DGNU09] Victor DeMiguel, Lorenzo Garlappi, Francisco J Nogales, and Raman Uppal. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812, 2009.
- [KBCK13] Anastasios Kyrillidis, Stephen Becker, Volkan Cevher, and Christoph Koch. Sparse projections onto the simplex. In *Proceedings of The 30th International Conference on Machine Learning*, pages 235–243, 2013.
- [Kyr14] Anastasios Kyrillidis. *Rigorous optimization recipes for sparse and low rank inverse problems with applications in data sciences*. PhD thesis, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, 2014.

- owitz1952] [Mar52] Harry Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.
- irani1996] [Tib96] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- zhao2015] [ZXL15] ZhiHua Zhao, FengMin Xu, and XiangYang Li. Adaptive projected gradient thresholding methods for constrained l0 problems. *Science China Mathematics*, 58(10):1–20, 2015.