A Novel Optimization Approach to Sparse Mean-Reverting Portfolios

Somaya Sadik[a], Mohamed Et-tolba[b], Benayad Nsiri[a]

[a] *Ecole Nationale Supérieure d'Arts et Métiers, Rabat, Morocco*
[b] *Institut National des Postes et Télécommunications, Rabat, Morocco*

Abstract

Mean-reverting portfolios, with fewer assets, but enough volatility are a real challenge for financial investors. Although they offer an ideal investment opportunity, they are very difficult to construct with real time data. To design such portfolios, one has to optimize their mean-reverting strength while maintaining sparsity constraints and a volatility threshold. Most of the existing approaches are framed as an eigenvector issue with a sparsity constraint. In this paper, we propose two methods to design a sparse mean-reverting portfolio. The idea is to optimize the predictability using a regularization technique that combines l_1 and l_2-norms. Computer simulations are performed on market data extracted from SP500. The obtained numerical results prove the effectiveness of the proposed methods compared with the existing approaches.

Keywords: Mean-reversion, Sparse portfolios, VAR (1) model, convex relaxations, l_p-norm

---

1. Introduction

Portfolio selection is one of the most important concepts in financial data analysis. The foundations of modern portfolio theory can be traced to Markowitz's early papers [1]. Accordingly, the investment returns should be maximized for a given degree of risk. As a result, to create an optimal portfolio, one has

Email addresses: somaya_sadik@um5.ac.ma (Somaya Sadik), ettolba@inpt.ac.ma (Mohamed Et-tolba)

to guarantee a greater return-to-risk ratio. Additionally, in the last decade, mean-reversion strategy has gained a considerable attention, especially for predictability measurement in portfolios management. It has been established that mean-reverting portfolio assets oscillate around their long-term mean [2, 3]. Hence, systematic trading strategies can be developed using portfolios with relatively predictable short-term behavior. Yet, such mean-reverting portfolios, with two or more assets, may be hard to build due to non-stationarity and combinatorial constraints. Moreover, finding the appropriate mean-reverting assets combination is insufficient for a portfolio to be realistic. What is more, its long-term returns should be able to compensate for transaction costs.

Currently, these assets are used in cointegration-based statistical arbitrage strategies [4]. This is the classical approach to pairs trading that identifies a stationary linear function of non-stationary time series. In [5], the authors presented a general survey of the cointegration analysis based on the vector autoregressive model. Their approaches use essentially statistical tests to guarantee that the time series are cointegrated. Unfortunately, such tests are very sensitive to the data's modeling assumptions. Besides, they are marginally profitable only if the capital is high, since they may incur additional costs. In other words, cointegration is optimal while working with a pair of assets, but not multi-asset baskets since it cannot properly handle too many of them [4].

To deal with the cointegration considering multi-assets portfolios, several approaches have been proposed recently. The methods proposed in [6, 7, 8] are based on the portfolio predictability analysis to measure the portfolio mean-reversion. The approaches of the authors in [6, 8] are based on the Box and Tiao canonical decomposition hypothesis [9], which models the pricing process independently from the historical data. In addition, it has been stated that the assets follow a vector autoregressive model under the condition that the portfolio's covariance matrix has to be stationary. In [7], the authors assumed that the generated portfolio follows an Orstein-Uhlenbeck process, which was then used to establish the link between the mean-reversion parameter and the predictability statistic.

In [6], the authors used the predictability ratio and defined a mean-reverting portfolio (MRP) to be the one with a low predictability, as opposed to a momentum portfolio, which has a high predictability. They modeled the MRP formulation as an optimization problem with constraints, and compared it to some alternative approaches such as portmanteau minimization test and crossing minimization statistic. They highlighted the need of a sparse MRP and recommended the addition of a minimum variance constraint to identify an optimal portfolio. In fact, the optimization problem is reformulated as a generalized eigenvalue problem with a sparsity constraint leading to NP-Hard problem. Therefore, several solutions were proposed using semi-definite relaxations. In [10, 11], the convex relaxation was altered to take into consideration additional constraints on the investment budget and leverage. Mousavi et al. [12] proposed an improved greedy approach by making it a two-steps process and adding a penalty decomposition algorithm. Despite the diversified strategies, their results have proven to be computationally expensive, and not scalable.

Taking into account the aforementioned limitations, in this paper, we propose a novel approach to sparse mean-reverting portfolio (SMRP) selection. Inspired by the method suggested in [6], our idea is to optimize the portfolio predictability while enforcing its sparsity using a combination of the l_1-norm and l_2-norm. We suggest two optimization methods to select the optimal SMRP. The first method is to assign a regularizing parameter to the l_1-norm and another one to the l_2-norm, assuming that their sum equals one. In the second method, we set two regularizing parameters that are different from each other. When applying these approaches, each one results in an optimized portfolio and regularizing parameters.

The rest of this paper is structured as follows. Section 2 introduces mean-reverting portfolios and their sparse optimization. Section 3 details the existing generalized algorithms to deal with optimal portfolios selection. Section 4 is concerned with the proposed. Section 5 gives a list of the metrics used to measure the model's performance. Section 6 presents the numerical results and their analysis. Finally, the paper is concluded in section 7.

2. Mean-Reverting Portfolios

In finance, investors make investments with the expectation of increasing financial returns. This is achieved by either buying or selling the assets before a fall in their prices occurs. The asset price, on the other hand, might be difficult to anticipate in many circumstances. In fact, for most people, the choice of the best moment to invest in an item is challenging [10].

Rather than investing in a single asset, people in statistical arbitrage strategy invest simultaneously in different assets assembled in a portfolio. Because such a portfolio and its volatility remain fixed, it is simple to select the optimum timing to invest. In practice, this volatility can be naturally stationary, manufactured using technical or fundamental analysis, or built using statistical models. As an example, a portfolio volatility made up of two security assets, denoted by y_1 and y_2 respectively is presented in figure 1.

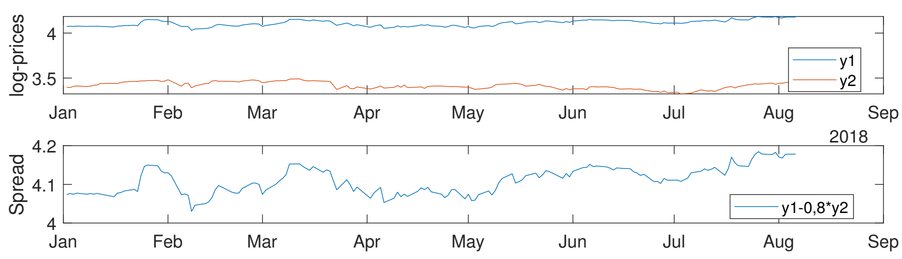

Figure 1: An example of the log-prices of two assets and their modeled spread.

Note that the log-price of a financial asset is expressed as,

$$\boldsymbol{y}_t = \log \boldsymbol{p}_t \tag{1}$$

where $\boldsymbol{p}_t$ denotes the price of a financial asset at time t.

The focus of this work is to construct an improved mean-reverting portfolio by combining these spreads. Let $\boldsymbol{s}_{i,t}$, where $i \in \{1, \ldots, n\}$, $t \in \{1, \ldots, m\}$, be the price of the i^{th} asset in the portfolio at time t, and x_i is its associated weight. Then, the portfolio value, at time t, is given by,

$$\boldsymbol{p}_t = x_i * \boldsymbol{s}_{i,t} \tag{2}$$

It is assumed that the portfolio values follow an Ornstein-Uhlenbeck process, which is expressed as,

$$d\boldsymbol{p}(t) = \lambda(\mu - \boldsymbol{p}(t))dt + \sigma \boldsymbol{w}(t) \tag{3}$$

where $\boldsymbol{w}(t)$ is a standard Brownian movement, λ denotes the mean-reversion coefficient, μ is the long-term mean, and σ represents the portfolio volatility. This model offers a straight representation of mean-reversion speed and could be used as a benchmark for mean-reversion time series. However, there is no obvious connection to the portfolio weights, which are crucial for portfolio optimization. To link this process to the portfolio weights, we use the predictability variable.

Let us assume that the asset prices follow a first order vector autoregressive VAR(1) process. We also take into consideration that the portfolio's estimated prices at time$t$ are generated using past data of the prices up to time $t-1$, which are affected by an additive white Gaussian noise (AWGN). Then, the estimated vector of prices at time t is expressed as,

$$\boldsymbol{s}_t = \boldsymbol{A}\boldsymbol{s}_{t-1} + \boldsymbol{w}_t \tag{4}$$

where $\boldsymbol{s}_t^T = (s_{1,t}, s_{2,t}, \ldots, s_{n,t})$ is the vector of prices, $\boldsymbol{A}$ is an $n * n$ matrix and $\boldsymbol{w}_t \sim N(0, \sigma I)$. By multiplying both sides, in (4), by $\boldsymbol{x}^T = (x_1, \ldots, x_n)$ whose i^{th} element represents the amount invested in the i^{th} asset, we get,

$$\boldsymbol{x}^T \boldsymbol{s}_t = \boldsymbol{x}^T \boldsymbol{s}_{t-1} \boldsymbol{A} + \boldsymbol{x}^T \boldsymbol{w}_t \tag{5}$$

Note that the negative values of $\boldsymbol{x}^T \boldsymbol{s}_t$ correspond to the case of short-selling [13].

The predictability is defined to be the statistic ratio that measures how a random process is close to the white noise. Let $\hat{\sigma}^2 = \mathrm{Var}\left(\hat{\boldsymbol{x}}_{t+1}\right)$ denotes the variance of estimated weight vector at time $t+1$, and let $\sigma^2 = \mathrm{Var}(\boldsymbol{x}_t)$ represents the variance of the past weight vector. Then, the predictability can be expressed as [9],

$$v = \frac{\mathrm{Var}(\widehat{\boldsymbol{x}}_{t+1})}{\mathrm{Var}(\boldsymbol{x}_t)} = \frac{\hat{\sigma}^2}{\sigma^2} \tag{6}$$

When v is high, $\hat{\sigma}^2$ dominates the noise, thus the process can be predicted. In the other hand, if v is low, the process prediction is difficult since it looks like a Gaussian noise. In other words, as long as predictability is low, the mean-reverting speed of the portfolio will be higher.

The predictability can be rewritten in terms of the covariance matrix of the asset prices, Γ, as follows [6],

$$v = \frac{\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{x}}{\boldsymbol{x}^T \Gamma \boldsymbol{x}} \tag{7}$$

Under the assumption that each asset price has zero mean so as to be normalized, the optimization problem becomes a minimization of the predictability. This is equivalent to solving the generalized eigenvalue problem, which is expressed as,

$$\det\left(\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^T - \lambda\boldsymbol{\Gamma}\right) = 0 \tag{8}$$

Other mean-reversion proxies. Instead of using predictability, mean-reversion can be modeled using other statistics such as Portmanteau and crossing statistics. Portmanteau statistic measures how much a process approaches a white noise. This statistic metric is given by [14]:

$$\widehat{\Phi}_{\mathrm{p}} = \frac{1}{\mathrm{p}} \sum_{\mathrm{i}=1}^{\mathrm{p}} \left(\frac{\mathbf{x}^{\mathrm{T}} \boldsymbol{\Gamma}_{\mathrm{i}} \mathbf{x}}{\mathbf{x}^{\mathrm{T}} \boldsymbol{\Gamma} \mathbf{x}} \right)^2 \tag{9}$$

where Γ_I is the lag-I autocovariance of the asset prices. This statistic promotes stronger mean-reversion in the portfolio if it is sufficiently low. The Crossing statistic evaluates the expected number of crosses around zero per unit of time of a univariate process $y(n=1)$ [15]. It is written as,

$$\gamma(y) = \mathrm{E}\left[\frac{\sum_{\mathrm{i}=2}^{\mathrm{t}} 1_{\{\mathrm{y_i y_{i-1}} \leq 0\}}}{\mathrm{t}-1}\right] \tag{10}$$

As stated by the cosine formula, if $\mathrm{y_i}$ is an autoregressive process of order one AR(1), $|a| < 1$, ϵ_i is i.i.d standard Gaussian noise and $\mathrm{y_i} = a\mathrm{y_{i-1}} + \epsilon_i$ then,

$$\gamma(y) = \frac{\arccos(a)}{\pi} \tag{11}$$

It is worth noting that the crossing statistic is only valid for stationary AR(1) process with $n = 1$ where the minimization of the first order autocorrelation can maximize the crossing rate of the process x [16]. In the case $n > 1$, we minimize the first order autocorrelation and maintain all the absolute autocorrelations small.

3. Sparse Mean-Reverting Portfolio Selection

In the previous section, we discussed how to handle a generalized eigenvalue problem and identify a portfolio that optimizes the predictability. However, our aim is to find the best portfolio vector with a strong mean-reversion under a sparseness condition. Accordingly, our optimization problem can be written as follows,

$$\text{minimize} \; \frac{\boldsymbol{x}^T \boldsymbol{A} \Gamma \boldsymbol{A}^T \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{\Gamma} \boldsymbol{x}} \; \text{s.t.} \; \|\boldsymbol{x}\|_0 = k \tag{12}$$

Where k is the sparsity rate. By adding a sparsity constraint to the generalized eigenvalue optimization, it becomes an NP-hard problem that is very difficult to solve in reality. Since an optimal solution is not available, we search for an approximate sub-optimal solution. There are three known classes of algorithms to solve this problem, namely LASSO, greedy algorithms and convex relaxations [6, 7].

Clustering via LASSO keeps the asset universe small, therefore easier to manage. Even the initial minimization problem is manageable. This method exploits the conditional dependance between asset prices in order to identify small asset clusters via LASSO. The application of greedy algorithms to the optimization process is equivalent to constructing a sparse solution from scratch. According to the numerical results presented in [7], this approach is the fastest in its computation and yields the best theoretical result in more than 50 percent of the cases in comparison to other algorithms applied to the same data set. Regarding convex optimization, it is appropriate to relax the l_0-norm to anl_1-norm constraint. It makes the problem more complex but also gives much needed leeway in the control of the portfolio's volatility.

3.1. Clustering via LASSO

LASSO clustering is a sparsifying method for the predictability minimization problem. It ensures the dimension reduction of the investing universe by considering only the combinations found in the clusters for building the portfolio [17]. Figure 2 presents the bloc diagram of LASSO clustering

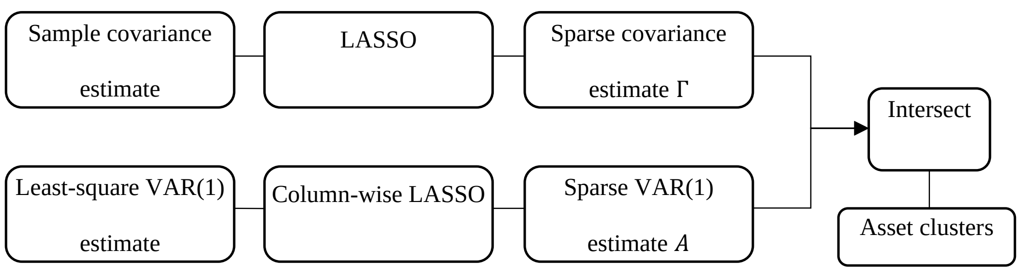

Figure 2: Steps of clustering asset via LASSO.

The LASSO is applied to both covariance selection and VAR model estimation to get the sparse estimates. Their intersection provides the asset clusters containing the assets to invest in.

3.1.1. Covariance estimate

In this bloc, the inverse covariance matrix is estimated using LASSO based on l_1-norm and penalized maximum likelihood. This estimation is a maximization problem, which is formulated as,

$$\max_{x} \log \det \boldsymbol{x} - \mathrm{Tr}(\boldsymbol{\Sigma x}) - \alpha \|\boldsymbol{x}\|_1 \tag{13}$$

where α is a parameter that represents the number of zeros in the final sparse estimation. The larger α is, the sparser the estimation of the covariance will be.

3.1.2. Structured VAR estimate

Recall that the evolution of the asset prices follows a VAR(1) model, given in (4). The least square estimate of $\boldsymbol{A}$ is given by the following minimization problem:

$$\arg\min_{A} \|\boldsymbol{s}_t - \boldsymbol{s}_{t-1}\boldsymbol{A}\|^2 \tag{14}$$

This optimization problem can be made sparse by adding an l_1-norm penalty. The strength of LASSO is its ability to modify the final structure of the sparse estimate of A by appending alternative forms ofl_1-norm. We may use a column-wise univariate LASSO to maximize the following objective function for every asset a:

$$\begin{gathered} \arg\min_{a} \|s_{it} - s_{t-1}a\|^2 + \lambda \|a\|_1 \\ \text{for every } i \in \{1, \ldots, n\} \end{gathered} \tag{15}$$

A multi-task LASSO can be also applied to get the sparse estimate of $\boldsymbol{A}$ by setting its entire columns to zero:

$$\arg\min_{A} \|\boldsymbol{s}_t - \boldsymbol{s}_{t-1}\boldsymbol{A}\|^2 + \lambda \sum_{i} \sqrt{\sum_{j} a_{ij}^2} \tag{16}$$

where a_{ij} is an element of the matrix $\boldsymbol{A}$.

3.1.3. Intersected clusters

Once the sparse estimates of the covariance matrix and the VAR coefficient are generated, the objective will be grouping the two clusters into one subset of data existing in them both. The mean-reverting portfolio selection can, afterwards, be made using components from the subset of the found data [18].

3.2. Greedy Algorithms

Greedy algorithms are used for their fast implementation of portfolio optimization [6]. The key idea is to build the portfolio recursively based on the previous local optimal solution. Let $I_k = \{i \in [1,n] \mid x_i \neq 0\}$ be the set of indices belonging to the k non-zero components of x. When $k = 1$, we set I_1 as

$$I_1 = \arg\min_{i\in[1,n]} \frac{\left(A^T \mathbf{\Gamma} \boldsymbol{A}\right)_{ii}}{\mathbf{\Gamma}_{ii}} \tag{17}$$

Once we have the first term, the recursion goes through the rest of the k indices. For every i not in I_k, the minimization problem is equivalent to finding the vector corresponding to the smallest eigenvalue of the matrix $\boldsymbol{E}$ given by,

$$\mathbf{E} = \mathbf{\Gamma}^{-1/2} \left(\boldsymbol{A}^T \Gamma \boldsymbol{A}\right)^T \Gamma \left(\boldsymbol{A}^T \Gamma \boldsymbol{A}\right) \mathbf{\Gamma}^{-1/2} \tag{18}$$

The index i of the smallest expected eigenvalue is then added to I_k. The iteration is repeated until the cardinality constraint is satisfied, i.e. $(i = k)$. Although this method is somewhat faster to implement, its main disadvantage is that many asset combinations are missed through the recursive portfolio construction. Therefore, the final result may not be the most optimal [19].

3.3. Convex Relaxations

The standard minimization techniques cannot be directly applied to mean-reverting portfolio optimization problem since it is non-convex and contains a cardinality condition. This issue has been addressed in [8] using semi-definite programming. They have first transformed the problem into a convex semi-definite optimization program (SDP), and then solved it using a minimum eigenvalue solver. To do this, the l_0-norm was relaxed to an l_1-norm constraint. By introducing the matrix $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^T$, the portfolio selection problem is rewritten as:

$$
\begin{gathered}
\min \frac{\operatorname{Tr}\left(\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^T\boldsymbol{X}\right)}{\operatorname{Tr}(\boldsymbol{\Gamma}\boldsymbol{X})} \\
s \cdot t. \|\boldsymbol{X}\|_1 \leq k \\
\operatorname{Tr}(\boldsymbol{X}) = 1 \\
| \ \boldsymbol{X} \geq 0
\end{gathered} \tag{19}
$$

The l_1-norm is the sum of the absolute values of all elements in the matrix $\boldsymbol{X}$. $\operatorname{Tr}(\boldsymbol{X}) = 1$ is added as an extra constraint to help with numerical stability. It is equivalent to normalizing the weight vector to a unit vector. Nonetheless, the minimization function is a ratio of two matrix traces, which makes it quasi-convex and not solvable by convex optimization methods. To cope with this problem, the l_1-norm can be added as a regularization instead of a constraint. The optimization problem is then given as,

$$
\begin{gathered}
\min \operatorname{Tr}\left(\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^T\boldsymbol{X}\right) + \rho\|\boldsymbol{X}\|_1 \\
\text{s.t.} \ \ \operatorname{Tr}(\boldsymbol{\Gamma}\boldsymbol{X}) \geq \sigma^2 \\
\operatorname{Tr}(\boldsymbol{X}) = 1 \\
\boldsymbol{X} \geq 0
\end{gathered} \tag{20}
$$

In this version, the numerator of the ratio in (19) is minimized while the denominator is kept over a threshold. Sparsity is induced by the addition of the l_1-norm penalty. Knowing that $\operatorname{Tr}(\boldsymbol{\Gamma}\boldsymbol{X})$ expresses the volatility of the portfolio, this optimization problem guarantees the minimization of the portfolio predictability with a sufficient volatility. Another advantage of this form is the flexibility and interpretability regarding the constraints gained [16]. The cardinality constraint is applied at various stages of the algorithm to produce the closest approximation to the optimal solution of this non-convex problem. Unfortunately, all of these algorithms barely estimate the optimal weight vector and are computationally intensive. Moreover, in some rare instances, only the convex relaxation method may exhibit optimum convergence [20]. Hereafter, we provide two new formulations of the MRP optimization problem as well as a process for computing the weight vector that gives satisfactory results.

4. New formulations of the MRP with a l_1-norm and l_2-norm combination

4.1. l_1-norm and l_2-norm description

To solve the MRP optimization problem, we propose to combine an l_1-norm and l_2-norm added to the convex relaxation algorithm. These measurements are known to enforce sparsity, which is the main objective of this work. Generally, an l_p norm is expressed as,

$$l_p : \|x\|_{l_p} = \sqrt[p]{\sum_i |x|^p} \tag{21}$$

Note that the l_1-norm is the mostly used in solving sparsity problems since it offers the sparsest solution, but we can't neglect the l_2-norm because it also offers a sparse option to choose from. To clearly explain the process for solving different norms, we assume that the optimization problem is two-dimensional and use the graphics in figure 3 to expose the solutions for the l_0-norm, l_1-norm and l_2-norm minimizations. Let $\boldsymbol{S}$ represent all possible solutions x^* given by,

$$\boldsymbol{S} = \{\boldsymbol{x}^* : \boldsymbol{y} = \boldsymbol{\Phi x}\} \tag{22}$$

where $\boldsymbol{y}$ is a measurement vector, $\boldsymbol{x}$ is the acquired data vector and $\boldsymbol{\Phi}$ is the measurement matrix.

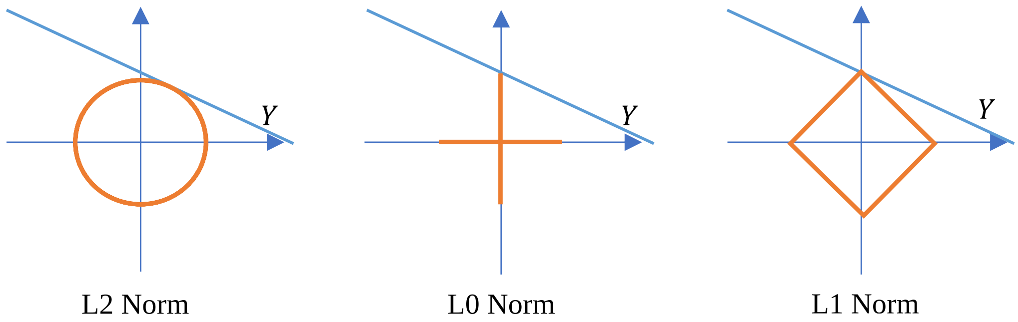

Figure 3: l_0, l_1 and l_2 norm solutions representation.

The equation system (22) is expressed by a blue line in a two-dimensional space in figure 3. Each norm solutions results in a different type of graph; however, the sparsest solutions are the intersection points of the line $\boldsymbol{S}$ and the graphs l_0, l_1 and l_2. The fact that the sparsest solutions are those localized on the coordinate axis, shows the advantage of choosing between the l_0-norm, l_1-norm even the l_2-norm produces a sparse solution. Unfortunately, the l_0-norm is non-computable so the l_1-norm is the mostly used in sparsity inducing algorithms [21].

4.2. MRP reformulated with one regularizing parameter

Our MRP reformulation will be based on the model developed in [6]. It rewrites the portfolio selection problem into a manageable algorithm by applying convex relaxations as shown in equation (16). We extend this model to add the l_2-norm as a regularization instead of a constraint in addition to the l_1-norm. the regularization with one parameter is given by,

$$\begin{aligned} \min \operatorname{Tr}\left(\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^T\boldsymbol{X}\right) + \mu\|\boldsymbol{X}\|_1 + (1-\mu)\|\boldsymbol{X}\|_2 \\ \text{s.t. } \operatorname{Tr}(\boldsymbol{\Gamma}\boldsymbol{X}) \geq \sigma^2 \\ \operatorname{Tr}(\boldsymbol{X}) = 1 \\ \boldsymbol{X} \geq 0 \end{aligned} \tag{23}$$

This becomes a minimization problem with both l_1-norm and l_2-norm as regularizations and a minimum variance constraint. To solve it, we use the historical data to estimate the price process knowing that it is a VAR(1) model. Once we have the estimates, we can extract the autoregressive covariance $\boldsymbol{A}$ and the estimated covariance matrix $\boldsymbol{\Gamma}$. These two matrices are the main inputs for the proposed algorithms. The following algorithm in table1 is used to solve the minimization problem,

Table 1: Algorithm 1 for solving the reformulated MRP with one regularizing parameter.

Solve (23)
Input $\boldsymbol{A}$, $\boldsymbol{\Gamma}$
Parameter μ
Set $\boldsymbol{X}_0$ to the equal weighted portfolio, introduce the constraints
For μ from 0 to 1 increment of 0.01
Solve the quadratic problem (23) to find the local minimum $\boldsymbol{X}_{opt}$
If $\boldsymbol{X}_{opt}$ is sparser than $\boldsymbol{X}_0$
Update $\boldsymbol{X}_0 = \boldsymbol{X}_{opt}$
End if
End for

4.3. MRP reformulated with two regularizing parameters

Our second MRP design differs from the first in that it combines l_1-norm and l_2-norm using two regularizing parameters. The optimization problem is then reformulated as,

$$\begin{aligned} &\min \operatorname{Tr}\left(\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^T\boldsymbol{X}\right) + \theta_1|\boldsymbol{X}\|_1 + \theta_2\|\boldsymbol{X}\|_2 \\ &\text{s.t. } \operatorname{Tr}(\boldsymbol{\Gamma}\boldsymbol{X}) \geq \sigma^2 \\ &\operatorname{Tr}(\boldsymbol{X}) = 1 \\ &\boldsymbol{X} \geq 0 \end{aligned} \tag{24}$$

where θ_1 and θ_2 are the regularizing parameters. It is worth mentionning that the use of two regularizing parameters increases the optimization computational complexity compared to the case of one regularizing parameter. To find the optimum weight's vector, we use the following algorithm in table2:

Table 2: Algorithm 2 for solving the reformulated MRP with two regularizing parameters.

Solve (24)
Input $\boldsymbol{A}$, $\boldsymbol{\Gamma}$
Parameters θ_1 and θ_2
Set $\boldsymbol{X}_0$ to the equal weighted portfolio, introduce the constraints
For θ_1 from 0 to 1 increment of 0.01
For θ_2 from 0 to 1 increment of 0.01
Solve the quadratic problem (24) to find the local minimum $\boldsymbol{X}_{opt}$
If $\boldsymbol{X}_{opt}$ is sparser than $\boldsymbol{X}_0$
Update $\boldsymbol{X}_0 = \boldsymbol{X}_{opt}$
End if
End for
End for

5. Performance Metrics

In order to measure the performance of the proposed algorithms, we use the three common metrics namely cumulative profit and loss (PL), return on investment (ROI) and sharp ratio (SR) [12].

5.1. Cumulative Profit and Loss

This metric is defined as the cumulative return of a mean-reverting portfolio (MRP) in one trading period from t_1 to t_2. We first define the profit and loss of an MRP as follows,

$$PL_t \triangleq \boldsymbol{x}_p^T * \boldsymbol{r}_t \tag{25}$$

Let $\boldsymbol{r}_t$ be the asset's returns given by,

$$\boldsymbol{r}_t = y_t - y_{t-1} = \log \boldsymbol{p_t} - \log \boldsymbol{p}_{t-1} \tag{26}$$

where $\boldsymbol{p_t}$ are the prices and $\boldsymbol{x}_p$ is the weights vector. The profit and loss (in dollars) of an investment in a portfolio for a single holding cycle is evaluated by the PL. if we want to determine the cumulative return performance, the cumulative PL can be defined as,

$$\text{Cum. } PL\,(t_1, t_2) \triangleq \sum_{t_1}^{t_2} P\&L_t \tag{27}$$

To further explain, the PL_t depends on whether we are holding a long or short position regarding the mean price. If it is a long position, opened at time t_a and closed at time t_b, the PL at time $t \in [t_a, t_b]$ is:

$$PL_t = \boldsymbol{x}_p^T * \boldsymbol{r}_t * (t - t_a) - \boldsymbol{x}_p^T * \boldsymbol{r_{t-1}} * (t - 1 - t_a) \tag{28}$$

If it is a short position instead, we have

$$PL_t = \boldsymbol{x}_p^T * \boldsymbol{r}_{t-1} * (t - 1 - t_a) - \boldsymbol{x}_p^T * \boldsymbol{r}_t * (t - t_a) \tag{29}$$

5.2. Return On Investment

Return on investment is another portfolio return metric that we implement. At time t, the ROI is the single-period PL normalized by the gross investment made, which is $\|\boldsymbol{x}_p\|_1$. It is written as:

$$ROI_t = \frac{PL_t}{\|\boldsymbol{x}_p\|_1} \tag{30}$$

5.3. Sharpe Ratio

The Sharpe ratio as introduced in [22], can be used for measuring risk-adjusted return. It determines the amount of additional profit that can be gained in exchange for additional volatility. The SR of ROI for trading time from t_1 to t_2 is defined as follows:

$$SR_{ROI}\left(t_1, t_2\right) = \sqrt{252} * \frac{\mu_{ROI}}{\sigma_{ROI}} \tag{31}$$

where

$$\mu_{ROI} = 1/\left(t_2 - t_1\right) \sum_{t_1}^{t_2} ROI_t \tag{32}$$

and

$$\sigma_{ROI} = \left[1/\left(t_2 - t_1\right) \sum_{t_1}^{t_2} \left(ROI_t - \mu_{ROI}\right)^2\right]^{\frac{1}{2}} \tag{33}$$

6. Numerical results

In this section, we evaluate the effectiveness of the two proposed methods for selecting mean-reverting portfolios with sufficient volatility and minimal sparsity from a universe of tradable assets. In order to analyze its performance, we applied a trading strategy designed specifically for mean-reverting processes.

The numerical simulations on real market data are carried on 49 selected stocks picked from the U.S stock market. The best representative of this stock market is the Standard and Poor's 500 Index (SP500) from which we selected the best 49 stocks in terms of weight factor in the index. The daily prices of the selection are retrieved from yahoo finance for the trading period from January 2, 2014 to December 28, 2018. The benchmark used for comparison is the equal weighted portfolio of these 49 assets, where each asset $i \in \{1, \ldots, 49\}$ has the same weight $w_i = 1/49$. In figure 4, we represent its performance during the trading period. In the short term (30 to 90 days), the prices' evolution is sufficently mean-reverting to make profits.

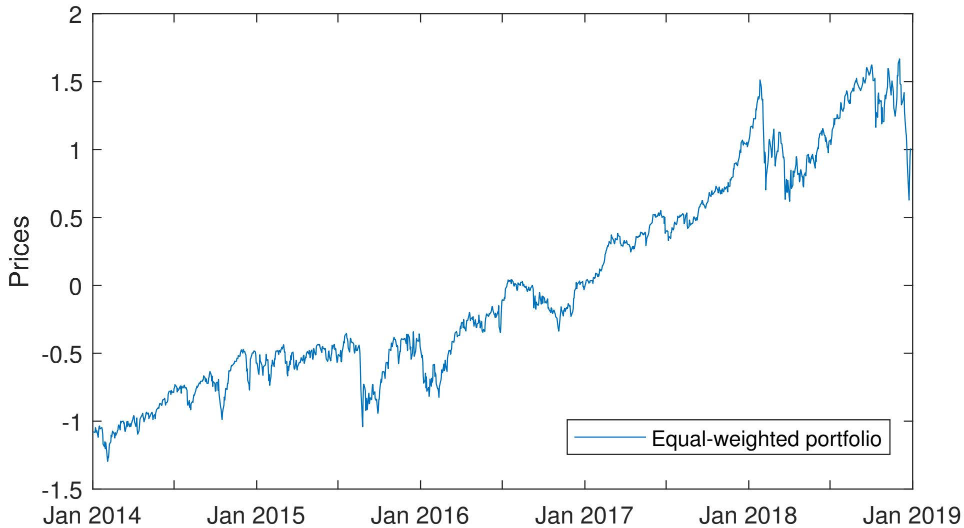

Figure 4: Equal weighted portfolio's performance during the trading time period.

To compare the models presented in section 4, we generate s_t the prices process. It is a VAR(1) model estimated by using the daily log-prices of the 49 selected stocks. We can then extract the autoregressive covariance matrix $\boldsymbol{A}$ and the estimated covariance matrix $\boldsymbol{\Gamma}$. These two matrices, whose size is $49 * 49$, are then used to compute the studied algorithms. The minimum variance is determined based on the idea given in [8], such that it should be greater than one fifth of the median variance of all assets in the pool. In figure 5, the estimated VAR(1) prices were simulated as an equal weighted portfolio. We can already see an improvement in the case of the portfolio's mean-reversion compared to the benshmark.

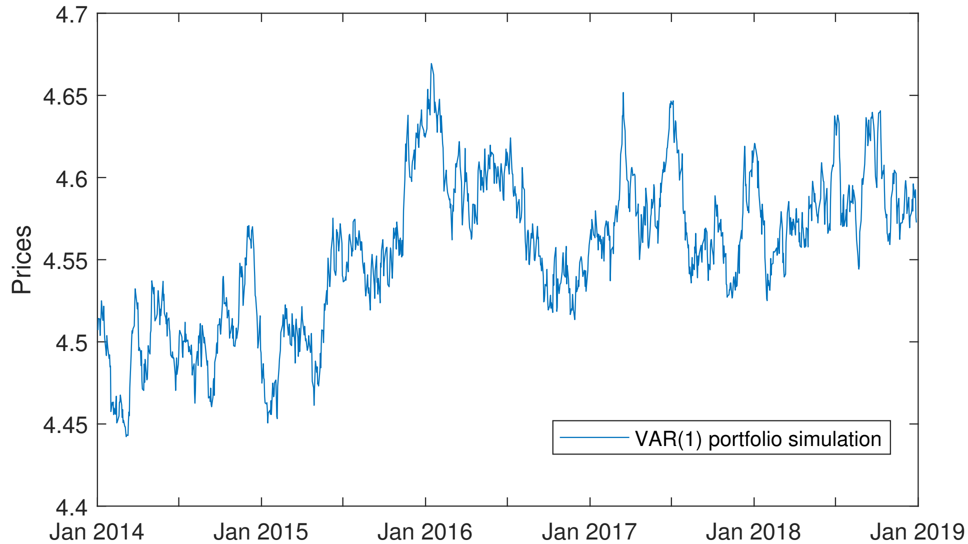

Figure 5: Estimated VAR(1) prices' simulation during the trading time period.

6.1. Designed Models Comparison

330 We test our models in this part by using the daily closing prices of the retrieved data. The equal weighted portfolio (MRP-eqw) was generated to be a benchmark for its advantages. Three MRP designs are applied, consisting first, of the model of minimizing predictability (MRP-pre) presented in [6], the proposed MRP design with one regularizing parameter (MRP-des1) and the second MRP design with two regularizing parameters (MRP-des2). Figure 6 shows the prices evolution after the estimation process. The portfolio's value increases in the case of MRP-des1 with time, while MRP-pre and MRP-des2 have a close evolution.

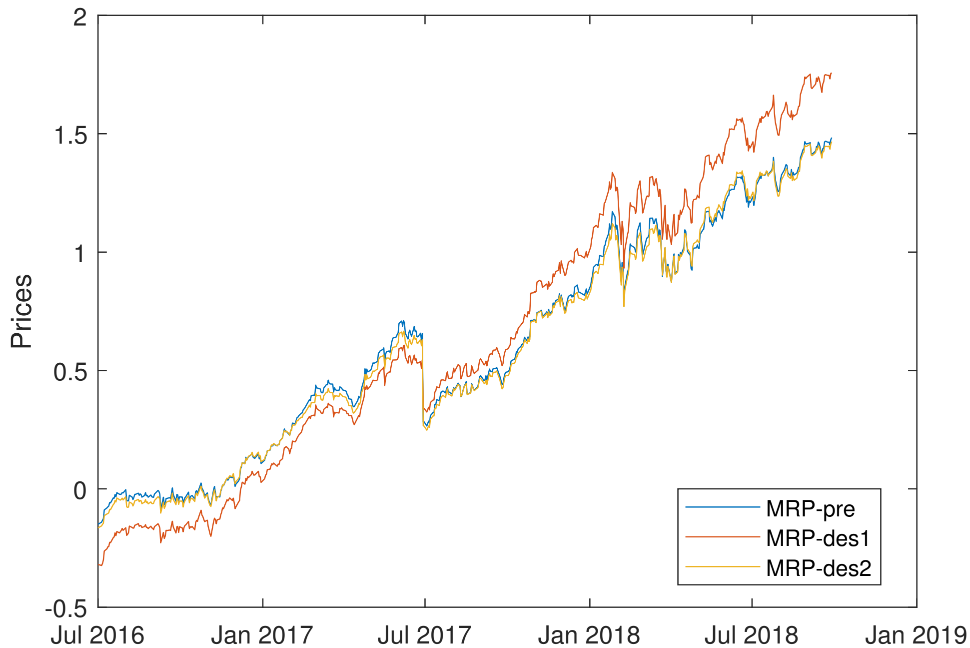

Figure 6: Prices evolution for MRP-pre, MRP-des1 and MRP-des2 after estimation during the trading time period.

In figure 7 and table 3, the performance of our developed MRPs are compared to MRP-pre model. Out-of-sample results such as ROIs, Sharpe ratios of ROIs, CPU time and cumulative PLs are presented. We can see clearly that the first designed MRP model (MRP-des1) can get a stronger Sharpe Ratio with a low CPU time. Moreover, its final cumulative returns is relatively high compared to the others. However, even if the average ROI of MRP-des1 is the highest, it is still low for gaining fast and easy profits from the portfolio. Conclusively, all the performance metrics show that the adventage of the designed model with one parametrizing parameter is better exposed when the investment's value is high and its period is long.

Table 3: CPU time (in seconds), Sharpe Ratio and average ROI for MRP-pre, MRP-des1 and MRP-des2 during the trading time period.

Model	**CPU time (in seconds)**	**Sharpe ratio**	**AverageROI**
MRP-pre	196.5918	3.4147	2.8%
MRP-des1	307.5506	3.5031	3.7%
MRP-des2	4290.6465	2.9249	3.0%

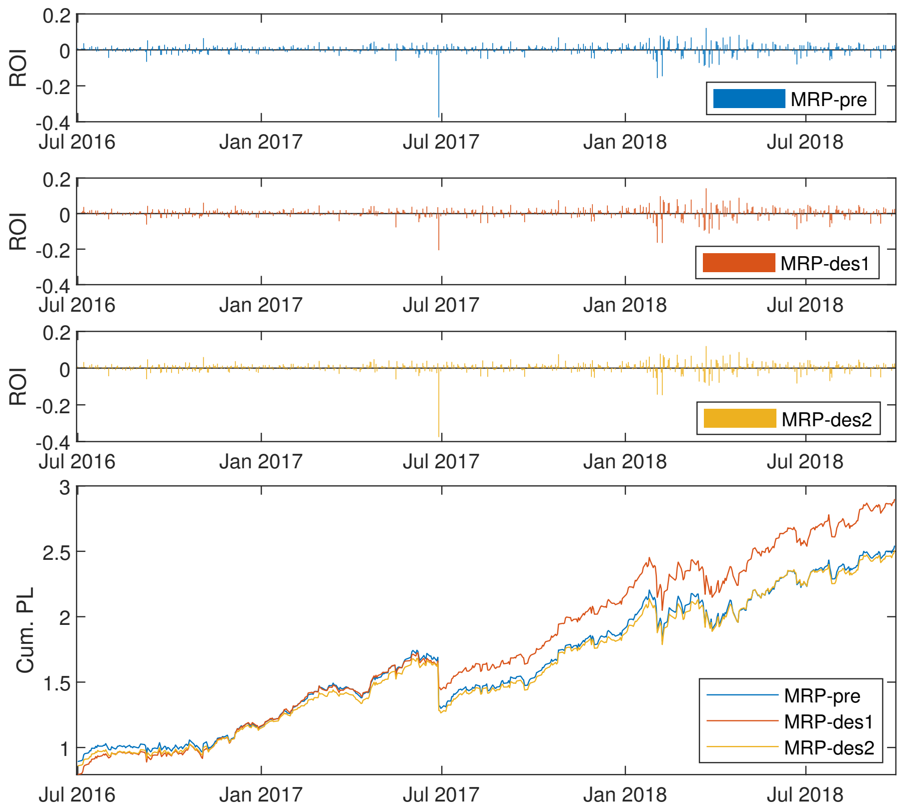

Figure 7: Estimated ROIs and cumulative PLs for MRP-pre, MRP-des1 and MRP-des2 during the trading time period.

We also compare our MRP-des1 and MRP-eqw models' performance. Figure 8 shows that our proposed MRP model with one regularizing parameter performs better than the benchmark method with better Sharpe Ratio and greater final cumulative returns.

Table 4: Sharpe Ratio and average ROI for MRP-eqw and MRP-des1 during the trading time period.

Model	**Sharpe ratio**	**AverageROI**
MRP-eqw	3.2683	3.1%
MRP-des1	3.5031	3.7%

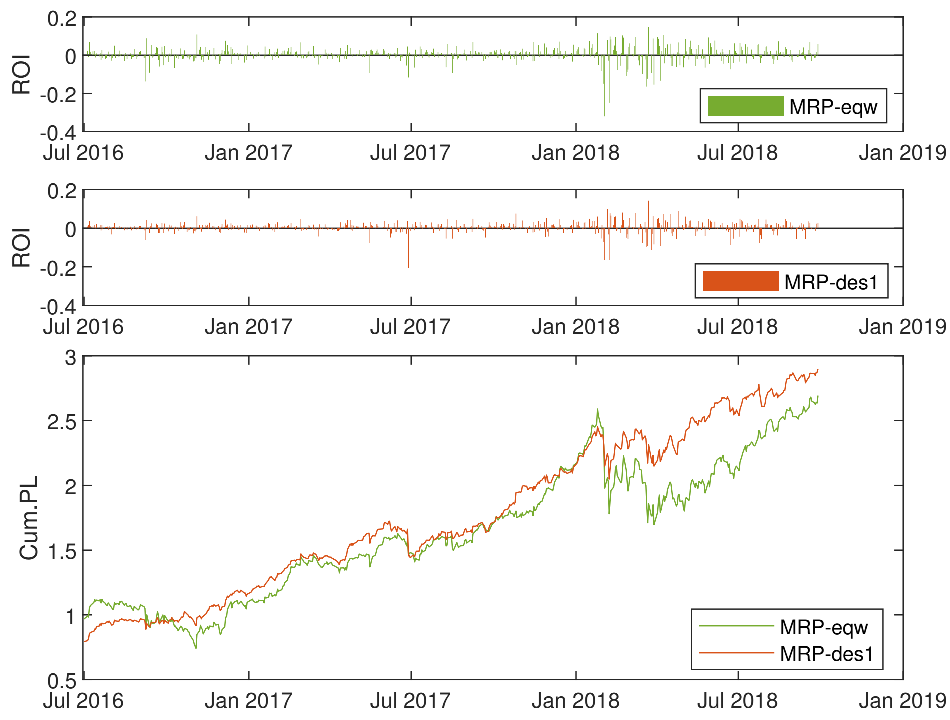

Figure 8: Estimated ROIs and cumulative PLs for MRP-eqw and MRP-des1 during the trading time period.

7. Conclusion

In this paper, we proposed two new methods for constructing optimal sparse mean-reverting portfolios. Our approaches aim to extract a sparse subset of assets from a larger pool of time series so it can be computed as a mean-reverting linear combination. For this, they combine l_1 and l_2 norms with one or two regularization parameters. Additionally, the proposed designs deal with optimal portfolios selection even with more than two assets. Numerical results show that the proposed methods provide better performance in terms of several metrics. It is also noticed that the use of one regularizing parameter improves mean-reverting trading strategies.

References

[1] H. Markowitz, Portfolio selection (1952).

[2] E. F. Fama, K. R. French, Permanent and temporary components of stock prices (1988).
URL `https://www.jstor.org/stable/1833108`

[3] J. M. Poterba, L. H. Summers, Mean reversion in stock prices. evidence and implications, Journal of Financial Economics 22 (1988) 27–59. `doi:10.1016/0304-405X(88)90021-9`.

[4] S. Johansen, Interpretation of cointegrating coefficients in the cointegrated vector autoregressive model, Oxford Bulletin of Economics and Statistics 67. `doi:10.1111/j.1468-0084.2005.00111.x`.

[5] S. Johansen, Cointegration: Overview and development (2009). `doi:10.1007/978-3-540-71297-8_29`.

[6] A. D'aspremont, Identifying small mean-reverting portfolios, Quantitative Finance 11. `doi:10.1080/14697688.2010.481634`.

[7] N. Fogarasi, J. Levendovszky, Sparse, mean reverting portfolio selection using simulated annealing, Algorithmic Finance 2 (2013) 197–211. doi:10.3233/AF-13026.

[8] M. Cuturi, A. d'Aspremont, Mean-reverting portfolios: Tradeoffs between sparsity and volatility.
URL http://arxiv.org/abs/1509.05954

[9] G. E. Box, G. C. Tiao, A canonical analysis of multiple time series, Biometrika 64. doi:10.1093/biomet/64.2.355.

[10] Z. Zhao, D. P. Palomar, Mean-reverting portfolio with budget constraint, IEEE Transactions on Signal Processing 66. doi:10.1109/TSP.2018.2799193.

[11] Z. Zhao, R. Zhou, D. P. Palomar, Optimal mean-reverting portfolio with leverage constraint for statistical arbitrage in finance, Vol. 67, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 1681–1695. doi:10.1109/TSP.2019.2893862.

[12] A. Mousavi, J. Shen, A penalty decomposition algorithm with greedy improvement for mean-reverting portfolios with sparsity and volatility constraints.
URL http://arxiv.org/abs/2104.07887

[13] T. Griveau-Billion, B. Calderhead, Efficient computation of mean reverting portfolios using cyclical coordinate descent.
URL http://arxiv.org/abs/1905.05841

[14] G. M. Ljung, G. E. Box, On a measure of lack of fit in time series models, Biometrika 65. doi:10.1093/biomet/65.2.297.

[15] E. McKenzie, B. Kedem, Time series analysis by higher order crossings., Journal of the Royal Statistical Society. Series A (Statistics in Society) 159. doi:10.2307/2983188.

[16] M. Cuturi, A. D'Aspremont, Mean reversion with a variance threshold, 2013.

[17] Y. Koike, De-biased graphical lasso for high-frequency data, Entropy 22. doi:10.3390/E22040456.

[18] F. G. Duarte, L. N. D. Castro, A framework to perform asset allocation based on partitional clustering, IEEE Access 8. doi:10.1109/ACCESS.2020.3001944.

[19] N. Fogarasi, J. Levendovszky, Improved parameter estimation and simple trading algorithm for sparse, mean reverting verting portfolios, Selected papers of 9th Joint Conference on Mathematics and Computer Science MaCS 2012 February 9-12, 2012, Siófok, Hungary 37.

[20] A. Dehghan, B. Lcsee, S. Hassan, A. Lcsee, P. W. Lcsee, Critique-mean reverting portfolio.

[21] X. Zhao, W. Li, M. Zhang, R. Tao, P. Ma, Adaptive iterated shrinkage thresholding-based lp-norm sparse representation for hyperspectral imagery target detection, Remote Sensing 12. doi:10.3390/rs12233991.

[22] W. F. Sharpe, The sharpe ratio (fall 1994) (2021). doi:10.1515/9781400829408-022.