# Split Bregman iteration for multi-period mean variance portfolio optimization

Stefania Corsaro [a], Valentina De Simone [b,*], Zelda Marino [a]

[a] Dipartimento di Studi Aziendali e Quantitativi, Università di Napoli "Parthenope", Via Generale Parisi, 13, Napoli I-80133, Italy
[b] Dipartimento di Matematica e Fisica, Università della Campania "Luigi Vanvitelli", Viale Lincoln, 5, Caserta I-81100, Italy

**A R T I C L E   I N F O**

**A B S T R A C T**

This paper investigates the problem of defining an optimal long-term investment strategy, where the investor can exit the investment before maturity without severe loss. Our setting is a multi-period one, where the aim is to make a plan for allocating all of wealth among the $n$ assets within a time horizon of $m$ periods. In addition, the investor can rebalance the portfolio at the beginning of each period. We develop a model in Markowitz context, based on a fused lasso approach. According to it, both wealth and its variation across periods are penalized using the $l_1$ norm, so to produce sparse portfolios, with limited number of transactions. The model leads to a non-smooth constrained optimization problem, where the inequality constraints are aimed to guarantee at least a minimum level of expected wealth at each date. We solve it by using split Bregman method, that has proved to be efficient in the solution of this type of problems. Due to the additive structure of the objective function, the alternating split Bregman at each iteration yields to easier subproblems to be solved, which either admit closed form solutions or can be solved very quickly. Numerical results on data sets generated using real-world price values show the effectiveness of the proposed model.

© 2020 Elsevier Inc. All rights reserved.

## 1. Introduction

We consider the regularized multi-period mean variance optimization problem and its solution using the split Bregman method.

In recent years there has been a growing interest in the solution of fused lasso models [1], that are linearly constrained minimization of functions given as:

$$E(\mathbf{w}) = f(\mathbf{w}) + \tau_1 \|\mathbf{w}\|_1 + \tau_2 \|L\mathbf{w}\|_1, \tag{1}$$

where $f: \Re^n \to \Re$ is a closed convex function at least twice continuously differentiable, and $L$ is the difference operator. The $l_1$ penalty in (1) promotes sparse solutions, and the term $\|L\mathbf{w}\|_1$ is included to produce smooth solutions. Problems described by fused lasso models arise, e.g., in image processing [2], classification [3], finance [4]. The nonsmoothness of the $l_1$-type regularization terms precludes the use of standard descent methods for smooth objective functions. Problems of this kind can be solved either by smoothing the $l_1$ terms, e.g., [5,6], and applying optimization solvers for differentiable problems

---

* Corresponding author.

*E-mail addresses:* stefania.corsaro@uniparthenope.it (S. Corsaro), valentina.desimone@unicampania.it (V. De Simone), zelda.marino@uniparthenope.it (Z. Marino).

such as gradient methods [7–9] or by using directly optimization solvers for nondifferentiable problems, such as Bregman, proximal and ADMM methods [2,10–12]. Due to the additive structure in (1), splitting methods have became popular because they yield algorithms which consist at each iteration of subproblems that are easier to solve [13,14]. These subproblems often either admit closed form solutions or can be solved very quickly with specialized methods. In this context, methods based on Split Bregman iteration have proved to be efficient in different fields [2,15–17]. At each iteration of the Bregman method an $l_1$ regularized unconstrained optimization subproblem is solved. Auxiliary variables allow to separate the two $l_1$ regularization terms making the use of splitting methods an easy task. In the field of portfolio optimization, both in the static and the dynamic case, Bregman iteration has been used to solve efficiently $l_1$ regularized models in Markowitz's framework. The use of $l_1$ penalty terms in portfolio modelling has become popular for several reasons. Transaction costs are well described by the $l_1$ norm of the portfolio, especially in moderate size trades. The properties of this norm allow one to obtain sparse solutions, that is, small portfolios. Small portfolios are often considered preferable to large ones. They are feasible to small investors, for the reduced holding cost. Moreover, it seems that the estimation errors for variances and covariances are reduced in this case [18,19]. Finally, it has been observed that the application of $l_1$ regularization has the effect of penalizing short positions [20,21], forbidden in several markets. The structure of Bregman iteration has been exploited both in the static and the dynamic case, to develop procedures which adaptively fix the value of the regularization parameter; this value is chosen as one that provides solutions with certain financial properties, while preserving fidelity to data [21,22].

In [4] a model based on the fused lasso approach is presented in a multi-period setting. This seems to be very promising, since the fusion term is a penalty on portfolio turnover, which produces strategies with low transaction cost by preserving the pattern of active position over time. The fused lasso model is also considered in this paper. The novelty is the focus on strategies in which the investor can exit before the end without incurring severe loss; this is particularly significant to protect investment from financial market crisis. At this purpose, control at intermediate periods on wealth is introduced in the model. According to the multi-period approach, the so called rebalancing dates are defined to divide the invest-ment period into subperiods; at these dates, the investment strategy is revised according to evolving information. We fix a minimum level of expected wealth at intermediate dates. From the mathematical point of view, this leads to a nonsmooth optimization problem with equality and inequality constraints. We reformulate the problem as one with equality constraints only and apply the alternating Split Bregman method for its solution. The resulting algorithm requires, at each iteration, the solution of unconstrained subproblems which are easily solved exactly.

In order to test our model, we introduce some perfomance measures. In particular, we assess performance with respect to two benchmarks built on real market data; this task is usually referred to as the Enhanced Index Tracking problem [23,24], where one aims at selecting a portfolio that outperforms a reference index.

In Section 2 we describe the mathematical model. In Section 3 we present the procedure for the numerical solution of the model, based on alternating Split Bregman iteration. In Section 4 we show results of our tests.

## 2. Mathematical model

In this section we extend the fused lasso model presented in [4] with the aim of guaranteeing the investor in the case of early exit. The model refers to either a medium or long-term investment, in which the investor could exit before the term; thus, at this purpose, the expected minimum wealth is fixed at each rebalancing date.

Let $m$ be the number of rebalancing dates, $n$ the number of assets and $(\mathbf{w}_j)_i$ the portion of the investor's total wealth invested in security $i$ at time $j$; the vector

$$\mathbf{w} = [\mathbf{w}_1^T, \ldots, \mathbf{w}_m^T]^T \in \Re^N, \tag{2}$$

with $N = m \cdot n$, defines the portfolio. At each rebalancing date $j = 1, \ldots, m$, the wealth is thus given by $\mathbf{w}_j^T \mathbf{1}_n$, where $\mathbf{1}_n$ is the column vector with $n$ elements, all equal to one. We assume that $j = 1$ is the starting date. We denote with $\mathbf{r}_j \in \Re^n$ the expected return vector and with $C_j \in \Re^{n \times n}$ the covariance matrix, that we assume positive definite. The model is stated as follows:

$$\min_{\mathbf{w}} \ \sum_{j=1}^m \left[ \mathbf{w}_j^T C_j \mathbf{w}_j + \tau_1 \|\mathbf{w}_j\|_1 \right] + \tau_2 \sum_{j=1}^{m-1} \|\mathbf{w}_{j+1} - \mathbf{w}_j\|_1$$

$$\begin{aligned}
\text{s.t.} \quad & \mathbf{w}_1^T \mathbf{1}_n = \xi_{\text{init}} \\
& \mathbf{w}_j^T \mathbf{1}_n = \left(\mathbf{1}_n + \mathbf{r}_{j-1}\right)^T \mathbf{w}_{j-1}, \quad j = 2, \ldots, m \\
& \mathbf{w}_j^T \mathbf{1}_n \geq (\mathbf{w}_{\text{min}})_{j-1}, \quad j = 2, \ldots, m \\
& (\mathbf{1}_n + \mathbf{r}_m)^T \mathbf{w}_m \geq (\mathbf{w}_{\text{min}})_m
\end{aligned} \tag{3}$$

where $\xi_{\text{init}}$ is the initial wealth, $\mathbf{w}_{\text{min}}$ is the vector of expected minimum wealth.

The quadratic term in the objective function represents the portfolio risk. This is obtained by summing single-period variances. The successive terms are regularization ones introduced by fused lasso approach. The $l_1$ norm is well-known to promote sparsity. We apply it to $\mathbf{w}$, weighted by $\tau_1 > 0$, in order to have small portfolios. This improves the control on the investment and reduces holding cost. Moreover, we have observed that it penalizes negative components, thus resulting in a penalty on shorting [21,22]. This turns out to be useful when short positions are not allowed, since one can obtain positive

weights by properly tuning $\tau_1$. The fusion term, that is the $l_1$ norm applied to the time variation $\Delta\mathbf{w}$, has a smoothing effect on solution, preserving patterns of non-zero values across time. Its financial interpretation is a penalty on the portfolio turnover; it has then the effect of reducing transaction cost [4]. This term is weighted by $\tau_2 > 0$. The values $\tau_1, \tau_2$ are referred to as the regularization parameters.

Both equality and inequality constraints are imposed on wealth. The first wealth must be equal to the initial investment, as the first constraint establishes. Constraints from 2 to $m$ state the self-financing property, that is, money is neither added to the portfolio nor withdrawn from it for $j > 1$. Thus, at the end of each period the wealth is given by the revaluation of the previous one. Inequality constraints state that a lower bound is defined on the portfolio wealth at the end of each subperiod, that is, at all rebalancing dates. Finally, the last component $(\mathbf{w}_{\min})_m$ is the minimum wealth expected by the overall investment. It is given by the revaluation of the last weight sequence produced by the investment strategy on the last rebalancing date. For sake of notation simplicity, we formulate problem (3) in compact form. We introduce the matrix $C \in \Re^{N \times N}$ as the $m \times m$ diagonal block matrix $C = diag(C_1, C_2, \ldots, C_m)$. The blocks are the covariance matrices estimated at the beginning of each subperiod. The matrix $C$ is symmetric positive definite (SPD) with $\left(1 - \frac{1}{m}\right)\%$ sparsity degree. Let $L \in \Re^{(N-n) \times N}$ be the discrete difference operator; it can be viewed as the $(m-1) \times n$ upper bidiagonal block matrix, with blocks of dimension $n \times n$, defined by:

$$diag(L) = (-I, -I, \ldots, -I)$$
$$supdiag(L) = (I, I, \ldots, I)$$

where $I$ is the identity matrix of dimension $n$. $L$ has $\left(1 - \frac{2}{N}\right)\%$ sparsity degree. The $m \times m$ lower bidiagonal block equality constraint matrix $A$, with blocks of dimension $1 \times n$ and $\left(1 - \frac{2}{m}\right)\%$ sparsity degree, is defined as follows:

$$diag(A) = \left(\mathbf{1}_n^T, \mathbf{1}_n^T, \ldots, \mathbf{1}_n^T\right)$$
$$subdiag(A) = \left(-(\mathbf{1}_n + \mathbf{r}_1)^T, \ldots, -(\mathbf{1}_n + \mathbf{r}_{m-1})^T\right).$$

Finally, let $G$ be the inequality constraint matrix; it is a $m \times m$ upper bidiagonal block matrix with blocks of dimension $1 \times n$ and $\left(1 - \frac{1}{N}\right)\%$ sparsity degree where

$$diag(G) = \left(\mathbf{0}_n^T, \mathbf{0}_n^T, \ldots, \mathbf{0}_n^T, (\mathbf{1}_n + \mathbf{r}_m)^T\right),$$
$$supdiag(G) = \left(\mathbf{1}_n^T, \mathbf{1}_n^T, \ldots, \mathbf{1}_n^T\right).$$

Then, problem (3) admits the following compact formulation:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\mathbf{w}^T C\mathbf{w} + \tau_1\|\mathbf{w}\|_1 + \tau_2\|L\mathbf{w}\|_1 \\ \text{s.t.} \quad & A\mathbf{w} = \mathbf{b} \\ & G\mathbf{w} \geq \mathbf{w}_{\min} \end{aligned} \tag{4}$$

where $\mathbf{b} = (\xi_{\text{init}}, 0, 0, \ldots, 0)^T \in \Re^m$.

## 3. Split Bregman iteration

We solve (4) using the split Bregman scheme. Therefore, we first provide a short description of it. It is based on Bregman iteration method for convex optimization problems with equality constraints

$$\min_{\mathbf{u}} E(\mathbf{u}) \quad \text{s.t.} H(u) = 0. \tag{5}$$

The main idea of Bregman iteration is to transform the problem into a sequence of easier ones which are constructed by adding a "cost-to-move" term to the original objective function [14]. This term penalizes the distance between two iterates defining the Bregman distance at point $\mathbf{u}$, as:

$$D_E^{\mathbf{p}}(\mathbf{v}, \mathbf{u}) = E(\mathbf{v}) - E(\mathbf{u}) - <\mathbf{p}, \mathbf{v} - \mathbf{u}>, \tag{6}$$

where $\mathbf{p} \in \partial E(\mathbf{u})$ is a subgradient in the subdifferential of $E$. Note that if $E$ is smooth, then (6) is the difference between $E(\mathbf{v})$ and the first-order Taylor expansion of $E$ at $\mathbf{v}$. Bregman iteration applied to problem (5) produces the following scheme:

$$\begin{cases} \mathbf{u}_{k+1} = \text{argmin}_{\mathbf{u}} D_E^{\mathbf{p_k}}(\mathbf{u}, \mathbf{u}_k) + \lambda H(\mathbf{u}), \\ \mathbf{p}_{k+1} = \mathbf{p}_k - \lambda \nabla H(\mathbf{u}_{k+1}) \in \partial E(\mathbf{u}_{k+1}), \end{cases} \tag{7}$$

with $\lambda > 0$. The split Bregman method was introduced in [2], where authors proposed to use an auxiliary variable $\mathbf{d}$ before applying Bregman iteration to solve problem (5). The introduction of $\mathbf{d}$ is aimed at replacing the original problem with an equivalent one in which the smooth and nonsmooth portions of objective function are separated. Then, a further constraint is added, which forces the equality between $\mathbf{d}$ and the nonsmooth term.

In order to apply split Bregman method to problem (4) the first step is to reformulate it in terms of equality constraints only. We introduce the slack variable $\mathbf{s} \in \Re^m$ in order to transform the inequality constraint in (4) into an equality one. We

rewrite the minimization problem using the indicator function $\mathcal{I}_D(\mathbf{s})$ to incorporate the non-negativity constraint on the slack variable into the objective function:

$$
\begin{aligned}
\min_{\mathbf{w,s}} \quad & \frac{1}{2}\mathbf{w}^T C\mathbf{w} + \tau_1\|\mathbf{w}\|_1 + \tau_2\|L\mathbf{w}\|_1 + \mathcal{I}_D(\mathbf{s}) \\
\text{s.t.} \quad & A\mathbf{w} = \mathbf{b} \\
& G\mathbf{w} - \mathbf{s} = \mathbf{w_{min}},
\end{aligned}
\tag{8}
$$

where

$$
D = \{\mathbf{s} \in \Re^m \;:\; \mathbf{s} \geq \mathbf{0}\}.
$$

According to split Bregman method, we introduce the auxiliary variables $\mathbf{d}$ and $\mathbf{z}$ such that $\mathbf{d} = L\mathbf{w}$ and $\mathbf{z} = \mathbf{w}$. Problem (8) is then reformulated in the following way:

$$
\begin{aligned}
\min_{\mathbf{w,s,d,z}} \quad & \frac{1}{2}\mathbf{w}^T C\mathbf{w} + \tau_1\|\mathbf{z}\|_1 + \tau_2\|\mathbf{d}\|_1 + \mathcal{I}_D(\mathbf{s}) \\
s.t. \quad & A\mathbf{w} = \mathbf{b} \\
& G\mathbf{w} - \mathbf{s} = \mathbf{w_{min}} \\
& L\mathbf{w} = \mathbf{d} \\
& \mathbf{w} = \mathbf{z}.
\end{aligned}
\tag{9}
$$

Alternating split Bregman method proposes to minimize the function in (9) with respect to the variables $\mathbf{w}$, $\mathbf{s}$, $\mathbf{d}$ and $\mathbf{z}$ alternately. Note that this algorithm is equivalent to other well-known methods such as the Douglas-Rachford splitting and the alternating direction method of minimizers [14]. Due to linearity of constraints in (9), a simplified version of split Bregman iteration can be used. In this version, the Bregman vectors allow one to use the function $E$ rather than its Bregman distance [25]. This leads to the following simplified alternating minimization algorithm:

$$
\begin{aligned}
\mathbf{w}_{k+1} &= \mathbf{argmin_w}\, Q(\mathbf{w}; \mathbf{s}_k, \mathbf{d}_k, \mathbf{z}_k) \\
\mathbf{s}_{k+1} &= \mathbf{argmin_s}\, \mathcal{I}_D(\mathbf{s}) + \frac{\lambda_2}{2}\|s - G\mathbf{w}_{k+1} - \mathbf{b}_k^{\mathbf{s}}\|^2 \\
\mathbf{d}_{k+1} &= \mathbf{argmin_d}\, \tau_2\|\mathbf{d}\|_1 + \frac{\lambda_3}{2}\|d - L\mathbf{w}_{k+1} - \mathbf{b}_k^{\mathbf{d}}\|^2 \\
\mathbf{z}_{k+1} &= \mathbf{argmin_z}\, \tau_1\|\mathbf{z}\|_1 + \frac{\lambda_4}{2}\|z - \mathbf{w}_{k+1} - \mathbf{b}_k^{\mathbf{z}}\|^2 \\
\mathbf{b}_{k+1}^{\mathbf{w}} &= \mathbf{b}_k^{\mathbf{w}} + \lambda_1(b - A\mathbf{w}_{k+1}) \\
\mathbf{b}_{k+1}^{\mathbf{s}} &= \mathbf{b}_k^{\mathbf{s}} + \lambda_2(G\mathbf{w}_{k+1} - \mathbf{w_{min}} - \mathbf{s}_{k+1}) \\
\mathbf{b}_{k+1}^{\mathbf{d}} &= \mathbf{b}_k^{\mathbf{d}} + \lambda_3(L\mathbf{w}_{k+1} - \mathbf{d}_{k+1}) \\
\mathbf{b}_{k+1}^{\mathbf{z}} &= \mathbf{b}_k^{\mathbf{z}} + \lambda_4(\mathbf{w}_{k+1} - \mathbf{z}_{k+1}).
\end{aligned}
\tag{10}
$$

with quadratic function $Q$ defined as

$$
\begin{aligned}
Q(\mathbf{w}; \mathbf{s}_k, \mathbf{d}_k, \mathbf{z}_k) = {} & \frac{1}{2}\mathbf{w}^T C\mathbf{w} + \frac{\lambda_1}{2}\|A\mathbf{w} - \mathbf{b}_k^{\mathbf{w}}\|_2^2 + \frac{\lambda_2}{2}\|\mathbf{s}_k - G\mathbf{w} - \mathbf{b}_k^{\mathbf{s}}\|_2^2 \\
& + \frac{\lambda_3}{2}\|\mathbf{d}_k - L\mathbf{w} - \mathbf{b}_k^{\mathbf{d}}\|_2^2 + \frac{\lambda_4}{2}\|\mathbf{z}_k - \mathbf{w} - \mathbf{b}_k^{\mathbf{z}}\|_2^2.
\end{aligned}
\tag{11}
$$

Closed form solutions can be obtained for minimization with respect to $\mathbf{s}$, $\mathbf{d}$ and $\mathbf{z}$. Minimization with respect to $\mathbf{d}$ and $\mathbf{z}$ can be done efficiently using soft thresholding operator, defined as:

$$
(S(\mathbf{x}, \gamma))_i = \frac{\mathbf{x}_i}{|\mathbf{x}_i|} \max(|\mathbf{x}_i| - \gamma, 0),
$$

where $\mathbf{x}$ is real vector and $\gamma > 0$, while the proximal mapping of the indicator function on a given set is the orthogonal projection operator onto the same set. Regarding the quadratic minimization with respect to $\mathbf{w}$, we note that at each step $k$ the optimal value can be obtained by solving the system

$$
H\mathbf{w} = \mathbf{rhs}_k,
\tag{12}
$$

with

$$
H = C + \lambda_1 A^T A + \lambda_2 G^T G + \lambda_3 L^T L + \lambda_4 I
\tag{13}
$$

and

$$
\mathbf{rhs}_k = \lambda_1 A^T \mathbf{b}_k^{\mathbf{w}} + \lambda_2 G^T(\mathbf{s} - \mathbf{b}_k^{\mathbf{s}}) + \lambda_3 L^T(\mathbf{d} - \mathbf{b}_k^{\mathbf{d}}) + \lambda_4(\mathbf{z} - \mathbf{b}_k^{\mathbf{z}}).
\tag{14}
$$

We observe that the coefficient matrix defined in (13) does not depend on the iteration and it is SPD and sparse, as shown in the next proposition.

**Proposition 1.** *Matrix H defined in (13) is a SPD, $m \times m$ tridiagonal block matrix with blocks of dimension n.*

**Proof.** *H is the sum of C and I, that are SPD, and of other matrices that are semi-positive definite; then the first statement follows. We analyze the structure of matrices in (13). L, A and G are bidiagonal block matrices with blocks of dimension $1 \times n$, then for each of them the product of its transpose by the matrix itself is a tridiagonal block matrix with blocks of dimension $n \times n$. It follows that the sum has a block tridiagonal structure.* $\square$

In the next proposition we give bounds on the eigenvalues of H.

**Proposition 2.** *Let $\lambda(H)$ be the set of eigenvalues of matrix H; we have $\lambda(H) \subseteq [\chi^-, \chi^+]$ with*

$$\chi^- = \lambda_{\min}(C) + \lambda_4$$
$$\chi^+ = \lambda_{\max}(C) + \lambda_1 \lambda_{\max}(A^T A) + \lambda_2 \lambda_{\max}(G^T G) + \lambda_3 \lambda_{\max}(L^T L) + \lambda_4.$$

**Proof.** *We recall that if A and B are real, symmetric matrices, then $A + B$ has real eigenvalues, and the following inequalities hold:*

$$\lambda_{min}(A) + \lambda_{min}(B) \le \lambda_{min}(A + B) \le \lambda_{max}(A + B) \le \lambda_{max}(A) + \lambda_{max}(B).$$

*The matrix H is the sum of real and symmetrix matrices, then*

$$\chi^- \le \lambda_{min}(H) \le \lambda_{max}(H) \le \chi^+,$$

*with*

$$\chi^- = \lambda_{min}(C) + \lambda_1 \lambda_{min}(A^T A) + \lambda_2 \lambda_{min}(G^T G) + \lambda_3 \lambda_{min}(L^T L) + \lambda_4$$
$$\chi^+ = \lambda_{max}(C) + \lambda_1 \lambda_{max}(A^T A) + \lambda_2 \lambda_{max}(G^T G) + \lambda_3 \lambda_{max}(L^T L) + \lambda_4.$$

*The matrices A, G and L have rank smaller then N, so, for each one of them, the product between the matrix and its transpose is rank deficient. This completes the proof.* $\square$

Proposition 1 suggests to use sparse Cholesky factorization to solve the systems (12). Sparse direct methods are a combination of techniques from numerical linear algebra, graph theory, graph algorithms, permutations, and other topics in discrete mathematics. They exploit the sparsity of a matrix to solve problems much faster and using far less memory than if all the entries of a matrix were stored and took part in explicit computations (see [26,27] and references therein). We note that the factorization is computed only once, while at each step two triangular linear systems are solved. The resulting method is outlined in Algorithm 1.

---

**Algorithm 1** Alternating Split Bregman for Portfolio Optimization.

---

Choose $\lambda_1, \lambda_2, \lambda_3, \lambda_4 > 0$;
Choose $\mathbf{w}_0 = \mathbf{0}, \mathbf{s}_0 = \mathbf{0}, \mathbf{d}_0 = \mathbf{0}, \mathbf{z}_0 = \mathbf{0}$
Choose $\mathbf{b}_0^{\mathbf{w}} = \mathbf{0}, \mathbf{b}_0^{\mathbf{s}} = \mathbf{0}, \mathbf{b}_0^{\mathbf{d}} = \mathbf{0}, \mathbf{b}_0^{\mathbf{z}} = \mathbf{0}$
Compute $H$ according to (13)
Compute Cholesky factorization of $H$
**for** $k = 0, 1, 2, \ldots$ **do**
    $\mathbf{w}_{k+1} = H^{-1}\mathbf{rhs}_k$
    $\mathbf{s}_{k+1} = \max(\mathbf{0}, G\mathbf{w}_{k+1} + \mathbf{b}_k^{\mathbf{s}})$
    $\mathbf{d}_{k+1} = \mathcal{S}(L\mathbf{w}_{k+1} + \mathbf{b}_k^{\mathbf{d}}, \tau_2/\lambda_3)$
    $\mathbf{z}_{k+1} = \mathcal{S}(\mathbf{w}_{k+1} + \mathbf{b}_k^{\mathbf{z}}, \tau_1/\lambda_4)$
    Update $\mathbf{b}_{k+1}^{\mathbf{w}}, \mathbf{b}_{k+1}^{\mathbf{s}}, \mathbf{b}_{k+1}^{\mathbf{d}}, \mathbf{b}_{k+1}^{\mathbf{z}}$ according to (10)
**end for**

---

## 4. Numerical experiments

In this section, we show some results of tests that we perform on real market data. Algorithm 1 is applied to several data sets generated using real-world price values provided in [28]. The datasets contain weekly return time series of assets belonging to several major stock markets across the world, cleaned from errors as much as possible. We simulate 10 years investment strategies, where the investor revises decisions once a year. In Table 1 we summarise information on the datasets.

The problem requires an estimate of the covariance matrix. It is well known that the sample covariance matrix is affected by estimation error; this is particularly severe when the number of stocks, which is the dimension of the matrix, is large relative to the number of historical returns, that is the sample size. We thus apply the linear shrinkage estimator proposed in [29] to C. This method acts on the eigenvalues of C, reducing their dispersion by shrinking them towards their grand

**Table 1**
Some characteristics of the datasets.

|   | Data set | Label | # of assets | Time interval |
|---|----------|-------|-------------|---------------|
| 1 | Dow Jones Industrial | DowJones | 28 | Feb1990-Apr2016 |
| 2 | NASDAQ 100 | NASDAQ100 | 82 | Nov2004-Apr2016 |
| 3 | FTSE 100 | FTSE100 | 83 | Jul2002-Apr2016 |
| 4 | S&P 500 | SP500 | 442 | Nov2004-Apr2016 |
| 5 | NASDAQ Composite | NASDAQComp | 1203 | Feb2003-Apr2016 |
| 6 | Fama and French 49 | FF49 | 49 | Jul1969-Jul2015 |

**Table 2**
Condition number of matrix $C$: effect of the shrinkage.

| Data Set | no shrinkage | shrinkage |
|----------|--------------|-----------|
| DowJones | $2.1053E+04$ | $2.1431E+03$ |
| NASDAQ100 | $7.0926E+20$ | $1.0375E+04$ |
| FTSE100 | $2.0993E+21$ | $2.4305E+04$ |
| SP500 | $2.4678E+23$ | $3.6621E+05$ |
| NASDAQComp | $1.2587E+23$ | $7.6804E+05$ |
| FF49 | $3.5266E+07$ | $6.3079E+04$ |

mean: as a consequence, it strongly improves the conditioning of the covariance matrix, as Table 2 shows. In this table we report the condition number of the sample matrix and the condition number of its estimator.

We compare optimal portfolios produced by the investment strategy with two benchmarks. The first one is the *naive* strategy, that extends the classical $1/n$ [30] to the multi-period case. The investor splits the money evenly among available assets, at the beginning of the investment as well as at rebalancing dates. Thus, at each rebalancing date we set as expected minimum wealth the expected value produced by the recursive application of the $1/n$ allocation strategy:

$$\begin{aligned} (\mathbf{w_{min}})_1 &= \tfrac{\xi_{\text{init}}}{n}\mathbf{1}_n^T(1+\mathbf{r}_1) \\ (\mathbf{w_{min}})_i &= \tfrac{(\mathbf{w_{min}})_{i-1}}{n}\mathbf{1}_n^T(1+\mathbf{r}_{i-1}), \quad i=2,\ldots,m \end{aligned} \tag{15}$$

Moreover, for datasets $1-5$ in Table 1 also weekly return time series for the index are reported in [28]. So for this datasets the market index is also used as benchmark. In this case the expected minimum wealth is the one of the market index:

$$\mathbf{w_{min}} = \mathbf{r}_{index} \tag{16}$$

where $\mathbf{r}_{index} \in \mathcal{R}^m$ is the index return.

We introduce six performance measures to evaluate the goodness of results:

**M1** we estimate the risk reduction when the naive strategy is taken as benchmark, by means of the following quantity:

$$RR = \frac{\mathbf{w}_{naive}^T C \mathbf{w}_{naive}}{\mathbf{w}_{opt}^T C \mathbf{w}_{opt}},$$

where the numerator is the variance of the portfolio produced by the naive strategy, the denominator is the variance of the optimal portfolio produced by Algorithm 1;

**M2** the number of nonzero elements in the solution (percentage), that gives an estimation of holding cost;

**M3** the transaction costs. In order to evaluate them we define the matrix $V \in \Re^{n \times (m-1)}$, with:

$$V_{i,j} = \begin{cases} 0 & \text{if } |(\mathbf{w}_{j+1})_i - (\mathbf{w}_j)_i)| > 0 \\ 1 & \text{otherwise} \end{cases}$$

for $i=1,\ldots,n$ and $j=1,\ldots,m-1$. Note that, as in [4], in order to discard variations not significant from the financial point of view, we neglect the differences below $10^{-6} * \xi_{\text{init}}$. The number of estimated transactions of the optimal strategy is therefore:

$$T_{opt} = \sum_{i=1}^{n} \sum_{j=1}^{m-1} V_{i,j}.$$

We then relate $T_{opt}$ to the maximum value that it can assume, that is, to the number of transactions of the portfolio with full turnover, which is the size $N$ of the portfolio:

$$\mathcal{T} = \frac{T_{opt}}{N}; \tag{17}$$

**Table 3**
Condition number of matrix $H$.

| Data Set | H |
|---|---|
| DowJones | $4.4016E + 02$ |
| NASDAQ100 | $5.3936E + 03$ |
| FTSE100 | $2.0430E + 03$ |
| SP500 | $4.8893E + 04$ |
| NASDAQComp | $7.0823E + 05$ |
| FF49 | $1.0054E + 03$ |

**M4** the excess return (ER) calculated in comparison to the benchmark:

$$ER = \frac{\mathbf{w}_m^T(\mathbf{1}_n + \mathbf{r}_m) - (\mathbf{w_{min}})_m}{(\mathbf{w_{min}})_m};$$

**M5** the Sharpe Ratio (SR) [31] is the ratio between the average of the expected return of the portfolio and its standard deviation. In the multi-period framework we compute the SR in the following way:

$$SR = \frac{\frac{1}{m}\sum_{j=1}^{m} R_j}{\sigma(\mathbf{R})},$$

where $\mathbf{R} = (R_1, \ldots . R_m)$, and

$$R_j = \mathbf{r}_j^T \frac{\mathbf{w}_j}{\mathbf{w}_j^T \mathbf{1}_n}, \quad j = 1, \ldots, m;$$

**M6** the Information Ratio (IR) [32] is the average excess return per unit of volatility in excess return. We compute it as:

$$IR = \frac{\frac{1}{m}\sum_{j=1}^{m} AER_j}{\sigma(\mathbf{AER})},$$

where $\mathbf{AER} = (AER_1, \ldots, AER_m)$, and

$$AER_j = \frac{\mathbf{w}_j^T \mathbf{1}_n - (\mathbf{w_{min}})_j}{(\mathbf{w_{min}})_j}, \quad j = 1, \ldots, m.$$

We assume that one unit of wealth is invested, so we fix $\xi_{init} = 1$. We set $\lambda_i = 1, \forall i = 1, \ldots, 4$. The stopping criterion is based on constraint violation, so the algorithm stops when all constraints are satisfied within $Tol = 10^{-6}$. In Algorithm 1 we set a lower bound on $\mathbf{w}_{min}$ to guarantee the investment:

$$\mathbf{w_{min}} = \max(\bar{\xi}, \mathbf{w_{min}}). \tag{18}$$

In our experiments we set $\bar{\xi} = \xi_{init} \cdot \mathbf{1}_n$, to preserve the initially invested amount.

In Table 3 we report the condition number of $H$. We observe that in almost cases the conditioning of $H$ is better with respect to $C$. This is motivated by the shift of the spectrum to the right, according to Proposition 2.

Regularization parameters $\tau_1$ and $\tau_2$ are chosen in the set $\{10^{-4}, 10^{-3}, 10^{-2}\}$; the combined choice of them depends on the specific metric one focuses on.

We start by analyzing the behaviour of the wealth produced by the optimum portfolio over time, and compare it with the benchmarks. Figs. 1 and 2 show the values of wealth at rebalanging dates on datasets FTSE 100 and Dow Jones, but similar results are observed also for the other datasets. The first row of figures refers to dataset FTSE 100, compared with the naive strategy. The second row of the figure refers to dataset DowJones, compared with the index return. In Fig. 1 parameter $\tau_2$ has been set to $10^{-3}$ while $\tau_1$ values range in the set $\{10^{-4}, 10^{-3}, 10^{-2}\}$ from left to right. According to the analysis in [4] for both the datasets we have that increasing $\tau_1$ results in a reduction of the number of short positions, from 102 to 0 for FTSE 100 and from 30 to 0 for Dow Jones, and the number of transactions, from 15% to 9% for FTSE 100 and from the 23% to the 16% for Dow Jones. This is due to the reduced size of the portfolio, as result of the sparsification effect of regularization on portfolio wealth. Focusing on wealth, Fig. 1 shows that the excess return is penalized as $\tau_1$ increases. Indeed, our model outperforms the benchmark, showing an expected wealth higher than the one of the index, in almost all cases, but the difference is higher for smaller values of $\tau_1$, while the gain in terms of cost reduction is valuable. All experiments exhibit a breakdown in the benchmark at the fourth year of the simulation period. This is probably due to the fact that $t = 4$ corresponds to year 2008, so the lower return can be ascribed to the Crisis. Thus, at this date the excess return of the optimal portfolio is higher, as effect of (18).

In Fig. 2 the behaviour of wealth over time is shown for $\tau_1$ equal to $10^{-3}$ and for $\tau_2$ varying in $\{10^{-4}, 10^{-3}, 10^{-2}\}$. For both the datasets, we have, as expected, a reduction on the number of transactions, from 22% to 5% for FTSE 100 and from 47% to 10% for Dow Jones. Limitation on transactions results in a slight increase in risk, and thus higher return at intermediate periods is observed.
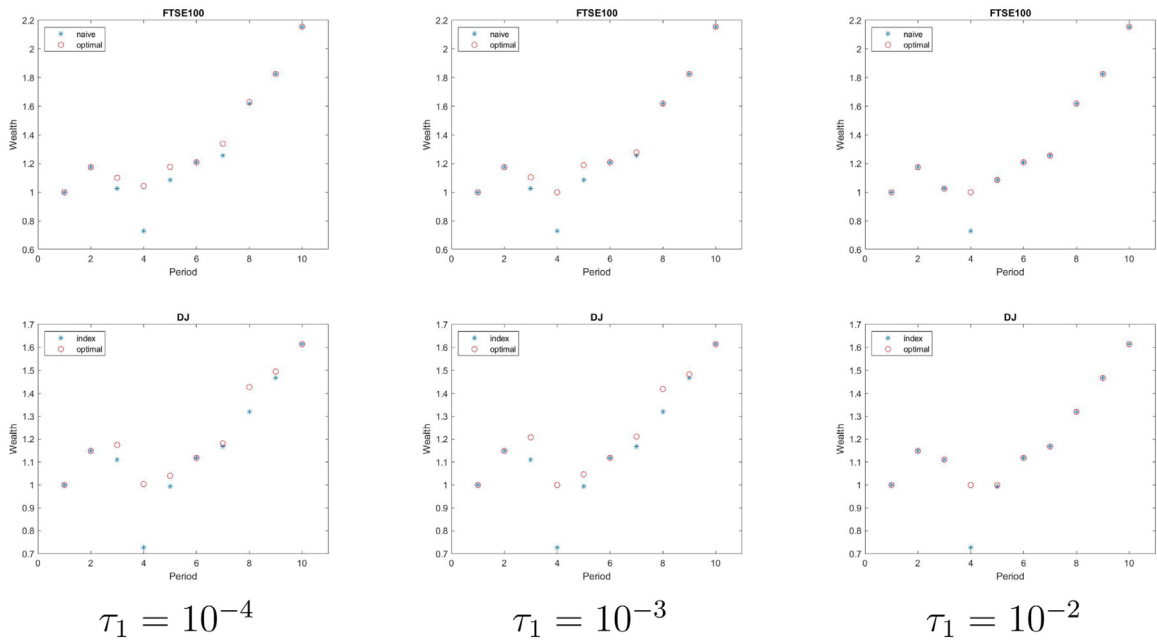
**Fig. 1.** Behaviour of the wealth over time. First row: FTSE 100 compared with the naive strategy. Second row: DowJones compared with the index return. In all cases $\tau_2 = 10^{-3}$, $\tau_1$ ranges among $10^{-4}$ (left), $10^{-3}$ (center), $10^{-2}$ (right).
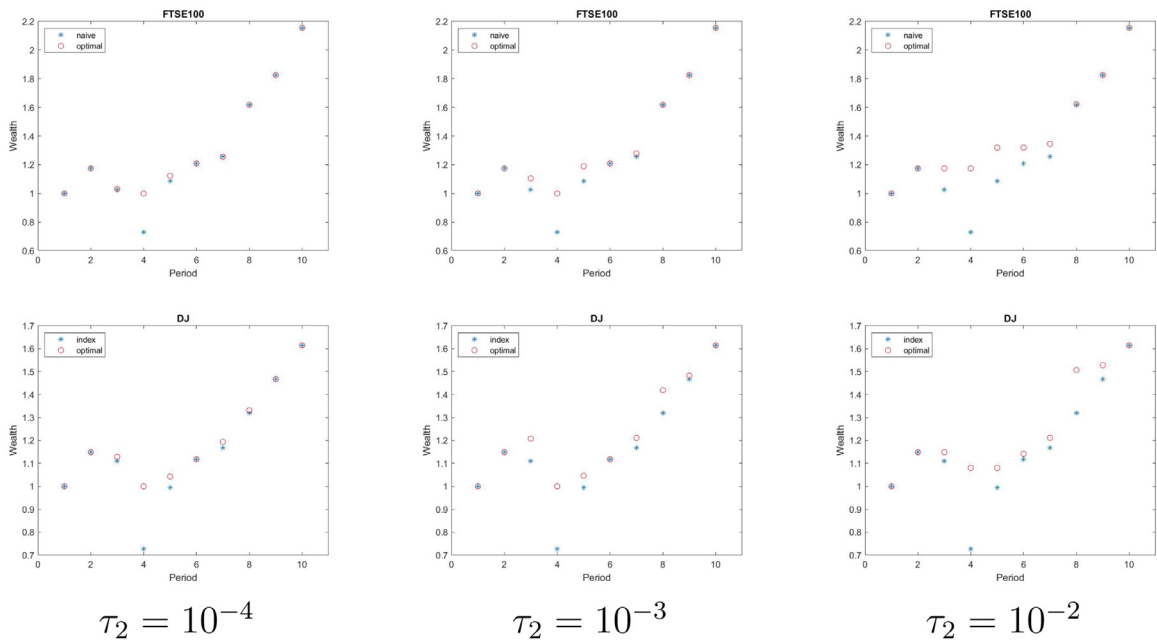


**Fig. 2.** Behaviour of the wealth over time. First row: FTSE 100 compared with the naive strategy. Second row: DowJones compared with the index return. $\tau_1 = 10^{-3}$, $\tau_2$ takes values $10^{-4}$ (left), $10^{-3}$ (center), $10^{-2}$ (right).

As already pointed out, financial targets of the investment and market conditions drive the choice of the regularization parameters. In the following tests $\tau_1$ is chosen as the smallest number that guarantees the minimum number of short positions. This is an important issue as in some cases stock market regulators can impose bans on short-selling like in the two European countries hardest hit by the Covid-19 in March 2020, Italy and Spain. Once fixed $\tau_1$, $\tau_2$ is chosen so to have a good trade-off among the other performance metrics.

In Table 4 we report on all the datasets the comparison with the index in terms of excess return and Information Ratio. However for completeness we report also the other metrics of the optimal portfolios, that are the amount of shorting, the

**Table 4**

Comparison with the index. Columns contain in order: the test case label, parameters $\tau_1$ and $\tau_2$, the number of short positions, the percentage of active positions, the percentage of transactions, the Sharpe Ratio, the excess return and the Information Ratio.

| TEST | $\tau_1$ | $\tau_2$ | short | density | $\mathcal{T}$ | SR | ER | IR |
|------|------|------|-------|---------|------|------|------|------|
| DJ | $10^{-2}$ | $10^{-3}$ | 0 | 37% | 16% | 0.859 | 2% | 0.302 |
| NASDAQ100 | $10^{-2}$ | $10^{-4}$ | 0 | 22% | 16% | 0.933 | 2% | 0.301 |
| FTSE100 | $10^{-2}$ | $10^{-3}$ | 0 | 16% | 6% | 0.460 | 14% | 0.301 |
| SP500 | $10^{-3}$ | $10^{-3}$ | 4 | 6% | 2% | 0.723 | 13% | 0.453 |
| NASDAQComp | $10^{-2}$ | $10^{-4}$ | 0 | 5% | 2% | 0.581 | 6% | 0.301 |

**Table 5**

Comparison with the naive strategy. Columns contain in order: test case label, parameters $\tau_1$ and $\tau_2$, the number of short positions, the percentage of active positions, the percentage of transactions, the Sharpe Ratio, the excess return, the Information Ratio and the risk reduction factor.

| TEST | $\tau_1$ | $\tau_2$ | short | density | $\mathcal{T}$ | SR | ER | IR | RR |
|------|------|------|-------|---------|------|------|------|------|------|
| DJ | $10^{-2}$ | $10^{-2}$ | 0 | 46% | 16% | 1.032 | 8% | 0.440 | 1.510 |
| NASDAQ100 | $10^{-2}$ | $10^{-2}$ | 0 | 20% | 7% | 1.278 | 9% | 0.407 | 1.932 |
| FTSE100 | $10^{-2}$ | $10^{-3}$ | 0 | 23% | 9% | 0.609 | 10% | 0.302 | 1.701 |
| SP500 | $10^{-2}$ | $10^{-3}$ | 0 | 6% | 2% | 0.821 | 17% | 0.302 | 3.184 |
| NASDAQComp | $10^{-2}$ | $10^{-2}$ | 0 | 4% | 1% | 0.740 | 13% | 0.324 | 4.962 |
| FF49 | $10^{-2}$ | $10^{-4}$ | 0 | 17% | 14% | 0.806 | 16% | 0.302 | 2.273 |

percentage of active positions and of transactions, and the Sharpe Ratio. In all tests the optimal portfolio outperforms the benchmark. We achieve an excess return of about the 7% on the average, varying between the 2% and the 14%. The IR ranges between 0.301 and 0.453; SP500 exhibits the highest value of IR, coherently with the observed value of the excess return. On the other hand, FSTSE100 provides the highest excess return but its IR is slightly affected by the dispersion of excess returns.

In Table 5 we show comparisons with the naive strategy. In this case we furthermore report the value of RR defined in **M1**. The reduction of transaction costs is about the 90% on the average. Note that the naive portfolio is a full-turnover one, thus the value of $T$ represents the percentage of transactions made by the optimal strategy with respect to the benchmark. We also observe an higher excess return, varying between the 8% and the 17%, with an average value of 12%. The IR ranges between 0.302 and 0.440; DJ exhibits the highest value of IR, while the excess return is maximum for SP500 and FF49. We note that the SR values are greater than the corresponding values obtained when the benchmark is the index. Finally, our strategy produces optimal portfolios that outperform the benchmark in terms of final wealth with a lower risk, as shown by RR, that varies between 1.510 to 4.962.

## 5. Conclusion

In this work we use split Bregman method for the problem of defining an optimal long-term investment strategy, where the investor can exit the investment before maturity without severe loss. We propose a model in a multi-period Markowitz framework, which extends the fused lasso model proposed in [4]. The inequality constraints on expected minimum wealth at each rebalancing date are introduced to guarantee the investment throughout the period, especially during the unforeseen events such as market crisis. Alternating Split Bregman produces an algorithm that yields subproblems that are solved by fast methods. Numerical comparisons with respect different benchmarks on real databeses show the its effectiveness.

## Acknowledgments

## References

[1] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, Sparsity and mmoothness via the fused lasso, J. Royal Stat. Soc. Series B (Stat. Methodol.) 67 (1) (2005) 91–108.
[2] T. Goldstein, S. Osher, The split Bregman for l1-regularization problems, SIAM J. Imag. Sci. 2 (2) (2009) 323–343.
[3] D. Yu, S. Lee, W. Lee, S. Kim, J. Lim, S. Know, Classification of spectral data using fused lasso logistic regression, Chemom. Intell. Lab. Syst. 142 (15) (2015) 70–77.
[4] S. Corsaro, V. De Simone, Z. Marino, Fused lasso approach in portfolio selection, Ann. Oper. Res. (2019), doi:10.1007/s10479-019-03289-w.
[5] L. Antonelli, V. De Simone, Comparison of minimization methods for nonsmooth image segmentation, Commun. Appl. Ind. Math. 9 (2018) 68–96.
[6] D. di Serafino, G. Landi, M. Viola, ACQUIRE: An inexact iteratively reweighted norm approach for TV-based Poisson image restoration, Appl. Math. Comput. 364 (2020) 124678.
[7] R. De Asmundis, D. di Serafino, W. Hager, G. Toraldo, H. Zhang, An efficient gradient method using the yuan steplength, Comput. Optim. Appl. 59 (3) (2014) 541–563.

[8] L. Antonelli, V. De Simone, D. di Serafino, On the application of the spectral projected gradient method in image segmentation, J. Math. Imag. Vis. 54 (2016) 106–116.

[9] D. di Serafino, V. Ruggiero, G. Toraldo, L. Zanni, On the steplength selection in gradient methods for unconstrained optimization, Appl. Math. Comput. 318 (2018) 176–195.

[10] A. Beck, M. Teboulle, Gradient-based algorithms with applications to signal recovery, Convex Optim. Signal Process. Commun. (2009) 42–88.

[11] B. O'Donoghue, G. Stathopoulos, S. Boyd, A splitting method for optimal control, IEEE Trans. Control Sys. Technol. 21 (6) (2013) 2432–2442.

[12] N. Parikh, S. Boyd, Proximal algorithms, Found. Trends Optim. 1 (3) (2014) 127–239.

[13] E. Rui, S. Boyd, A primer on monotone operator methods, Appl. Comput. Math. 15 (1) (2016) 3–43.

[14] S. Setzer, Operator splittings, bregman methods and frame shrinkage in image processing, Int. J. Computer Vis. 9 (2011) 265–280.

[15] V. De Simone, D. di Serafino, M. Viola, A subspace-accelerated split Bregman method for sparse data recovery with joint $\ell_1$-type regularizers, Electron. Trans. Numer. Anal. 53 (2020) 406–425.

[16] T. Goldstein, X. Bresson, S. Osher, Geometric applications of the split Bregman method: segmentation and surface reconstruction, J. Sci. Comput. 45 (1–3) (2010) 272–293.

[17] G. Ye, X. Xie, Split Bregman method for large scale fused lasso, Comput. Stat. Data Anal. 55 (2011) 1552–1569.

[18] F. Cesarone, A. Scozzari, F. Tardella, Linear vs. quadratic portfolio selection models with hard real-world constraints, Comput. Manag. Sci. 12 (2015) 345–370.

[19] F. Cesarone, J. Moretti, F. Tardella, Optimally chosen small portfolios are better than large ones, Econ. Bull. 36 (4) (2016) 1876–1891.

[20] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, I. Loris, Sparse and stable Markowitz portfolios, PNAS 30 (106) (2009) 12267–12272.

[21] S. Corsaro, V. De Simone, Z. Marino, F. Perla, $L_1$-regularization for multi-period portfolio selection, Ann. Oper. Res. (2019), doi:10.1007/s10479-019-03308-w.

[22] S. Corsaro, V. De Simone, Adaptive $l_1$-regularization for short-selling control in portfolio selection, Comput. Optim. Appl. 72 (2) (2019) 457–478.

[23] R. Bruni, F. Cesarone, A. Scozzari, F. Tardella, On exact and approximate stochastic dominance strategies for portfolio selection, Eur. J. Oper. Res. 259 (1) (2017) 322–329.

[24] G. Guastaroba, R. Mansini, W. Ogryczak, M. Speranza, Linear programming models based on omega ratio for the enhanced index tracking problem, Eur. J. Oper. Res. 251 (3) (2016) 938–956.

[25] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for l1-minimization with applications to compressed sensing, SIAM J. Imag. Sci. 1 (1) (2008) 143–168.

[26] T.A. Davis, Direct Methods for Sparse Linear Systems, Society for Industrial and Applied Mathematics, 2006, doi:10.1137/1.9780898718881.

[27] T.A. Davis, S. Rajamanickam, W.M. Sid-Lakhdar, A survey of direct methods for sparse linear systems, Acta Numer. 25 (2016) 383–566.

[28] R. Bruni, F. Cesarone, A. Scozzari, F. Tardella, Real-world datasets for portfolio selection and solutions of some stochastic dominance portfolio models, Data Brief 8 (2016) 858–862.

[29] O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivar. Anal. 88 (2) (2004) 365–411.

[30] V. DeMiguel, L. Garlappi, R. Uppal, Optimal versus naive diversification: how inefficient is the 1-n portfolio strategy? Rev. Financ. Stud. 22 (5) (2009) 1915–1953.

[31] W.F. Sharpe, The sharpe ratio, J. Portf. Manag. 21 (1) (1994) 49–58.

[32] T.H. Goodwin, The information ratio, Financ. Anal. J. 54 (4) (1998) 34–43.