# Multi-Period Portfolio Optimization*

Edmond Lezmi
Amundi Quantitative Research
edmond.lezmi@amundi.com

Thierry Roncalli
Amundi Quantitative Research
thierry.roncalli@amundi.com

Jiali Xu
Amundi Quantitative Research
jiali.xu@amundi.com

March 2022

**Abstract**

In this article, we consider a multi-period portfolio optimization problem, which is an extension of the single-period mean-variance model. We discuss several formulations of the objective function, constraints and coupling relationships. We then derive three numerical algorithms that can be used to solve such problems: the alternating direction method of multipliers, the block coordinate descent algorithm and the augmented quadratic programming method. We illustrate the differences between single-period and multi-period solutions by considering three asset allocation problems: transition management (Rattray, 2003), total variation regularized portfolio (Corsaro *et al.*, 2020) and trading trajectory modeling (Gârleanu and Pedersen, 2013). Finally, we solve the portfolio alignment problem of Le Guenedal and Roncalli (2022) when the fund manager has to dynamically control the carbon footprint of his investment portfolio by integrating a carbon reduction scenario. Comparing the single-period and multi-period solutions shows that the active share between the two portfolios may be greater than 25%. This figure can also reach 40% if we include carbon trends and they are increasing.

**Keywords:** Multi-period optimization, portfolio allocation, ADMM, block coordinate descent, quadratic programming, coupling variables, transition management, total variation regularization, optimal trading trajectory problem, portfolio decarbonization, net zero alignment.

**JEL Classification:** C61, G11, Q54.

## 1 Introduction

Multi-period portfolio optimization is a natural extension of the mean-variance optimization (MVO) model developed by Harry Markowitz in 1952. The goal is to find the dynamic asset allocation policy by considering inter-temporal effects such as rebalancing costs, trading impacts, time-varying constraints, price trends, etc. Since such models include feedback features, we might think that they are commonly used by the asset management indus-

*The authors are grateful to Mickael Bellaïche for his helpful comments.

try. However, while mean-variance optimization was very successful among investors and portfolio managers, multi-period optimization is mainly an academic research field[1]:

> "*In practice, multi-period models are seldom used. There are several practical reasons for that. First, it is often very difficult to accurately estimate return/risk for multiple periods, let alone for a single period. Second, multi-period models are in general computationally intensive, especially if the universe of assets considered is large. Third, the most common existing multi-period models do not handle real-world constraints. [...] For these reasons, practitioners typically use single-period models to rebalance the portfolio from one period to another*" (Kolm *et al.*, 2014).

Recently, developments in computing capacity have renewed the interest in such models. For instance, we can cite the research by Boyd *et al.* (2017), Calafiore (2009), Corsaro *et al.* (2021), Huang *et al.* (2021), Li *et al.* (2022), Rosenberg *et al.* (2016), Skaf and Boyd (2009) and Wahlberg *et al.* (2012). Moreover, alongside transition management and trading trajectory, which are the two most famous multi-period problems[2], a new problem has emerged these last two years in climate investing. Indeed, portfolio alignment can be viewed as the dynamic version of portfolio decarbonization:

> "*While portfolio decarbonization is a static problem, portfolio alignment involves a dynamic strategy in order to comply with a given climate policy. Therefore, the dynamic problem is trickier since it involves several rebalancing decisions and depends on the future behavior of corporate issuers*" (Le Guenedal and Roncalli, 2022).

The primary objective of our study is to solve the multi-period portfolio alignment problem defined by Le Guenedal and Roncalli (2022, pages 36-37). Indeed, if the investor decarbonizes its current portfolio by $\boldsymbol{\mathcal{R}}$ and he knows that he will decarbonize it by $\boldsymbol{\mathcal{R}} + \Delta\boldsymbol{\mathcal{R}}$ during the next period, then it is obvious that the current optimal portfolio is contingent on the additional reduction $\Delta\boldsymbol{\mathcal{R}}$ during the next period.

This research project fits into our previous works on portfolio optimization and the development of efficient numerical algorithms for solving asset allocation problems. In particular, we can cite augmented quadratic programming (Bruder *et al.*, 2013; Roncalli, 2013; Bourgeron *et al.*, 2018), coordinate descent (Griveau-Billion *et al.*, 2013; Roncalli, 2015; Richard and Roncalli, 2015, 2019) and alternating direction method of multipliers (Bourgeron *et al.*, 2018; Chen *et al.*, 2019). These three numerical algorithms are extensively explained in the survey of Perrin and Roncalli (2020) in the context of portfolio optimization, e.g., risk parity portfolios, strategic asset allocation, smart beta portfolios, minimum-variance strategies, regularized allocation problems and turnover management. In this work, we show how to apply these algorithms to multi-period portfolio optimization and solve the portfolio alignment problem.

This article is organized as follows. In section two, we present the multi-period portfolio optimization problem. We discuss some special cases of the objective function, constraints and coupling relationships. We develop numerical algorithms and apply them to three asset allocation problems. The third section is dedicated to the portfolio decarbonization problem. Finally, section four offers some concluding remarks.

[1] For instance, multi-period portfolio optimization is not available in asset management software such as MSCI Barra Optimizer or Axioma Portfolio Optimizer.

[2] Here, we do not consider Merton-like continuous-time models, whose solution follows a Hamilton-Jacobi-Bellman equation. Indeed, these models mainly concern two assets, but they are not adapted to deal with a large universe of assets. Nevertheless, they have been successful in solving liability-driven investment problems or retirement strategies.

# 2 Multi-period portfolio optimization

We consider a multi-period optimization problem that we encounter in asset allocation. After defining the objective problem, we discuss some special cases. Then, we show how we can solve these multi-period optimization problems using ADMM and QP algorithms.

## 2.1 General problem with linear and non-linear constraints

Let us consider a universe of $n$ assets. We define the following $h$-period optimization problem:

$$\begin{aligned} x^{\star} &= \arg \max_{x_{t+1},x_{t+2},\ldots} \mathbb{E}\left[\mathcal{U}\left(x_{t+1},\ldots,x_{t+h}\right) \mid \mathcal{F}_t\right] \\ &\text{s.t.} \quad x \in \Omega \end{aligned} \tag{1}$$

where $x_t = \left(x_{1,t},\ldots,x_{n,t}\right)$ is the vector of the portfolio weights at the $t^{\text{th}}$ period, $x = \left(x_{t+1},x_{t+2},\ldots,x_{t+h}\right)$ is the vector of the $h$ allocations, $\mathcal{U}\left(x_{t+1},\ldots,x_{t+h}\right)$ is the inter-temporal utility function, $\mathcal{F}_t$ is the filtration associated to the probability space[3], and $x \in \Omega$ is a set of linear and non-linear constraints.

In order to obtain a tractable objective function, we assume that the utility function is separable in time:

$$-\mathbb{E}\left[\mathcal{U}\left(x_{t+1},\ldots,x_{t+h}\right) \mid \mathcal{F}_t\right] = \sum_{s=t+1}^{t+h} \left\{g_s\left(x_s\right) + h_s\left(x_{s-1},x_s\right)\right\} \tag{2}$$

where $g_s$ and $h_s$ are two convex functions. While $g_s\left(x_s\right)$ only depends on the current portfolio $x_s$, $h_s\left(x_{s-1},x_s\right)$ is a convex function that depends on both the current portfolio $x_s$ and the previous portfolio $x_{s-1}$. Therefore, $g_s\left(x_s\right)$ is the static part of the objective function whereas the dynamic part is modeled by the coupling function $h_s\left(x_{s-1},x_s\right)$. Similarly, we split the set of constraints as $\Omega = \Omega^{(g)} \cap \Omega^{(h)}$ where $\Omega^{(g)} = \bigcap_{s=t+1}^{t+h} \Omega_s$ and $\Omega_s$ corresponds to the constraints that only relies on $x_s$ and not on the other variables $x_u$ $(u \neq s)$. Therefore, Problem (1) becomes:

$$\begin{aligned} x^{\star} &= \arg\min_x \left\{g\left(x\right) + h\left(x\right)\right\} \\ &\text{s.t.} \quad x \in \Omega^{(g)} \cap \Omega^{(h)} \end{aligned} \tag{3}$$

where $g\left(x\right) = \sum_{s=t+1}^{t+h} g_s\left(x_s\right)$, $h\left(x\right) = \sum_{s=t+1}^{t+h} h_s\left(x_{s-1},x_s\right)$ and $x^{\star} = \left(x_{t+1}^{\star},x_{t+2}^{\star},\ldots,x_{t+h}^{\star}\right)$. Although Problem (3) defines the optimal solution $x^{\star}$, we are only interested in $x_{t+1}^{\star}$. Indeed, since the filtration at time $t+1$ will be updated, the optimal solution $x_{t+2}^{\star}$ at time $t+1$ is no longer valid. The right formulation of Problem (3) is then:

$$\begin{aligned} x_{t+1}^{\star} &= \arg\min_x \left\{g\left(x\right) + h\left(x\right)\right\} \\ &\text{s.t.} \quad x \in \Omega^{(g)} \cap \Omega^{(h)} \end{aligned} \tag{4}$$

**Remark 1.** *We could discuss what the goal is when writing the objective function as $f\left(x\right) = g\left(x\right) + h\left(x\right)$. Indeed, Problem (4) is equivalent to the traditional non-linear optimization problem $x_{t+1}^{\star} = \arg\min f\left(x\right)$ s.t. $x \in \Omega$. In fact, the underlying idea is to separate the coupling and non-coupling parts. Therefore, we notice that Problem (4) is the overlapping of two problems:*

$$\left\{\begin{array}{ll} x_{t+1}^{\star} = \arg\min g\left(x\right) & \text{s.t.} \quad x \in \Omega^{(g)} \\ x_{t+1}^{\star} = \arg\min h\left(x\right) & \text{s.t.} \quad x \in \Omega^{(h)} \end{array}\right. \tag{5}$$

[3] The stochastic process will be defined later.

*The first problem is static and corresponds to a traditional single-period optimization problem since it is equivalent to:*

$$x_{t+1}^{\star} = \arg\min g_{t+1}\left(x_{t+1}\right) \quad \text{s.t.} \quad x_{t+1} \in \Omega_{t+1}^{(g)} \tag{6}$$

*The second problem is a dynamic feedback problem. Knowing the optimal solution at time* $t+2$*, it modifies the solution at time* $t+1$ *because of the feedback effects. In asset allocation,* $h(x)$ *is generally a penalty function and not really an objective function. In what follows, we extensively use the previous breakdown to find the numerical solution.*

## 2.2 Some special cases

In this section, we discuss some special cases. First, we consider different objective functions that are used in portfolio management. Second, we specify some penalty functions. Finally, we list the main constraints that are specified when we perform portfolio optimization.

### 2.2.1 Objective function

**Single-period optimization problem** When $h$ is equal to 1, the problem reduces to:

$$\begin{aligned} x_{t+1}^{\star} &= \arg\min_{x} \left\{g_{t+1}\left(x_{t+1}\right) + h_{t+1}\left(x_t, x_{t+1}\right)\right\} \\ &\text{s.t.} \quad x_{t+1} \in \Omega \end{aligned} \tag{7}$$

**Mean-variance optimization** In the mean-variance optimization problem, the objective function $g_s\left(x_s\right)$ is equal to (Roncalli, 2013, Section 1.1.1, page 7):

$$g_s\left(x_s\right) = \frac{1}{2} x_s^{\top} \Sigma_s x_s - \gamma x_s^{\top} \mu_s \tag{8}$$

where $\Sigma_s$ is the covariance matrix and $\mu_s$ is the vector of expected returns. The parameter $\gamma$ is a coefficient that controls the trade-off between the portfolio's volatility and its expected return. Let $\boldsymbol{\mathcal{R}}_s = \left(\boldsymbol{\mathcal{R}}_{1,s}, \ldots, \boldsymbol{\mathcal{R}}_{n,s}\right)$ be the vector of asset returns at time $s$. Since we have:

$$\mathbb{E}\left[\boldsymbol{\mathcal{R}}_s \mid \mathcal{F}_t\right] = \mathbb{E}\left[\boldsymbol{\mathcal{R}}_t \mid \mathcal{F}_t\right] = \mu_t \qquad \text{for } s \geq t \tag{9}$$

and:

$$\operatorname{var}\left(\boldsymbol{\mathcal{R}}_s \mid \mathcal{F}_t\right) = \operatorname{var}\left(\boldsymbol{\mathcal{R}}_t \mid \mathcal{F}_t\right) = \Sigma_t \qquad \text{for } s \geq t \tag{10}$$

we obtain:

$$g\left(x\right) = \sum_{s=t+1}^{t+h} \left\{\frac{1}{2} x_s^{\top} \Sigma_t x_s - \gamma x_s^{\top} \mu_t\right\} \tag{11}$$

**Tracking-error minimization** We recall that the tracking error variance of the portfolio $x_s$ with respect to the benchmark $b_s$ is equal to:

$$\sigma^2\left(x_s \mid b_s\right) = \left(x_s - b_s\right)^{\top} \Sigma_s \left(x_s - b_s\right) \tag{12}$$

Therefore, we can show that the objective function $g_s\left(x_s\right)$ is equal to (Roncalli, 2013, Section 1.2.4, page 60):

$$g_s\left(x_s\right) = \frac{1}{2} x_s^{\top} \Sigma_s x_s - x_s^{\top} \Sigma_s b_s \tag{13}$$

Finally, we obtain:

$$g(x) = \sum_{s=t+1}^{t+h} \left\{ \frac{1}{2} x_s^\top \Sigma_t x_s - x_s^\top \Sigma_t b_s \right\} \tag{14}$$

If we assume that we do not know the future composition of the benchmark at time $s > t$, Equation (14) becomes:

$$g(x) = \sum_{s=t+1}^{t+h} \left\{ \frac{1}{2} x_s^\top \Sigma_t x_s - x_s^\top \Sigma_t b_t \right\} \tag{15}$$

**Portfolio optimization with a benchmark** We can mix the two approaches. In this case, the investor would like to maximize the expected excess return of the portfolio with respect to the benchmark and control the level of the tracking error volatility. The multi-period objective function becomes (Roncalli, 2013, Section 1.1.4, page 19):

$$g(x) = \sum_{s=t+1}^{t+h} \left\{ \frac{1}{2} x_s^\top \Sigma_t x_s - x_s^\top (\Sigma_t b_t + \gamma \mu_t) \right\} \tag{16}$$

**Remark 2.** *We notice that mean-variance, tracking-error and benchmark optimization problems can be cast into a quadratic programming problem:*

$$g_s(x_s) = \frac{1}{2} x_s^\top Q_s x_s - x_s^\top R_s \tag{17}$$

*where $Q_s = \Sigma_s$ and $R_s$ is respectively equal to $\gamma \mu_s$, $\Sigma_s b_s$ and $\Sigma_s b_s + \gamma \mu_s$. In what follows, we use this notation and the term 'mean-variance' to name these three problems.*

**Other objective functions** Perrin and Roncalli (2020, Table 1, page 29) reviewed the different objective functions used in portfolio optimization. It includes minimum variance, most diversified, risk budgeting or Kullback-Leibler portfolios.

### 2.2.2 Penalty function

Perrin and Roncalli (2020) observed that four regularization penalties are mainly used in portfolio management: ridge, lasso, log-barrier and entropy.

**Ridge penalization** In the case of the ridge penalty, we have:

$$h_s(x_{s-1}, x_s) = \frac{\lambda_s}{2} \|x_s - x_{s-1}\|_2^2 \tag{18}$$

where $\lambda_s$ is the scalar penalty value. Gârleanu and Pedersen (2013) used quadratic transaction costs:

$$h_s(x_{s-1}, x_s) = \frac{1}{2} (x_s - x_{s-1})^\top \Lambda_s (x_s - x_{s-1}) \tag{19}$$

where $\Lambda_s$ is the Kyle's matrix for temporary price impact. We notice that the penalization with quadratic transaction costs generalizes the ridge penalty where $\Lambda_s = \lambda_s I_n$.

**Lasso penalization** Instead of using the $\boldsymbol{\ell}_2$-norm, we can use the $\boldsymbol{\ell}_1$-norm:

$$h_s(x_{s-1}, x_s) = \lambda_s \|x_s - x_{s-1}\|_1 \tag{20}$$

This regularization can be viewed as a turnover penalization problem.

### 2.2.3 Portfolio constraints

**Linear constraints** If the constraints are linear, we have:

$$x \in \Omega \Leftrightarrow \left\{ \begin{array}{l} Ax = B \\ Cx \leq D \\ \underline{x} \leq x \leq \overline{x} \end{array} \right. \tag{21}$$

It follows that $\Omega = \Omega^{(h)}$ and $\Omega^{(g)} = \left\{ x \in \mathbb{R}^{nh} \right\}$. In the case where constraints are separable, we obtain $\Omega = \Omega^{(g)}$ where:

$$x_s \in \Omega_s \Leftrightarrow \left\{ \begin{array}{l} A_s x_s = B_s \\ C_s x_s \leq D_s \\ \underline{x}_s \leq x_s \leq \overline{x}_s \end{array} \right. \tag{22}$$

**Turnover constraint** The turnover constraint is defined as:

$$\Omega^{(h)} = \left\{ \forall s = t+1, \ldots, t+h : \left\| x_s - x_{s-1} \right\|_1 \leq \tau_s \right\} \tag{23}$$

where $\tau_s$ is the turnover limit at time $s$. In the single-period optimization problem, imposing a turnover constraint is equivalent to add a lasso penalization. Therefore, we have a relationship between $\tau_s$ and $\lambda_s$. In the multi-period optimization problem, we lose the strict equivalence.

**Other constraints** We can specify other constraints such as asset class limits, sector limits, number of active bets, etc. (Perrin and Roncalli, 2020, Table 3, page 30).

## 2.3 Numerical algorithms

### 2.3.1 ADMM approach

**Derivation of the algorithm** Following Perrin and Roncalli (2020), we use the alternating direction method of multipliers (ADMM) introduced by Gabay and Mercier (1976) to propose a numerical solution. We note:

$$\mathbb{1}_{\Omega}(x) = \left\{ \begin{array}{lll} 0 & \text{if} & x \in \Omega \\ +\infty & \text{if} & x \notin \Omega \end{array} \right. \tag{24}$$

Then, we can rewrite the optimization problem as follows:

$$x^{\star}_{t+1} = \arg\min_{x} \left\{ g(x) + \mathbb{1}_{\Omega^{(g)}}(x) + h(x) + \mathbb{1}_{\Omega^{(h)}}(x) \right\} \tag{25}$$

Using the first and third tricks given by Perrin and Roncalli (2020), the equivalent ADMM form is:

$$\begin{aligned} \left\{ x^{\star}, y^{\star} \right\} &= \arg\min_{(x,y)} \left\{ f_x(x) + f_y(y) \right\} \\ & \text{s.t.} \quad \left\{ \begin{array}{l} x = y \\ f_x(x) = g(x) + \mathbb{1}_{\Omega^{(g)}}(x) \\ f_y(y) = h(y) + \mathbb{1}_{\Omega^{(h)}}(y) \end{array} \right. \end{aligned} \tag{26}$$

Boyd *et al.* (2010) showed that the associated ADMM algorithm consists of the following three steps:

1. The $x$-update is:

$$x^{(k+1)} = \arg\min_{x} \left\{ f_x^{(k+1)}(x) = f_x(x) + \frac{\varphi}{2} \left\| x - y^{(k)} + u^{(k)} \right\|_2^2 \right\} \tag{27}$$

2. The $y$-update is:

$$y^{(k+1)} = \arg\min_{y} \left\{ f_y^{(k+1)}(y) = f_y(y) + \frac{\varphi}{2} \left\| x^{(k+1)} - y + u^{(k)} \right\|_2^2 \right\} \tag{28}$$

3. The $u$-update is:

$$u^{(k+1)} = u^{(k)} + \left( x^{(k+1)} - y^{(k+1)} \right) \tag{29}$$

In this approach, $u^{(k)}$ is the dual variable of the primal residual $r = x - y$ and $\varphi$ is the $\ell_2$-norm penalty variable. The parameter $\varphi$ can be constant or may change at each iteration. The ADMM algorithm benefits from the dual ascent principle and the method of multipliers. The difference with the latter is that the $x$- and $y$-updates are performed in an alternating way. In practice, ADMM may be slow to converge with high accuracy, but is fast to converge if we consider modest accuracy. This is why ADMM is a good candidate for solving large-scale optimization problems, where high accuracy does not necessarily lead to a better solution.

We notice that:

$$\begin{aligned} f_x(x) &= \sum_{s=t+1}^{t+h} g_s(x_s) + \sum_{s=t+1}^{t+h} \mathbb{1}_{\Omega_s}(x_s) \\ &= \sum_{s=t+1}^{t+h} f_s(x_s) \end{aligned} \tag{30}$$

where:

$$f_s(x_s) = g_s(x_s) + \mathbb{1}_{\Omega_s}(x_s) \tag{31}$$

Using the same partition for $y$ and $u$ than for $x$, we have:

$$y = (y_{t+1}, y_{t+2}, \ldots, y_{t+h}) \tag{32}$$

and:

$$u = (u_{t+1}, u_{t+2}, \ldots, u_{t+h}) \tag{33}$$

We deduce that:

$$\begin{aligned} \left\| x - y^{(k)} + u^{(k)} \right\|_2^2 &= \left( x - y^{(k)} + u^{(k)} \right)^\top \left( x - y^{(k)} + u^{(k)} \right) \\ &= \sum_{s=t+1}^{t+h} \left( x_s - y_s^{(k)} + u_s^{(k)} \right)^\top \left( x_s - y_s^{(k)} + u_s^{(k)} \right) \end{aligned}$$

It follows that the solution $x^{(k+1)}$ is equal to:

$$x^{(k+1)} = \begin{pmatrix} x_{t+1}^{(k+1)} \\ \vdots \\ x_{t+h}^{(k+1)} \end{pmatrix} \tag{34}$$

where:

$$x_s^{(k+1)} = \arg\min_{x_s} \left\{ g_s(x_s) + \mathbb{1}_{\Omega_s}(x_s) + \frac{\varphi}{2} \left\| x_s - y_s^{(k)} + u_s^{(k)} \right\|_2^2 \right\} \tag{35}$$

Here, we exploit the separability property of $f_x(x)$. Instead of solving a problem of dimension $nh$, we solve $h$ problems of dimension $n$. We conclude that the ADMM algorithm becomes:

1. The $x$-update is:

$$x_s^{(k+1)} = \arg\min_{x_s} \left\{ g_s(x_s) + \mathbb{1}_{\Omega_s}(x_s) + \frac{\varphi}{2} \left\| x_s - y_s^{(k)} + u_s^{(k)} \right\|_2^2 \right\} \tag{36}$$

   where $s = t+1, t+2, \ldots, t+h$.

2. The $y$-update is:

$$y^{(k+1)} = \arg\min_{y} \left\{ f_y^{(k+1)}(y) = f_y(y) + \frac{\varphi}{2} \left\| x^{(k+1)} - y + u^{(k)} \right\|_2^2 \right\} \tag{37}$$

3. The $u$-update is:

$$u^{(k+1)} = u^{(k)} + \left( x^{(k+1)} - y^{(k+1)} \right) \tag{38}$$

**The case $f_s(x_s)$ is a QP problem** $f_s(x_s)$ is a QP problem when the objective function $g_s(x_s) = \frac{1}{2} x_s^\top Q_s x_s - x_s^\top R_s$ is quadratic and the constraints $x_s \in \Omega_s$ are linear. This is for example the case when we perform mean-variance or tracking-error optimization[4]. It follows that the $x$-update reduces to solve a standard QP problem:

$$\begin{aligned} x_s^{(k+1)} &= \arg\min_{x_s} \left\{ \frac{1}{2} x_s^\top (Q_s + \varphi I_n) x_s - x_s^\top \left( R_s + \varphi \left( y_s^{(k)} - u_s^{(k)} \right) \right) \right\} \\ \text{s.t.} & \quad \left\{ \begin{array}{l} A_s x_s = B_s \\ C_s x_s \leq D_s \\ \underline{x}_s \leq x_s \leq \overline{x}_s \end{array} \right. \end{aligned} \tag{39}$$

**The case $\mathbb{1}_{\Omega^{(h)}}(x) = 0$** Since there is no coupling constraint, the function $f_y(y)$ reduces to $h(y)$ and we have:

$$\begin{aligned} f_y^{(k+1)}(y) &= h(y) + \frac{\varphi}{2} \left\| x^{(k+1)} - y + u^{(k)} \right\|_2^2 \\ &= h(y) + \frac{\varphi}{2} \left\| y - v^{(k+1)} \right\|_2^2 \end{aligned} \tag{40}$$

where $v^{(k+1)} = x^{(k+1)} + u^{(k)}$. We deduce that:

$$\begin{aligned} y^{(k+1)} &= \arg\min_{y} \left\{ \varphi^{-1} h(y) + \frac{1}{2} \left\| y - v^{(k+1)} \right\|_2^2 \right\} \\ &= \mathbf{prox}_{\varphi^{-1} h} \left( v^{(k+1)} \right) \end{aligned} \tag{41}$$

[4] In what follows, we only consider this specification because it corresponds to the standard approach of portfolio allocation.

where $\mathbf{prox}_f(v)$ is the proximal operator of the function $f$ at the point $v$ (Perrin and Roncalli, 2020).

We consider the ridge penalization $h_s(y_{s-1}, y_s) = \frac{1}{2}(y_s - y_{s-1})^\top \Lambda_s (y_s - y_{s-1})$. We have:

$$f_y^{(k+1)}(y) = \frac{1}{2}\sum_{s=t+1}^{t+h}\left\{(y_s - y_{s-1})^\top \Lambda_s (y_s - y_{s-1}) + \varphi \left\| y_s - v_s^{(k+1)} \right\|_2^2\right\}$$

The first-order condition is:

$$\frac{\partial\, f_y^{(k+1)}(y)}{\partial\, y_s} = \Lambda_s (y_s - y_{s-1}) - \Lambda_{s+1}(y_{s+1} - y_s) + \varphi\left(y_s - v_s^{(k+1)}\right) = \mathbf{0}_n \tag{42}$$

with the convention $\Lambda_{t+h+1} = \mathbf{0}_{n,n}$. We deduce that:

$$\alpha_s y_{s-1} + \beta_s y_s + \gamma_s y_{s+1} = \varphi v_s^{(k+1)} \tag{43}$$

where $\alpha_s = -\Lambda_s$, $\beta_s = \Lambda_s + \Lambda_{s+1} + \varphi$ and $\gamma_s = -\Lambda_{s+1}$. We obtain a block-tridiagonal system:

$$\begin{pmatrix} \beta_{t+1} & \gamma_{t+1} & \mathbf{0}_{n,n} & \cdots & \mathbf{0}_{n,n} \\ \alpha_{t+2} & \beta_{t+2} & \gamma_{t+2} & & \\ & & & \ddots & \\ \mathbf{0}_{n,n} & \cdots & \mathbf{0}_{n,n} & \alpha_{t+h} & \beta_{t+h} \end{pmatrix} \begin{pmatrix} y_{t+1} \\ y_{t+2} \\ \vdots \\ y_{t+h} \end{pmatrix} = \begin{pmatrix} \delta_{t+1} \\ \delta_{t+2} \\ \vdots \\ \delta_{t+h} \end{pmatrix} \tag{44}$$

where $\delta_s = \varphi v_s^{(k+1)} + \mathbb{1}\{s = t+1\} \cdot \Lambda_{t+1} y_t$. We can easily solve Equation (44) by using the recurrence algorithm of block-Gaussian elimination.

**Remark 3.** *In the case where the matrices $\Lambda_s$ are diagonal, we can exploit their structure to obtain a better efficient algorithm. Indeed, we notice that $f_y^{(k+1)}(y)$ becomes separable:*

$$f_y^{(k+1)}(y) = \frac{1}{2}\sum_{i=1}^{n}\sum_{s=t+1}^{t+h}\left\{\lambda_{i,s}(y_{i,s} - y_{i,s-1})^2 + \varphi\left(y_{i,s} - v_{i,s}^{(k+1)}\right)^2\right\}$$

*where $\lambda_{i,s}$ is the $i^{\text{th}}$ element of the diagonal matrix $\Lambda_s$. Using the same analysis as previously, we obtain a tridiagonal system:*

$$\begin{pmatrix} \beta_{i,t+1} & \gamma_{i,t+1} & 0 & \cdots & 0 \\ \alpha_{i,t+2} & \beta_{i,t+2} & \gamma_{i,t+2} & & \\ & & & \ddots & \\ 0 & \cdots & 0 & \alpha_{i,t+h} & \beta_{i,t+h} \end{pmatrix} \begin{pmatrix} y_{i,t+1} \\ y_{i,t+2} \\ \vdots \\ y_{i,t+h} \end{pmatrix} = \begin{pmatrix} \delta_{i,t+1} \\ \delta_{i,t+2} \\ \\ \delta_{i,t+h} \end{pmatrix} \tag{45}$$

*where $\alpha_{i,s} = -\lambda_{i,s}$, $\beta_{i,s} = \lambda_{i,s} + \lambda_{i,s+1} + \varphi$, $\gamma_{i,s} = -\lambda_{i,s+1}$ and $\delta_{i,s} = \varphi v_{i,s}^{(k+1)} + \mathbb{1}\{s = t+1\} \cdot \lambda_{i,t+1} y_{i,t}$.*

**The case $h(y)$ is an additively separable function** If we assume that:

$$h_s(y_{s-1}, y_s) = \sum_{i=1}^{n} h_s(y_{i,s-1}, y_{i,s}) \tag{46}$$

we deduce that:

$$f_y^{(k+1)}(y) = \sum_{i=1}^{n} \sum_{s=t+1}^{t+h} \left\{ h_s\left(y_{i,s-1}, y_{i,s}\right) + \frac{\varphi}{2}\left(y_{i,s} - v_{i,s}^{(k+1)}\right)^2 \right\} \tag{47}$$

Let $y_{(i)} = \left(y_{i,t+1}, \ldots, y_{i,t+h}\right)$ be the $h \times 1$ vector that collects the weights of asset $i$. The first-order condition is:

$$\frac{\partial\, f_y^{(k+1)}(y)}{\partial\, y_{(i)}} = \frac{\partial\, h_s\left(y_{i,s-1}, y_{i,s}\right)}{\partial\, y_{(i)}} + \varphi\left(y_{(i)} - v_{(i)}^{(k+1)}\right) = \mathbf{0}_h \tag{48}$$

Therefore, we have to solve $n$ problems of dimension $h$ instead of one problem of dimension $nh$.

For instance, the lasso penalization $h_s(y_{s-1}, y_s) = \lambda_s \left\|y_s - y_{s-1}\right\|_1$ is an additively separable function:

$$h_s(y_{s-1}, y_s) = \sum_{i=1}^{n} \lambda_s \left|y_{i,s} - y_{i,s-1}\right| \tag{49}$$

We deduce that:

$$\begin{aligned} h(y) &= \sum_{s=t+1}^{t+h} \sum_{i=1}^{n} \lambda_s \left|y_{i,s} - y_{i,s-1}\right| \\ &= \sum_{i=1}^{n} \sum_{s=t+1}^{t+h} \lambda_s \left|y_{i,s} - y_{i,s-1}\right| \\ &= \sum_{i=1}^{n} \left\|\mathcal{D}(\boldsymbol{\lambda})\, y_{(i)} - \lambda_{t+1}\mathbf{e}_1 y_{i,t}\right\|_1 \end{aligned} \tag{50}$$

where $\boldsymbol{e}_1 = (1, 0, \ldots, 0)$ and $\mathcal{D}(\boldsymbol{\lambda})$ is a $h \times h$ matrix with diagonal and sub-diagonal entries:

$$\mathcal{D}(\boldsymbol{\lambda}) = \begin{pmatrix} \lambda_{t+1} & & & \\ -\lambda_{t+2} & \lambda_{t+2} & & \\ & & \ddots & \\ & & -\lambda_{t+h} & \lambda_{t+h} \end{pmatrix} \tag{51}$$

Therefore, we split the $y$-update problem and the solution for $y_{(i)}^{(k+1)}$ is given by the following optimization problem:

$$\begin{aligned} y_{(i)}^{(k+1)} &= \arg\min_{y_{(i)}} \left\{ \left\|\mathcal{D}(\boldsymbol{\lambda})\, y_{(i)} - \lambda_{t+1}\mathbf{e}_1 y_{i,t}\right\|_1 + \frac{\varphi}{2}\left\|y_{(i)} - v_{(i)}^{(k+1)}\right\|_2^2 \right\} \\ &= \mathbf{prox}_{\zeta_\varphi\left(y_{(i)}; y_{i,t}, \boldsymbol{\lambda}\right)}\left(v_{(i)}^{(k+1)}\right) \end{aligned} \tag{52}$$

where the function $\zeta_\varphi(x; x_0, \boldsymbol{\lambda})$ is defined as:

$$\zeta_\varphi(x; x_0, \boldsymbol{\lambda}) := \varphi^{-1}\left\|\mathcal{D}(\boldsymbol{\lambda})\, x - \lambda_1 \boldsymbol{e}_1 x_0\right\|_1 \tag{53}$$

This proximal solution can be found by using the augmented QP problem or other algorithms (see Appendix C.1 on page 42).

### 2.3.2 Block coordinate descent

**Definition** The goal of the coordinate descent algorithm is to find the solution $x^{\star} = \arg\min f(x)$ by using a series of optimization problems that are more simple to solve. For that, we choose a coordinate $i \in \{1, n\}$, we solve the uni-dimensional problem:

$$x_i^{(k+1)} = \arg\min_{x_i} f\left(x_1^{(k)}, \ldots, x_{i-1}^{(k)}, x_i, x_{i+1}^{(k)}, \ldots, x_n^{(k)}\right) \tag{54}$$

we update the solution such that $x_j^{(k+1)} \leftarrow x_j^{(k)}$ if $j \neq i$ and we iterate the algorithm until convergence. Coordinate descent is then a variant of the gradient descent and minimizes the function along one coordinate at each step. The simplest way to choose the coordinate is to consider cyclic coordinates. In this case, we have:

$$x_i^{(k+1)} = \arg\min_{x_i} f\left(x_1^{(k+1)}, \ldots, x_{i-1}^{(k+1)}, x_i, x_{i+1}^{(k)}, \ldots, x_n^{(k)}\right) \tag{55}$$

The previous algorithm can be extended to the case where $x_i$ is not a scalar, but a block of coordinates. In this case, we solve the minimization problem with respect to $x_1$ by considering the blocks $x_2, \ldots, x_n$ as given. Then, we update the solution $x_1$ and we solve the minimization problem with respect to $x_2$ by considering the blocks $x_1, x_3, \ldots, x_n$ as given. Since each iteration involves a block of coordinates, this algorithm is called "*block coordinate descent*" (BCD).

**Application to the multi-period optimization problem** The function to minimize is equal to:

$$f(x_{t+1}, \ldots, x_{t+h}) = \sum_{s=t+1}^{t+h} \left(g_s(x_s) + h_s(x_{s-1}, x_s)\right) + \mathbb{1}_{\Omega}(x_{t+1}, \ldots, x_{t+h}) \tag{56}$$

We consider each vector $x_s$ as a block of coordinates. We deduce that the BCD algorithm consists in the following iterations:

$$\begin{aligned} x_s^{(k+1)} &= \arg\min_{x_s} f(x_{t+1}, \ldots, x_{s-1}, x_s, x_{s+1}, \ldots, x_{t+h}) \\ &= \arg\min f_s\left(x_s \mid x_{(-s)}\right) \end{aligned} \tag{57}$$

where $x_{(-s)} = (x_{t+1}, \ldots, x_{s-1}, x_{s+1}, \ldots, x_{t+h})$ and[5]:

$$\begin{aligned} f_s\left(x_s \mid x_{(-s)}\right) = \; & g_s(x_s) + h_s(x_{s-1}, x_s) + h_{s+1}(x_s, x_{s+1}) + \\ & \mathbb{1}_{\Omega_s}(x_s) + \mathbb{1}_{\Omega^{(h)}}(x_{t+1}, \ldots, x_{t+h}) \end{aligned} \tag{58}$$

In the case where $\mathbb{1}_{\Omega^{(h)}}(x) = 0$, the function $f_s\left(x_s \mid x_{(-s)}\right)$ reduces to $f_s(x_s \mid x_{s-1}, x_{s+1})$:

$$f_s(x_s \mid x_{s-1}, x_{s+1}) = g_s(x_s) + h_s(x_{s-1}, x_s) + h_{s+1}(x_s, x_{s+1}) + \mathbb{1}_{\Omega_s}(x_s) \tag{59}$$

Therefore, the case $\mathbf{1}_{\Omega^{(h)}}(x) = 0$ is appealing since it can be considered as a single-period regularized optimization problem with two penalty functions[6] $h_s(x_{s-1}, x_s)$ and $h_{s+1}(x_s, x_{s+1})$.

[5] Because we have $\Omega = \Omega^{(g)} \cap \Omega^{(h)}$ and $\Omega^{(g)} = \bigcap_{s=t+1}^{t+h} \Omega_s$.

[6] If $s = t + h$, the function $h_{s+1}(x_s, x_{s+1})$ is set to 0 and we have only one penalty function.

**The mean-variance-ridge problem** We formulate the function $g_s(x_s)$ as follows:

$$g_s(x_s) = \frac{1}{2} x_s^\top Q_s x_s - x_s^\top R_s \tag{60}$$

This general formulation encompasses the mean-variance optimization problem, the tracking-error minimization problem and the portfolio optimization with a benchmark. We have:

$$\begin{aligned} f_s(x_s \mid x_{s-1}, x_{s+1}) &= \frac{1}{2} x_s^\top Q_s x_s - x_s^\top R_s + \mathbb{1}_{\Omega_s}(x_s) \\ &\quad \frac{1}{2}(x_s - x_{s-1})^\top \Lambda_s (x_s - x_{s-1}) + \\ &\quad \frac{1}{2}(x_{s+1} - x_s)^\top \Lambda_{s+1} (x_{s+1} - x_s) \\ &= \frac{1}{2} x_s^\top (Q_s + \Lambda_s + \Lambda_{s+1}) x_s - x_s^\top (R_s + \Lambda_s x_{s-1} + \Lambda_{s+1} x_{s+1}) + \\ &\quad \mathbb{1}_{\Omega_s}(x_s) + \frac{1}{2} x_{s-1}^\top \Lambda_s x_{s-1} + \frac{1}{2} x_{s+1}^\top \Lambda_{s+1} x_{s+1} \end{aligned} \tag{61}$$

We deduce that:

$$x_s^{(k+1)} = \arg\min_{x_s} \left\{ \frac{1}{2} x_s^\top (Q_s + \Lambda_s + \Lambda_{s+1}) x_s - x_s^\top (R_s + \Lambda_s x_{s-1} + \Lambda_{s+1} x_{s+1}) + \mathbb{1}_{\Omega_s}(x_s) \right\} \tag{62}$$

If the constraints $x_s \in \Omega_s$ are linear, we obtain a QP problem. The BCD algorithm consists then in solving a series of QP problems. While the dimension of the original problem is $nh$, the dimension of the BCD algorithm reduces to $n$.

**The mean-variance-lasso problem** In this case, we have:

$$x_s^{(k+1)} = \arg\min_{x_s} \left\{ \frac{1}{2} x_s^\top Q_s x_s - x_s^\top R_s + \lambda_s \|x_s - x_{s-1}\|_1 + \lambda_{s+1} \|x_{s+1} - x_s\|_1 + \mathbb{1}_{\Omega_s}(x_s) \right\} \tag{63}$$

If the constraints $x_s \in \Omega_s$ are linear, we can transform this optimization problem into an augmented QP problem (see Appendix C.4 on page 51).

### 2.3.3 Quadratic programming

If $g(x) + h(x)$ can be written as a quadratic function and if the constraints $x \in \Omega$ are linear, we obtain a quadratic programming problem:

$$\begin{aligned} x^\star &= \arg\min_x \left\{ \frac{1}{2} x^\top Q x - x^\top R \right\} \\ &\text{s.t.} \left\{ \begin{array}{l} Ax = B \\ Cx \leq D \\ \underline{x} \leq x \leq \overline{x} \end{array} \right. \end{aligned} \tag{64}$$

This is for example the case of the multi-period mean-variance-ridge problem. For this problem, the matrices $Q$ and $R$ are given in Appendix C.5 on page 53. If the constraints are separable with respect to time $s$, we obtain block linear equality and inequality constraints (see Appendix C.6 on page 54).

The multi-period mean-variance-lasso problem is not a QP problem since we have:

$$h(x) = \sum_{s=t+1}^{t+h} h_s(x_{s-1}, x_s) = \sum_{s=t+1}^{t+h} \lambda_s \|x_s - x_{s-1}\|_1 \tag{65}$$

Nevertheless, we can use the trick of augmented variables:

$$x_s = x_{s-1} - x_s^- + x_s^+ \tag{66}$$

in order to transform the multi-period mean-variance-lasso problem into an augmented QP problem. The details are given in Appendix C.7 on page 55.

## 2.4 Convergence and efficiency of the algorithms

The choice of an algorithm depends on the problem specification and the efficiency of the algorithm. For instance, in the case where $g_s(x_s)$ is a mean-variance function, $h_s(x_{s-1}, x_s)$ is a ridge penalty and the constraints $\Omega$ are linear, the QP algorithm is certainly the best choice. Nevertheless, the dimension of this problem is equal to $nh$ where $n$ is the number of assets and $h$ is the number of periods. For instance, if $n = 200$ and $h = 5$, the $Q$ matrix of the QP algorithm has a dimension $1\,000 \times 1\,000$. If the investment universe corresponds to the MSCI world index, $n$ is greater than $1\,500$, and the dimension of the QP problem is larger than $7\,500$. However, most of QP algorithms are valid for a dimension lower than $2\,000$. Therefore, we need to use sparse QP algorithms, which are not always implemented in programming languages. The alternative solution is to use the block coordinate descent algorithm. For instance, in the case of the mean-variance-ridge problem, it consists in solving a series of QP problems, whose dimension is equal to $n$ whatever the value of $h$. From a theoretical point of view, we can show that this algorithm will converge (Tseng, 2001; Xu and Yin, 2013).

If we consider that $g_s(x_s)$ is a mean-variance function, $h_s(x_{s-1}, x_s)$ is a lasso penalty and the constraints $\Omega$ are linear, the best choice is the augmented QP algorithm. Nevertheless, the dimension of the problem is now equal to $3nh$. If we consider an investment universe of 200 assets and 5 periods, the $Q$ matrix of the augmented QP algorithm has a dimension $3\,000 \times 3\,000$. Therefore, it is better to use the block coordinate descent algorithm, because the dimension of each iterated problem is equal to 600. Nevertheless, we are not sure that the algorithm will converge (Tseng, 2001). An alternative solution is to use the ADMM algorithm. Again, the convergence of the algorithm is not guaranteed. Moreover, the convergence depends on the algorithm used for solving the $y$-step. This can also be the case for non-linear problems (objective function and/or constraints), which require to use Dykstra algorithms (Dykstra, 1983; Bauschke and Borwein, 1994; Perrin and Roncalli, 2020).

In this context, we must be careful with the solutions obtained by those algorithms. We must test several starting values and several algorithms when it is possible before deciding whether or not the numerical solution is acceptable. In all cases, we must consider how small changes of the problem impact the numerical solution. Moreover, it is also important to test the algorithm in degenerate and simplified cases in order to understand its behavior.

## 2.5 Some examples

### 2.5.1 Transition management

**Example 1.** *We consider a strategic asset allocation problem. The investment universe is made up of seven asset classes: three fixed-income classes, three equity classes and one commodity class. The parameters are given in Appendix A.1 on page 30. The current portfolio is equal to* $x_t = \left(20\%, 10\%, 15\%, 20\%, 10\%, 15\%, 10\%\right)$.

Using the expected returns $\mu_t$ and the covariance matrix $\Sigma_t$ defined on page 30, we compute the long-only mean-variance portfolio: $x^\star = \arg\min \frac{1}{2} x^\top \Sigma_t x - \gamma x^\top \mu_t$ with $\gamma = 1\%$.

The solution is $x^{\star} = \left(46.21\%, 38.21\%, 0\%, 4.09\%, 4.09\%, 7.11\%, 0.30\%\right)$. Since we notice that the turnover $\tau\left(x_t, x^{\star}\right) = \left\|x^{\star} - x_t\right\|_1$ is equal to 108.82%, it is not possible to directly implement this strategic asset allocation policy. Indeed, there is a large discrepancy between the current portfolio $x_t$ and the target portfolio $x^{\star}$. Therefore, we generally consider a transition management approach, whose goal is to change the assets of the portfolio by minimizing the transaction costs and mitigating the market risks associated with these changes (Rattray, 2003).

Let us assume that the transition management process consists in changing the composition of the portfolio in $h$ periods and limiting the turnover at each rebalancing date. We can proceed in an iterative way and solve the following mean-variance problem:

$$\begin{aligned} x_s^{\star} &= \arg\min \left\{\frac{1}{2} x_s^{\top} \Sigma_t x_s - \gamma x_s^{\top} \mu_t\right\} \qquad \text{for } s = t+1, \ldots, t+h \qquad (67) \\ &\text{s.t.} \quad \left\{\begin{array}{l} \mathbf{1}_n^{\top} x_s = 1 \\ \left\|x_s - x_{s-1}^{\star}\right\|_1 \leq \tau_s \\ x_s \geq \mathbf{0}_n \end{array}\right. \end{aligned}$$

where $x_t^{\star} = x_t$ and $\tau_s$ is the maximum turnover at time $s$. For instance, in the previous example, we had $\tau\left(x_t, x^{\star}\right) = 108.82\%$, implying that we can impose $\tau_s = 25\%$ if we consider $h = 5$ rebalancing periods. Results are given in Table 1. We verify that the optimal solution at time $t + 5$ is equal to the mean-variance portfolio.

Table 1: Iterative solution in % (transition management)

| $s$ | $t$ | $t+1$ | $t+2$ | $t+3$ | $t+4$ | $t+5$ |
|---|---|---|---|---|---|---|
| $x_{1,s}$ | 20.00 | 20.00 | 20.04 | 27.25 | 39.74 | 46.21 |
| $x_{2,s}$ | 10.00 | 22.50 | 34.96 | 40.25 | 40.26 | 38.21 |
| $x_{3,s}$ | 15.00 | 15.00 | 15.00 | 15.00 | 3.92 | 0.00 |
| $x_{4,s}$ | 20.00 | 20.00 | 11.62 | 5.65 | 5.57 | 4.09 |
| $x_{5,s}$ | 10.00 | 10.00 | 9.99 | 5.65 | 4.70 | 4.09 |
| $x_{6,s}$ | 15.00 | 7.77 | 5.56 | 5.16 | 5.16 | 7.11 |
| $x_{7,s}$ | 10.00 | 4.74 | 2.82 | 1.04 | 0.64 | 0.30 |
| $\tau_s$ | | 25.00 | 25.00 | 25.00 | 25.00 | 16.83 |

Another method is to consider the multi-period mean-variance-lasso optimization problem, which is described on page 58. We have:

$$\begin{aligned} \left(x_{t+1}^{\star}, \ldots, x_{t+h}^{\star}\right) &= \arg\min \sum_{s=t+1}^{t+h} \left\{\frac{1}{2} x_s^{\top} \Sigma_t x_s - \gamma x_s^{\top} \mu_t\right\} \qquad (68) \\ &\text{s.t.} \quad \left\{\begin{array}{l} \mathbf{1}_n^{\top} x_s = 1 \\ \left\|x_s - x_{s-1}^{\star}\right\|_1 \leq \tau_s \\ x_s \geq \mathbf{0}_n \\ x_t^{\star} = x_t \end{array}\right. \qquad \text{for } s = t+1, \ldots, t+h \end{aligned}$$

Using the augmented QP (or a-QP) algorithm and the block coordinate descent method[7] (or BCD), we obtain the results given in Tables 2 and 3.

We notice that we obtain the same solution for $x_{t+5}^{\star}$, but the path $\left(x_{t+1}^{\star}, \ldots, x_{t+h}^{\star}\right)$ is not the same if we compare the iterative, a-QP and BCD solutions. Moreover, the total

[7] The BCD algorithm is initialized with $x^{(0)} = (x_t, \ldots, x_t)$.

Table 2: Augmented QP solution in % (transition management)

| $s$ | $t$ | $t+1$ | $t+2$ | $t+3$ | $t+4$ | $t+5$ |
|---|---|---|---|---|---|---|
| $x_{1,s}$ | 20.00 | 20.00 | 20.10 | 28.05 | 40.52 | 46.21 |
| $x_{2,s}$ | 10.00 | 22.50 | 34.90 | 39.45 | 39.48 | 38.21 |
| $x_{3,s}$ | 15.00 | 15.00 | 15.00 | 15.00 | 3.83 | 0.00 |
| $x_{4,s}$ | 20.00 | 20.00 | 11.43 | 5.17 | 5.10 | 4.09 |
| $x_{5,s}$ | 10.00 | 10.00 | 10.00 | 5.57 | 4.70 | 4.09 |
| $x_{6,s}$ | 15.00 | 7.76 | 5.78 | 5.78 | 5.78 | 7.11 |
| $x_{7,s}$ | 10.00 | 4.74 | 2.80 | 0.98 | 0.59 | 0.30 |
| $\tau_s$ | | 25.00 | 25.00 | 25.00 | 25.00 | 14.02 |

Table 3: Block CD solution in % (transition management)

| $s$ | $t$ | $t+1$ | $t+2$ | $t+3$ | $t+4$ | $t+5$ |
|---|---|---|---|---|---|---|
| $x_{1,s}$ | 20.00 | 20.00 | 20.05 | 27.72 | 40.18 | 46.21 |
| $x_{2,s}$ | 10.00 | 22.50 | 34.95 | 39.78 | 39.82 | 38.21 |
| $x_{3,s}$ | 15.00 | 15.00 | 15.00 | 15.00 | 3.90 | 0.00 |
| $x_{4,s}$ | 20.00 | 20.00 | 11.61 | 5.54 | 5.37 | 4.09 |
| $x_{5,s}$ | 10.00 | 10.00 | 10.00 | 5.61 | 4.77 | 4.09 |
| $x_{6,s}$ | 15.00 | 7.76 | 5.56 | 5.33 | 5.33 | 7.11 |
| $x_{7,s}$ | 10.00 | 4.74 | 2.82 | 1.03 | 0.63 | 0.30 |
| $\tau_s$ | | 25.00 | 25.00 | 25.00 | 25.00 | 16.52 |

turnover $\sum_{s=t+1}^{t+h} \left\| x_s - x_{s-1}^{\star} \right\|_1$ is respectively equal to 116.83%, 114.02% and 116.52%. It is therefore larger than the figure 108.82% computed previously. This is because we can buy more than the target at a rebalancing date $s$ and then sell a part at the next rebalancing date $s+1$. As a result, the total turnover may be greater than 108.82%. In order to find an optimal path, we can introduce transaction costs. For instance, if we consider quadratic transaction costs, the objective function becomes:

$$f_s(x_s) = \frac{1}{2} x_s^\top \Sigma_t x_s - \gamma x_s^\top \mu_t + \frac{1}{2} (x_s - x_{s-1})^\top \Lambda_s (x_s - x_{s-1}) \tag{69}$$

Nevertheless, by modifying the mean-variance function, we are not certain to obtain the right solution. In particular, we can face the following two extreme situations:

1. If the matrix $\Lambda_s$ is too small, we obtain the previous solutions because the penalization is too low;

2. If the matrix $\Lambda_s$ is too big, we obtain a solution which is far from optimal. In fact, the risk is that the solution sticks to the current allocation ($x_{t+1}^{\star} = \ldots = x_{t+h}^{\star} = x_t$) because the penalization is too prohibitive.

Let $x_t$ and $x^{\star}$ be the current and target portfolios. A transition management process requires that the allocation to an asset can only be increasing or decreasing:

$$\begin{cases} x_i^{\star} \geq x_{i,t} \Rightarrow x_{i,s} \geq x_{i,s-1} \\ x_i^{\star} \leq x_{i,t} \Rightarrow x_{i,s} \leq x_{i,s-1} \end{cases} \tag{70}$$

This means that we cannot decrease the allocation to asset $i$ at a given rebalancing date $s$ if its weight in the target portfolio is larger than its weight in the current portfolio. In

Appendix C.8 on page 60, we show how to translate the monotonic property into inequality constraints. We report the results in Tables 11, 12 and 13 on page 31. In this case, we obtain similar results with each of the three methods.

### 2.5.2 Total variation regularized portfolio

**Example 2.** *We consider an investment universe of* 7 *stocks. The values of their idiosyncratic risk* $\tilde{\sigma}_i$ *and beta* $\beta_i$ *are equal to:*

| *Stock i* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $\tilde{\sigma}_i$ | 3% | 5% | 15% | 16% | 10% | 8% | 10% |
| $\beta_i$ | −0.50 | −0.50 | 0.00 | 0.50 | 1.00 | 1.75 | 2.00 |

*We also assume that the market risk volatility* $\sigma_m$ *is equal to* 20%. *The current portfolio corresponds to the global minimum variance (GMV) portfolio*[8]*:* $x_t = (54.15\%, 19.50\%, 2.30\%, 2.14\%, 5.78\%, 9.74\%, 6.39\%)$.

We consider the following $\boldsymbol{\ell}_1$-regularization problem[9]:

$$\begin{aligned} x_{t+1}^{\star} &= \arg\min \sum_{s=t+1}^{t+h} \left\{ \frac{1}{2} x_s^{\top} \Sigma x_s + \lambda_s \left\| x_s - x_{s-1} \right\|_1 \right\} \\ & \text{s.t.} \quad \left\{ \begin{array}{l} \mathbf{1}_n^{\top} x_s = 1 \\ \beta\left(x_s\right) \geq \beta_s^{-} \end{array} \right. \end{aligned} \tag{72}$$

where $\Sigma = \beta\beta^{\top}\sigma_m^2 + \operatorname{diag}\left(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_n^2\right)$ is the one-factor covariance matrix, $\lambda_s$ is the penalization factor at time $s$, $\beta\left(x_s\right)$ is the beta of the portfolio $x_s$ and $\beta_s^{-}$ is the minimum beta at time $s$. Since we have $\beta\left(x_s\right) = \sum_{i=1}^{n} x_{i,s}\beta_i = x_s^{\top}\beta$, the beta constraint is equivalent to the system $C_s x_s \leq D_s$ where $C_s = -\beta^{\top}$ and $D_s = -\beta_s^{-}$. Problem (72) can be solved using the ADMM algorithm where the $x$-update corresponds to a series of QP problems and the $y$-update is the proximal operator of the function $\zeta_{\varphi}\left(x; x_t, \boldsymbol{\lambda}\right)$.

Let us assume that $\beta_s^{-} = \beta_0 \cdot \big(s - (t+1)\big)$ and $\lambda_s = \lambda_0 \cdot (s - t)$. We solve Problem (72) when $\beta_0 = 0.125$ and $\lambda_0$ takes different values (1%, 2%, 3%, 4%). Moreover, we study the impact of the time horizon $h$ on the solution. The optimal weights $x_{i,t+1}^{\star}$ are reported in Figure 1 for the first, second, sixth and seventh assets. If $\lambda_0$ is equal to zero, we can verify that the optimal solution $x_{t+1}^{\star}$ does not depend on the time horizon $h$ because this is not a coupling problem[10]. If $\lambda_0 > 0$, the optimal solution $x_{t+1}^{\star}$ depends on the time horizon $h$. For instance, the weight of the seventh stock increases when $h$ is small and then decreases when $h$ is high. The weight of the sixth stock does not change and then increases whereas the weight of the first stock decreases initially and then reaches a floor.

[8] The GMV portfolio is given by minimizing the portfolio variance without any constraints:

$$\begin{aligned} x_t &= \arg\min \left\{ \frac{1}{2} x^{\top} \Sigma x \right\} \\ & \text{s.t.} \quad \mathbf{1}_n^{\top} x = 1 \end{aligned} \tag{71}$$

where $\Sigma = \beta\beta^{\top}\sigma_m^2 + \operatorname{diag}\left(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_n^2\right)$ is the one-factor covariance matrix.

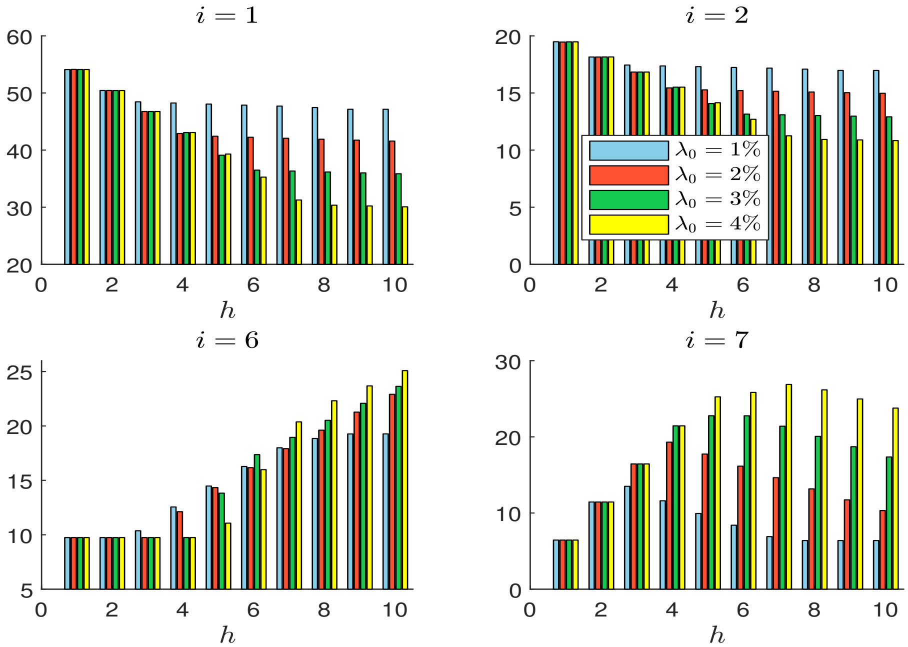
[9] This is a special case of total variation regularization problems (Corsaro *et al.*, 2020, 2021).

[10] Indeed, Problem (72) reduces to $h$ problems:

$$\begin{aligned} x_s^{\star} &= \arg\min \left\{ \frac{1}{2} x_s^{\top} \Sigma x_s \right\} \\ & \text{s.t.} \left\{ \begin{array}{l} \mathbf{1}_n^{\top} x_s = 1 \\ \beta\left(x_s\right) \geq \beta_s^{-} \end{array} \right. \end{aligned} \tag{73}$$

when $\lambda_s = 0$.

Figure 1: Regularized minimum variance portfolio $x_{t+1}^{\star}$ in % ($\beta_0 = 0.125$)




### 2.5.3 Trading trajectory problem

We consider the trading trajectory problem, which consists in finding the optimal trading strategy while considering trading costs. If we use a mean-variance framework, the objective function becomes:

$$\begin{aligned} f_s(x_s) &= \frac{1}{2} x_s^\top \Sigma_s x_s - x_s^\top \gamma \mu_s + \\ & \quad \mathbb{TC}_s(x_{s-1}, x_s) + \mathbb{PI}_s(x_{s-1}, x_s) \end{aligned} \tag{74}$$

where $\mathbb{TC}_s(x_{s-1}, x_s)$ and $\mathbb{PI}_s(x_{s-1}, x_s)$ are the transaction costs and price impacts. Following Gârleanu and Pedersen (2013, page 2320), we have:

$$\mathbb{TC}_s(x_{s-1}, x_s) = \frac{1}{2} (x_s - x_{s-1})^\top \Lambda_s (x_s - x_{s-1}) \tag{75}$$

and:

$$\begin{aligned} \mathbb{PI}_s(x_{s-1}, x_s) &= \phi x_s^\top \Gamma_s (x_s - x_{s-1}) - x_{s-1}^\top \Gamma_s (x_s - x_{s-1}) - \\ & \quad \frac{1}{2} (x_s - x_{s-1})^\top \Gamma_s (x_s - x_{s-1}) \end{aligned} \tag{76}$$

where $\phi$ is the mean-reversion parameter of the price distortion, $\Lambda_s$ and $\Gamma_s$ are Kyle's matrices for temporary trading costs and permanent price impacts.

**Remark 4.** *To understand the previous formula, we assume that the number $n$ of assets is equal to one and a linear price impact model (Roncalli et al., 2021). If we continuously sell or buy the security between $s-1$ and $s$, the average change is equal to $\frac{1}{2}\Delta x_s$. The unit* *transaction cost is then equal to* $c_s = \frac{1}{2}\lambda_s \Delta x_s$ *whereas the total transaction cost is:*

$$\mathbb{TC}_s\left(x_{s-1}, x_s\right) = c_s \cdot \Delta x_s = \frac{1}{2}\lambda_s \left(\Delta x_s\right)^2 \tag{77}$$

*At the same time, we observe a price distortion during the period* $[s-1, s+1]$*. If* $\Delta x_s > 0$*, this implies that the price increases between* $s-1$ *and* $s$*. By assuming that the return increases by* $\Delta\mu_s = \gamma_s \Delta x_s$*, we obtain the following potential gain:*

$$\mathbb{G}_s\left(x_{s-1}, x_s\right) = \Delta\mu_s \cdot \left(x_{s-1} + \frac{1}{2}\Delta x_s\right) = \gamma_s x_{s-1}\Delta x_s + \frac{1}{2}\gamma_s \left(\Delta x_s\right)^2 \tag{78}$$

*Nevertheless, the price distortion can mean-revert, implying a potential loss between* $s$ *and* $s+1$*:*

$$\mathbb{L}_s\left(x_{s-1}, x_s\right) = -\Delta\mu_{s+1} \cdot x_s = \phi\Delta\mu_s \cdot x_s = \phi\gamma_s x_s \Delta x_s \tag{79}$$

*Finally, we obtain:*

$$\begin{aligned}\mathbb{PI}_s\left(x_{s-1}, x_s\right) &= \mathbb{L}_s\left(x_{s-1}, x_s\right) - \mathbb{G}_s\left(x_{s-1}, x_s\right) \\ &= \phi\gamma_s x_s \Delta x_s - \gamma_s x_{s-1}\Delta x_s - \frac{1}{2}\gamma_s\left(\Delta x_s\right)^2 \end{aligned} \tag{80}$$

*Formula (76) is a generalization of this equation in the multi-dimensional case.*

Rosenberg *et al.* (2016) proposed another formulation[11]:

$$\begin{aligned} f_s\left(x_s\right) &= \frac{1}{2}x_s^\top \Sigma_s x_s - x_s^\top \gamma\mu_s + \\ & \left(x_s - x_{s-1}\right)^\top \Lambda_s'\left(x_s - x_{s-1}\right) - x_s^\top \Gamma_s'\left(x_s - x_{s-1}\right) \end{aligned} \tag{81}$$

The transaction costs and the price impacts differ slightly. Indeed, the authors did not breakdown the effects between $s-1$ and $s+1$, assumed discrete trading instead of continuous trading[12] and used a net impact[13]. The two formulations can be cast into the following function:

$$\begin{aligned} f_s\left(x_s\right) &= \frac{1}{2}x_s^\top \Sigma_s x_s - x_s^\top\left(\gamma\mu_s - \phi\Gamma_s\Delta x_s\right) + \\ & \frac{1}{2}\Delta x_s^\top \Lambda_s \Delta x_s - \varepsilon\left(x_{s-1}^\top \Gamma_s \Delta x_s + \frac{1}{2}\Delta x_s^\top \Gamma_s \Delta x_s\right) \end{aligned} \tag{83}$$

The formulation of Gârleanu and Pedersen (2013) is obtained with $\varepsilon = 1$ whereas the formulation of Rosenberg *et al.* (2016) corresponds to $\phi = -1$, $\varepsilon = 0$, $\Lambda_s = 2\Lambda_s'$ and $\Gamma_s = \Gamma_s'$.

[11] Steinhauer *et al.* (2020) used an optimization procedure based on simulated bifurcation to solve this problem, which can be an alternative approach to the quantum annealer suggested by Rosenberg *et al.* (2016).

[12] This explains that the factor 1/2 vanishes.

[13] We have:

$$\begin{aligned}\mathbb{PI}_s\left(x_{s-1}, x_s\right) &= \phi x_s^\top \Gamma_s \Delta x_s - x_{s-1}^\top \Gamma_s \Delta x_s - \Delta x_s^\top \Gamma_s \Delta x_s \\ &= \phi x_s^\top \Gamma_s \Delta x_s - x_s^\top \Gamma_s \Delta x_s \\ &= -\left(1-\phi\right) x_s^\top \Gamma_s \Delta x_s \\ &= -x_s^\top \Gamma_s' \Delta x_s \end{aligned} \tag{82}$$

where $\Gamma_s' = \left(1-\phi\right)\Gamma_s$ is the Kyle's matrix for net price impacts, $\left(1-\phi\right)$ is the net factor and $\Gamma_s$ is the Kyle's matrix for gross price impacts.

In Appendix C.9 on page 61, we show that:

$$f_s(x_s) = \frac{1}{2} x_s^\top Q_s^{0,0} x_s - x_s^\top Q_s^{0,1} x_{s-1} + \frac{1}{2} x_{s-1}^\top Q_s^{1,1} x_{s-1} - x_s^\top R_s \tag{84}$$

where $Q_s^{0,0} = Q_s + \Lambda_s + (2\phi - \varepsilon)\Gamma_s$, $Q_s^{0,1} = \Lambda_s + \phi\Gamma_s$, $Q_s^{1,1} = \Lambda_s + \varepsilon\Gamma_s$ and $R_s = \gamma_s \mu_s$. To solve the trading trajectory problem, we can use the block coordinate descent. Let us introduce the notation $f_s(x_{s-1}, x_s) := f_s(x_s)$. A first idea is to consider the following iterative step:

$$x_s^{(k+1)} = \arg\min_{x_s} f_s\left(x_{s-1}^{(k+1)}, x_s\right) \tag{85}$$

Nevertheless, the cyclical BCD algorithm stops after the first cycle and does not converge because the coupling is between $x_{s-1}$ and $x_s$, implying that the new coordinate $x_s^{(k+1)}$ has no impact on the previous coordinates $x_u^{(k+1)}$ for $u < s$. Empirical experiments show that this issue remains if we use the random BCD algorithm. Let us consider the function $f_s(x_{s-1}, x_s, x_{s+1})$, whose coupling concerns $x_{s-1}$, $x_s$ and $x_{s+1}$:

$$\begin{aligned} f_s(x_{s-1}, x_s, x_{s+1}) \quad := \quad & \frac{1}{2} x_s^\top Q_s^{0,0} x_s - x_s^\top Q_s^{0,1} x_{s-1} + \frac{1}{2} x_{s-1}^\top Q_s^{1,1} x_{s-1} - x_s^\top R_s + \\ & \frac{1}{2} x_{s+1}^\top Q_{s+1}^{0,0} x_{s+1} - x_{s+1}^\top Q_{s+1}^{0,1} x_s + \frac{1}{2} x_s^\top Q_{s+1}^{1,1} x_s - x_{s+1}^\top R_{s+1} \end{aligned} \tag{86}$$

The iterative step becomes:

$$x_s^{(k+1)} = \arg\min_{x_s} f_s\left(x_{s-1}^{(k+1)}, x_s, x_{s+1}^{(k)}\right) = \arg\min_{x_s} f_s^\star\left(x_{s-1}^{(k+1)}, x_s, x_{s+1}^{(k)}\right) \tag{87}$$

where:

$$f_s^\star\left(x_{s-1}^{(k+1)}, x_s, x_{s+1}^{(k)}\right) = \frac{1}{2} x_s^\top \left(Q_s^{0,0} + Q_{s+1}^{1,1}\right) x_s - x_s^\top \left(R_s + Q_s^{0,1} x_{s-1}^{(k+1)} + Q_{s+1}^{0,1} x_{s+1}^{(k)}\right) \tag{88}$$

If the constraints are linear, $x_s^{(k+1)}$ is the solution of a QP problem. Concerning the ADMM algorithm, this approach is not really appropriate because the objective function is not separable[14]. Finally, we can also use the QP problem if the constraints are linear[15].

**Example 3.** *We consider a universe of four assets. Their expected returns are equal to* 5%, 6%, 7% *and* 8% *while their volatilities are equal to* 15%, 20%, 25% *and* 30%. *The correlation matrix of asset returns is given by the following matrix:*

$$C = \begin{pmatrix} 1.00 & & & \\ 0.10 & 1.00 & & \\ 0.40 & 0.70 & 1.00 & \\ 0.50 & 0.40 & 0.40 & 1.00 \end{pmatrix}$$

[14] Indeed, we have:

$$f_x(x) = \sum_{s=t+1}^{t+h} \left\{ \frac{1}{2} x_s^\top Q_s^{0,0} x_s - x_s^\top R_s \right\} \tag{89}$$

and:

$$f_y(y) = \sum_{s=t+1}^{t+h} \left\{ \frac{1}{2} y_{s-1}^\top Q_s^{1,1} y_{s-1} - y_s^\top Q_s^{0,1} y_{s-1} \right\} \tag{90}$$

$f_x(x)$ is separable, but not $f_y(y)$.

[15] The derivation of the QP problem is given on page 61.

In what follows, we use the model of Gârleanu and Pedersen (2013). We assume that the transaction costs and price impacts are proportional to asset volatilities. We consider the stationary model: $\Sigma_s = \Sigma$, $\mu_s = \mu$, $\gamma = 1$, $\Lambda_s = \varrho_\Lambda S$ and $\Gamma_s = \varrho_\Gamma S$, where $S = \operatorname{diag}(\sigma_1, \ldots, \sigma_4)$ is the diagonal matrix of volatilities. We estimate the optimal trading trajectory $\left\{x^*_{t+1}, \ldots, x^*_{t+h}\right\}$ by considering that the current portfolio $x_t$ is the equally-weighted portfolio. The mean-variance optimized portfolio is equal to $x_{\text{mvo}} = \left(20.39\%, 23.11\%, 24.74\%, 31.76\%\right)$. If we set $\varrho_\Lambda = \varrho_\Gamma = 0$, we obtain $x^*_{t+1} = \ldots = x^*_{t+h} = x_{\text{mvo}}$. If $\varrho_\Lambda = \infty$, we verify that $x^*_{t+1} = \ldots = x^*_{t+h} = x_t$. In Table 4, we report the optimal trajectory for $\varrho_\Lambda = 5\%$ and two values of $\varrho_\Gamma$. Moreover, we consider three values of the mean-reversion parameter $\phi$: 0%, 50% and 100%. It is interesting to notice that the solution at time $t+1$ is largely the same, but the solution at time $t+5$ is very different. In particular, we notice that the turnover is a decreasing function of the mean-reversion parameter $\phi$ and an increasing function of the price impact parameter $\varrho_\Gamma$. The mean-reversion effect disappears when $\phi = 0$, implying that the trader can create a "*momentum ignition*". This is why we observe a monotonic allocation increase in the third and fourth assets between $t+1$ and $t+5$. This momentum pattern also depends on the price impact magnitude, which explains why it is stronger for $\varrho_\Gamma = 10\%$ than for $\varrho_\Gamma = 1\%$. When $\phi = 1$, the price momentum at time $s$ is offset by the price reversal at time $s+1$. Therefore, the trading gains are limited and the turnover is reduced.

Table 4: Optimal trading trajectory in % ($\gamma = 1$, $\varrho_\Lambda = 5\%$)

| $\phi$ | $s$ | $\varrho_\Gamma = 1\%$ | | | | $\varrho_\Gamma = 10\%$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $x^\star_{1,s}$ | $x^\star_{2,s}$ | $x^\star_{3,s}$ | $x^\star_{4,s}$ | $x^\star_{1,s}$ | $x^\star_{2,s}$ | $x^\star_{3,s}$ | $x^\star_{4,s}$ |
| 0% | $t+1$ | 21.48 | 23.60 | 24.53 | 30.40 | 21.40 | 23.34 | 24.81 | 30.46 |
| | $t+2$ | 20.63 | 23.24 | 24.64 | 31.48 | 20.32 | 22.38 | 25.58 | 31.73 |
| | $t+3$ | 20.40 | 23.11 | 24.75 | 31.74 | 19.20 | 20.48 | 27.66 | 32.66 |
| | $t+4$ | 20.20 | 22.97 | 24.91 | 31.92 | 15.45 | 15.18 | 33.77 | 35.60 |
| | $t+5$ | 19.48 | 22.69 | 25.26 | 32.56 | 0.00 | 0.00 | 52.13 | 47.87 |
| 50% | $t+1$ | 21.54 | 23.62 | 24.52 | 30.31 | 21.93 | 23.69 | 24.63 | 29.75 |
| | $t+2$ | 20.67 | 23.27 | 24.62 | 31.44 | 20.74 | 23.12 | 24.86 | 31.28 |
| | $t+3$ | 20.44 | 23.14 | 24.72 | 31.71 | 20.00 | 22.62 | 25.35 | 32.03 |
| | $t+4$ | 20.30 | 23.06 | 24.81 | 31.83 | 18.77 | 21.85 | 26.26 | 33.12 |
| | $t+5$ | 19.96 | 22.92 | 24.97 | 32.14 | 15.45 | 20.36 | 28.07 | 36.11 |
| 100% | $t+1$ | 21.61 | 23.65 | 24.52 | 30.23 | 22.31 | 23.92 | 24.52 | 29.25 |
| | $t+2$ | 20.71 | 23.29 | 24.60 | 31.40 | 21.19 | 23.49 | 24.51 | 30.81 |
| | $t+3$ | 20.47 | 23.18 | 24.68 | 31.67 | 20.72 | 23.30 | 24.58 | 31.39 |
| | $t+4$ | 20.41 | 23.14 | 24.71 | 31.73 | 20.54 | 23.22 | 24.63 | 31.61 |
| | $t+5$ | 20.40 | 23.12 | 24.73 | 31.75 | 20.47 | 23.19 | 24.66 | 31.68 |

**Remark 5.** *We notice that the momentum ignition implies an arbitrage opportunity in the last period when $\phi < 1$. Indeed, the multi-period optimization does not consider profits and losses after the period $t+h$. In order to eliminate this free lunch, we can impose solving the problem by considering the extended period $[t+1, t+h+1]$ and imposing the constraint $x_{t+h} = x_{t+h+1}$. This last restriction means that trading is stopped after $s > t+h$. Since we include the period $s = t+h+1$ in the multi-period optimization problem, the objective function takes into account profits and losses after the period $t+h$. This boundary condition is particularly relevant when the mean-reversion parameter $\phi$ is close to zero. Let us consider the previous example. Results obtained with the boundary condition are reported in Table 14 on page 32. A comparison of the two approaches is provided in Table 5 for the case*

$\varrho_\Gamma = 10\%$ *and* $\phi = 0$. *Whereas the solution* $x^\star_{t+5}$ *is equal to* $(0\%, 0\%, 52.13\%, 47.87\%)$ *without the boundary condition, it becomes* $(12.82\%, 16.81\%, 32.08.13\%, 38.29\%)$ *when the correction is implemented. The two solutions are very different, because the performance of the trader is not impacted by the P&L in the period* $t+6$ *in the first approach.*

Table 5: Impact of the boundary condition on the optimal solution ($\gamma = 1$, $\varrho_\Lambda = 5\%$, $\varrho_\Gamma = 10\%$, $\phi = 0$)

| $s$ | Without the boundary condition | | | | With the boundary condition | | | |
|---|---|---|---|---|---|---|---|---|
| | $x^\star_{1,s}$ | $x^\star_{2,s}$ | $x^\star_{3,s}$ | $x^\star_{4,s}$ | $x^\star_{1,s}$ | $x^\star_{2,s}$ | $x^\star_{3,s}$ | $x^\star_{4,s}$ |
| $t+1$ | 21.40 | 23.34 | 24.81 | 30.46 | 21.45 | 23.53 | 24.60 | 30.42 |
| $t+2$ | 20.32 | 22.38 | 25.58 | 31.73 | 20.53 | 23.01 | 24.89 | 31.56 |
| $t+3$ | 19.20 | 20.48 | 27.66 | 32.66 | 20.01 | 22.42 | 25.52 | 32.05 |
| $t+4$ | 15.45 | 15.18 | 33.77 | 35.60 | 18.61 | 20.95 | 27.21 | 33.23 |
| $t+5$ | 0.00 | 0.00 | 52.13 | 47.87 | 12.82 | 16.81 | 32.08 | 38.29 |

**Remark 6.** *In practice, the mean-reversion parameter* $\phi$ *is stochastic because we do not really know how the market will react. Moreover,* $\phi$ *also depends on the size of the trade. Therefore, it is better to consider several values of* $\phi$ *in order to analyze the several trading scenarios and the different possible outcomes.*

# 3 Application to portfolio decarbonization

The aim of portfolio decarbonization is to construct an investment portfolio that tracks a benchmark portfolio but with a lower carbon risk metric, which is generally the carbon intensity (Le Guenedal and Roncalli, 2022). The concept of portfolio decarbonization has been extended by taking into account a carbon trajectory (Le Guenedal *et al.*, 2022). While portfolio decarbonization is a static problem, portfolio alignment implies a dynamic approach in order to comply with a given climate policy (e.g. Paris-based benchmark approach or net zero carbon objective approach). Therefore, the portfolio construction becomes a multi-period portfolio optimization problem with time-varying constraints. The constraints impose both a decarbonization path for the dynamic portfolio and a minimum financing level for sectors that are essential to the transition to a low-carbon economy. Nevertheless, these constraints are not always coherent since we know that there is a negative correlation between carbon intensities and green revenues for instance. Therefore, the dynamic portfolio may involve rebalancing allocations that are not always optimal. For example, the weight of an asset may increase in a first period and then decrease in a second period because one constraint becomes tighter with time.

## 3.1 Definition of the optimization problem

Le Guenedal and Roncalli (2022) consider the following optimization problem:

$$x^\star_{t+1} = \arg\min \sum_{s=t+1}^{t+h} \left\{ e^{-\varrho_1(s-t-1)} \frac{1}{2} \sigma^2 \left(x_s \mid b_s\right) + \lambda_s e^{-\varrho_2(s-t-1)} \tau\left(x_{s-1}, x_s\right) \right\} \quad (91)$$

$$\text{s.t.} \quad \begin{cases} \mathbf{1}_n^\top x_s = 1 \\ x_s \geq \mathbf{0}_n \\ \boldsymbol{\mathcal{CI}}\left(x_s\right) \leq \left(1 - \boldsymbol{\mathcal{R}}\left(t, s\right)\right) \cdot \boldsymbol{\mathcal{CI}}\left(b_t\right) \\ \mathcal{CIS}_{\mathcal{H}igh}\left(x_s\right) \geq \varphi_{\mathcal{CIS}} \cdot \mathcal{CIS}_{\mathcal{H}igh}\left(b_t\right) \end{cases}$$

where $\varrho_1$ and $\varrho_2$ are the discount rates, $\sigma^2\left(x_s \mid b_s\right)$ is the tracking error variance of the investment portfolio $x_s$ with respect to the investment benchmark $b_s$:

$$\sigma^2\left(x_s \mid b_s\right) = \left(x_s - b_s\right)^\top \Sigma_s \left(x_s - b_s\right) \tag{92}$$

and $\tau\left(x_{s-1}, x_s\right)$ is the turnover ratio between $s-1$ and $s$:

$$\tau\left(x_{s-1}, x_s\right) = \left\| x_s - x_{s-1} \right\|_1 \tag{93}$$

The objective function defines a classical tracking problem where we would like to minimize the tracking risk and limit the turnover. However, Problem (91) has two "*climate investing*" constraints. The first one imposes that the carbon intensity $\boldsymbol{\mathcal{CI}}\left(x_s\right)$ of the portfolio at time $s > t$ is less than the carbon intensity $\boldsymbol{\mathcal{CI}}\left(b_t\right)$ of the benchmark and the reduction between $t$ and $s$ is denoted by $\boldsymbol{\mathcal{R}}\left(t, s\right)$. This constraint defines a decarbonization pathway and we have:

$$\boldsymbol{\mathcal{R}}\left(t, t+1\right) < \boldsymbol{\mathcal{R}}\left(t, t+2\right) < \cdots < \boldsymbol{\mathcal{R}}\left(t, t+h\right) \tag{94}$$

This implies that the carbon intensity of Portfolio $x_s$ must decrease with the time. The second constraint imposes that the portfolio has a minimum exposure on high climate impact sectors (CIS).

## 3.2 Numerical solution of the optimization problem

The decarbonization and CIS constraints can be written as:

$$\sum_{i=1}^{n} \boldsymbol{\mathcal{CI}}_i \cdot x_{i,s} \leq \left(1 - \boldsymbol{\mathcal{R}}\left(t, s\right)\right) \cdot \boldsymbol{\mathcal{CI}}\left(b_t\right) \tag{95}$$

and:

$$\sum_{i=1}^{n} \mathbb{1}\left\{i \in \mathcal{CIS}_{\mathcal{H}igh}\right\} \cdot x_{i,s} \geq \varphi_{\mathcal{CIS}} \cdot \mathcal{CIS}_{\mathcal{H}igh}\left(b_t\right) \tag{96}$$

Let $\xi_i = \mathbb{1}\left\{i \in \mathcal{CIS}_{\mathcal{H}igh}\right\}$ be the indicator function which is equal to 1 if the asset $i$ belongs the high CIS class or 0 otherwise. We can combine the two constraints in order to obtain the inequality system $C_s x_s \leq D_s$ where:

$$C_s = \begin{pmatrix} \boldsymbol{\mathcal{CI}}_1 & \boldsymbol{\mathcal{CI}}_2 & \cdots & \boldsymbol{\mathcal{CI}}_n \\ -\xi_1 & -\xi_2 & \cdots & -\xi_n \end{pmatrix} \tag{97}$$

and:

$$D_s = \begin{pmatrix} \left(1 - \boldsymbol{\mathcal{R}}\left(t, s\right)\right) \cdot \boldsymbol{\mathcal{CI}}\left(b_t\right) \\ -\varphi_{\mathcal{CIS}} \cdot \mathcal{CIS}_{\mathcal{H}igh}\left(b_t\right) \end{pmatrix} \tag{98}$$

Problem (91) becomes[16]:

$$\begin{aligned} x_{t+1}^\star &= \arg\min \sum_{s=t+1}^{t+h} \left\{g_s\left(x_s\right) + h_s\left(x_{s-1}, x_s\right)\right\} \\ &\text{s.t.} \quad \begin{cases} A_s x_s = B_s \\ C_s x_s \leq D_s \\ \mathbf{0}_n \leq x_s \end{cases} \end{aligned} \tag{99}$$

[16] We have $A_s = \mathbf{1}_n^\top$ and $B_s = 1$.

The function $g_s(x_s)$ has a quadratic form:

$$g_s(x_s) = \frac{1}{2} x_s^\top Q_s x_s - x_s^\top R_s \tag{100}$$

where $Q_s = e^{-\varrho_1(s-t-1)}\Sigma_s$ and $R_s = e^{-\varrho_1(s-t-1)}\Sigma_s b_s$. The coupling function $h_s(x_{s-1}, x_s)$ is a $\ell_1$-norm penalty function:

$$h_s(x_{s-1}, x_s) = \lambda'_s \|x_s - x_{s-1}\|_1 \tag{101}$$

where $\lambda'_s = \lambda_s e^{-\varrho_2(s-t-1)}$. Therefore, we can solve Problem (91) using the ADMM, BCD or a-QP algorithms.

## 3.3 Illustration

### 3.3.1 A toy example

**Example 4.** *We consider an investment universe of* 10 *stocks with the following characteristics:*

| *Stock i* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\tilde{\sigma}_i$ *(in %)* | 15 | 31 | 21 | 19 | 27 | 23 | 41 | 28 | 22 | 21 |
| $\beta_i$ | 0.52 | 1.15 | 1.06 | 0.29 | 0.44 | 1.06 | 1.39 | 1.51 | 0.67 | 0.29 |
| $b_i$ *(in %)* | 17.25 | 15.75 | 13.68 | 11.40 | 10.29 | 9.56 | 7.56 | 5.39 | 5.85 | 3.27 |
| $\mathcal{CI}_i$ | 747.7 | 30.05 | 500.6 | 58.87 | 111.7 | 1 082 | 408 | 29.0 | 80.1 | 45.7 |
| $\xi_i$ | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |

*where $\tilde{\sigma}_i$ is the idiosyncratic volatility (expressed in %), $\beta_i$ is the beta, $b_i$ is the weight in the benchmark (expressed in %), $\mathcal{CI}_i$ is the carbon intensity (expressed in $tCO_2e$/\$ mn) and $\xi_i$ is the high CIS indicator function. We also assume that the market risk volatility $\sigma_m$ is equal to* 25%*.*

At time $t$, the current portfolio $x_t$ exactly replicates the benchmark $b_t$. Its carbon intensity is then equal to 362.4 $tCO_2e$/\$ mn, whereas high climate impact sectors represent 46.34% of the allocation. At time $t+1, t+2, \ldots$, we would like to implement the following decarbonization path: $\mathcal{R}(t, t+h) = 15\% \cdot h$. This means that the carbon intensity is reduced by 15% every year from the base year. Moreover, we assume that $\varphi_{\mathcal{CIS}}$ is equal to 1, implying that the exposure to high climate impact sectors must not decrease. The discount rates $\varrho_1$ and $\varrho_2$ are set to zero. Results are reported in Tables 6 and 7. The solution $x_s$, the tracking error $\sigma(x_s \mid b_s)$, the turnover $\tau(x_{s-1}, x_s)$, the weight $\mathcal{CIS}_{\mathcal{H}igh}(x_s)$ of high climate impact sectors and the reduction rate $\mathcal{R}(t, s)$ of the carbon intensity are expressed in %, whereas the carbon intensity $\mathcal{CI}(x_s)$ of the portfolio is measured in $tCO_2e$/\$ mn. If we consider the single-period solution without turnover control ($h = 1$, $\lambda_s = 0$), the tracking-error volatility is respectively equal to 1.59% at time $t+1$, 3.18% at time $t+2$ and 4.81% at time $t+3$. In order to reduce the turnover, we add the $\ell_1$-norm penalty with $\lambda_s = 0.5\%$. In this case, the tracking-error volatility is increased by 23 bps on average[17], but the turnover is decreased by 403 bps for the period $[t+1, t+3]$.

If we analyze the multi-period solution, we notice that the optimal portfolio depends on two parameters: the penalty factor $\lambda_s$ and the number $h$ of periods. We observe that the tracking-error volatility $\sigma(x_s \mid b_s)$ for $h > 1$ and a given value of $\lambda_s$ is greater than the value obtained for $h = 1$ without the penalty function and less than the value obtained for $h = 1$ with the penalty function. In fact, the optimal solution $x_s$ anticipates the future reductions of the carbon intensity for the period $[s+1, s+h-1]$, which explains that the multi-period solution better exploits the trade-off between tracking-error volatility and turnover.

[17] We have $\tau(x_t, x_{t+1}) = 1.81\%$, $\tau(x_{t+1}, x_{t+2}) = 3.47\%$ and $\tau(x_{t+2}, x_{t+3}) = 5.01\%$.

Table 6: Single-period solution ($h = 1$)

| | $\lambda_s = 0$ | | | $\lambda_s = 0.5\%$ | | |
|---|---|---|---|---|---|---|
| $s$ | $t+1$ | $t+2$ | $t+3$ | $t+1$ | $t+2$ | $t+3$ |
| $x_{1,s}$ | 14.45 | 11.65 | 6.40 | 17.25 | 15.31 | 7.69 |
| $x_{2,s}$ | 16.12 | 16.49 | 16.83 | 15.75 | 15.75 | 15.86 |
| $x_{3,s}$ | 15.16 | 16.65 | 17.54 | 13.68 | 13.68 | 13.68 |
| $x_{4,s}$ | 11.40 | 11.40 | 11.68 | 11.40 | 11.40 | 11.40 |
| $x_{5,s}$ | 10.01 | 9.72 | 9.42 | 10.29 | 10.29 | 10.29 |
| $x_{6,s}$ | 5.70 | 1.84 | 0.00 | 4.13 | 0.00 | 0.00 |
| $x_{7,s}$ | 6.76 | 5.97 | 4.77 | 7.56 | 7.56 | 6.63 |
| $x_{8,s}$ | 5.96 | 6.54 | 7.00 | 5.39 | 5.39 | 6.21 |
| $x_{9,s}$ | 11.03 | 16.20 | 22.40 | 11.28 | 17.35 | 24.97 |
| $x_{10,s}$ | 3.41 | 3.55 | 3.96 | 3.27 | 3.27 | 3.27 |
| $\sigma\left(x_s \mid b_s\right)$ | 1.59 | 3.18 | 4.81 | 1.81 | 3.47 | 5.01 |
| $\tau\left(x_{s-1}, x_s\right)$ | 15.48 | 15.48 | 17.18 | 10.85 | 12.15 | 17.09 |
| $\mathcal{CIS}_{\mathcal{H}igh}\left(x_s\right)$ | 46.34 | 46.34 | 46.34 | 46.34 | 46.34 | 46.34 |
| $\boldsymbol{\mathcal{CI}}\left(x_s\right)$ | 308.1 | 253.7 | 199.3 | 308.1 | 253.7 | 199.3 |
| $\boldsymbol{\mathcal{R}}\left(t, s\right)$ | $-15$ | $-30$ | $-45$ | $-15$ | $-30$ | $-45$ |

Table 7: Multi-period solution

| | ($h = 2$, $\lambda_s = 0.5\%$) | | | ($h = 3$, $\lambda_s = 0.5\%$) | | | ($h = 3$, $\lambda_s = 5\%$) | | |
|---|---|---|---|---|---|---|---|---|---|
| $s$ | $t+1$ | $t+2$ | $t+3$ | $t+1$ | $t+2$ | $t+3$ | $t+1$ | $t+2$ | $t+3$ |
| $x_{1,s}$ | 15.43 | 12.40 | 8.62 | 14.86 | 12.29 | 8.38 | 14.70 | 12.46 | 8.74 |
| $x_{2,s}$ | 15.75 | 16.06 | 16.65 | 16.01 | 16.29 | 16.70 | 15.75 | 16.23 | 16.67 |
| $x_{3,s}$ | 13.68 | 13.68 | 13.68 | 13.68 | 14.25 | 14.25 | 13.68 | 13.68 | 13.68 |
| $x_{4,s}$ | 11.40 | 11.40 | 11.40 | 11.40 | 11.40 | 11.40 | 11.40 | 11.40 | 11.40 |
| $x_{5,s}$ | 10.29 | 10.29 | 10.29 | 10.29 | 10.29 | 10.29 | 10.29 | 10.29 | 10.29 |
| $x_{6,s}$ | 5.43 | 2.44 | 0.00 | 6.06 | 2.43 | 0.00 | 5.83 | 2.54 | 0.00 |
| $x_{7,s}$ | 7.36 | 6.23 | 5.00 | 6.69 | 5.83 | 4.79 | 7.56 | 5.86 | 4.79 |
| $x_{8,s}$ | 5.59 | 6.41 | 7.06 | 6.00 | 6.58 | 7.21 | 5.39 | 6.62 | 7.24 |
| $x_{9,s}$ | 11.81 | 17.82 | 24.04 | 11.74 | 17.37 | 23.71 | 12.13 | 17.66 | 23.92 |
| $x_{10,s}$ | 3.27 | 3.27 | 3.27 | 3.27 | 3.27 | 3.27 | 3.27 | 3.27 | 3.27 |
| $\sigma\left(x_s \mid b_s\right)$ | 1.68 | 3.28 | 4.93 | 1.64 | 3.24 | 4.90 | 1.69 | 3.28 | 4.93 |
| $\tau\left(x_{s-1}, x_s\right)$ | 12.32 | 14.27 | 14.91 | 13.53 | 14.11 | 14.76 | 12.55 | 14.46 | 14.67 |
| $\mathcal{CIS}_{\mathcal{H}igh}\left(x_s\right)$ | 46.34 | 46.34 | 46.34 | 46.34 | 46.34 | 46.34 | 46.34 | 46.34 | 46.34 |
| $\boldsymbol{\mathcal{CI}}\left(x_s\right)$ | 308.1 | 253.7 | 199.3 | 308.1 | 253.7 | 199.3 | 308.1 | 253.7 | 199.3 |
| $\boldsymbol{\mathcal{R}}\left(t, s\right)$ | $-15$ | $-30$ | $-45$ | $-15$ | $-30$ | $-45$ | $-15$ | $-30$ | $-45$ |

### 3.3.2 Application to the Eurostoxx 50 index

We consider the Eurostoxx 50 index at the end of 2019 and implement the following carbon reduction pathway:

$$\begin{cases} \mathcal{CI}(x_s) = \left(1 - \mathcal{R}(t,s)\right) \cdot \mathcal{CI}(b_t) \\ \mathcal{R}(t,s) = 1 - 0.93^{(s-t)} \end{cases} \tag{102}$$

where the base year $t$ corresponds to January $1^{\text{st}}$, 2020. Therefore, we impose a year-on-year decarbonization of 7% per annum. We test several values of the parameters $h$ and $\lambda$. Moreover, we consider two rebalancing schemes: yearly and quarterly. We note the optimized portfolio by $x_s^\star(h,\lambda,\Delta_s)$, where $h$ is the period, $\lambda_s$ is the penalty factor and $\Delta_s$ is the rebalancing frequency. Results are reported in Figures 4–19 on pages 33–40. The first panel represents the tracking-error volatility $\sigma\left(x_s^\star(h,\lambda_s,\Delta_s) \mid b_t\right)$ between the optimized portfolio $x_s^\star(h,\lambda_s,\Delta_s)$ and the current benchmark $b_t$. In the second panel, we compute the turnover $\tau\left(x_{s-1}^\star(h,\lambda_s,\Delta_s), x_s^\star(h,\lambda_s,\Delta_s)\right)$ between two consecutive optimized portfolios. Finally, the third panel measures the active share between the multi-period optimized portfolio and the single-period optimized portfolio:

$$\mathcal{AS} = \frac{1}{2}\sum_{i=1}^{n}\left|x_{s,i}^\star(h,\lambda_s,\Delta_s) - x_{s,i}^\star(1,0,\Delta_s)\right| \tag{103}$$

We verify that there is a trade-off between the tracking-error volatility and the turnover. Indeed, the multi-period solution increases the tracking risk, but reduces the turnover. The active share measures how the multi-period solution is different from the single-period solution. Figure 2 shows that the relationship between $\mathcal{AS}$ and the parameters is not obvious. For instance, the active share is not necessarily an increasing function of the period. In this example, the active share is larger when $h$ is equal to two years and $\lambda$ is set to 1%. Similarly, the active share is not monotonous with respect to the time $s$. For example, it can increase and then decrease.

### 3.3.3 Extension to net zero carbon metrics

The previous analysis assumes that the carbon intensities are constant over time. In terms of portfolio alignment, we can relax this hypothesis. The carbon footprint constraint becomes:

$$\widehat{\mathcal{CI}}_s(x_s) \leq \left(1 - \mathcal{R}(t,s)\right) \cdot \mathcal{CI}(b_t) \tag{104}$$

where $\widehat{\mathcal{CI}}_s(x)$ is the estimated carbon intensity of the portfolio $x$ at time $s$. In this case, we can use the net zero carbon metrics proposed by Le Guenedal *et al.* (2022) to define $\widehat{\mathcal{CI}}_s(x)$:

$$\widehat{\mathcal{CI}}_s(x) = \sum_{i=1}^{n} x_i \cdot \widehat{\mathcal{CI}}_{i,s} \tag{105}$$

where $\widehat{\mathcal{CI}}_{i,s}$ is the estimated carbon intensity of the issuer $i$ at time $s$. For example, we can use the carbon trend or the carbon target. Since carbon intensities are not the same in each period, multi-period optimization makes more sense.

We can face two situations. In the first case, overall carbon intensities are decreasing over time. Therefore, it will be easier to decarbonize the portfolio in the future. This implies that the tracking-error volatility will be reduced. This results in a lower active share between the single-period solution and the multi-period solution. In the second case, overall carbon intensities are increasing over time, implying an increase in the tracking-error volatility, and

Figure 2: Active share in % of decarbonized portfolios (Eurostoxx 50 index)

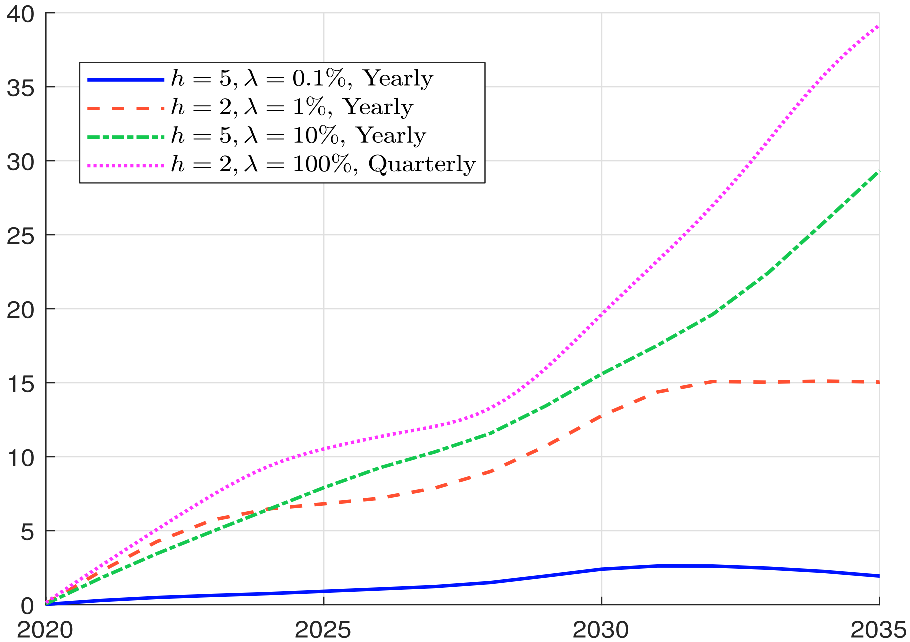
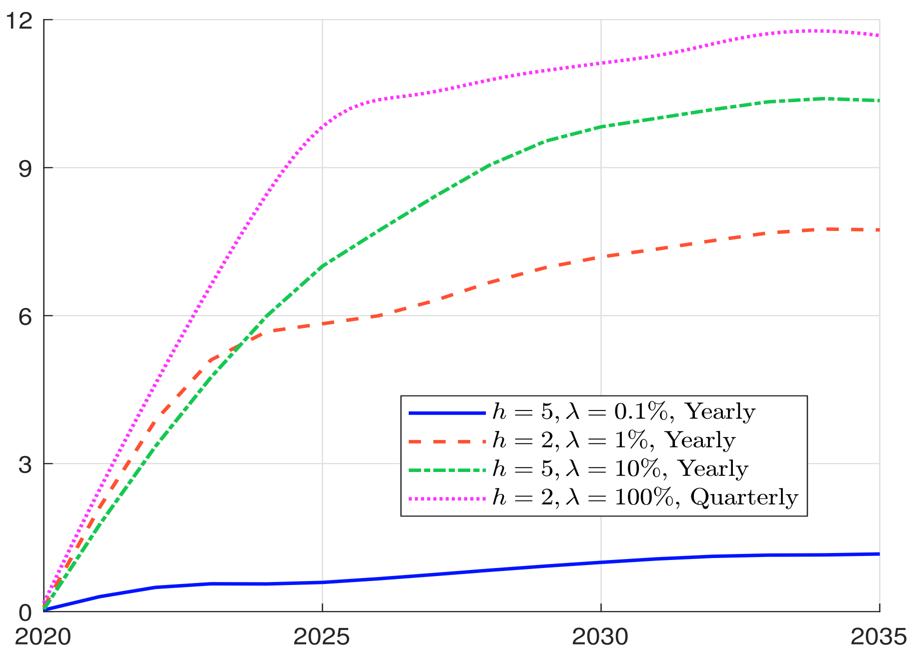



Figure 3: Active share in % of aligned portfolios using 2013-2020 carbon trends (Eurostoxx 50 index)



a larger active share between the single-period solution and the multi-period solution. For instance, we are in the first situation with the Eurostoxx 50 Index[18] as illustrated in Figure 3 when we use the 2013-2020 carbon trends to estimate $\widehat{\mathcal{CI}}_{i,s}$. On the contrary, this second situation is observed for the S&P 500 index (Le Guenedal and Roncalli, 2022). In this case, the active share between the single-period solution and the multi-period solution can reach 40%.

### 3.3.4 Dimension of the problem and choice of the algorithm

The choice of the algorithm highly depends on the dimension of the portfolio decarbonization problem. In Table 8, we have reported the complexity of the three algorithms. In the case of the ADMM algorithm, we must distinguish the $x$-update and the $y$-update. Let $n$ and $h$ be the number of assets and periods. The $x$-step requires to solve $h$ QP problems of dimension $n$, whereas we solve $n$ QP problems of dimension $3h$ in the $y$-step. If we consider the block coordinate descent, each iteration is made up of $h$ QP problems of dimension $5n$. Finally, the augmented QP algorithm consists in solving one QP problem of dimension $3nh$. For a given value of $h$, the choice of the algorithm depends then on the number $n$ of assets. If $h$ is set to 5 periods, we obtain the following results. For a small universe (e.g., $n \leq 100$), we can use the a-QP algorithm. For a medium universe (e.g., $100 \leq n \leq 200$), the BCD algorithm is appropriate. Finally, for a large universe (e.g., $n \geq 200$), we can only use the ADMM algorithm.

Table 8: Complexity of the algorithms

| Algorithm | ADMM-$x$ | ADMM-$y$ | BCD | a-QP |
|---|---|---|---|---|
| dim(QP) | $n$ | $3h$ | $5n$ | $3nh$ |
| card(QP) | $h$ | $n$ | $h$ | 1 |
| Iterations | $\gg 1$ | $\gg 1$ | $\gg 1$ | 1 |

## 4 Conclusion

Multi-period portfolio optimization is the extension of the mean-variance optimization problem when we consider a dynamic strategy. Nevertheless, it has never been popular for two reasons. First, it is a tricky problem and solving this optimization problem is not obvious (Kolm *et al.*, 2014). Second, Merton (1969, 1971) has established equivalence between the single-period and multi-period solutions under some specific hypotheses on the utility function and the trend in asset prices. However, the optimization problem is not always stationary, and the solution may depend on the finite horizon in some cases. For instance, we face this situation when we consider net zero portfolio alignment or portfolio decarbonization with a carbon reduction pathway.

In this article, we consider a simple formulation of multi-period optimization problems that asset managers encounter in real life. By exploiting some separability properties, we derive three numerical algorithms to optimize the multi-period objective function. They correspond to the alternating direction method of multipliers (ADMM), the block coordinate descent (BCD) approach and the augmented quadratic programming (a-QP). We apply these algorithms to three dynamic problems: transition management, total variation regularized portfolio and trading trajectory modeling. Finally, we consider the multi-period portfolio alignment problem of Le Guenedal and Roncalli (2022) and demonstrate how to solve it.

[18]See for example Figures 20 and 21 on page 41.

# References


BAUSCHKE, H.H., and BORWEIN, J.M. (1994), Dykstra's Alternating Projection Algorithm for Two Sets, *Journal of Approximation Theory*, 79(3), pp. 418-443.

BOURGERON, T., LEZMI, E., and RONCALLI, T. (2018), Robust Asset Allocation for Robo-Advisors, *arXiv*, 1902.05710.

BOYD, S., BUSSETI, E., DIAMOND, S., KAHN, R.N., KOH, K., NYSTRUP, P., and SPETH, J. (2017), Multi-period Trading via Convex Optimization, *Foundations and Trends® in Optimization*, 3(1), pp. 1-76.

BRUDER, B., GAUSSEL, N., RICHARD, J-C., and RONCALLI, T. (2013), Regularization of Portfolio Allocation, *SSRN*, www.ssrn.com/abstract=2767358.

BOYD, S., PARIKH, N., CHU, E., PELEATO, B., and ECKSTEIN, J. (2010), Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, *Foundations and Trends® in Machine learning*, 3(1), pp. 1-122.

CALAFIORE, G.C. (2009), An Affine Control Method for Optimal Dynamic Asset Allocation with Transaction Costs, *SIAM Journal on Control and Optimization*, 48(4), pp. 2254-2274.

CHEN, P., LEZMI, E., RONCALLI, T. and XU, J. (2019), A Note on Portfolio Optimization with Quadratic Transaction Costs, *arXiv*, 2001.01612.

CORSARO, S., DE SIMONE, V., MARINO, Z., and PERLA (2020), $\ell_1$-Regularization for Multi-period Portfolio Selection, *Annals of Operations Research*, 294(1), pp. 75-86.

CORSARO, S., DE SIMONE, V., and MARINO, Z. (2021), Split Bregman Iteration for Multi-period Mean Variance Portfolio Optimization, *Applied Mathematics and Computation*, 392, 125715.

DYKSTRA, R.L. (1983), An Algorithm for Restricted Least Squares Regression, *Journal of the American Statistical Association*, 78(384), pp. 837-842.

GABAY, D., and MERCIER, B. (1976), A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation, *Computers & Mathematics with Applications*, 2(1), pp. 17-40.

GÂRLEANU, N., and PEDERSEN, L.H. (2013), Dynamic Trading with Predictable Returns and Transaction Costs, *Journal of Finance*, 68(6), pp. 2309-2340.

GRIVEAU-BILLION, T., RICHARD, J-C., and RONCALLI, T. (2013), A Fast Algorithm for Computing High-dimensional Risk Parity Portfolios, *SSRN*, www.ssrn.com/abstract=2325255.

HUANG, X., ZHANG, Z., and ZHAO, Z. (2021), Multi-Period Portfolio Optimization for Index Tracking in Finance, *Asilomar Conference on Signals, Systems, and Computers*, 55, pp. 1383-1387.

KOLM, P.N., TÜTÜNCÜ, R., and FABOZZI, F.J. (2014), 60 Years of Portfolio Optimization: Practical Challenges and Current Trends, *European Journal of Operational Research*, 234(2), pp. 356-371.

Le Guenedal, T., Lombard, F., Roncalli, T., and Sekine, T. (2022), Net Zero Carbon Metrics, *Amundi Working Paper*.

Le Guenedal, T., and Roncalli, T. (2022), Portfolio Construction with Climate Risk Measures, *Amundi Working Paper*.

Li, X., Uysal, A.S., and Mulvey, J.M. (2022), Multi-period Portfolio Optimization using Model Predictive Control with Mean-variance and Risk Parity Frameworks, *European Journal of Operational Research*, 299(3), pp. 1158-1176.

Merton R.C. (1969), Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case, *Review of Economics and Statistics*, 51(3), pp. 247-257.

Merton R.C. (1971), Optimum Consumption and Portfolio Rules in a Continuous-Time Model, *Journal of Economic Theory*, 3(4), pp. 373-413.

Perrin, S., and Roncalli, T. (2020), Machine Learning Optimization Algorithms & Portfolio Allocation, in Jurczenko, E. (Ed.), *Machine Learning for Asset Management: New Developments and Financial Applications*, Chapter 8, Wiley, pp. 261-328.

Richard, J-C., and Roncalli, T. (2015), Smart Beta: Managing Diversification of Minimum Variance Portfolios, in Jurczenko, E. (Ed.), *Risk-based and Factor Investing*, ISTE Press – Elsevier.

Richard, J-C., and Roncalli, T. (2019), Constrained Risk Budgeting Portfolios: Theory, Algorithms, Applications & Puzzles, *arXiv*, 1902.05710.

Roncalli, T. (2013), *Introduction to Risk Parity and Budgeting*, Chapman & Hall/CRC Financial Mathematics Series.

Roncalli, T. (2015), Introducing Expected Returns into Risk Parity Portfolios: A New Framework for Asset Allocation, *Bankers, Markets & Investors*, 138, pp. 18-28.

Roncalli, T., Cherief, A., Karray-Meziou, F., and Regnault, M. (2021), Liquidity Stress Testing in Asset Management — Part 2. Modeling the Asset Liquidity Risk, *arXiv*, 2105.08377.

Rosenberg, G., Haghnegahdar, P., Goddard, P., Carr, P., Wu, K., and De Prado, M.L. (2016), Solving the Optimal Trading Trajectory Problem using a Quantum Annealer, *IEEE Journal of Selected Topics in Signal Processing*, 10(6), pp. 1053-1060.

Rattray, S. (2003), A Guide to Transition Management, *Trading*, 1, pp. 9-32.

Skaf, J., and Boyd, S. (2009), Multi-period Portfolio Optimization with Constraints and Transaction Costs, *Working paper*.

Steinhauer, K., Fukadai, T., and Yoshida, S. (2020), Solving the Optimal Trading Trajectory Problem Using Simulated Bifurcation, *arXiv*, 2009.08412.

Tseng, P. (2001), Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization, *Journal of Optimization Theory and Applications*, 109(3), pp. 475-494.

Wahlberg, B., Boyd, S., Annergren, M., and Wang, Y. (2012), An ADMM Algorithm for a Class of Total Variation Regularized Estimation Problems, *IFAC Proceedings Volumes*, 45(16), pp. 83-88.

Xu, Y., and Yin, W. (2013), A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Applications to Nonnegative Tensor Factorization and Completion, *SIAM Journal on Imaging Sciences*, 6(3), pp. 1758-1789.

# A Data

## A.1 Example 1

We use the example of Roncalli (2013, Section 6.2.2, page 286-288). The investment universe is made up of seven asset classes: (1) US Bonds 10Y, (2) EUR Bonds, (3) IG Bonds, (4) US Equities, (5) European Equities, (6) EM Equities and (7) Commodities. In Tables 9 and 10, we report the long-run statistics used to compute the strategic asset allocation.

Table 9: Expected return and volatility (in %)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| $\mu_i$ | 4.2 | 3.8 | 5.3 | 9.2 | 8.6 | 11.0 | 8.8 |
| $\sigma_i$ | 5.0 | 5.0 | 7.0 | 15.0 | 15.0 | 18.0 | 30.0 |

Table 10: Correlation matrix of asset returns (in %)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| (1) | 100 | | | | | | |
| (2) | 80 | 100 | | | | | |
| (3) | 60 | 40 | 100 | | | | |
| (4) | −10 | −20 | 30 | 100 | | | |
| (5) | −20 | −10 | 20 | 90 | 100 | | |
| (6) | −20 | −20 | 30 | 70 | 70 | 100 | |
| (7) | 0 | 0 | 10 | 20 | 20 | 30 | 100 |

# B Additional results

## B.1 Transition management

Table 11: Iterative solution (Example 1, monotonic constraint)

| $s$ | $t$ | $t+1$ | $t+2$ | $t+3$ | $t+4$ | $t+5$ |
|---|---|---|---|---|---|---|
| $x_{1,s}$ | 20.00 | 20.00 | 20.82 | 29.29 | 41.79 | 46.20 |
| $x_{2,s}$ | 10.00 | 22.50 | 34.18 | 38.21 | 38.21 | 38.21 |
| $x_{3,s}$ | 15.00 | 15.00 | 15.00 | 15.00 | 3.66 | 0.00 |
| $x_{4,s}$ | 20.00 | 20.00 | 10.24 | 4.09 | 4.09 | 4.09 |
| $x_{5,s}$ | 10.00 | 10.00 | 10.00 | 5.45 | 4.66 | 4.09 |
| $x_{6,s}$ | 15.00 | 7.76 | 7.11 | 7.11 | 7.11 | 7.11 |
| $x_{7,s}$ | 10.00 | 4.74 | 2.65 | 0.84 | 0.48 | 0.30 |
| $\tau_s$ | | 25.00 | 25.00 | 25.00 | 25.00 | 8.81 |

Table 12: Augmented QP solution (Example 1, monotonic constraint)

| $s$ | $t$ | $t+1$ | $t+2$ | $t+3$ | $t+4$ | $t+5$ |
|---|---|---|---|---|---|---|
| $x_{1,s}$ | 20.00 | 20.00 | 20.82 | 29.29 | 41.79 | 46.21 |
| $x_{2,s}$ | 10.00 | 22.50 | 34.18 | 38.21 | 38.21 | 38.21 |
| $x_{3,s}$ | 15.00 | 15.00 | 15.00 | 15.00 | 3.65 | 0.00 |
| $x_{4,s}$ | 20.00 | 20.00 | 10.23 | 4.18 | 4.18 | 4.09 |
| $x_{5,s}$ | 10.00 | 10.00 | 10.00 | 5.37 | 4.58 | 4.09 |
| $x_{6,s}$ | 15.00 | 7.76 | 7.11 | 7.11 | 7.11 | 7.11 |
| $x_{7,s}$ | 10.00 | 4.74 | 2.65 | 0.84 | 0.48 | 0.30 |
| $\tau_s$ | | 25.00 | 25.00 | 25.00 | 25.00 | 8.82 |

Table 13: Block CD solution (Example 1, monotonic constraint)

| $s$ | $t$ | $t+1$ | $t+2$ | $t+3$ | $t+4$ | $t+5$ |
|---|---|---|---|---|---|---|
| $x_{1,s}$ | 20.00 | 20.00 | 20.82 | 29.29 | 41.79 | 46.21 |
| $x_{2,s}$ | 10.00 | 22.50 | 34.18 | 38.21 | 38.21 | 38.21 |
| $x_{3,s}$ | 15.00 | 15.00 | 15.00 | 15.00 | 3.65 | 0.00 |
| $x_{4,s}$ | 20.00 | 20.00 | 10.23 | 4.09 | 4.09 | 4.09 |
| $x_{5,s}$ | 10.00 | 10.00 | 10.00 | 5.45 | 4.66 | 4.09 |
| $x_{6,s}$ | 15.00 | 7.76 | 7.11 | 7.11 | 7.11 | 7.11 |
| $x_{7,s}$ | 10.00 | 4.74 | 2.65 | 0.84 | 0.48 | 0.30 |
| $\tau_s$ | | 25.00 | 25.00 | 25.00 | 25.00 | 8.82 |

## B.2 Trading trajectory problem

Table 14: Optimal trading trajectory in % with the boundary condition ($\gamma = 1$, $\varrho_\Lambda = 5\%$)

| $\phi$ | $s$ | $\varrho_\Gamma = 1\%$ | | | | $\varrho_\Gamma = 10\%$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $x^\star_{1,s}$ | $x^\star_{2,s}$ | $x^\star_{3,s}$ | $x^\star_{4,s}$ | $x^\star_{1,s}$ | $x^\star_{2,s}$ | $x^\star_{3,s}$ | $x^\star_{4,s}$ |
| 0% | $t+1$ | 21.48 | 23.60 | 24.53 | 30.40 | 21.45 | 23.53 | 24.60 | 30.42 |
| | $t+2$ | 20.64 | 23.25 | 24.63 | 31.48 | 20.53 | 23.01 | 24.89 | 31.56 |
| | $t+3$ | 20.42 | 23.13 | 24.73 | 31.72 | 20.01 | 22.42 | 25.52 | 32.05 |
| | $t+4$ | 20.29 | 23.04 | 24.83 | 31.84 | 18.61 | 20.95 | 27.21 | 33.23 |
| | $t+5$ | 19.89 | 22.87 | 25.04 | 32.20 | 12.82 | 16.81 | 32.08 | 38.29 |
| | $t+6$ | 19.89 | 22.87 | 25.04 | 32.20 | 12.82 | 16.81 | 32.08 | 38.29 |
| 50% | $t+1$ | 21.54 | 23.62 | 24.52 | 30.31 | 21.96 | 23.74 | 24.57 | 29.73 |
| | $t+2$ | 20.67 | 23.27 | 24.62 | 31.44 | 20.84 | 23.25 | 24.71 | 31.21 |
| | $t+3$ | 20.45 | 23.15 | 24.70 | 31.70 | 20.27 | 22.91 | 25.00 | 31.81 |
| | $t+4$ | 20.35 | 23.09 | 24.77 | 31.79 | 19.56 | 22.48 | 25.51 | 32.45 |
| | $t+5$ | 20.15 | 23.00 | 24.88 | 31.97 | 17.78 | 21.69 | 26.47 | 34.06 |
| | $t+6$ | 20.15 | 23.00 | 24.88 | 31.97 | 17.78 | 21.69 | 26.47 | 34.06 |
| 100% | $t+1$ | 21.61 | 23.65 | 24.52 | 30.23 | 22.31 | 23.92 | 24.52 | 29.25 |
| | $t+2$ | 20.71 | 23.29 | 24.60 | 31.40 | 21.18 | 23.49 | 24.51 | 30.81 |
| | $t+3$ | 20.47 | 23.18 | 24.68 | 31.67 | 20.72 | 23.30 | 24.59 | 31.40 |
| | $t+4$ | 20.41 | 23.14 | 24.72 | 31.74 | 20.52 | 23.21 | 24.65 | 31.62 |
| | $t+5$ | 20.40 | 23.12 | 24.73 | 31.75 | 20.44 | 23.16 | 24.69 | 31.71 |
| | $t+6$ | 20.40 | 23.12 | 24.73 | 31.75 | 20.44 | 23.16 | 24.69 | 31.71 |

## B.3 Portfolio decarbonization

Figure 4: Dynamic portfolio decarbonization ($h = 2$, $\lambda = 0.1\%$, yearly rebalancing)

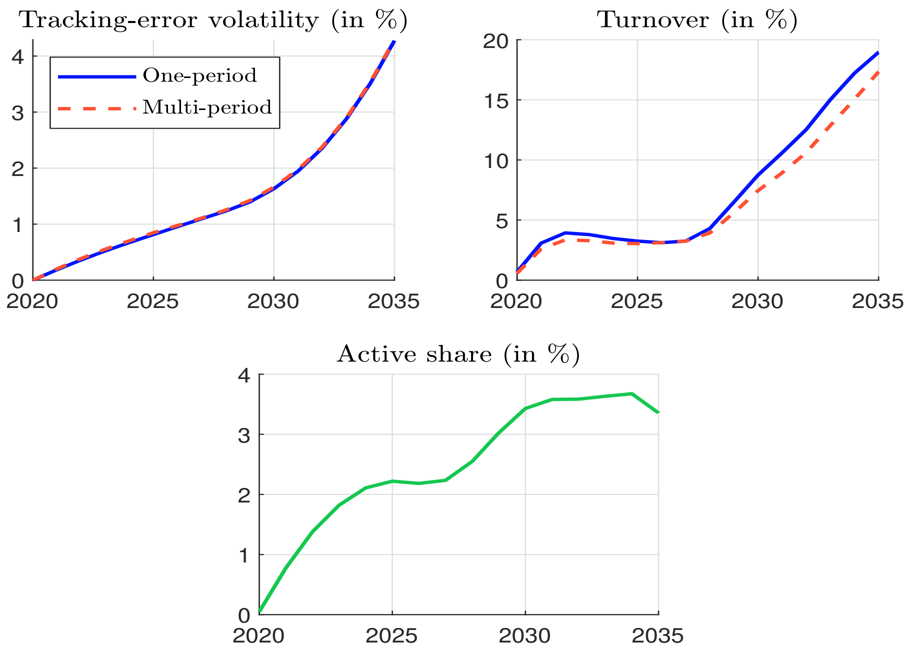



Figure 5: Dynamic portfolio decarbonization ($h = 2$, $\lambda = 1\%$, yearly rebalancing)

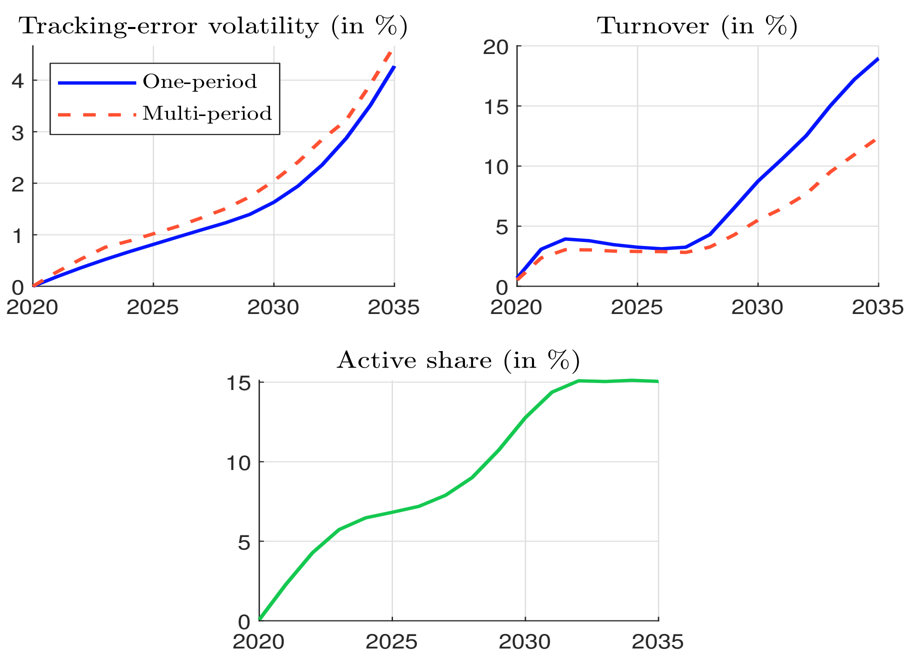


Figure 6: Dynamic portfolio decarbonization ($h = 2$, $\lambda = 10\%$, yearly rebalancing)

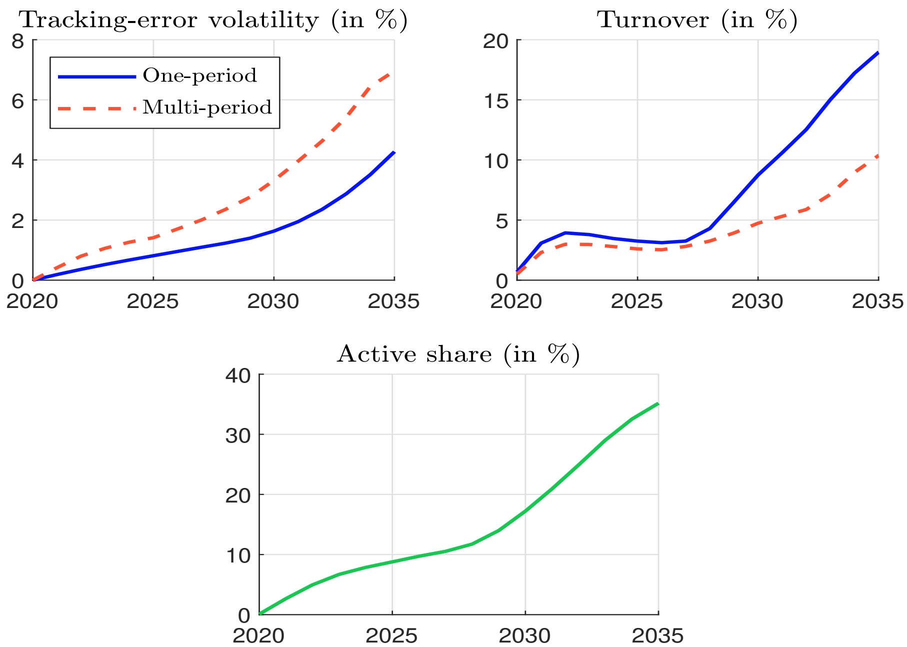



Figure 7: Dynamic portfolio decarbonization ($h = 2$, $\lambda = 100\%$, yearly rebalancing)

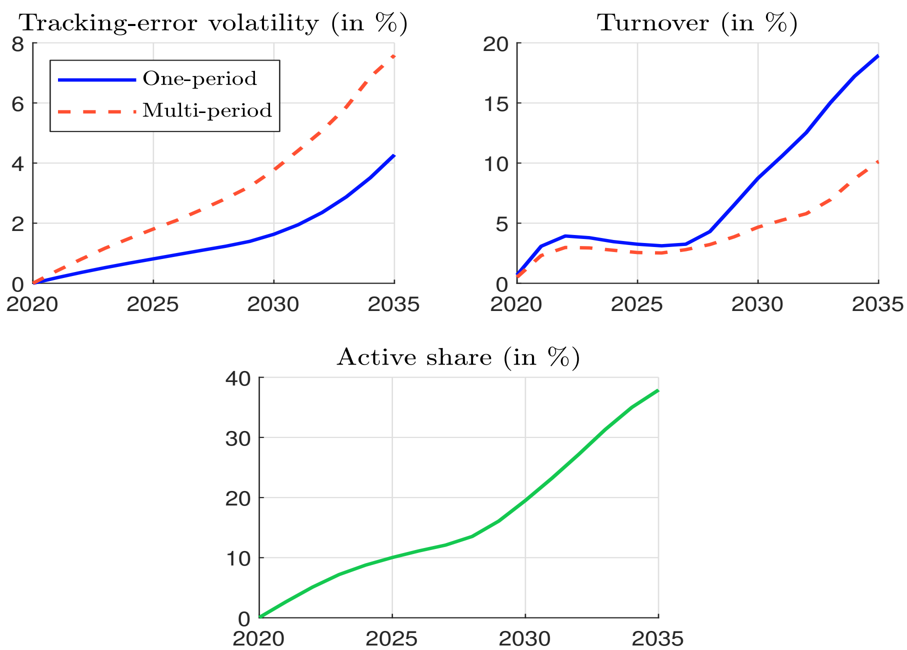


Figure 8: Dynamic portfolio decarbonization ($h = 5$, $\lambda = 0.1\%$, yearly rebalancing)

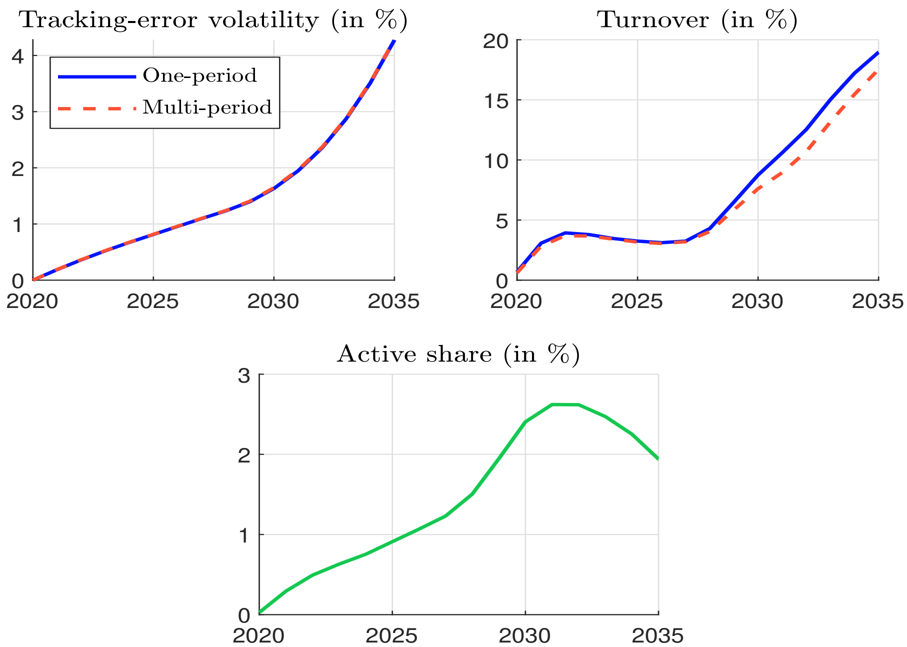



Figure 9: Dynamic portfolio decarbonization ($h = 5$, $\lambda = 1\%$, yearly rebalancing)

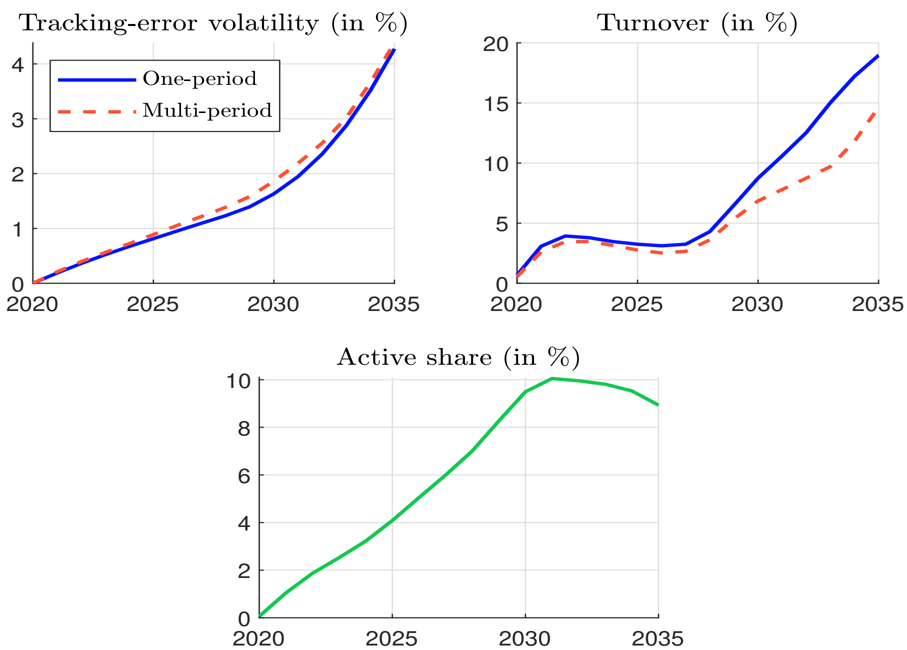


Figure 10: Dynamic portfolio decarbonization ($h = 5$, $\lambda = 10\%$, yearly rebalancing)

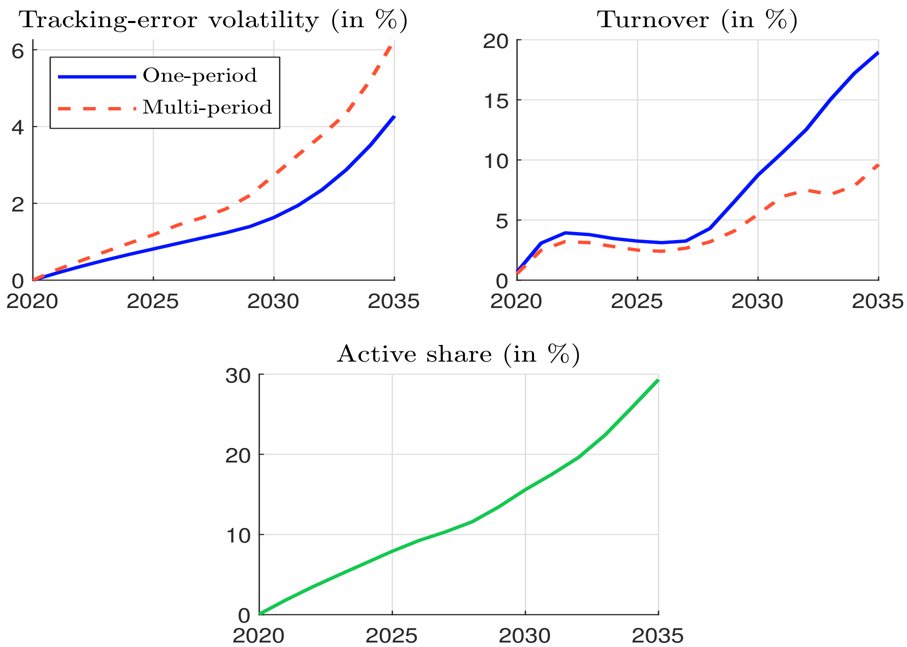



Figure 11: Dynamic portfolio decarbonization ($h = 5$, $\lambda = 100\%$, yearly rebalancing)

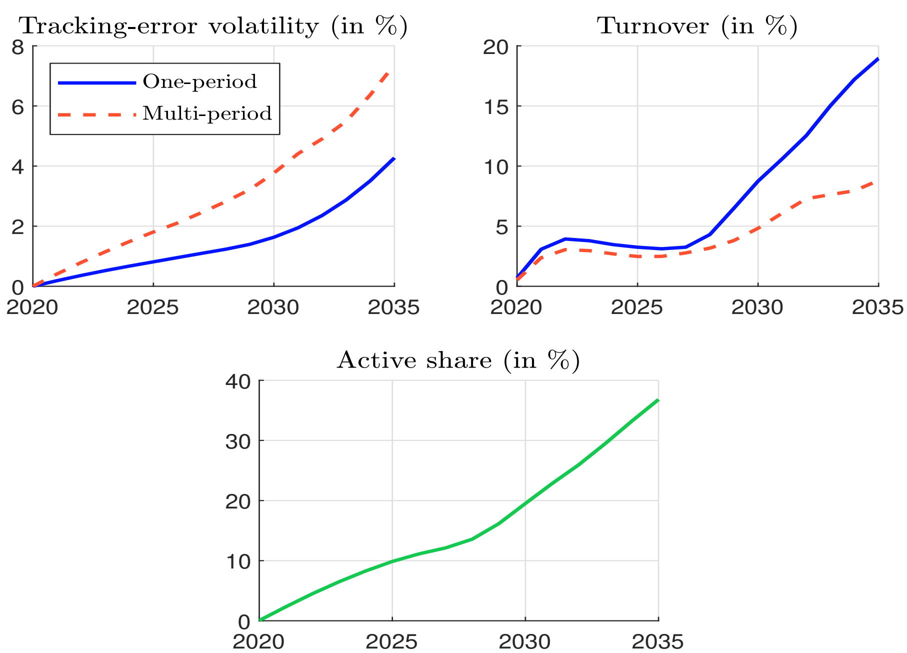


Figure 12: Dynamic portfolio decarbonization ($h = 2$, $\lambda = 0.1\%$, quarterly rebalancing)

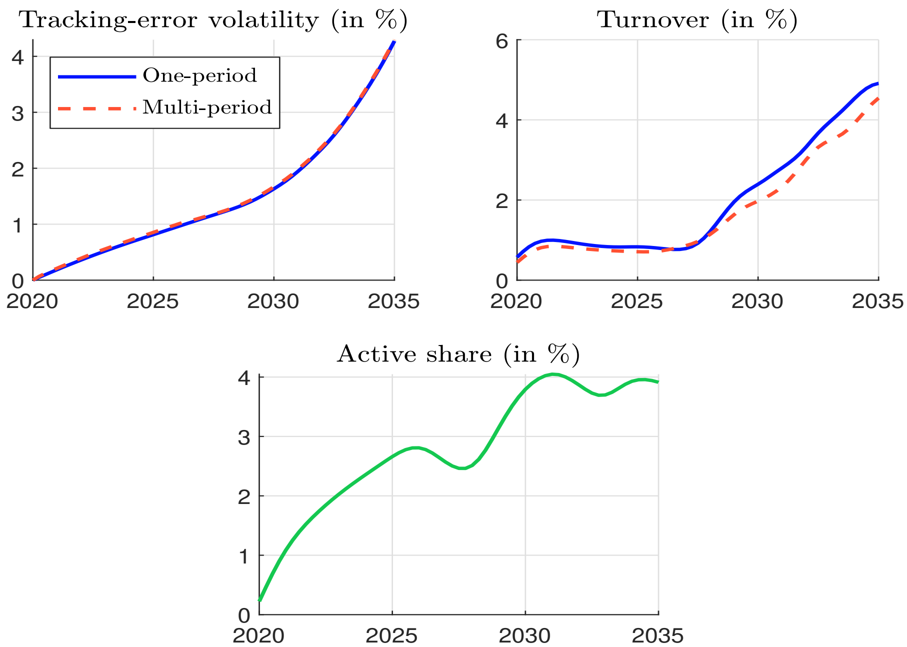



Figure 13: Dynamic portfolio decarbonization ($h = 2$, $\lambda = 1\%$, quarterly rebalancing)

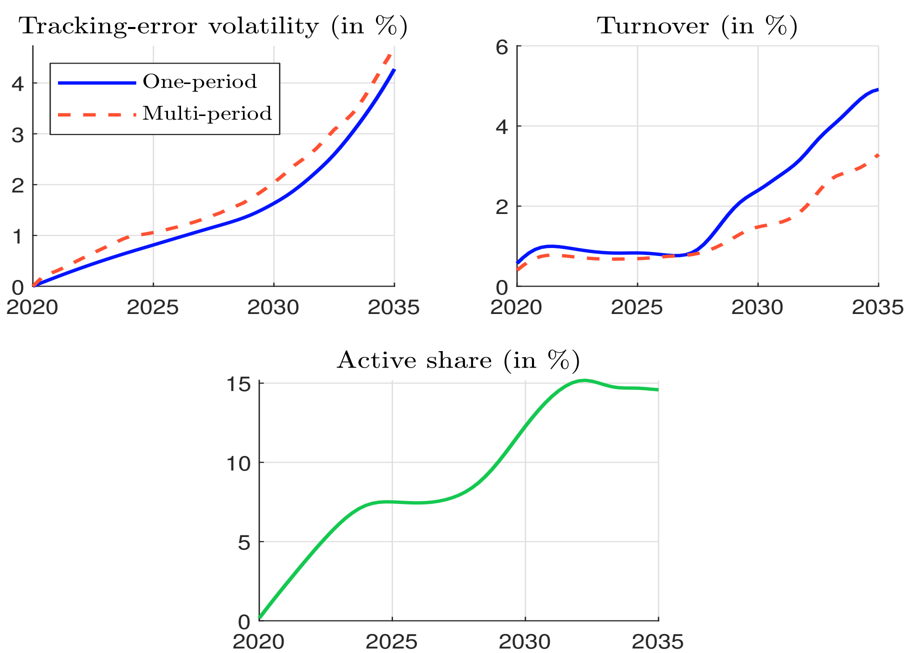


Figure 14: Dynamic portfolio decarbonization ($h = 2$, $\lambda = 10\%$, quarterly rebalancing)

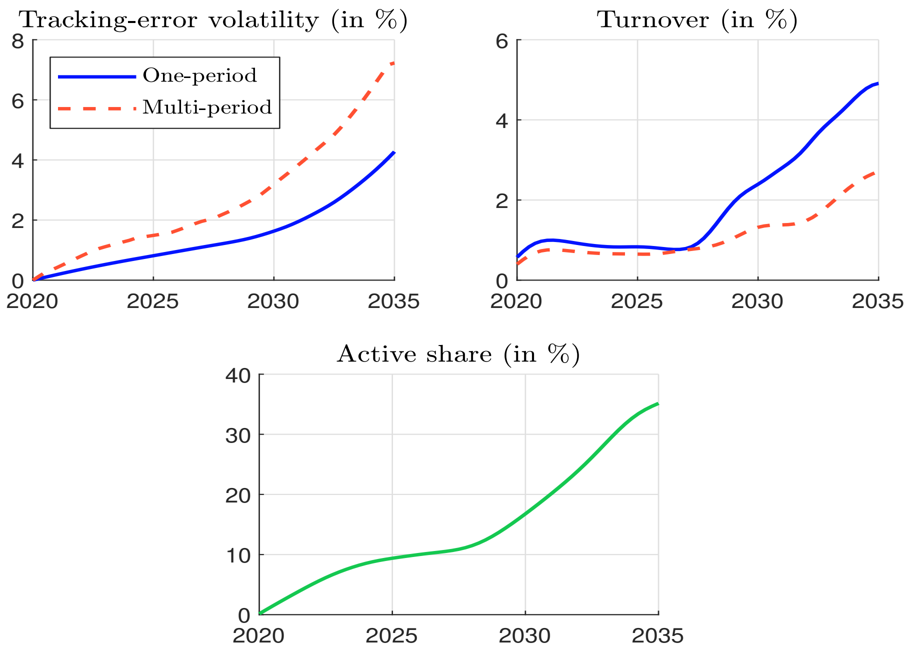



Figure 15: Dynamic portfolio decarbonization ($h = 2$, $\lambda = 100\%$, quarterly rebalancing)

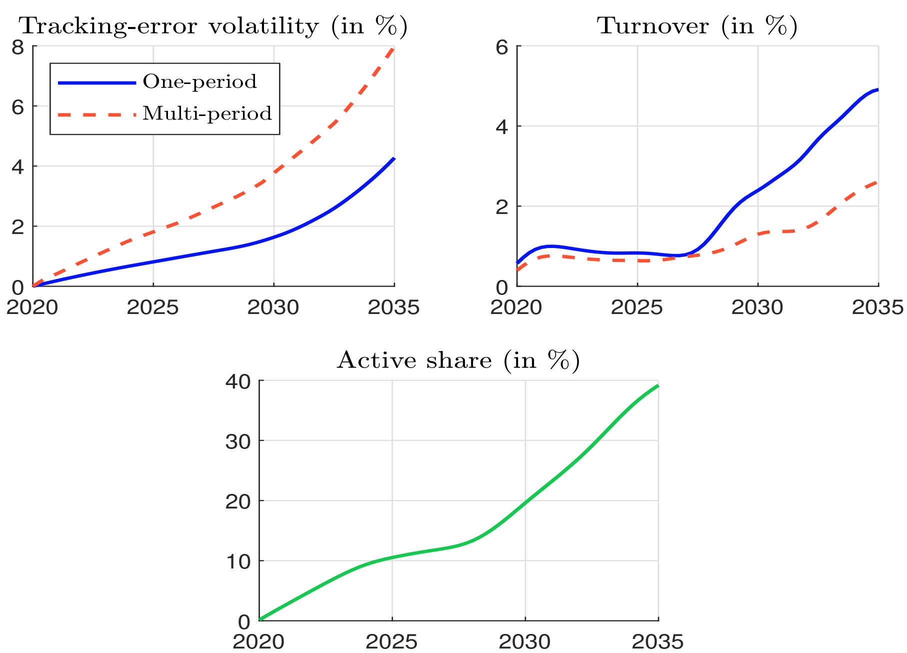


Figure 16: Dynamic portfolio decarbonization ($h = 5$, $\lambda = 0.1\%$, quarterly rebalancing)

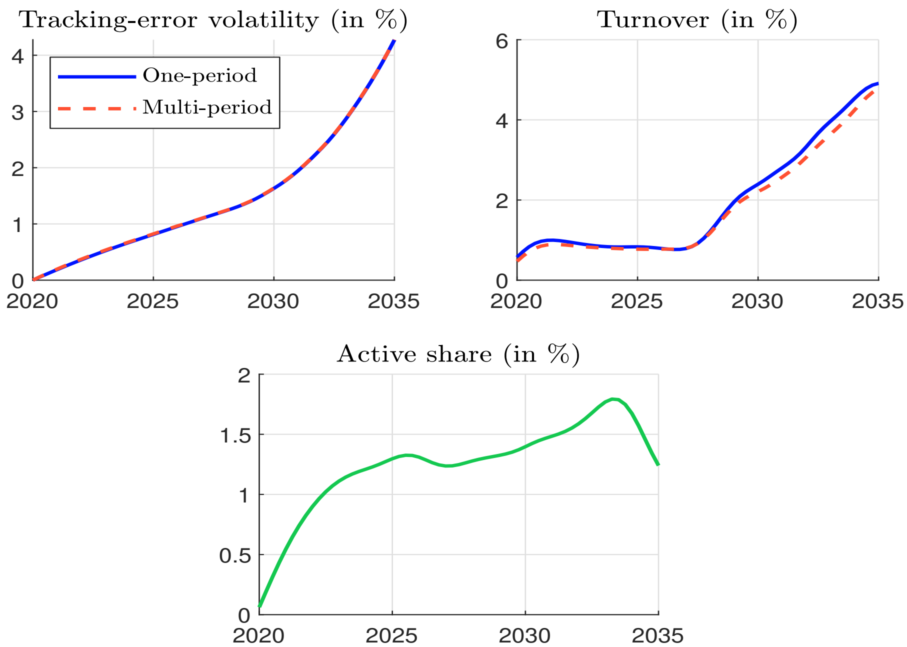



Figure 17: Dynamic portfolio decarbonization ($h = 5$, $\lambda = 1\%$, quarterly rebalancing)

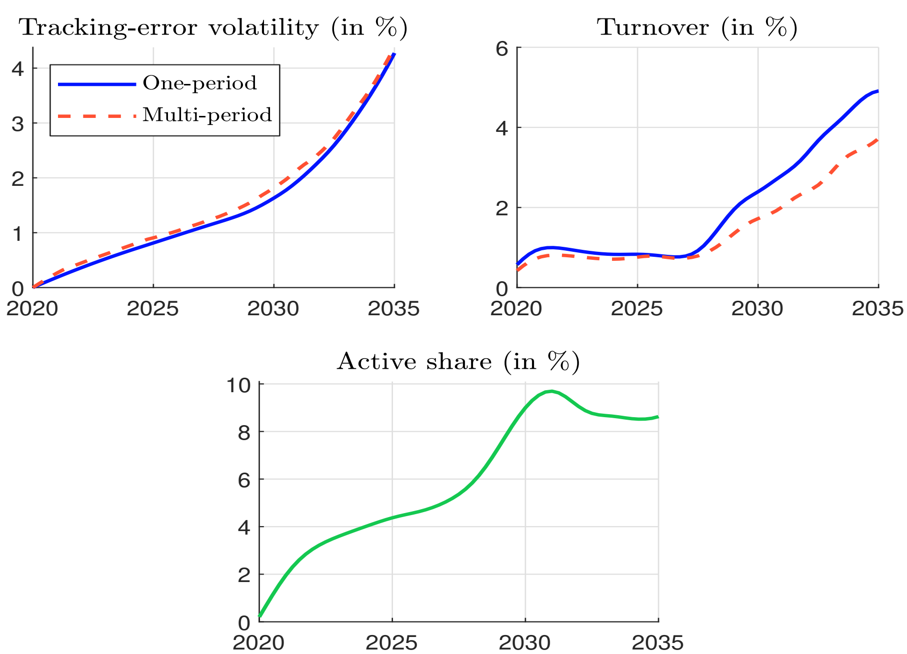


Figure 18: Dynamic portfolio decarbonization ($h = 5$, $\lambda = 10\%$, quarterly rebalancing)

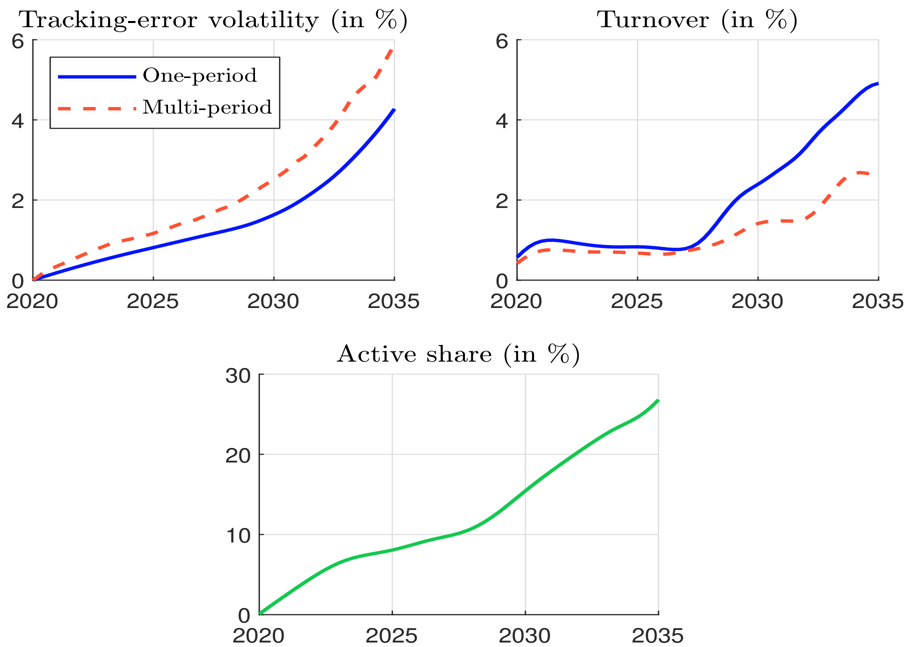



Figure 19: Dynamic portfolio decarbonization ($h = 5$, $\lambda = 100\%$, quarterly rebalancing)

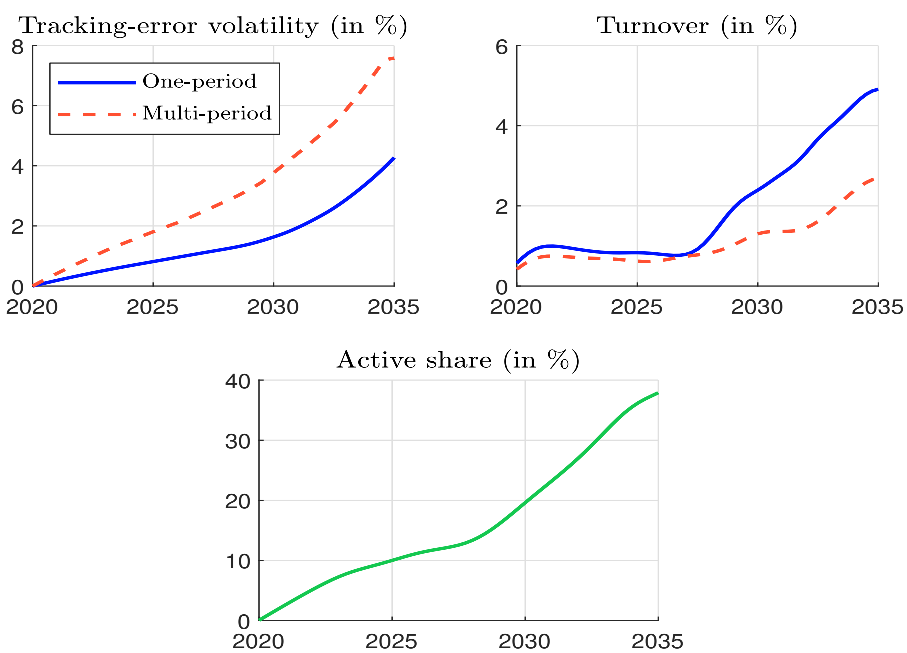


Figure 20: Portfolio alignment ($h = 2$, $\lambda = 100\%$, yearly rebalancing)

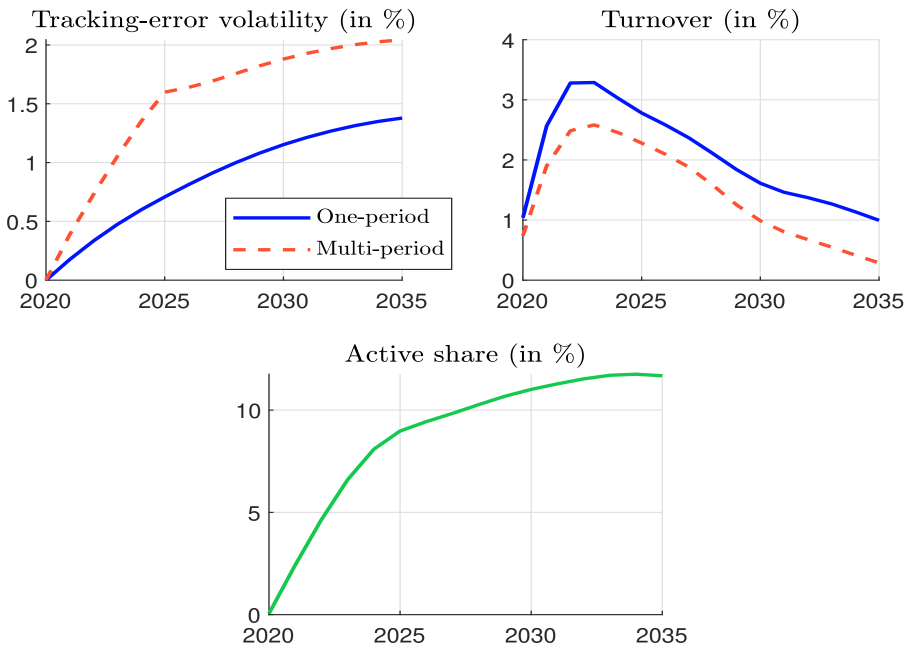
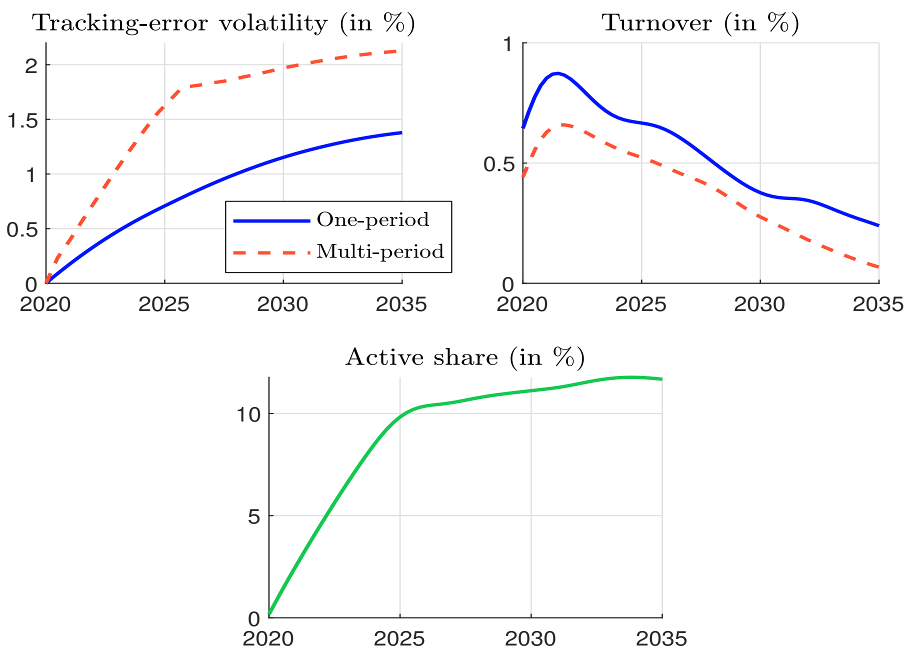



Figure 21: Portfolio alignment ($h = 2$, $\lambda = 100\%$, quarterly rebalancing)



# C Technical appendix

## C.1 Proximal operator of the function $\zeta_\varphi(x; x_0, \boldsymbol{\lambda})$

We note $x$ and $\boldsymbol{\lambda} \geq \mathbf{0}_n$ the vectors $(x_1, \ldots, x_n)$ and $(\lambda_1, \ldots, \lambda_n)$ of dimension $n$. Let $x_0$ be a scalar. We have:

$$\zeta_\varphi(x; x_0, \boldsymbol{\lambda}) := \varphi^{-1} \left\| \mathcal{D}(\boldsymbol{\lambda}) x - \lambda_1 \boldsymbol{e}_1 x_0 \right\|_1 \tag{106}$$

where $\boldsymbol{e}_1 = (1, 0, \ldots, 0)$ and:

$$\mathcal{D}(\boldsymbol{\lambda}) = \begin{pmatrix} \lambda_1 & & & \\ -\lambda_2 & \lambda_2 & & \\ & & \ddots & \\ & & -\lambda_n & \lambda_n \end{pmatrix} \tag{107}$$

The proximal operator is defined as:

$$x^\star = \mathbf{prox}_{\zeta_\varphi}(v) = \arg\min_x \left\{ \left\| \mathcal{D}(\boldsymbol{\lambda}) x - \lambda_1 \boldsymbol{e}_1 x_0 \right\|_1 + \frac{\varphi}{2} \left\| x - v \right\|_2^2 \right\} \tag{108}$$

### C.1.1 Augmented QP solution

Following Perrin and Roncalli (2020), we introduce the additional variables:

$$x_i = x_{i-1} - x_i^- + x_i^+ \tag{109}$$

where $x^- \geq 0$ and $x^+ \geq 0$. Let $z \geq \mathbf{0}_{3n}$ be the $3n \times 1$ vector $\left(x, x^-, x^+\right)$. Equation (109) is equivalent to:

$$\begin{aligned} (*) \quad & \Leftrightarrow \quad \mathcal{D}(\mathbf{1}_n) x + I_n x^- - I_n x^+ = \boldsymbol{e}_1 x_0 \\ & \Leftrightarrow \quad Az = B \end{aligned} \tag{110}$$

where $A = \begin{pmatrix} \mathcal{D}(\mathbf{1}_n) & I_n & -I_n \end{pmatrix}$ and $B = \boldsymbol{e}_1 x_0$. We also have:

$$\begin{aligned} \left\| \mathcal{D}(\boldsymbol{\lambda}) x - \lambda_1 \boldsymbol{e}_1 x_0 \right\|_1 &= \sum_{i=1}^n \lambda_i \left| x_i - x_{i-1} \right| \\ &= \sum_{i=1}^n \lambda_i \left| x_i^+ - x_i^- \right| \\ &= \boldsymbol{\lambda}^\top x^- + \boldsymbol{\lambda}^\top x^+ \end{aligned} \tag{111}$$

The objective function becomes:

$$\begin{aligned} (*) &= \left\| \mathcal{D}(\boldsymbol{\lambda}) x - \lambda_1 \boldsymbol{e}_1 x_0 \right\|_1 + \frac{\varphi}{2} \left\| x - v \right\|_2^2 \\ &= \boldsymbol{\lambda}^\top x^- + \boldsymbol{\lambda}^\top x^+ + \frac{\varphi}{2} (x - v)^\top (x - v) \\ &= \frac{1}{2} x^\top (\varphi I_n) x - \left( \varphi v^\top x - \boldsymbol{\lambda}^\top x^- - \boldsymbol{\lambda}^\top x^+ \right) + \frac{1}{2} v^\top (\varphi I_n) v \end{aligned} \tag{112}$$

It follows that:

$$x^\star = T z^\star \tag{113}$$

where $T = \begin{pmatrix} I_n & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \end{pmatrix}$ and:

$$\begin{aligned} z^\star &= \arg\min_z \left\{ \frac{1}{2} z^\top Q z - z^\top R \right\} \\ \text{s.t.} & \left\{ \begin{array}{l} Az = B \\ z \geq \mathbf{0}_{3n} \end{array} \right. \end{aligned} \tag{114}$$

The $3n \times 3n$ matrix $Q$ and the $3n \times 1$ vector $R$ are given by:

$$Q = \begin{pmatrix} \varphi I_n & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \end{pmatrix} \tag{115}$$

and:

$$R = \begin{pmatrix} \varphi v \\ -\boldsymbol{\lambda} \\ -\boldsymbol{\lambda} \end{pmatrix} \tag{116}$$

### C.1.2 CCD solution

We can also use the CCD algorithm to find the solution. We note:

$$\begin{aligned} f(x) &= \left\| \mathcal{D}(\boldsymbol{\lambda}) x - \lambda_1 \boldsymbol{e}_1 x_0 \right\|_1 + \frac{\varphi}{2} \|x - v\|_2^2 \\ &= \sum_{i=1}^n \lambda_i |x_i - x_{i-1}| + \frac{\varphi}{2} \sum_{i=1}^n (x_i - v_i)^2 \end{aligned} \tag{117}$$

We notice that:

$$\arg\min_{x_i} f(x) = \arg\min_{x_i} \left\{ \lambda_i |x_i - x_{i-1}| + \lambda_{i+1} |x_{i+1} - x_i| + \frac{\varphi}{2} (x_i - v_i)^2 \right\} \tag{118}$$

Therefore, each step of the CCD algorithm consists in solving the following optimization problem:

$$x_i^\star = \arg\min_{x_i} \eta(x_i; \lambda_i, \lambda_{i+1}, \varphi, x_{i-1}, x_{i+1}, v_i) \qquad \text{for } i < n \tag{119}$$

where:

$$\eta(x; a_1, a_2, a_3, b_1, b_2, b_3) = a_1 |x - b_1| + a_2 |x - b_2| + \frac{a_3}{2} (x - b_3)^2 \tag{120}$$

and $a_1 > 0$, $a_2 > 0$ and $a_3 > 0$. For $i = n$, we have:

$$x_n^\star = \arg\min_{x_n} \eta(x_n; \lambda_n, 0, \varphi, x_{n-1}, 0, v_n) \tag{121}$$

We consider the scalar problem:

$$x^\star = \arg\min_x \left\{ a_1 |x - b_1| + a_2 |x - b_2| + \frac{a_3}{2} (x - b_3)^2 \right\} \tag{122}$$

Let us assume that $b_2 > b_1$. The optimality condition is:

$$\begin{aligned} & \{0\} \in \partial \left( a_1 |x - b_1| \right) + \partial \left( a_2 |x - b_2| \right) + \nabla \left( \frac{a_3}{2} (x - b_3)^2 \right) \\ \Leftrightarrow \quad & \{0\} \in a_1 \partial |x - b_1| + a_2 \partial |x - b_2| + a_3 (x - b_3) \end{aligned} \tag{123}$$

where $\partial |x - b|$ is the subgradient of $|x - b|$. Then, we deduce that:

- If $x > b_2 > b_1$, we have $\partial|x-b_1| = 1$, $\partial|x-b_2| = 1$ and:

$$a_1 + a_2 + a_3(x^\star - b_3) = 0 \Leftrightarrow x^\star = b_3 - \frac{a_1 + a_2}{a_3} \tag{124}$$

  We must verify that:

$$x^\star = b_3 - \frac{a_1 + a_2}{a_3} > b_2 \tag{125}$$

- If $x = b_2 > b_1$, we have $\partial|x-b_1| = 1$, $\partial|x-b_2| \in [-1,1]$ and $\left|a_2\partial|x-b_2|\right| \leq a_2$. We deduce that $\left|a_1\partial|x-b_1| + a_3(x-b_3)\right| = \left|a_2\partial|x-b_2|\right| \leq a_2$ and $\left|a_1 + a_3(x-b_3)\right| \leq a_2$. We must verify that $-a_2 \leq a_1 + a_3(x-b_3) \leq a_2$ or:

$$b_3 - \frac{a_1 + a_2}{a_3} \leq x^\star = b_2 \leq b_3 - \frac{a_1 - a_2}{a_3} \tag{126}$$

- If $b_1 < x < b_2$, we have $\partial|x-b_1| = 1$, $\partial|x-b_2| = -1$ and:

$$a_1 - a_2 + a_3(x^\star - b_3) = 0 \Leftrightarrow x^\star = b_3 - \frac{a_1 - a_2}{a_3} \tag{127}$$

  We must verify that:

$$b_1 < x^\star = b_3 - \frac{a_1 - a_2}{a_3} < b_2 \tag{128}$$

- If $x = b_1$, we have $\partial|x-b_1| \in [-1,1]$, $\partial|x-b_2| = -1$ and $\left|a_1\partial|x-b_1|\right| \leq a_1$. We deduce that $\left|a_2\partial|x-b_2| + a_3(x-b_3)\right| = \left|a_1\partial|x-b_1|\right| \leq a_1$ and $\left|-a_2 + a_3(x-b_3)\right| \leq a_1$. We must verify that $-a_1 \leq -a_2 + a_3(x-b_3) \leq a_1$ or:

$$b_3 - \frac{a_1 - a_2}{a_3} \leq x^\star = b_1 \leq b_3 + \frac{a_1 + a_2}{a_3} \tag{129}$$

- If $x < b_1$, we have $\partial|x-b_1| = -1$, $\partial|x-b_2| = -1$ and:

$$-a_1 - a_2 + a_3(x^\star - b_3) = 0 \Leftrightarrow x^\star = b_3 + \frac{a_1 + a_2}{a_3} \tag{130}$$

  We must verify that:

$$x^\star = b_3 + \frac{a_1 + a_2}{a_3} < b_1 \tag{131}$$

Putting all the several cases together, we get the following solution:

$$x^\star = \begin{cases} x_1^\star = b_3 - \dfrac{a_1+a_2}{a_3} & \text{if} \quad x_1^\star > b_2 = x_2^\star \\ x_2^\star = b_2 & \text{if} \quad x_1^\star = b_3 - \dfrac{a_1+a_2}{a_3} \leq x_2^\star \leq b_3 - \dfrac{a_1-a_2}{a_3} = x_3^\star \\ x_3^\star = b_3 - \dfrac{a_1-a_2}{a_3} & \text{if} \quad x_4^\star = b_1 < x_3^\star < b_2 = x_2^\star \\ x_4^\star = b_1 & \text{if} \quad x_3^\star = b_3 - \dfrac{a_1-a_2}{a_3} \leq x_4^\star \leq b_3 + \dfrac{a_1+a_2}{a_3} = x_5^\star \\ x_5^\star = b_3 + \dfrac{a_1+a_2}{a_3} & \text{if} \quad x_5^\star < b_1 = x_4^\star \end{cases} \tag{132}$$

If $b_2 \leq b_1$, we use the relationship:

$$\eta(x; a_1, a_2, a_3, b_1, b_2, b_3) = \eta(x; a_2, a_1, a_3, b_2, b_1, b_3) \tag{133}$$

The case $b_1 = b_2$ requires a new result given in the next remark.

**Remark 7.** *If $a_2 = 0$, the solution reduces to the soft-thresholding operator:*

$$\begin{aligned} & \{0\} \in a_1 \partial \left| x - b_1 \right| + a_3 \left( x - b_3 \right) \\ \Leftrightarrow \quad & \{0\} \in \left( x - b_1 \right) - \left( b_3 - b_1 \right) + \frac{a_1}{a_3} \partial \left| x - b_1 \right| \end{aligned} \tag{134}$$

*Following Perrin and Roncalli (2020, Appendix A.5 & A.8.3), we deduce that:*

$$\begin{aligned} x^\star &= \mathbf{prox}_{a_1 a_3^{-1} |y|} \left( b_3 - b_1 \right) + b_1 \\ &= \mathcal{S} \left( b_3 - b_1 ; \frac{a_1}{a_3} \right) + b_1 \\ &= b_1 + \operatorname{sign} \left( b_3 - b_1 \right) \cdot \left( \left| b_3 - b_1 \right| - \frac{a_1}{a_3} \right)_+ \qquad (135) \\ &= \left\{ \begin{array}{lll} b_3 - \dfrac{a_1}{a_3} & \text{if} & \left| b_3 - b_1 \right| > \dfrac{a_1}{a_3} \\ b_1 & \text{if} & \left| b_3 - b_1 \right| \leq \dfrac{a_1}{a_3} \end{array} \right. \qquad (136) \end{aligned}$$

*where $\mathcal{S}(v;\lambda)$ is the soft-thresholding operator. The direct computation gives:*

$$x^\star = \left\{ \begin{array}{lll} x_1^\star = b_3 - \dfrac{a_1}{a_3} & \text{if} & x_2^\star = b_1 < x_1^\star \\ x_2^\star = b_1 & \text{if} & x_1^\star = b_3 - \dfrac{a_1}{a_3} \leq x_2^\star \leq b_3 + \dfrac{a_1}{a_3} = x_3^\star \\ x_3^\star = b_3 + \dfrac{a_1}{a_3} & \text{if} & x_3^\star < b_1 = x_2^\star \end{array} \right. \tag{137}$$

We can now consider the case $b_1 = b_2$. The first-order condition becomes:

$$\{0\} \in \left( a_1 + a_2 \right) \partial \left| x - b_1 \right| + a_3 \left( x - b_3 \right) \tag{138}$$

The solution corresponds to the case above, and we have:

$$x^\star = b_1 + \operatorname{sign} \left( b_3 - b_1 \right) \cdot \left( \left| b_3 - b_1 \right| - \frac{a_1 + a_2}{a_3} \right)_+ \tag{139}$$

### C.1.3 ADMM solution

We consider the change of variable $y = \mathcal{D}(\boldsymbol{\lambda}) x - \lambda_1 \mathbf{e}_1 x_0$. The ADMM form of the optimization problem (108) is:

$$\begin{aligned} \left\{ x^\star, y^\star \right\} &= \arg \min_{(x,y)} \left\{ \frac{\varphi}{2} \left\| x - v \right\|_2^2 + \left\| y \right\|_1 \right\} \\ & \text{s.t.} \quad \mathcal{D}(\boldsymbol{\lambda}) x - y = \lambda_1 \mathbf{e}_1 x_0 \end{aligned} \tag{140}$$

Using the notations of Perrin and Roncalli (2020), we have $f_x(x) = \frac{\varphi}{2} \left\| x - v \right\|_2^2$, $f_y(y) = \left\| y \right\|_1$, $A = \mathcal{D}(\boldsymbol{\lambda})$, $B = -I_n$ and $c = \lambda_1 \mathbf{e}_1 x_0$. The associated ADMM algorithm consists in the following steps:

1. The $x$-update is:

$$x^{(k+1)} = \arg \min_x \left\{ f_x^{(k+1)}(x) = \frac{\varphi}{2} \left\| x - v \right\|_2^2 + \frac{\varphi'}{2} \left\| \mathcal{D}(\boldsymbol{\lambda}) x - y^{(k)} - \lambda_1 \mathbf{e}_1 x_0 + u^{(k)} \right\|_2^2 \right\} \tag{141}$$

It follows that:

$$\begin{aligned}\partial_x f_x^{(k+1)}(x) &= \varphi(x-v)+\varphi'\mathcal{D}(\boldsymbol{\lambda})^\top\left(\mathcal{D}(\boldsymbol{\lambda})x-y^{(k)}-\lambda_1\mathbf{e}_1x_0+u^{(k)}\right)\\ &= \left(\varphi I_n+\varphi'\mathcal{D}(\boldsymbol{\lambda})^\top\mathcal{D}(\boldsymbol{\lambda})\right)x-\\ &\quad\left(\varphi v+\varphi'\mathcal{D}(\boldsymbol{\lambda})^\top\left(y^{(k)}+\lambda_1\mathbf{e}_1x_0-u^{(k)}\right)\right)\end{aligned}\tag{142}$$

Solving the first-order condition gives:

$$x^{(k+1)}=\left(\varphi I_n+\varphi'\mathcal{D}(\boldsymbol{\lambda})^\top\mathcal{D}(\boldsymbol{\lambda})\right)^{-1}\left(\varphi v+\varphi'\mathcal{D}(\boldsymbol{\lambda})^\top\left(y^{(k)}+\lambda_1\mathbf{e}_1x_0-u^{(k)}\right)\right)\tag{143}$$

2. The $y$-update is:

$$\begin{aligned}y^{(k+1)} &= \arg\min_y\left\{\|y\|_1+\frac{\varphi'}{2}\left\|\mathcal{D}(\boldsymbol{\lambda})x^{(k+1)}-y-\lambda_1\mathbf{e}_1x_0+u^{(k)}\right\|_2^2\right\}\\ &= \mathbf{prox}_{\varphi'^{-1}\|y\|}\left(\mathcal{D}(\boldsymbol{\lambda})x^{(k+1)}-\lambda_1\mathbf{e}_1x_0+u^{(k)}\right)\\ &= \mathcal{S}\left(\mathcal{D}(\boldsymbol{\lambda})x^{(k+1)}-\lambda_1\mathbf{e}_1x_0+u^{(k)};\frac{1}{\varphi'}\right)\end{aligned}\tag{144}$$

where $\mathcal{S}(v;\lambda)=\operatorname{sign}(v)\odot\left(|v|-\lambda\mathbf{1}_n\right)_+$ is the soft-thresholding operator.

3. The $u$-update is:

$$u^{(k+1)}=u^{(k)}+\left(\mathcal{D}(\boldsymbol{\lambda})x^{(k+1)}-y^{(k+1)}-\lambda_1\mathbf{e}_1x_0\right)\tag{145}$$

## C.2 Single-period mean-variance-ridge optimization

We consider the following optimization problem:

$$\begin{aligned} x_s^\star &= \arg\min_{x_s} \left\{ \frac{1}{2} x_s^\top Q_s x_s - x_s^\top R_s + \frac{1}{2} \left(x_s - x_{s-1}\right)^\top \Lambda_s \left(x_s - x_{s-1}\right) \right\} \qquad (146) \\ & \text{s.t.} \left\{ \begin{array}{l} A_s x_s = B_s \\ C_s x_s \leq D_s \\ \mathbf{0}_n \leq \underline{x}_s \leq x_s \leq \overline{x}_s \leq \mathbf{1}_n \end{array} \right. \end{aligned}$$

The objective function is equal to:

$$f_s\left(x_s\right) = \frac{1}{2} x_s^\top \left(Q_s + \Lambda_s\right) x_s - x_s^\top \left(R_s - \Lambda_s x_{s-1}\right) + \frac{1}{2} x_{s-1}^\top \Lambda_s x_{s-1} \qquad (147)$$

We obtain a QP problem:

$$\begin{aligned} x_s^\star &= \arg\min_{x_s} \left\{ \frac{1}{2} x_s^\top \tilde{Q}_s x_s - x_s^\top \tilde{R}_s \right\} \qquad (148) \\ & \text{s.t.} \left\{ \begin{array}{l} A_s x_s = B_s \\ C_s x_s \leq D_s \\ \underline{x}_s \leq x_s \leq \overline{x}_s \end{array} \right. \end{aligned}$$

where $\tilde{Q}_s = Q_s + \Lambda_s$ and $\tilde{R}_s = R_s - \Lambda_s x_{s-1}$. This is equivalent to regularize the covariance matrix $Q_s$ with the matrix $\Lambda_s$ and penalize the vector of expected returns $R_s$ by the '*maximal marginal*' transaction cost $\mathcal{MC}^+ = \Lambda_s x_{s-1}$ of the existing portfolio $x_{s-1}$. Indeed, we have:

$$x_s^\top \tilde{R}_s = x_s^\top R_s - x_s^\top \Lambda_s x_{s-1} \qquad (149)$$

and:

$$\begin{aligned} \mathcal{MC}\left(x_s\right) &= \frac{\partial}{\partial\, x_s} \left( \frac{1}{2} \left(x_s - x_{s-1}\right)^\top \Lambda_s \left(x_s - x_{s-1}\right) \right) \\ &= \Lambda_s \left(x_s - x_{s-1}\right) \qquad (150) \end{aligned}$$

We deduce that[19]:

$$\mathcal{MC}^+ = \sup_{x_s} \left|\mathcal{MC}\left(x_s\right)\right| = \Lambda_s x_{s-1} \qquad (151)$$

[19]The total transaction cost $\mathcal{TC}\left(x_s\right)$ is then bounded by $x_s^\top \Lambda_s x_{s-1}$.

## C.3 Single-period mean-variance-lasso optimization

### C.3.1 Definition of the optimization problem

We consider the following optimization problem:

$$\begin{aligned} x_s^\star &= \arg\min_{x_s} \left\{ \frac{1}{2} x_s^\top Q_s x_s - x_s^\top R_s + \lambda_s \left\| x_s - x_{s-1} \right\|_1 \right\} \\ & \text{s.t. } \left\{ \begin{array}{l} A_s x_s = B_s \\ C_s x_s \leq D_s \\ \mathbf{0}_n \leq \underline{x}_s \leq x_s \leq \overline{x}_s \leq \mathbf{1}_n \end{array} \right. \end{aligned} \tag{152}$$

We assume that $\dim(x_s) = n$, $\dim(B_s) = n_A$ and $\dim(D_s) = n_C$, meaning that we have $n$ variables, $n_A$ equality constraints and $n_C$ inequality constraints. Following Roncalli (2013); Perrin and Roncalli (2020), we introduce the additional variables $x_s^-$ and $x_s^+$ such that:

$$x_s = x_{s-1} - x_s^- + x_s^+ \tag{153}$$

where $x_s^- \geq \mathbf{0}_n$ is the vector of negative weight changes with respect to the previous allocation $x_{s-1}$ and $x_s^+ \geq \mathbf{0}_n$ is the vector of positive weight changes. $x_s^-$ and $x_s^+$ correspond then to selling and buying reallocations. The expression of the lasso penalty becomes:

$$\begin{aligned} \left\| x_s - x_{s-1} \right\|_1 &= \sum_{i=1}^n \left| x_{i,s} - x_{i,s-1} \right| \\ &= \sum_{i=1}^n x_{i,s}^- + \sum_{i=1}^n x_{i,s}^+ \\ &= \mathbf{1}_n^\top x_s^- + \mathbf{1}_n^\top x_s^+ \end{aligned} \tag{154}$$

Since $x_s \geq \underline{x}_s$ and $x_s \leq \overline{x}_s$, we obtain:

$$\begin{aligned} & x_{s-1} - x_s^- + x_s^+ \geq \underline{x}_s \\ \Leftrightarrow\ & x_s^+ - x_s^- \geq \underline{x}_s - x_{s-1} \\ \Leftrightarrow\ & \left\{ \begin{array}{l} x_s^- \leq \max\left(\mathbf{0}_n, x_{s-1} - \underline{x}_s\right) \\ x_s^+ \geq \max\left(\mathbf{0}_n, \underline{x}_s - x_{s-1}\right) \end{array} \right. \end{aligned} \tag{155}$$

and:

$$\begin{aligned} & x_{s-1} - x_s^- + x_s^+ \leq \overline{x}_s \\ \Leftrightarrow\ & x_s^+ - x_s^- \leq \overline{x}_s - x_{s-1} \\ \Leftrightarrow\ & \left\{ \begin{array}{l} x_s^- \geq \max\left(\mathbf{0}_n, x_{s-1} - \overline{x}_s\right) \\ x_s^+ \leq \max\left(\mathbf{0}_n, \overline{x}_s - x_{s-1}\right) \end{array} \right. \end{aligned} \tag{156}$$

We obtain an augmented QP problem:

$$\begin{aligned} x_s^\star &= \arg\min_{x_s} \left\{ \frac{1}{2} x_s^\top Q_s x_s - \left( x_s^\top R_s - \lambda_s \mathbf{1}_n^\top x_s^- - \lambda_s \mathbf{1}_n^\top x_s^+ \right) \right\} \\ & \text{s.t. } \left\{ \begin{array}{l} A_s x_s = B_s \\ x_s + x_s^- - x_s^+ = x_{s-1} \\ C_s x_s \leq D_s \\ \underline{x}_s \leq x_s \leq \overline{x}_s \\ \underline{x}_s^- \leq x_s^- \leq \overline{x}_s^- \\ \underline{x}_s^+ \leq x_s^+ \leq \overline{x}_s^+ \end{array} \right. \end{aligned} \tag{157}$$

where $\underline{x}_s^- = \max\left(\mathbf{0}_n, x_{s-1} - \overline{x}_s\right)$, $\overline{x}_s^- = \max\left(\mathbf{0}_n, x_{s-1} - \underline{x}_s\right)$, $\underline{x}_s^+ = \max\left(\mathbf{0}_n, \underline{x}_s - x_{s-1}\right)$ and $\overline{x}_s^+ = \max\left(\mathbf{0}_n, \overline{x}_s - x_{s-1}\right)$.

### C.3.2 Augmented QP formulation

The augmented QP problem is defined by:

$$\begin{aligned} x^{\star} &= \arg\min_{x} \left\{ \frac{1}{2} x^{\top} Q x - x^{\top} R \right\} \\ & \text{s.t.} \quad \left\{ \begin{array}{l} Ax = B \\ Cx \leq D \\ \underline{x} \leq x \leq \overline{x} \end{array} \right. \end{aligned} \tag{158}$$

where $m = 3n$ and $x = \left(x_s, x_s^-, x_s^+\right)$ is a $m \times 1$ vector. The $Q$ matrix is a $m \times m$ matrix that depends on the quadratic matrix $Q_s$:

$$Q = \left( \begin{array}{ccc} Q_s & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \end{array} \right) \tag{159}$$

For the $R$ vector, we have:

$$R = \left( \begin{array}{c} R_s \\ -\lambda_s \mathbf{1}_n \\ -\lambda_s \mathbf{1}_n \end{array} \right) \tag{160}$$

For the linear equation $Ax = B$, we have:

$$A = \left( \begin{array}{ccc} A_s & \mathbf{0}_{n_A,n} & \mathbf{0}_{n_A,n} \\ I_n & I_n & -I_n \end{array} \right) \tag{161}$$

and

$$B = \left( \begin{array}{c} B_s \\ x_{s-1} \end{array} \right) \tag{162}$$

For the linear inequality $Cx \leq D$, we have:

$$C = \left( \begin{array}{ccc} C_s & \mathbf{0}_{n_C,n} & \mathbf{0}_{n_C,n} \end{array} \right) \tag{163}$$

and $D = D_s$. Finally, the lower and upper bounds are $\underline{x} = \left(\underline{x}_s, \underline{x}_s^-, \underline{x}_s^+\right)$ and $\overline{x} = \left(\overline{x}_s, \overline{x}_s^-, \overline{x}_s^+\right)$.

### C.3.3 The case $\left\|x_s - x_{s-1}\right\|_1 \leq \tau_s$

We consider a variant of the optimization problem (152):

$$\begin{aligned} x_s^{\star} &= \arg\min_{x_s} \left\{ \frac{1}{2} x_s^{\top} Q_s x_s - x_s^{\top} R_s \right\} \\ & \text{s.t.} \quad \left\{ \begin{array}{l} A_s x_s = B_s \\ C_s x_s \leq D_s \\ \left\|x_s - x_{s-1}\right\|_1 \leq \tau_s \\ \mathbf{0}_n \leq \underline{x}_s \leq x_s \leq \overline{x}_s \leq \mathbf{1}_n \end{array} \right. \end{aligned} \tag{164}$$

We use the previous approach for finding the solution. The only differences concern the specification of the vector $R$ and the inequality constraints $Cx \leq D$. We have:

$$R = \left( \begin{array}{c} R_s \\ \mathbf{0}_n \\ \mathbf{0}_n \end{array} \right) \tag{165}$$

The turnover constraint $\|x_s - x_{s-1}\|_1 \leq \tau_s$ is equivalent to $\mathbf{1}_n^\top x_s^- + \mathbf{1}_n^\top x_s^+ \leq \tau_s$. We deduce that:

$$C = \left( \begin{array}{ccc} C_s & \mathbf{0}_{n_C,n} & \mathbf{0}_{n_C,n} \\ \mathbf{0}_n^\top & \mathbf{1}_n^\top & \mathbf{1}_n^\top \end{array} \right) \tag{166}$$

and:

$$D = \left( \begin{array}{c} D_s \\ \tau_s \end{array} \right) \tag{167}$$

## C.4 Double-lasso penalized problem

### C.4.1 Definition of the optimization problem

We consider the following optimization problem:

$$\begin{aligned} x_s^\star &= \arg\min_{x_s} \left\{ \frac{1}{2} x_s^\top Q_s x_s - x_s^\top R_s + \lambda_s \left\| x_s - x_{s-1} \right\|_1 + \lambda_{s+1} \left\| x_{s+1} - x_s \right\|_1 \right\} \\ &\text{s.t.} \quad \left\{ \begin{array}{l} A_s x_s = B_s \\ C_s x_s \leq D_s \\ \mathbf{0}_n \leq \underline{x}_s \leq x_s \leq \overline{x}_s \leq \mathbf{1}_n \end{array} \right. \end{aligned} \tag{168}$$

We assume that $\dim(x_s) = n$, $\dim(B_s) = n_A$ and $\dim(D_s) = n_C$, meaning that we have $n$ variables, $n_A$ equality constraints and $n_C$ inequality constraints. We introduce the additional variables $x_s^-$, $x_s^+$, $x_{s+1}^-$, $x_{s+1}^+$ such that $x_s = x_{s-1} - x_s^- + x_s^+$ and $x_{s+1} = x_s - x_{s+1}^- + x_{s+1}^+$. We have $\left\| x_s - x_{s-1} \right\|_1 = \mathbf{1}_n^\top x_s^- + \mathbf{1}_n^\top x_s^+$ and $\left\| x_{s+1} - x_s \right\|_1 = \mathbf{1}_n^\top x_{s+1}^- + \mathbf{1}_n^\top x_{s+1}^+$. We obtain the augmented QP problem:

$$\begin{aligned} x_s^\star &= \arg\min_{x_s} \left\{ \begin{array}{l} \frac{1}{2} x_s^\top Q_s x_s - \left( x_s^\top R_s - \lambda_s \mathbf{1}_n^\top x_s^- - \lambda_s \mathbf{1}_n^\top x_s^+ \right) + \\ \lambda_{s+1} \mathbf{1}_n^\top x_{s+1}^- + \lambda_{s+1} \mathbf{1}_n^\top x_{s+1}^+ \end{array} \right\} \\ &\text{s.t.} \quad \left\{ \begin{array}{l} A_s x_s = B_s \\ x_s + x_s^- - x_s^+ = x_{s-1} \\ x_s - x_{s+1}^- + x_{s+1}^+ = x_{s+1} \\ C_s x_s \leq D_s \\ \underline{x}_s \leq x_s \leq \overline{x}_s \\ \underline{x}_s^- \leq x_s^- \leq \overline{x}_s^- \\ \underline{x}_s^+ \leq x_s^+ \leq \overline{x}_s^+ \\ \underline{x}_{s+1}^- \leq x_{s+1}^- \leq \overline{x}_{s+1}^- \\ \underline{x}_{s+1}^+ \leq x_{s+1}^+ \leq \overline{x}_{s+1}^+ \end{array} \right. \end{aligned} \tag{169}$$

### C.4.2 Augmented QP formulation

The augmented QP problem is defined by:

$$\begin{aligned} x^\star &= \arg\min_{x} \left\{ \frac{1}{2} x^\top Q x - x^\top R \right\} \\ &\text{s.t.} \quad \left\{ \begin{array}{l} Ax = B \\ Cx \leq D \\ \underline{x} \leq x \leq \overline{x} \end{array} \right. \end{aligned} \tag{170}$$

where $m = 5n$ and $x = \left( x_s, x_s^-, x_s^+, x_{s+1}^-, x_{s+1}^+ \right)$ is a $m \times 1$ vector. The $Q$ matrix is a $m \times m$ matrix that depends on the quadratic matrix $Q_s$:

$$Q = \left( \begin{array}{cc} Q_s & \mathbf{0}_{n,4n} \\ \mathbf{0}_{4n,n} & \mathbf{0}_{4n,4n} \end{array} \right) \tag{171}$$

For the $R$ vector, we have:

$$R = \left( \begin{array}{c} R_s \\ -\lambda_s \mathbf{1}_n \\ -\lambda_s \mathbf{1}_n \\ -\lambda_{s+1} \mathbf{1}_n \\ -\lambda_{s+1} \mathbf{1}_n \end{array} \right) \tag{172}$$

For the linear equation $Ax = B$, we have:

$$A = \begin{pmatrix} A_s & \mathbf{0}_{n_A,n} & \mathbf{0}_{n_A,n} & \mathbf{0}_{n_A,n} & \mathbf{0}_{n_A,n} \\ I_n & I_n & -I_n & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \\ I_n & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & -I_n & I_n \end{pmatrix} \tag{173}$$

and

$$B = \begin{pmatrix} B_s \\ x_{s-1} \\ x_{s+1} \end{pmatrix} \tag{174}$$

For the linear inequality $Cx \leq D$, we have:

$$C = \begin{pmatrix} C_s & \mathbf{0}_{n_C,4n} \end{pmatrix} \tag{175}$$

and $D = D_s$. Finally, the lower and upper bounds are $\underline{x} = \left(\underline{x}_s, \underline{x}_s^-, \underline{x}_s^+, \underline{x}_{s+1}^-, \underline{x}_{s+1}^+\right)$ and $\overline{x} = \left(\overline{x}_s, \overline{x}_s^-, \overline{x}_s^+, \overline{x}_{s+1}^-, \overline{x}_{s+1}^+\right)$ where $\underline{x}_s^- = \max\left(\mathbf{0}_n, x_{s-1} - \overline{x}_s\right)$, $\overline{x}_s^- = \max\left(\mathbf{0}_n, x_{s-1} - \underline{x}_s\right)$, $\underline{x}_s^+ = \max\left(\mathbf{0}_n, \underline{x}_s - x_{s-1}\right)$, $\overline{x}_s^+ = \max\left(\mathbf{0}_n, \overline{x}_s - x_{s-1}\right)$, $\underline{x}_{s+1}^- = \mathbf{0}_n$, $\overline{x}_{s+1}^- = \mathbf{1}_n$, $\underline{x}_{s+1}^+ = \mathbf{0}_n$, $\overline{x}_{s+1}^+ = \mathbf{1}_n$.

### C.4.3 The case $\left\|x_s - x_{s-1}\right\|_1 \leq \tau_s$ and $\left\|x_{s+1} - x_s\right\|_1 \leq \tau_{s+1}$

We consider a variant of the optimization problem (168):

$$\begin{aligned} x_s^\star &= \arg\min_{x_s} \left\{ \frac{1}{2} x_s^\top Q_s x_s - x_s^\top R_s \right\} \\ &\text{s.t.} \left\{ \begin{array}{l} A_s x_s = B_s \\ C_s x_s \leq D_s \\ \left\|x_s - x_{s-1}\right\|_1 \leq \tau_s \\ \left\|x_{s+1} - x_s\right\|_1 \leq \tau_{s+1} \\ \mathbf{0}_n \leq \underline{x}_s \leq x_s \leq \overline{x}_s \leq \mathbf{1}_n \end{array} \right. \end{aligned} \tag{176}$$

We use the previous approach for finding the solution. The only differences concern the specification of the vector $R$ and the inequality constraints $Cx \leq D$. We have:

$$R = \begin{pmatrix} R_s \\ \mathbf{0}_n \\ \mathbf{0}_n \\ \mathbf{0}_n \\ \mathbf{0}_n \end{pmatrix} \tag{177}$$

The turnover constraints $\left\|x_s - x_{s-1}\right\|_1 \leq \tau_s$ and $\left\|x_{s+1} - x_s\right\|_1 \leq \tau_{s+1}$ are equivalent to $\mathbf{1}_n^\top x_s^- + \mathbf{1}_n^\top x_s^+ \leq \tau_s$ and $\mathbf{1}_n^\top x_{s+1}^- + \mathbf{1}_n^\top x_{s+1}^+ \leq \tau_{s+1}$. We deduce that:

$$C = \begin{pmatrix} C_s & \mathbf{0}_{n_C,n} & \mathbf{0}_{n_C,n} & \mathbf{0}_{n_C,n} & \mathbf{0}_{n_C,n} \\ \mathbf{0}_n^\top & \mathbf{1}_n^\top & \mathbf{1}_n^\top & \mathbf{0}_n^\top & \mathbf{0}_n^\top \\ \mathbf{0}_n^\top & \mathbf{0}_n^\top & \mathbf{0}_n^\top & \mathbf{1}_n^\top & \mathbf{1}_n^\top \end{pmatrix} \tag{178}$$

and:

$$D = \begin{pmatrix} D_s \\ \tau_s \\ \tau_{s+1} \end{pmatrix} \tag{179}$$

## C.5 Multi-period mean-variance-ridge optimization

We have:

$$g_s(x_s) = \frac{1}{2} x_s^\top Q_s x_s - x_s^\top R_s \tag{180}$$

and:

$$\begin{aligned} h_s(x_{s-1}, x_s) &= \frac{1}{2} (x_s - x_{s-1})^\top \Lambda_s (x_s - x_{s-1}) \\ &= \frac{1}{2} x_s^\top \Lambda_s x_s - x_s^\top \Lambda_s x_{s-1} + \frac{1}{2} x_{s-1}^\top \Lambda_s x_{s-1} \end{aligned} \tag{181}$$

We deduce that:

$$\begin{aligned} f(x) &= g(x) + h(x) \\ &= \frac{1}{2} \sum_{s=t+1}^{t+h} x_s^\top (Q_s + \Lambda_s) x_s - \sum_{s=t+1}^{t+h} x_s^\top \Lambda_s x_{s-1} + \frac{1}{2} \sum_{s=t+1}^{t+h} x_{s-1}^\top \Lambda_s x_{s-1} - \\ & \quad \sum_{s=t+1}^{t+h} x_s^\top R_s \\ &= \frac{1}{2} x^\top Q x - x^\top R + \text{constant} \end{aligned} \tag{182}$$

where $x = (x_{t+1}, \ldots, x_{t+h})$, the matrix $Q = Q_1 + Q_2$ is a block tridiagonal matrix and:

$$R = \begin{pmatrix} R_{t+1} \\ R_{t+2} \\ \vdots \\ R_{t+h} \end{pmatrix} + \begin{pmatrix} \Lambda_{t+1} x_t \\ \mathbf{0}_n \\ \vdots \\ \mathbf{0}_n \end{pmatrix} \tag{183}$$

For $Q = Q_1 + Q_2$, we have:

$$Q_1 = \begin{pmatrix} Q_{t+1} + \Lambda_{t+1} & -\Lambda_{t+2} & & & & \mathbf{0}_{n,n} \\ -\Lambda_{t+2} & Q_{t+2} + \Lambda_{t+2} & -\Lambda_{t+3} & & & \\ & -\Lambda_{t+3} & Q_{t+3} + \Lambda_{t+3} & & & \\ & & \ddots & & & \\ & & -\Lambda_{t+h-1} & Q_{t+h-1} + \Lambda_{t+h-1} & -\Lambda_{t+h} \\ \mathbf{0}_{n,n} & & & -\Lambda_{t+h} & Q_{t+h} + \Lambda_{t+h} \end{pmatrix} \tag{184}$$

and:

$$Q_2 = \begin{pmatrix} \Lambda_{t+2} & & & & \mathbf{0}_{n,n} \\ & \Lambda_{t+3} & & & \\ & & \ddots & & \\ & & & \Lambda_{t+h} & \\ \mathbf{0}_{n,n} & & & & \mathbf{0}_{n,n} \end{pmatrix} \tag{185}$$

The constant term is equal to $\frac{1}{2} x_t^\top \Lambda_{t+1} x_t$.

## C.6 Separable linear constraints in a multi-period optimization problem

If the constraints are fully separable:

$$x \in \Omega \Leftrightarrow \left\{ \begin{array}{l} A_s x_s = B_s \\ C_s x_s \leq D_s \\ \underline{x}_s \leq x_s \leq \overline{x}_s \end{array} \right. \tag{186}$$

we obtain:

$$x \in \Omega \Leftrightarrow \left\{ \begin{array}{l} Ax = B \\ Cx \leq D \\ \underline{x} \leq x \leq \overline{x} \end{array} \right. \tag{187}$$

The matrices $A$ and $C$ are block tridiagonal matrices. We have:

$$Ax = B \Leftrightarrow \begin{pmatrix} A_{t+1} & & & \mathbf{0}_{n_{A_{t+1}},n} \\ & A_{t+2} & & \\ & & \ddots & \\ \mathbf{0}_{n_{A_{t+h}},n} & & & A_{t+h} \end{pmatrix} \begin{pmatrix} x_{t+1} \\ x_{t+2} \\ \vdots \\ x_{t+h} \end{pmatrix} = \begin{pmatrix} B_{t+1} \\ B_{t+2} \\ \vdots \\ B_{t+h} \end{pmatrix} \tag{188}$$

and:

$$Cx \leq D \Leftrightarrow \begin{pmatrix} C_{t+1} & & & \mathbf{0}_{n_{C_{t+1}},n} \\ & C_{t+2} & & \\ & & \ddots & \\ \mathbf{0}_{n_{C_{t+h}},n} & & & C_{t+h} \end{pmatrix} \begin{pmatrix} x_{t+1} \\ x_{t+2} \\ \vdots \\ x_{t+h} \end{pmatrix} \leq \begin{pmatrix} D_{t+1} \\ D_{t+2} \\ \vdots \\ D_{t+h} \end{pmatrix} \tag{189}$$

For the bounds, we have:

$$\underline{x} \leq x \leq \overline{x} \Leftrightarrow \begin{pmatrix} \underline{x}_{t+1} \\ \underline{x}_{t+2} \\ \vdots \\ \underline{x}_{t+h} \end{pmatrix} \leq \begin{pmatrix} x_{t+1} \\ x_{t+2} \\ \vdots \\ x_{t+h} \end{pmatrix} \leq \begin{pmatrix} \overline{x}_{t+1} \\ \overline{x}_{t+2} \\ \vdots \\ \overline{x}_{t+h} \end{pmatrix} \tag{190}$$

## C.7 Multi-period mean-variance-lasso optimization

We extend Appendix C.4 to the multi-period framework.

### C.7.1 Definition of the optimization problem

We consider the following optimization problem:

$$\begin{aligned} x_{t+1}^{\star} &= \arg \min_{x_{t+1},x_{t+2},\ldots} \left\{ \sum_{s=t+1}^{t+h} \frac{1}{2} x_s^\top Q_s x_s - x_s^\top R_s + \lambda_s \left\| x_s - x_{s-1} \right\|_1 \right\} \\ & \text{s.t.} \quad \left\{ \begin{array}{l} A_s x_s = B_s \\ C_s x_s \leq D_s \\ \mathbf{0}_n \leq \underline{x}_s \leq x_s \leq \overline{x}_s \leq \mathbf{1}_n \end{array} \right. \end{aligned} \tag{191}$$

We assume that $\dim(x_s) = n$, $\dim(B_s) = n_A$ and $\dim(D_s) = n_C$, meaning that we have $n$ variables, $n_A$ equality constraints and $n_C$ inequality constraints. As previously, we introduce the additional variables $x_s^-$ and $x_s^+$ such that $x_s = x_{s-1} - x_s^- + x_s^+$. We reiterate that the expression of the turnover is $\left\| x_s - x_{s-1} \right\|_1 = \mathbf{1}_n^\top x_s^- + \mathbf{1}_n^\top x_s^+$. The treatment of the bounds is more complicated. For instance, the single-period constraint $x_s^- \leq \max(\mathbf{0}_n, x_{s-1} - \underline{x}_s)$ is equivalent to the upper bound $x_s^- \leq \overline{x}_s^- = \max(\mathbf{0}_n, x_{s-1} - \underline{x}_s)$ because $x_{s-1}$ is a given variable. In the multi-period problem, the constraint $x_s^- \leq \max(\mathbf{0}_n, x_{s-1} - \underline{x}_s)$ cannot be cast into an upper bound since $x_{s-1}$ is an endogenous variable (except for the case $s = t$). Therefore, we have:

$$\begin{aligned} x_s^- &\leq \max(\mathbf{0}_n, x_{s-1} - \underline{x}_s) \\ &\leq \max(\mathbf{0}_n, \overline{x}_{s-1} - \underline{x}_s) \\ &\leq \overline{x}_{s-1} \end{aligned} \tag{192}$$

and:

$$\begin{aligned} x_s^- &\geq \max(\mathbf{0}_n, x_{s-1} - \overline{x}_s) \\ &\geq \max\left(\mathbf{0}_n, \underline{x}_{s-1} - \overline{x}_s\right) \\ &\geq \mathbf{0}_n \end{aligned} \tag{193}$$

Similarly, we have $x_s^+ \leq \overline{x}_{s-1}$ and $x_s^+ \geq \mathbf{0}_n$. Nevertheless, the constraint $\underline{x}_s \leq x_s \leq \bar{x}_s$ imposes that:

$$\underline{x}_s - x_{s-1} \leq x_s^+ - x_s^- \leq \bar{x}_s - x_{s-1} \tag{194}$$

These constraints are difficult to manage. Therefore, we use less restrictive constraints: $x_s^- \leq x_{s-1}$ and $x_s^+ \leq \mathbf{1}_n - x_{s-1}$. Finally, we obtain an augmented QP problem:

$$\begin{aligned} x_{t+1}^{\star} &= \arg \min_{x_{t+1},x_{t+2},\ldots} \left\{ \sum_{s=t+1}^{t+h} \frac{1}{2} x_s^\top Q_s x_s - \left( x_s^\top R_s - \lambda_s \mathbf{1}_n^\top x_s^- - \lambda_s \mathbf{1}_n^\top x_s^+ \right) \right\} \\ & \text{s.t.} \quad \left\{ \begin{array}{l} A_s x_s = B_s \\ x_s + x_s^- - x_s^+ - x_{s-1} = \mathbf{0}_n \\ C_s x_s \leq D_s \\ x_s^- \leq x_{s-1} \\ x_s^+ \leq \mathbf{1}_n - x_{s-1} \\ \underline{x}_s \leq x_s \leq \bar{x}_s \\ \mathbf{0}_n \leq x_s^- \leq \overline{x}_{s-1} \\ \mathbf{0}_n \leq x_s^+ \leq \overline{x}_{s-1} \end{array} \right. \end{aligned} \tag{195}$$

### C.7.2 Augmented QP formulation

**Objective function** The augmented QP problem is defined by:

$$\begin{aligned} x^{\star} &= \arg\min_{x} \left\{ \frac{1}{2} x^{\top} Q x - x^{\top} R \right\} \\ \text{s.t.} & \left\{ \begin{array}{l} Ax = B \\ Cx \leq D \\ \underline{x} \leq x \leq \bar{x} \end{array} \right. \end{aligned} \tag{196}$$

where $m = 3nh$ and $x = \left( x_{t+1}, x_{t+1}^{-}, x_{t+1}^{+}, \ldots, x_{t+h}, x_{t+h}^{-}, x_{t+h}^{+} \right)$ is a $m \times 1$ vector. The $Q$ matrix is a $m \times m$ matrix that depends on the quadratic matrices $Q_{t+1}, \ldots, Q_{t+h}$:

$$Q = \begin{pmatrix} \tilde{Q}_{t+1} & \mathbf{0}_{3n,3n} & & \mathbf{0}_{3n,3n} \\ \mathbf{0}_{3n,3n} & \tilde{Q}_{t+2} & \ddots & \\ & \ddots & \ddots & \mathbf{0}_{3n,3n} \\ \mathbf{0}_{3n,3n} & & \mathbf{0}_{3n,3n} & \tilde{Q}_{t+h} \end{pmatrix} \tag{197}$$

and:

$$\tilde{Q}_s = \begin{pmatrix} Q_s & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \end{pmatrix} \tag{198}$$

For the $R$ vector, we have:

$$R = \begin{pmatrix} \tilde{R}_{t+1} \\ \tilde{R}_{t+2} \\ \vdots \\ \tilde{R}_{t+h} \end{pmatrix} \tag{199}$$

and:

$$\tilde{R}_s = \begin{pmatrix} R_s \\ -\lambda_s \mathbf{1}_n \\ -\lambda_s \mathbf{1}_n \end{pmatrix} \tag{200}$$

**Linear equality constraints** For the linear equation $Ax = B$, we have:

$$A = \begin{pmatrix} \tilde{A}_{t+1} & \mathbf{0}_{n_A,3n} & & & \mathbf{0}_{n_A,3n} \\ \breve{A}_{t+2} & \tilde{A}_{t+2} & \ddots & & \\ \mathbf{0}_{n_A,3n} & \ddots & \ddots & \mathbf{0}_{n_A,3n} & \\ & \ddots & \breve{A}_{t+h-1} & \tilde{A}_{t+h-1} & \mathbf{0}_{n_A,3n} \\ \mathbf{0}_{n_A,3n} & & \mathbf{0}_{n_A,3n} & \breve{A}_{t+h} & \tilde{A}_{t+h} \end{pmatrix} \tag{201}$$

and:

$$B = \begin{pmatrix} \tilde{B}_{t+1} \\ \tilde{B}_{t+2} \\ \vdots \\ \tilde{B}_{t+h} \end{pmatrix} \tag{202}$$

The dimension of the $A$ matrix is $m_A \times m$ where $m_A = (n_A + n)\, h$. The computation of $A$ requires the following matrices[20]:

$$\tilde{A}_s = \begin{pmatrix} A_s & \mathbf{0}_{n_A,n} & \mathbf{0}_{n_A,n} \\ I_n & I_n & -I_n \end{pmatrix} \tag{203}$$

and:

$$\breve{A}_s = \begin{pmatrix} \mathbf{0}_{n_A,n} & \mathbf{0}_{n_A,n} & \mathbf{0}_{n_A,n} \\ -I_n & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \end{pmatrix} \tag{204}$$

The dimension of the $B$ matrix is $m_A \times 1$. We have:

$$\tilde{B}_s = \begin{pmatrix} B_s \\ \mathbf{0}_n \end{pmatrix} \qquad \text{for } s > t+1 \tag{205}$$

and:

$$\tilde{B}_{t+1} = \begin{pmatrix} B_{t+1} \\ x_t \end{pmatrix} \tag{206}$$

**Linear inequality constraints** For the linear inequality $Cx \leq D$, we have:

$$C = \begin{pmatrix} \tilde{C}_{t+1} & \mathbf{0}_{n_C,3n} & & & \mathbf{0}_{n_C,3n} \\ \breve{C}_{t+2} & \tilde{C}_{t+2} & \ddots & & \\ \mathbf{0}_{n_C,3n} & \ddots & \ddots & \mathbf{0}_{n_C,3n} & \\ & \ddots & \breve{C}_{t+h-1} & \tilde{C}_{t+h-1} & \mathbf{0}_{n_C,3n} \\ \mathbf{0}_{n_C,3n} & & \mathbf{0}_{n_C,3n} & \breve{C}_{t+h} & \tilde{C}_{t+h} \end{pmatrix} \tag{207}$$

and:

$$D = \begin{pmatrix} \tilde{D}_{t+1} \\ \tilde{D}_{t+2} \\ \vdots \\ \tilde{D}_{t+h} \end{pmatrix} \tag{208}$$

The dimension of the $C$ matrix is $m_C \times m$ where $m_C = (n_C + 2n)\, h$. The computation of $C$ requires the following matrices[21]:

$$\tilde{C}_s = \begin{pmatrix} C_s & \mathbf{0}_{n_C,n} & \mathbf{0}_{n_C,n} \\ \mathbf{0}_{n,n} & I_n & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & I_n \end{pmatrix} \tag{209}$$

and:

$$\breve{C}_s = \begin{pmatrix} \mathbf{0}_{n_C,n} & \mathbf{0}_{n_C,n} & \mathbf{0}_{n_C,n} \\ -I_n & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \\ I_n & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \end{pmatrix} \tag{210}$$

The dimension of the $D$ vector is $m_C \times 1$. We have:

$$\tilde{D}_s = \begin{pmatrix} D_s \\ \mathbf{0}_n \\ \mathbf{1}_n \end{pmatrix} \qquad \text{for } s > t+1 \tag{211}$$

[20] $\tilde{A}_s$ and $\breve{A}_s$ are two $(n_A + n) \times 3n$ matrices.
[21] $\tilde{C}_s$ and $\breve{C}_s$ are two $(n_C + 2n) \times 3n$ matrices.

and:

$$\tilde{D}_{t+1} = \begin{pmatrix} D_{t+1} \\ x_t \\ \mathbf{1}_n - x_t \end{pmatrix} \tag{212}$$

**Bounds** Finally, the lower and upper bounds are:

$$\underline{x} = \left(\underline{x}_{t+1}, \mathbf{0}_n, \mathbf{0}_n, \ldots, \underline{x}_{t+h}, \mathbf{0}_n, \mathbf{0}_n\right) \tag{213}$$

and:

$$\overline{x} = \left(\overline{x}_{t+1}, \mathbf{1}_n, \mathbf{1}_n, \ldots, \overline{x}_{t+h}, \mathbf{1}_n, \mathbf{1}_n\right) \tag{214}$$

### C.7.3 The case $\|x_s - x_{s-1}\|_1 \leq \tau_s$

As previsouly, the only differences concern the specification of the vector $R$ and the inequality constraints $Cx \leq D$. We have:

$$\tilde{R}_s = \begin{pmatrix} R_s \\ \mathbf{0}_n \\ \mathbf{0}_n \end{pmatrix} \tag{215}$$

The turnover constraint $\|x_s - x_{s-1}\|_1 \leq \tau_s$ is equivalent to $\mathbf{1}_n^\top x_s^- + \mathbf{1}_n^\top x_s^+ \leq \tau_s$. The dimension of the $C$ matrix is $m_C \times m$ where $m_C = (n_C + 2n + 1)\,h$. The computation of $C$ requires the following matrices[22]:

$$\tilde{C}_s = \begin{pmatrix} C_s & \mathbf{0}_{n_C,n} & \mathbf{0}_{n_C,n} \\ \mathbf{0}_{n,n} & I_n & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & I_n \\ \mathbf{0}_{n,n} & \mathbf{1}_n^\top & \mathbf{1}_n^\top \end{pmatrix} \tag{216}$$

and:

$$\breve{C}_s = \begin{pmatrix} \mathbf{0}_{n_C,n} & \mathbf{0}_{n_C,n} & \mathbf{0}_{n_C,n} \\ -I_n & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \\ I_n & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} \\ \mathbf{0}_n^\top & \mathbf{0}_n^\top & \mathbf{0}_n^\top \end{pmatrix} \tag{217}$$

The dimension of the $D$ vector is $m_C \times 1$. We have:

$$\tilde{D}_s = \begin{pmatrix} D_s \\ \mathbf{0}_n \\ \mathbf{1}_n \\ \tau_s \end{pmatrix} \qquad \text{for } s > t+1 \tag{218}$$

and:

$$\tilde{D}_{t+1} = \begin{pmatrix} D_{t+1} \\ x_t \\ \mathbf{1}_n - x_t \\ \tau_{t+1} \end{pmatrix} \tag{219}$$

[22] $\tilde{C}_s$ and $\breve{C}_s$ are two $(n_C + 2n + 1) \times 3n$ matrices.

### C.7.4 The case $\sum_{s=t+1}^{t+h} \left\| x_s - x_{s-1} \right\|_1 \leq \tau$

Whereas the constraint $\left\| x_s - x_{s-1} \right\|_1 \leq \tau_s$ controls the turnover management between $s-1$ and $s$, the constraint $\sum_{s=t+1}^{t+h} \left\| x_s - x_{s-1} \right\|_1 \leq \tau$ limits the total turnover over the period $[t+1, t+h]$. We have:

$$\sum_{s=t+1}^{t+h} \left\| x_s - x_{s-1} \right\|_1 \leq \tau \Leftrightarrow \sum_{s=t+1}^{t+h} \left( \mathbf{1}_n^\top x_s^- + \mathbf{1}_n^\top x_s^+ \right) \leq \tau \tag{220}$$

In this case, the dimension of the $C$ matrix becomes $m_C \times m$ where $m_C = (n_C + 2n)\, h + 1$. We have to add a last constraint to the existing constraints $Cx \leq D$:

$$\left( \begin{array}{ccccccc} \mathbf{0}_n^\top & \mathbf{1}_n^\top & \mathbf{1}_n^\top & \cdots & \mathbf{0}_n^\top & \mathbf{1}_n^\top & \mathbf{1}_n^\top \end{array} \right) \leq \tau \tag{221}$$

We can also mix the single-period turnover constraint and the total turnover constraint. Again, we add the previous constraint to the existing linear system $Cx \leq D$ and the dimension of the matrix $C$ becomes $m_C \times m$ where $m_C = (n_C + 2n + 1)\, h + 1$.

### C.7.5 Remark about the constraints

In this section, we have assumed that the number of linear equality $A_s x_s = B_s$ and the number of linear inequality $C_s x_s \leq D_s$ are fixed and do not depend on the time $s = t+1, \ldots, t+h$. We can of course consider that the number of constraints vary with respect to the time $s$. The notations are more complex since $n_A$ and $n_C$ are time-varying. This only changes the dimension of the null matrices $\mathbf{0}_{n_A,3n}$ and $\mathbf{0}_{n_C,3n}$.

## C.8 Transition management constraints

Let $x_t$ and $x^\star$ be the current and target portfolios. The transition management process requires that the allocation in an asset can only be increasing or decreasing:

$$\begin{cases} x_i^\star > x_{i,t} \Leftrightarrow x_{i,s} \geq x_{i,s-1} \\ x_i^\star < x_{i,t} \Leftrightarrow x_{i,s} \leq x_{i,s-1} \end{cases} \tag{222}$$

We note $\epsilon = \operatorname{sign}(x^\star - x_t)$ the vector of $\pm 1$, where $\epsilon_i$ indicates whether $x_i^\star > x_{i,t}$ ($\epsilon_i = +1$) or $x_i^\star < x_{i,t}$ ($\epsilon_i = -1$). The condition (222) is equivalent to the system of inequalities $Cx \leq D$ where $x = (x_{t+1}, \ldots, x_{t+h})$ is the $nh \times 1$ vector of weights. We have:

$$C = \begin{pmatrix} \tilde{C} & & & \mathbf{0}_{n,n} \\ \breve{C} & \tilde{C} & & \\ & \ddots & \ddots & \\ \mathbf{0}_{n,n} & & \breve{C} & \tilde{C} \end{pmatrix} \tag{223}$$

where $\tilde{C} = \operatorname{diag}(-\epsilon)$, $\breve{C} = \operatorname{diag}(\epsilon)$ and:

$$D = \begin{pmatrix} -\epsilon \odot x_t \\ \mathbf{0}_n \\ \vdots \\ \mathbf{0}_n \end{pmatrix} \tag{224}$$

The conditions $x_i^\star > x_{i,t}$ and $x_i^\star < x_{i,t}$ also implies that $x_{i,s} \leq x_i^\star$ and $x_{i,s} \geq x_i^\star$. They can be implemented using the lower and upper bounds $\underline{x} \leq x \leq \overline{x}$.

## C.9 Trading trajectory problem

### C.9.1 Objective function

We consider the optimal trading trajectory problem:

$$\begin{aligned} x_{t+1}^{\star} &= \arg\min \sum_{s=t+1}^{t+h} f_s\left(x_s\right) \\ & \text{s.t.} \quad \left\{ \begin{array}{l} \mathbf{1}_n^{\top} x_s = 1 \\ x_s \geq \mathbf{0}_n \end{array} \right. \end{aligned} \tag{225}$$

where:

$$\begin{aligned} f_s\left(x_s\right) &= \frac{1}{2} x_s^{\top} Q_s x_s - x_s^{\top}\left(R_s - \phi \Gamma_s \Delta x_s\right) + \\ & \frac{1}{2} \Delta x_s^{\top} \Lambda_s \Delta x_s - \varepsilon\left(x_{s-1}^{\top} \Gamma_s \Delta x_s + \frac{1}{2} \Delta x_s^{\top} \Gamma_s \Delta x_s\right) \end{aligned} \tag{226}$$

Since $\Delta x_s = x_s - x_{s-1}$, it follows that:

$$\begin{aligned} f_s\left(x_s\right) &= \frac{1}{2} x_s^{\top} Q_s x_s - x_s^{\top}\left(R_s - \phi \Gamma_s\left(x_s - x_{s-1}\right)\right) + \\ & \frac{1}{2}\left(x_s - x_{s-1}\right)^{\top}\left(\Lambda_s - \varepsilon \Gamma_s\right)\left(x_s - x_{s-1}\right) - \varepsilon x_{s-1}^{\top} \Gamma_s\left(x_s - x_{s-1}\right) \\ &= \frac{1}{2} x_s^{\top} Q_s x_s - x_s^{\top} R_s + x_s^{\top}\left(\phi \Gamma_s\right) x_s - x_s^{\top}\left(\phi \Gamma_s\right) x_{s-1} + \\ & \frac{1}{2} x_s^{\top}\left(\Lambda_s - \varepsilon \Gamma_s\right) x_s - x_s^{\top}\left(\Lambda_s - \varepsilon \Gamma_s\right) x_{s-1} + \frac{1}{2} x_{s-1}^{\top}\left(\Lambda_s - \varepsilon \Gamma_s\right) x_{s-1} - \\ & x_s^{\top}\left(\varepsilon \Gamma_s\right) x_{s-1} + x_{s-1}^{\top}\left(\varepsilon \Gamma_s\right) x_{s-1} \\ &= \frac{1}{2} x_s^{\top} Q_s^{0,0} x_s - x_s^{\top} Q_s^{0,1} x_{s-1} + \frac{1}{2} x_{s-1}^{\top} Q_s^{1,1} x_{s-1} - x_s^{\top} R_s \end{aligned}$$

where:

$$\left\{ \begin{array}{lcl} Q_s^{0,0} &=& Q_s + \Lambda_s + (2\phi - \varepsilon)\Gamma_s \\ Q_s^{0,1} &=& \Lambda_s + \phi \Gamma_s \\ Q_s^{1,1} &=& \Lambda_s + \varepsilon \Gamma_s \end{array} \right. \tag{227}$$

### C.9.2 QP problem

The QP problem is defined by:

$$\begin{aligned} x^{\star} &= \arg\min_x \left\{\frac{1}{2} x^{\top} Q x - x^{\top} R\right\} \\ & \text{s.t.} \quad \left\{ \begin{array}{l} Ax = B \\ Cx \leq D \\ \underline{x} \leq x \leq \bar{x} \end{array} \right. \end{aligned}$$

where $x = \left(x_{t+1}, \ldots, x_{t+h}\right)$, the matrix $Q = Q_1 + Q_2$ is a block tridiagonal matrix and:

$$R = \begin{pmatrix} R_{t+1} \\ R_{t+2} \\ \vdots \\ R_{t+h} \end{pmatrix} + \begin{pmatrix} Q_{t+1}^{0,1} x_t \\ \mathbf{0}_n \\ \vdots \\ \mathbf{0}_n \end{pmatrix} \tag{228}$$

For $Q = Q_1 + Q_2$, we have:

$$Q_1 = \begin{pmatrix} Q_{t+1}^{0,0} & -Q_{t+2}^{0,1} & & & \mathbf{0}_{n,n} \\ -Q_{t+2}^{0,1} & Q_{t+2}^{0,0} & -Q_{t+3}^{0,1} & & \\ & -Q_{t+3}^{0,1} & Q_{t+3}^{0,0} & & \\ & & \ddots & & \\ & & -Q_{t+h-1}^{0,1} & Q_{t+h-1}^{0,0} & -Q_{t+h}^{0,1} \\ \mathbf{0}_{n,n} & & & -Q_{t+h}^{0,1} & Q_{t+h}^{0,0} \end{pmatrix} \tag{229}$$

and:

$$Q_2 = \begin{pmatrix} Q_{t+2}^{1,1} & & & & \mathbf{0}_{n,n} \\ & Q_{t+3}^{1,1} & & & \\ & & \ddots & & \\ & & & Q_{t+h}^{1,1} & \\ \mathbf{0}_{n,n} & & & & \mathbf{0}_{n,n} \end{pmatrix} \tag{230}$$

### C.9.3 Linear constraints

If the linear constraints are separable, we use the results given in Appendix C.6 on page 54.