Working Paper 108-2021 | **March 2021**

Bond Index Tracking with Genetic Algorithms

Confidence must be earned
Amundi
ASSET MANAGEMENT

Bond Index Tracking with Genetic Algorithms

Abstract

Mohamed BEN SLIMANE
Quantitative Research,
Amundi
mohamed.benslimane@amundi.com

Bond portfolio optimization is very different from equity portfolio optimization. Indeed, while continuous optimization is efficient when managing a portfolio of stocks, it is not always well-adapted to building a bond portfolio because the transformation of portfolio weights into numbers of shares may lead to significant rounding errors. Indeed, bond investors are often restricted to purchasing bonds in multiples of a minimum transaction unit, which can be expressed as a minimum number of bonds or a minimum amount of dollars. This is why discrete optimization has generally replaced continuous optimization when the investment universe concerns fixed-income securities, especially when we consider the business of passive management, such as bond index funds, exchange-traded funds, and dedicated fixed-income funds. Generally, bond portfolio optimization consists in tracking a benchmark index or a current portfolio with investment constraints, which mainly concern the transaction costs, the liquidity axes and more recently ESG and carbon risk restrictions.

Solving such problems requires using integer programming and combinatorial optimization algorithms. This article focuses on genetic algorithms that have proven to be efficient for large-scale optimization problems. We show how to implement them in the context of bond index tracking, where the goal is to match several tracking risk metrics of a bond portfolio facing subscriptions or redemptions under multiple practical constraints such as minimum tradable amounts, secondary market lot sizes, liquidity axes, and market-making constraints. We also present simulations of a real portfolio with several definitions of genetic operators and determine the best simulation according to matching and transaction cost criteria. Our experience shows that genetic algorithms are an industrial solution for passive bond management even though the entry cost to be familiar with these algorithms is high. The stability and convergence of investment solutions require thousands of tests before proposing an automation to fund managers.

Keywords: Bond indexation, portfolio replication, market making, index sampling, genetic algorithm, constraint handling, duration, spread, credit risk, liquidity

JEL classification: C61, G11, G12

Acknowledgement
The author is very grateful to Edmond Lezmi for his invaluable assistance, Thierry Roncalli and Takaya Sekine for their insightful comments and support, and Stéphanie Pless and Fabrice Degni Yace for their valuable input and advice.

About the author

Mohamed Ben Slimane

Mohamed Ben Slimane is Head of Fixed-Income Quantitative Research in the Amundi Quantitative Research team. Prior to his current position, he was Senior Quantitative Analyst (2015-2020), Head of Credit and Counterparty Risks (2013-2015) and Head of Regulatory Risk for Fixed-Income and Structured Funds (2010-2013). Before joining Amundi, he was Head of Regulatory Risk at Sociéte Générale AM (2005-2010), an IT Consultant with Altran (2002-2005) and a Software Engineer at Sungard Systems (2001-2002). He started his working career with Cap Gemini in 1999 as a Software Engineer.

Mohamed holds an Engineering Degree from the Ecole Nationale de l'Aviation Civile (1999). He is Certified International Investment Analyst CIIA (2010) and FRM Charterholder (2013).

1 Introduction

Passive fund managers' goal is to replicate a market index's performance by holding the same securities or a stratified sampling of the securities (Neyman, 1992) that comprise the index. Therefore, they track the index portfolio by exhibiting the same risk/return characteristics. In the fixed-income space, modified duration (MD) and duration-times-spread (DTS) are the most widely used risk metrics. Indeed, historical price volatility, which is used to measure equity portfolios' risk, is not a reliable predictor of bond volatility since bonds are less frequently traded and mature over time. Therefore, fund managers use MD, which is the sensitivity of the bond return to interest risk, and DTS, which measures the systematic exposure to credit risk by quantifying sensitivity to a shift in the yield spread (Ben Dor *et al.*, 2007).

Modified duration is frequently tracked per maturity bucket. Seven buckets are generally considered: 0Y, 2Y, 5Y, 7Y, 10Y, 15Y, and 20Y. For instance, the five-year bucket refers to the bonds whose term to maturity is around five years. A bond maturing in 3 years contributes two-thirds[1] of its MD to bucket 2Y and then one-third of its MD to bucket 5Y. The weight is the last metric to be considered. the market index is often tracked using a sector-neutral approach[2]. Fund managers do not take active positions, and hence the weights of sectors are matched within an allowed deviation range.

Passive fund managers track broad bond indices that generally include thousands of individual securities. For instance, the Bloomberg Barclays US Corporate Bond Index has 6 647 members as of 31 January 2021. Fund managers sample these indices preferring liquid bonds as their purchase or sale does not cause a drastic change in their price. The liquidity cost can be proxied by the bond price times the liquidity score developed by Ben Slimane and De Jong (2017). This score, which is highly correlated to bid-ask spreads, is based on seventeen characteristics of the bond indenture. The older a debt obligation, the higher the score typically, or by the same token, the longer the time-to-maturity, or the smaller the debt issue, the higher the score.

Besides liquidity, portfolio managers of bond exchange-traded funds (ETF) may have to comply with market-making constraints. Indeed, market makers create ETF units by delivering a basket of underlying securities to the ETF sponsor in exchange for a block of the ETF units with the same market value. In a redemption process, the market maker exchanges ETF units with the ETF provider for an equivalent basket of underlying securities from the ETF. Market makers may then provide fund managers with their constraints on the bonds to be bought or sold. These constraints, referred to hereafter as *axis constraints*, may impose, if possible, that bonds are part of an authorised list $\mathcal{A}$ and that traded quantities do not exceed an allowed quantity per bond.

Portfolio managers of fixed-income index funds match their portfolios with the associated benchmark at least once a month to diminish the tendency for "*portfolio drift*" and potentially reduce their exposure to risk relative to their target asset allocation. Bonds with a final maturity of less than one year are sold, and the proceeds are used, in conjunction with the available cash, to match the metrics of the new benchmark. The adjustment consists of buying bonds most of the time, as the new benchmark is riskier[3] from month to month.

---

[1] We have $\dfrac{5-3}{5-2} = \dfrac{2}{3}$.

[2] Country-neutral when the market index is a government bond index.

[3] As new bonds are included by the end of each month.

Like any mutual funds, fixed-income index funds experience subscriptions and redemptions between the two monthly rebalancing dates. It is common to handle them based on *in-kind baskets*. Fund managers produce these pro-forma portfolios daily, worth EUR 5 million, for example. Two baskets are typically produced per fund, one of purchasable bonds for potential inflows and one of saleable bonds for outflows. In the event of an in-or-outflow, if the size matches and all parties agree, the in-kind basket will in effect be traded; otherwise, a new basket is agreed upon and traded. Unlike the rebalancing process, the reference when building in-kind baskets is the current portfolio. The intention is to keep the risk metrics unaffected by the in-kind process activated only by the subscription-redemption flows.

The study's objective is to design a tool to automatically handle the excessive workload the fund managers are facing with daily flows. The tool must match the different metrics and consider the liquidity and axis constraints and generate a tradable basket of bonds. Bonds, unlike equities, cannot be traded at one unit of quantity. Each bond has its minimum tradable amount, under which one cannot buy or sell a quantity of the bond (for example EUR 100 000). Above this minimum tradable amount, lot size is the incremental nominal amount traded (for example EUR 1 000).

The automation could be seen as an optimization problem where the objective function is to minimize the tracking risks under the constraints of liquidity and axis. A significant challenge is then to handle the discrete quantities of bonds. This challenge can be taken up thanks to genetic algorithms. This article is structured as follows. Section Two is dedicated to presenting the genetic algorithms, while Section Three applies them to bond portfolio optimization. In Section Four, we discuss the results of two examples of in-kind adjustments performed thanks to genetic algorithms. Finally, Section Five offers some concluding remarks.

2 The concept of genetic algorithms

A genetic algorithm (GA) is a search-based optimization technique that mimics the Genetics and Natural Selection principles elaborated by Darwin (1859). The technique became popular through the ground-breaking works of Holland (1975) and since has been quite successfully applied to solve NP-hard optimization problems (Feng-Tse *et al.*, 1993), especially in machine learning.

Algorithm 1 illustrates the fundamental structures of a GA. In GA, we have a pool of potentially feasible solutions called "*population*". These solutions undergo recombination and mutation (like in natural genetics) to produce new solutions. The process is repeated over various "*generations*". Each candidate solution is assigned a fitness value and the fitter individuals are given a higher chance to breed and yield more "*fitter*" individuals[4]. Thereby, we keep "*evolving*" better individuals over generations until we reach a stopping criterion.

GAs have various advantages. They do not require any auxiliary information about the objective function value, such as derivatives, which may not be available. Therefore, they may be used to optimize both continuous and discrete functions and multi-objective problems, providing a list of "*good*" solutions and not just a single solution. Moreover, GAs have outstanding parallel capabilities (Wang *et al.*, 2005), which is an outstanding property when the search space is vast

---

[4] This is consistent with the Darwinian Theory of "survival of the fittest".

Algorithm 1 Fundamental structures of a GA

```
$g \leftarrow 0$
Generate initial population $P(g)$ by encoding routine
Evaluate initial population $P(g)$ by decoding routine
while Termination criteria not fulfilled do
    Create $C(g)$ from $P(g)$ by crossover routine
    Create $C(g)$ from $P(g)$ by mutation routine
    Evaluate $C(g)$ by decoding routine
    Select $P(g+1)$ from $P(g)$ and $C(g)$ by selection routine
    Find best
    $g \leftarrow g+1$
end while
return best
```

or when the number of parameters involved is considerable. GAs are also sufficiently randomized, but they perform much better than random local search. They try various random solutions, keeping track of the best so far, as they exploit historical information.

Like any technique, GAs also suffer from a few limitations. They are not suited to all problems, especially those for which derivative information is available. As the algorithm is stochastic, there is no guarantee of the optimality or the quality of the solution. GAs are not specialized algorithms as they are application dependent. Their success depends on the knowledge of the problem and the design of the evaluation function. The solution space must consider only the feasible solutions. If it is not adequately implemented, the GAs may not converge to the optimal solution.

2.1 Terminology

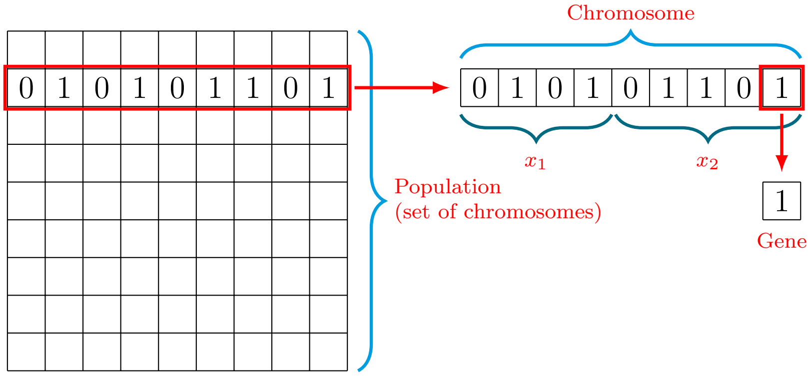


Figure 1: Genotype terminology

Typically, the population of potential solutions is made up of between 30 and 100 individuals. Each individual is **encoded** as a string, called a **chromosome**, in an apparent reference to biology. The chromosome is composed over some alphabet(s) so that the genotypes (chromosome values) are uniquely mapped onto the decision variable (phenotypic) domain. The most commonly used representation in GA is the binary alphabet {0,1}. Each decision variable is encoded as a binary string, and these are concatenated to form a chromosome. Figure 1 displays a problem with two decision variables x_1 and x_2, where x_1 is mapped with 4 bits (or **genes**) and x_2 with 5 bits reflecting their level of accuracy or their range. Examining the chromosome string in isolation yields no information about the problem we are trying to solve. Only with the chromosome's **decoding** into its phenotypic values can any meaning be applied to the representation. However, the search process will always operate on this encoding of the decision variables rather than the decision variables themselves.

2.2 Parent selection and ranking

2.2.1 Fitness function

The fitness function is the heart of the genetic algorithm. The fitness is a function that takes a candidate solution to the problem as input and produces as output how "*fit*" or "*well*" the solution fulfils whatever criteria the algorithm is optimizing. In the case of a minimization problem, the fittest individuals will have the lowest numerical value of the associated fitness function. The fitness is applied to each individual in the population and calculated repeatedly, and therefore it should be sufficiently fast. A slow computation of the fitness value can adversely affect the algorithm and make it exceptionally slow.

2.2.2 Parent selection

According to Darwin's theory of evolution, the best individuals survive to participate in reproduction. Potential parents should then be selected according to their fitness. There are many selection schemes for GAs, each with different characteristics. An ideal selection technique would be simple to code, efficient for both nonparallel and parallel architectures. Tournament selection satisfies all the above criteria, being in addition to being independent of the fitness function and can even work with negative fitness values.

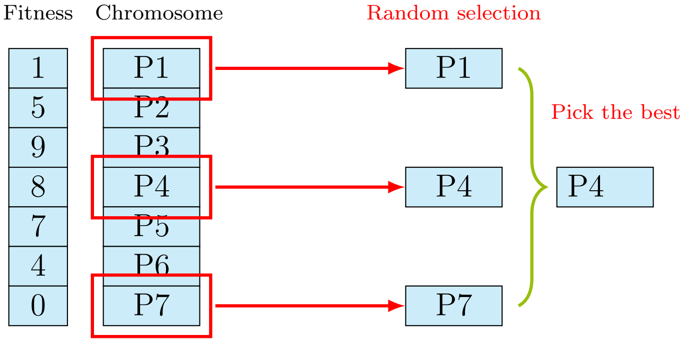


Figure 2: A 3-tournment selection

In this approach, a k-tournament is run between k individuals chosen at random from the population, and the one with the best fitness is selected as the winner. Figure 2 illustrates a 3-tournament selection where three entities are picked out of the pool, their fitnesses are compared, and the best is permitted to reproduce. In the example, we suppose that fitness is maximized.

2.3 Genetic operators

2.3.1 Crossover

Crossover or recombination is the main genetic operator. It operates on two chromosomes at a time and generates offspring by combining both chromosomes' features. A simple way to perform crossover is to choose a random cut-point and generate the offspring by combining one parent segment's to the left of the cut-point with the other parent's segment to the right of the cut-point. This process, called one-point crossover[5] and an often-used method for GAs operating on binary strings, is visualized in Figure 3.

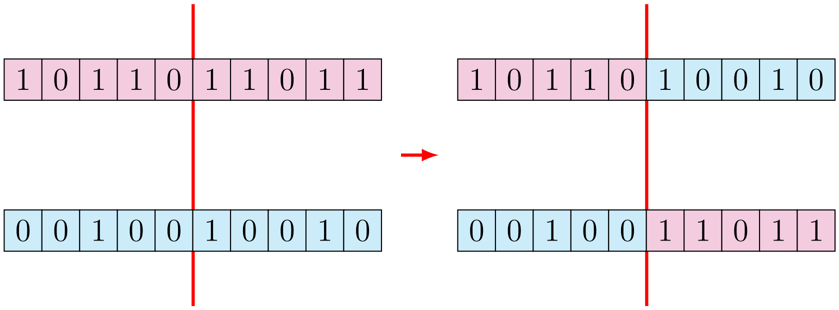


Figure 3: One-point crossover

The crossover is performed with a probability denoted by $\rho^{Crossover}$ when the pairs are chosen for breeding. This probability controls the expected number of chromosomes to undergo the crossover operation. A higher crossover probability allows exploitation of more of the solution space and reduces the chance of settling for a false optimum. However, a too high rate results in omputation time being wasted in exploring unpromising regions of the solution space.

2.3.2 Mutation

The genetic material can be changed randomly by erroneous reproduction or other genes' deformations in real evolution. In GA, a mutation can be realized by altering one or more genes of a selected chromosome. Figure 4 illustrates a technique called "*inversion of single bits*", where one randomly chosen bit is flipped.

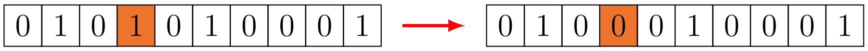


Figure 4: The inversion of single bits mutation technique

The mutation is the part of the GA related to the exploration of the search space. It serves to either replace the genes lost from the population during the selection process to be tried in a new context or provide the genes that were not present in the initial population. The mutation is considered as the operator that ensures that the probability of searching a particular subspace of the search space is never zero and thus tends to alleviate the possibility of converging to a local

---

[5] The one-point crossover can be generalized to k-point crossover for any integer k, picking k crossover points and swapping the bits in between the crossover points.

optimum rather than the global optimum. The mutation probability denoted by $\rho^{\mathcal{M}utation}$ governs the rate with which new genes are introduced into the population. When it is too low, many genes that would have been used are never tried out, while if it is too high, there will be much more random perturbation, the offspring will start losing their resemblance to the parents, and the GA will lose the ability to learn from the search history.

2.3.3 Survivor selection

This component determines which individuals are to be kicked out and which are to be kept. Traditional genetic algorithms put the new individuals directly into the next generation. Goldberg (1989) has introduced the elitist strategy (Figure 5). The latter propagates the current fittest member of the population to the next generation. Except for the fittest member, the older better individuals, which have been crossed and mutated, have no chance to access the future generation. Dong and Wu (2009) introduce the expansive optimal sampling where the new individuals are put together with the previous generation and where the best half of the individuals survive to the next generation.

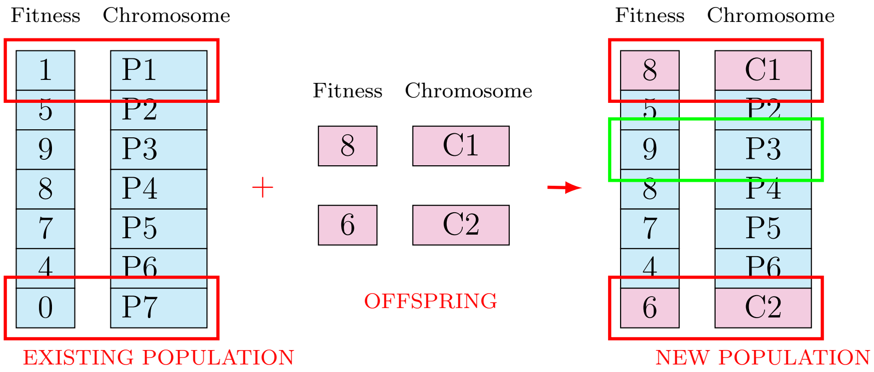


Figure 5: Survivor selection

2.4 Termination

Because GA is a stochastic search method (Spall, 2003), it is not easy to specify convergence criteria formally. As a population's fitness may remain static for several generations before a superior individual is found, the application of conventional termination criteria becomes problematic. Usually, the GA run comes to an end when given criteria are satisfied, e.g. no improvement in the best fitness value for a high number of iterations or a certain number of generations or a specific pre-defined value for the fitness.

3 Application of genetic algorithms to our basket creation problem

After studying the bonds held in the principal portfolios managed by the passive fund management team, we find a variety of used minimum tradable (MT), but two of them (100 000 and 1 000) stand out for EUR credit funds[6]. We also note that the quantity of the bonds to be bought or sold is mostly a multiple of the minimum tradable since the lot size is seldom used. The multiple of MT is generally lower than 10. When the MT equals 1 000, the multiple is frequently seen as $100 \cdot k$ where $k \in \{1, 2, \ldots, 10\}$ for EUR credit funds[7]. Based on these observations, we will seek quantities in the form of $100\,000 \cdot k$, where k is a natural number lower than exceed 10.

3.1 Chromosome encoding

Regarding a given sector, our decision variables in the problem are the quantities of each bond held from the reference portfolio's sector. As we use the binary representation, we must find a trick to reduce the number of bits representing the quantity of a bond in a chromosome. Indeed, the overall number of potential portfolios is $2^{N_{bonds} \cdot N_{bits}}$, where N_{bonds} and N_{bits} are respectively the number of bonds and the number of bits. Decreasing N_{bits} by 1 unit divides the number of potential solutions by $2^{N_{bonds}}$ and dramatically reduces the computation time.

Natural	Binary	Gray
10	1010	1111
11	1011	1110
12	1100	1010
13	1101	1011
14	1110	1001
15	1111	1000

Table 1: Binary and Gray codes

Academics use the standard binary or the Gray code[8] to represent a natural. Using 4 bits, 10 can be written as 1010 in binary or 1111 in Gray. However, with these codes, we go beyond our threshold of 10. Indeed, 15 can be written as 1111 in binary or 1000 in Gray (See Table 1). Table 2 shows that we can write any natural between 0 and 10 as a linear combination of 1, 2, 3, and 4 where the coefficients are 0 or 1. The combination is however not unique. For instance, if we take the example of 7, two encodings are possible: 1101 as we can write $7 = 1 \cdot 1 + 1 \cdot 2 + 1 \cdot 4$ and 1100 as we can write $7 = 1 \cdot 3 + 1 \cdot 4$. Nevertheless, the combination becomes unique if we force it to prioritize the first bits as shown in column Encoding 1.

As discussed above, the range of possible solutions remains huge, even with $N_{bits} = 4$. Performing a "*warm start*" initialization can speed up a GA's convergence to an optimal solution.

---

[6] In the case of USD credit funds, 2 000 seems to be the most used minimum tradable amount.

[7] This figure becomes $50 \cdot k$ for USD credit funds.

[8] A Gray Code, named after Frank Gray (Gray, 1953), represents numbers using a binary encoding scheme that groups a sequence of bits so that only one bit in the group changes from the number before and after.

Natural	Encoding 1	Encoding 2
0	0000	
1	0001	
2	0010	
3	0011	0100
4	0101	1000
5	0110	1001
6	0111	1010
7	1011	1100
8	1101	
9	1110	
10	1111	

Table 2: Yet another encoding method

We include in the initial population one chromosome encoded with the quantities of the current portfolio. Here we face an issue to encode quantities that cannot satisfy the linear combination. For instance, how do we cope with quantities that are not multiples of MT? A workaround consists of the adding of an extra gene to reflect the odd quantity. Figure 6 illustrates an example of a quantity of 3.5 MT. We decompose the quantity to use the 4 available bits and we add one gene to reflect the residual, keeping in mind that all bits should correspond to tradable quantities. The second bit is set to 1 and the residual 1.5 MT is reflected in the fifth bit since the residual cannot be less than 1 MT. As can be noted, the extra gene's value is not uniform and depends on the residual of each odd quantity.

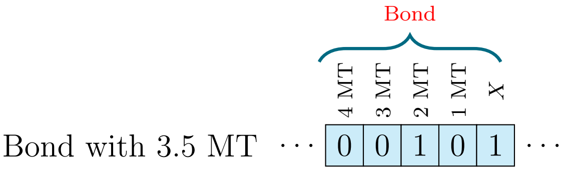


Figure 6: From quantity to binary

3.2 Chromosome decoding

The decoding process is the reverse of encoding. It takes the chromosome representation and transforms it back to the decision variable domain. Figure 7 shows the two different decoding cases whether an extra gene is added or not during the encoding process. Here, we suppose that only Bond a has an extra gene with X_a as the value. For both bonds, we calculate the associated quantity as $Q = \sum_{j=1}^{4} j\alpha_j \mathrm{MT} + \alpha_5 X$, where $(\alpha_1, \ldots, \alpha_4)$ are the values of their genes and X is the value of the eventual extra gene.

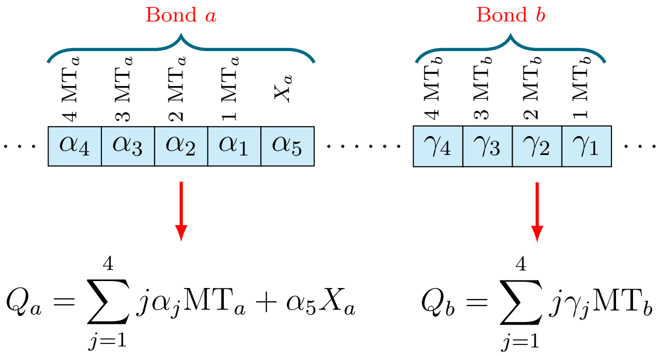


Figure 7: From binary to quantity

3.3 Definition of fitness

Our GA aims to find solutions that match the following risk metrics per sector[9]: the modified duration per bucket of maturity, the duration-times-spread[10], and the weight. Solutions are constrained in terms of liquidity cost and by the presence of desired bonds on the axis. The fitness should consider all these features in one single function.

If we note x as one possible solution, b as the reference portfolio, and s as one sector among the several sectors of b, the fitness is defined as the sum of three risk measures with two penalty terms:

$$\begin{aligned}\mathcal{F}itness\left(x,s\right) &= \mathcal{R}_{\mathrm{W}}\left(x \mid b,s\right) + \mathcal{R}_{\mathrm{MD}}\left(x \mid b,s\right) + \mathcal{R}_{\mathrm{DTS}}\left(x \mid b,s\right) + \\ & \lambda_L \cdot \mathcal{L}iquidity\left(x,s\right) + \lambda_A \cdot \mathcal{A}xis\left(x,s\right)\end{aligned}$$

Each of these components is defined below in such a way that the best solutions will minimize the fitness. Regarding the weight risk measure, we define $\mathcal{R}_{\mathrm{W}}\left(x \mid b,s\right)$ as the weight difference between Portfolio x and Portfolio b within the sector s:

$$\mathcal{R}_{\mathrm{W}}\left(x \mid b,s\right) = \left| \sum_{i \in \mathcal{S}ector(s)} \left(x_i - b_i\right) \right|$$

where x_i and b_i are the weights of Bond i in Portfolios x and b. We define $\mathcal{R}_{\mathrm{MD}}\left(x \mid b,s\right)$ as the modified duration risk of x with respect to b within the sector s:

$$\mathcal{R}_{\mathrm{MD}}\left(x \mid b,s\right) = \sum_{j=1}^{N_{\mathcal{B}uckets}} \left| \sum_{i \in \mathcal{S}ector(s)} \left(x_i - b_i\right) \cdot \mathrm{MD}_i\left(\mathcal{B}ucket_j\right) \right|$$

where $N_{\mathcal{B}uckets}$ is the number of maturity buckets and $\mathrm{MD}_i\left(\mathcal{B}ucket_j\right)$ is the modified duration contribution of Bond i to the maturity bucket j. The rationale of this definition is to track the difference in modified duration per bucket. We may remark that this measure of risk is higher

---

[9] Or per country if the reference portfolio is a government bond index. Any subsequent mention of "*sector*" refers to a sector or a country.

[10] When the reference portfolio bears credit risk.

than or equal[11] to the traditional measure of modified duration risk[12]. Finally, we define the DTS risk measure $\mathcal{R}_{\text{DTS}}(x \mid b, s)$ as the weighted DTS difference between x and b:

$$\mathcal{R}_{\text{DTS}}(x \mid b, s) = \left| \sum_{i \in \mathcal{S}ector(s)} (x_i - b_i) \cdot \text{DTS}_i \right|$$

where DTS_i is the duration-times-spread of Bond i. Unlike the three previous risk measures, $\mathcal{L}iquidity(x, s)$ is not used to match the liquidity cost of b but it indicates the preference in terms of liquid bonds:

$$\mathcal{L}iquidity(x, s) = \sum_{i \in \mathcal{S}ector(s)} x_i \cdot \text{LTP}_i$$

where LTP_i is the liquidity-times-price[13] of Bond i. According to Ben Slimane and De Jong (2017), the liquidity-times-price is a proxy of the bid-ask spread. The lower this value, the more liquid the bond. The axis component $\mathcal{A}xis(x, s)$ is the second penalty function with the liquidity component $\mathcal{L}iquidity(x, s)$. It is defined as the number of changes in the quantities of bonds that do not belong to axis $\mathcal{A}$:

$$\mathcal{A}xis(x, s) = \#\left\{ i \in \mathcal{S}ector(s) : (1 - \delta_i) \cdot \left| \underbrace{Q_i(x) - Q_i(x^0)}_{\Delta Q_i(x)} \right| > 0 \right\}$$

where $\delta_i = 1$ if Bond i is on the axis $\mathcal{A}$ and $\delta_i = 0$ otherwise. $Q_i(x)$ and $Q_i(x^0)$ are respectively the quantity of Bond i hold in Portfolio x and the current portfolio x^0. With this definition, Bond i, which is held in x and whose change in quantity with regard to the current portfolio (denoted $\Delta Q_i(x)$ afterwards) differs from zero, will penalize the fitness if it does not belong to axis $\mathcal{A}$. Both penalty function terms are multiplied by respectively the penalty factors λ_L and λ_A.

3.4 Handling of constraints

The penalty methods are the most common approaches for constraint handling. Penalty terms are added to the objective (or the fitness) function. The given problem is then converted from a constrained to an unconstrained problem. Our fitness uses this approach to handle constraints on liquidity or axis. The challenge with this approach is calibrating the penalty coefficients. According to Chehouri *et al.* (2016), "*the penalty term cannot be too high or else the algorithm will be locked inside the feasible domain and cannot move towards the border with the infeasible area. If it is too low, the term will be irrelevant with regard to the objective function and the search will remain in the infeasible region*". The fitness defined above does handle the constraints on liquidity and axis but does not consider the adjustment direction. Indeed, no constraints are set in the fitness to have only positive (negative) changes in quantities in case of an in-kind purchase (sale). Hereafter, we present two methods to handle these constraints without using any additional penalty functions.

---

[11] See Appendix A.3 on page 41.

[12] As a reminder, traditional modified duration risk is defined as $\left|\sum_{i \in \mathcal{S}ector(s)} (x_i - b_i) \cdot \text{MD}_i\right|$.

[13] LTP is calculated as the product of a liquidity score and the clean price

3.4.1 VCH method

VCH is the acronym for Violation-Constraint-Handling introduced by Chehouri *et al.* (2016). This method is free of any penalty parameters and can sort a population with feasible and infeasible individuals. For each sector s, the cardinal of violations is equal to:

$$\begin{aligned} CV(x,s): \ & \mathcal{Population} \longrightarrow \mathbb{R}_+ \\ & x \longmapsto \begin{cases} \#\left\{i \in \mathcal{Sector}(s) : \Delta Q_i(x) < 0\right\} & \text{if} \quad \text{purchases} \\ \#\left\{i \in \mathcal{Sector}(s) : \Delta Q_i(x) > 0\right\} & \text{if} \quad \text{sales} \end{cases} \end{aligned}$$

whereas the amount of violations is defined by:

$$\begin{aligned} AV(x,s): \ & \mathcal{Population} \longrightarrow \mathbb{R}_+ \\ & x \longmapsto \begin{cases} \sum_{i \in \mathcal{Sector}(s)} \max\left(0, -\Delta Q_i(x)\right) & \text{if} \quad \text{purchases} \\ \sum_{i \in \mathcal{Sector}(s)} \max\left(0, \Delta Q_i(x)\right) & \text{if} \quad \text{sales} \end{cases} \end{aligned}$$

The function $CV(x,s)$ takes one candidate solution among the population and returns the number of violations. Violations stand for the number of purchases when the process has to sell or for the number of sales when the process has to purchase. The function $AV(x,s)$ indicates the amount of violations measured as the sum of changes in quantities for each case. We then divide the population into two sets of feasible and infeasible solutions:

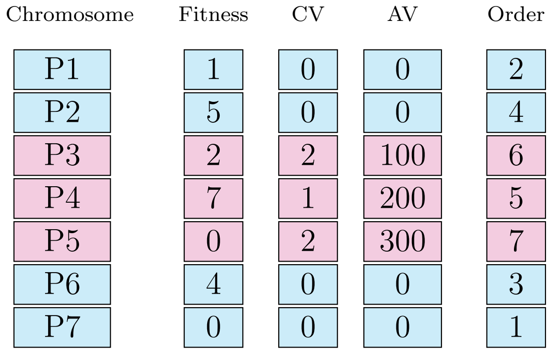
$$\begin{cases} \mathcal{P}_0 = \left\{x \in \mathcal{Population} : CV(x,s) = 0\right\} \\ \mathcal{P}_1 = \left\{x \in \mathcal{Population} : CV(x,s) > 0\right\} \end{cases}$$

We amend the GA by integrating functions $CV(x,s)$ and $AV(x,s)$ as shown in Algorithm 2. If we note $\mathcal{X} = \{x_1, x_2\}$ a set of two candidate solutions, the fitness function is called only if one of the two individuals is a feasible solution, i.e. when $\mathcal{X} \cap \mathcal{P}_0 \neq \emptyset$. In this case, the best individual is the one with the lowest fitness value. If one individual is feasible and the other is infeasible, the best individual will be the feasible solution. If both individuals are infeasible, the best individual will have the lowest $CV(x,s)$ if there are two different values of $CV(x,s)$ or the lowest value $AV(x,s)$ if both individuals have the same value $CV(x,s)$.

Algorithm 2 New ranking rules

```
Require: Portfolios $x_1$ and $x_2$ and Sector $s$
  Let $\mathcal{X} = \{x_1, x_2\}$
  if $\mathcal{X} \cap \mathcal{P}_0 \neq \emptyset$ then
    return $\text{argmin}_{x \in \mathcal{X} \cap \mathcal{P}_0} \mathcal{Fitness}(x,s)$
  else
    if $\#CV(\mathcal{X},s) = 2$ then
      return $\text{argmin}_{x \in \mathcal{X}} CV(x,s)$
    else
      return $\text{argmin}_{x \in \mathcal{X}} AV(x,s)$
    end if
  end if
```

Figure 8 illustrates the new ranking rules on a population of seven individuals where the fitness is minimized. The amended algorithm ensures that feasible solutions (displayed in blue) are always ahead of infeasible solutions[14]. Hence, if the algorithm is initiated with a feasible solution, it will always return a feasible solution. One possible initial guess could be the current portfolio x^0 since by construction $\Delta Q_i\left(x^0\right) = 0$ for each Bond i held.



Chromosome	Fitness	CV	AV	Order
P1	1	0	0	2
P2	5	0	0	4
P3	2	2	100	6
P4	7	1	200	5
P5	0	2	300	7
P6	4	0	0	3
P7	0	0	0	1

Figure 8: Illustration of the new ranking rules

3.4.2 Method of additional quantities

The three risk measures and the liquidity penalty functions listed above can be rearranged to emphasize the change in quantities compared to the current portfolio. The axis penalty function is already defined using the change in quantities. Expressions for both in-kind and standard processes are available in Appendix A.2 on page 39. The GA is then warm-started with a null chromosome and seeks a basket of purchasable or saleable bonds. This method ensures that all changed quantities are positive since genes' unique values are 0 or 1.

3.4.3 Constraints on axis

Let $\mathcal{A} = \{A_1, \ldots, A_{n_b}\}$ be the axis where n_b is the number of bonds in the reference portfolio b and A_i is the maximum change in quantity allowed for Bond i. We set $A_i = +\infty$ if Bond i is not on axis $\mathcal{A}$ and we note $\mathcal{A}(s) = \{i \in \mathcal{S}ector\,(s) : A_i < +\infty\}$ the subset of $\mathcal{A}$ composed by constrained bonds of sector s. We draw on again the VCH method to take care of these constraints and use the same approach as above by introducing for each sector s two additional functions. The new cardinal of violation is then defined as:

$$
\begin{aligned}
CW\left(x,s\right) : \; & \mathcal{P}opulation \longrightarrow \mathbb{R}_+ \\
& x \longmapsto \begin{cases} \#\left\{i \in \mathcal{S}ector\left(s\right) : \left|\Delta Q_i\left(x\right)\right| > A_i\right\} & \text{if } \#\left\{\mathcal{A}\left(s\right)\right\} > 0 \\ \max\left(0, \#\left\{i \in \mathcal{S}ector\left(s\right) : \left|\Delta Q_i\left(x\right)\right| > 0\right\} - N_{\max}\right) & \text{otherwise} \end{cases}
\end{aligned}
$$

whereas the new amount of violations is given by:

$$
\begin{aligned}
AW\left(x,s\right) : \; & \mathcal{P}opulation \longrightarrow \mathbb{R}_+ \\
& x \longmapsto CW\left(x,s\right) \cdot \left|\omega\left(x^0,s\right) \cdot \mathcal{A}djustment - \textstyle\sum_{i \in \mathcal{S}ector(s)} \Delta Q_i\left(x\right) \cdot P_i\right|
\end{aligned}
$$

---

[14] This is also the case when fitness is maximized.

where P_i is the dirty price of Bond i, $\omega\left(x^0, s\right)$ is the weight of Sector s in the current portfolio x^0 and $\mathcal{A}djustment$ is the signed amount of in-kind subscription or redemption.

The function $CW\left(x, s\right)$ returns the number of bonds whose quantities exceed the allowed quantity or the number of altered positions if no limit is set for the sector's bonds. A tolerance value $N_{\max}$ may be added, below which the function returns no violation. The function $AW\left(x, s\right)$ calculates the difference in mark-to-market between the portfolio of new trades and the adjustment times the sector's weight. The rationale of this definition is to obtain a basket of bonds whose weight does not differ much from the sector's weight in the current portfolio x^0. The multiplication by $CW\left(x, s\right)$ ensures only a value of zero for feasible solutions. Then, we separate the feasible set $\mathcal{P}_0$ defined above into $\mathcal{P}_{0,0}$ and $\mathcal{P}_{0,1}$:

$$\begin{cases} \mathcal{P}_{0,0} = \left\{x \in \mathcal{P}opulation : CV\left(x, s\right) = 0, CW\left(x, s\right) = 0\right\} \\ \mathcal{P}_{0,1} = \left\{x \in \mathcal{P}opulation : CV\left(x, s\right) = 0, CW\left(x, s\right) > 0\right\} \end{cases}$$

The amended ranking rules are shown in Algorithm 3. The same principle is reiterated. The fitness is called only in the presence of at least one feasible individual. Two individuals from the same infeasible set are ranked by the cardinal and amount functions. For instance, if they are part of $\mathcal{P}_{0,1}$, we compare their values $CW\left(x, s\right)$ or $AW\left(x, s\right)$. Otherwise, we compare their values $CV\left(x, s\right)$ or $AV\left(x, s\right)$. We note that this algorithm establishes a hierarchy when comparing two individuals from different sets. The best is the individual that comes from $\mathcal{P}_{0,0}$, then $\mathcal{P}_{0,1}$, and finally $\mathcal{P}_1$.

Algorithm 3 Amended ranking rules

```
Require: Portfolios $x_1$ and $x_2$ and Sector $s$
  Let $\mathcal{X} = \{x_1, x_2\}$
  if $\mathcal{X} \cap \mathcal{P}_{0,0} \neq \emptyset$ then
    return $\operatorname{argmin}_{x \in \mathcal{X} \cap \mathcal{P}_{0,0}} \mathcal{F}itness\left(x, s\right)$
  else
    if $\mathcal{X} \cap \mathcal{P}_{0,1} \neq \emptyset$ then
      if $\#\{CW\left(\mathcal{X}, s\right)\} = 2$ then
        return $\operatorname{argmin}_{x \in U} CW\left(x, s\right)$
      else
        return $\operatorname{argmin}_{x \in U} AW\left(x, s\right)$
      end if
    else
      if $\#\{CV\left(\mathcal{X}, s\right)\} = 2$ then
        return $\operatorname{argmin}_{x \in \mathcal{X}} CV\left(x, s\right)$
      else
        return $\operatorname{argmin}_{x \in \mathcal{X}} AV\left(x, s\right)$
      end if
    end if
  end if
```

3.5 Cash management

The GAs are performed per sector. They can run sequentially or in parallel, thus the overall amount of purchases/sales, constrained by the amount of the adjustment, can be tested only when all the GAs have terminated. Two scenarios regarding the overall amount of purchases/sales are possible: (1) it can be above the adjustment or (2) it can be below the adjustment by an amount exceeding a threshold[15]. In the first scenario, some sectors are selected and others are discarded. The choice can be achieved if we formulate the problem as a Knapsack problem[16] (KP). KP's items will be the sectors and the KP's limited capacity is the adjustment. Items' weights and values are the amount of purchases/sales of bonds in sectors and hence the KP will consider first the sectors with the highest values. In the second scenario, the basket found is added to the existing quantities. The GAs are restarted using the updated quantities and the adjustment decreased by the mark-to-market of the basket.

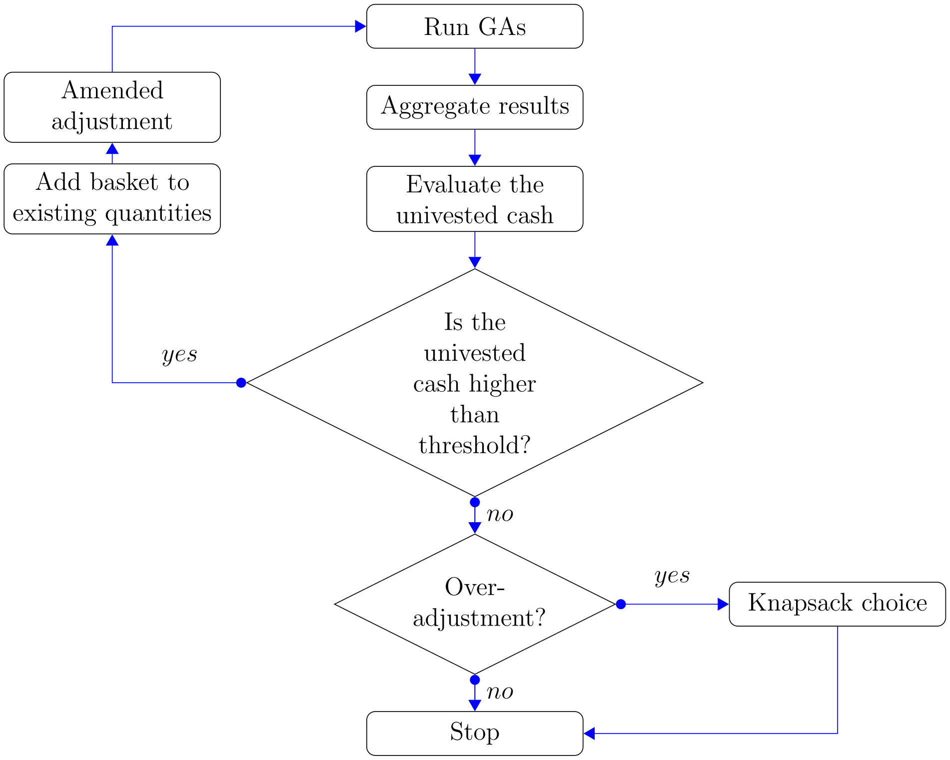


Figure 9: Cash management

3.6 Calibration of parameters

Any GA has four intrinsic parameters to be calibrated: (1) the population size, (2) the termination criteria, (3) the crossover rate and (4) the mutation rate. The size of the population indicates the total number of the population's individuals. On the one hand, if the size of the population is

---

[15] For instance, uninvested cash is higher than EUR 100 000.

[16] See Appendix A.4 on page 41.

small, the available search space is small too, and therefore it is possible to reach a local optimum. On the other hand, if the population size is very large, the search area is increased, and the computational load becomes high. Therefore, the size of the population must be reasonable. We set up this parameter to 50.

As termination criteria, we set up a threshold for the number of generations. The GA will terminate when the number of iterations reaches this threshold. This number depends on the problem type and complexity. In some cases, hundreds of loops are sufficient, but we might need more in other cases. Setting a higher number may increase the computation time. We set up this parameter to 500. The total number of generations may not be used if no improvement in the best fitness is observed for I iterations. We set I to 100.

Choosing the crossover and mutation rates is not an easy task. Usually, these rates are fixed, i.e. they do not change during the execution of the GA. These parameters' values are determined with a sensitivity analysis after carrying out multiple GA runs with different values and having compared the outcome. Since the GA is intrinsically a dynamic and adaptive process, constant parameters contrast with the evolutionary spirit. Therefore, the natural idea is to try to modify the GA's values while running it. This is possible to perform by using a particular rule or by employing a particular self-adaptive process. Dong and Wu (2009) and Hassanat *et al.* (2019) fall into these categories.

3.6.1 An example of self-adaptive mechanisms

Dong and Wu (2009) use a self-adaptive mechanism for both crossover and mutation rates. They define the crossover rate for each couple of chromosomes as the difference in fitness scaled by the difference between the greatest and lowest fitness values of the chromosomes in the population. The authors justify this approach by the fact that "*when the crossover rate is the same, all the individuals in the cross-operation are retained at the same probability, thus the current better individuals are selected several times during the choice operation in the next round and the poorer individuals are eliminated leading the population to quickly evolve towards the current optimal individual. If this current optimal individual is a local optimum, then the entire algorithm can fall into local optimum*". With this definition, very close individuals have small chances to generate offspring, contrary to very distant individuals. The mutation rate is defined as a quadratic function of the fitness. When the fitness is maximized, the mutation rate is highest for the fitness's highest values. Better individuals are then allowed to have a bigger rate to avoid better individuals occupying the entire population. We adopt this approach, and we define for each sector s the crossover rate function:

$$\rho^{\mathcal{C}rossover}(x_1, x_2, s) = \frac{|\mathcal{F}itness\,(x_1, s) - \mathcal{F}itness\,(x_2, s)|}{\max_x \mathcal{F}itness\,(x, s) - \min_x \mathcal{F}itness\,(x, s)}$$

and the mutation rate function:

$$\rho^{\mathcal{M}utation}(x_1, s) = \varphi \cdot \left(1 - \frac{\mathcal{F}itness\,(x_1, s) - \min_x \mathcal{F}itness\,(x, s)}{\max_x \mathcal{F}itness\,(x, s)}\right)^2$$

where x_1 and x_2 are two candidate portfolios, and φ is the mutation rate reached when the fitness is at its lowest value. To use the rates introduced above in our problem, we extend the defintion of fitness, thanks to the logistic function[17], to take into account the infeasible solutions:

$$\begin{aligned}
\mathcal{F}itness^{\star}(x,s) = {} & \mathcal{F}itness(x,s) + \\
& \left(1+\max_y \mathcal{F}itness(y,s)\right)\cdot\left(CW(x,s)+\mathcal{L}ogistic(AW(x,s))-\frac{1}{2}\right) + \\
& \left(1+\max_y \mathcal{F}itness(y,s)\right)\cdot\left(CV(x,s)+\mathcal{L}ogistic(AV(x,s))-\frac{1}{2}\right)\cdot \\
& \left(1+\max_y CW(y,s)\right)
\end{aligned}$$

This extended fitness function preserves the priority of feasible solutions over infeasible solutions. It also coincides with the original fitness function as far as feasible solutions are concerned since we have $CV(x,s) = AV(x,s) = CW(x,s) = AW(x,s) = 0$ when a candidate solution $x \in \mathcal{P}_{0,0}$. The coefficient $(1+\max_y \mathcal{F}itness(y,s))$ ensures that any infeasible solution of $\mathcal{P}_{0,1}$ lags any feasible solution. For instance, in the extreme case where its fitness value is equal to 0 and $CV(x,s) = 1$, then the associated extended fitness will be higher than $(1+\max_y \mathcal{F}itness(y,s))$. The same idea applies to the infeasible solutions of $\mathcal{P}_1$. The logistic function ensures the hierarchy between two infeasible solutions having the same value for $CV(x,s)$ or $CW(x,s)$.

3.6.2 An example of deterministic mechanisms

Hassanat *et al.* (2019) use a deterministic adaptation of the rates. A time-varying rule modifies crossover and mutation rates with respect to the number of generations. The rule gradually decreases (respectively increases) the crossover (respectively the mutation) rates as generations elapse[18]. The crossover then decreases linearly from 1 to 0 when the maximum number of generations is reached and, at the same time, the mutation increases linearly from 0 to 1. At each generation, the sum of both rates is 1.

An alternative to Hassanat *et al.* (2019) mutation curve is to cap the mutation rate to κ, where $0 \leq \kappa < 1$. The rationale is to prevent too high mutation rates as they reduce the GA's searchability to a simple random walk. We may also apply a floor to the mutation rate (for instance 5%) to ensure some mutation for the first generations. A fixed mutation is an extreme case where the floor equals the cap. These adaptations' profiles can be seen as similar to option payoffs: (1) a short put when the mutation is capped, and (2) a long call spread when a floor is added to the capped mutation. The original version of Hassanat *et al.* (2019) has a long stock payoff profile. Figure 10 illustrates the different profiles.

By analogy, we may apply a floor to the crossover rate (for instance 60%) to ensure a high crossover for the latest generations. The profile is then similar to a long put payoff. Similar to what we have done for the mutation rates, we may add a cap to prevent very high crossover rates. The profile then becomes a payoff of a long put spread option. The crossover version of Hassanat *et al.* (2019) has a short stock payoff profile. The profiles discussed are shown in Figure 11.

---

[17] The logistic function $\mathcal{L}ogistic(z) = \dfrac{1}{1+e^{-z}}$ is strictly increasing in $\mathbb{R}_+$ and takes its values in $[0.5, 1[$.

[18] Hassanat *et al.* (2019) also document an increasing (respectively decreasing) rule for the crossover (respectively the mutation).

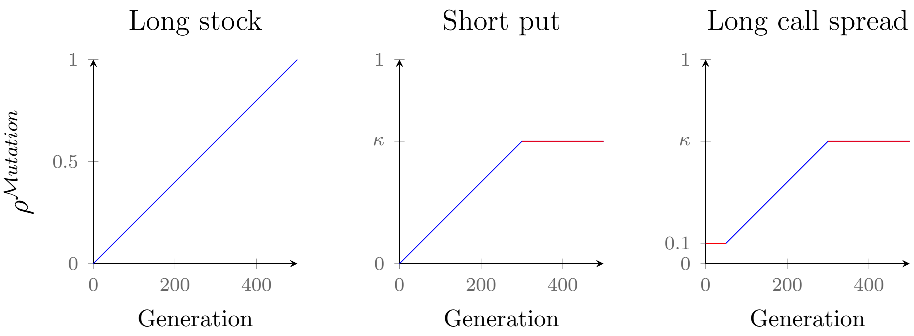


Figure 10: Deterministic mutation rates

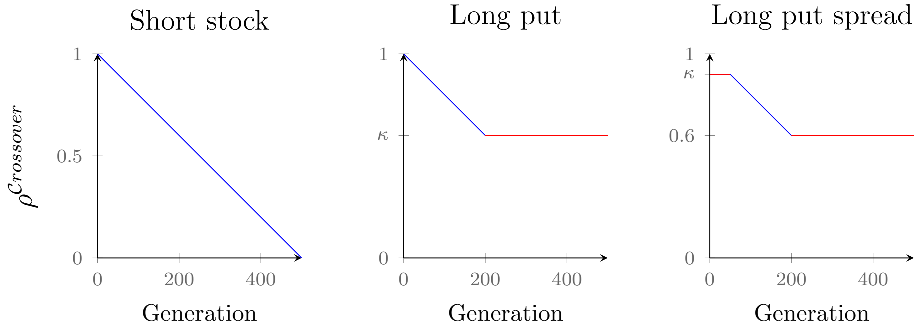


Figure 11: Deterministic crossover rates

4 Application

In this section, we explore the capabilities of GA in a case study. Given a portfolio and its benchmark, we generate in-kind basket portfolios using different genetic operators' definition to match two subscription and redemption amounts. We look closely at three metrics: (1) the number of bonds in the baskets, (2) the percentage of the bonds that belong to the counterparty's book and (3) the uninvested amount. We strive to find the best parameters that optimize these metrics first on a standalone basis and then combined.

4.1 Data

We consider a credit portfolio and its benchmark made up respectively of 1 408 and 2 879 of euro-denominated bonds belonging to 16 sectors. The main metrics and contributions per sector are reported in Table 3. The banking sector is the most represented sector, with 321 bonds in the portfolio and 632 in the benchmark. This sector will be the more time-consuming sector for the GA, especially when it comes to adjusting an in-kind subscription as the space search includes the benchmark's bonds. In Table 4, we also assume that the market maker has an axis list of 69 bonds, covering most sectors, where each traded bond cannot exceed a certain quantity. From the list, only 52 bonds are held in the portfolio and thus are constrained in the in-kind redemption process. We note that the banking sector is well represented among the constrained bonds. This adds another layer of difficulty for the sector.

	Portfolio				Benchmark			
Sector	Holdings	DTS (bps)	Duration (bps)	Weight (%)	Holdings	DTS (bps)	Duration (bps)	Weight (%)
Agencies	13	4	5	0.83	29	4	6	0.82
Auto	105	42	33	7.14	202	41	33	7.18
Banking	321	90	106	25.47	632	89	107	25.46
Capital Goods	69	21	23	4.21	139	22	24	4.26
Chemistry	131	46	49	7.97	238	47	49	7.90
Consum	128	44	50	8.84	278	45	50	8.88
Energy	74	31	30	5.27	138	30	29	5.35
Insurance	26	8	9	1.61	54	8	9	1.67
Insurance Sub	42	24	14	2.65	69	25	14	2.66
Lower Tier2	50	22	16	3.88	101	22	16	3.94
Others	26	9	9	1.72	62	10	9	1.90
Pfandbriefe	1	0	0	0.04	2	0	0	0.04
Spec finance	73	33	27	4.81	189	33	27	4.98
Telecom/Techno	150	66	64	11.23	305	65	65	11.28
Transportation	62	25	23	4.04	135	24	22	3.92
Utility	137	53	57	9.81	306	53	55	9.75
Total	1 408	517	516	99.51	2 879	519	516	100.00

Table 3: Sector breakdown

Two adjustment cases will be studied: an in-kind subscription of EUR 5 million and an in-kind redemption of the same amount. This amount represents roughly 2% of the portfolio's initial net asset value. To obtain comparable results between the subscriptions and the redemptions, we neutralize the liquidity component ($\lambda_L = 0$) and we set the axis penalty multiplier λ_A to 7. We will use the two self-adaptive and deterministic mechanisms for both crossover and mutation rates introduced above. In Table 5, we expose the definitions of genetic operators to be used. Regarding the self-adaptive mechanism, we adopt the dynamic processes introduced by Dong and Wu (2009). We note $\text{DW}_{\mathcal{C}rossover}$ and $\text{DW}_{\mathcal{M}utation}$ as the associated crossover and mutation methods afterwards. For deterministic mechanisms, we use the definition proposed by Hassanat *et al.* (2019) (H00 method) and our amended definitions (H01-H14 methods). Each definition

Sector	Constrained bonds	Held in portfolio
Auto	4	3
Banking	16	11
Capital Goods	1	1
Chemistry	5	4
Consum	9	6
Energy	1	1
Lower Tier 2	1	
Others	3	2
Spec finance	7	6
Telecom/Techno	11	8
Transportation	3	3
Utility	8	7
Total	69	52

Table 4: Axis constraints

of crossover or mutation refers to its payoff. When it is fixed, we study different values of the crossover rate in the set $\{0.60, 0.65, \ldots, 0.95, 0.99\}$. When the mutation is concerned, except in the cases of H00 and H04, we assume that $\kappa \in \{0.01, 0.02, \ldots, 0.07, 0.10, 0.15, \ldots, 0.40\}$. For the long call spread and long put spread methods, the floors are equal to 0.05 and 0.60. Two items are reported in the Selection column in Table 5. Traditional means that offspring are placed directly into the next generation. Expansive OS refers to the expansive optimal sampling introduced by Dong and Wu (2009) where new individuals have to struggle with the previous generation to access the next generation. We use both items when the deterministic mechanism is concerned.

Mechanism	Method	Crossover	Mutation	Selection
Self-Adaptive	D00	$\mathrm{DW}_{\mathcal{C}rossover}$	$\mathrm{DW}_{\mathcal{M}utation}$	Expansive OS
Deterministic	H00	Short stock	Long stock	Expansive OS | Traditional
	H01	Short stock	Short put	Expansive OS | Traditional
	H02	Short stock	Fixed	Expansive OS | Traditional
	H03	Short stock	Long call spread	Expansive OS | Traditional
	H04	Long put	Long stock	Expansive OS | Traditional
	H05	Long put	Short put	Expansive OS | Traditional
	H06	Long put	Long call spread	Expansive OS | Traditional
	H07	Long put	Fixed	Expansive OS | Traditional
	H08	Fixed	Long stock	Expansive OS | Traditional
	H09	Fixed	Short put	Expansive OS | Traditional
	H10	Fixed	Long call spread	Expansive OS | Traditional
	H11	Long put spread	Long stock	Expansive OS | Traditional
	H12	Long put spread	Short put	Expansive OS | Traditional
	H13	Long put spread	Long call spread	Expansive OS | Traditional
	H14	Long put spread	Fixed	Expansive OS | Traditional
	T00	Fixed	Fixed	Expansive OS | Traditional

Table 5: Studied methods

4.2 Statistics

We perform 200 random runs for each method and each adjustment for a total of 899 800 runs in total. Each run takes on average 5 seconds on a 1.7 GHz Intel Core i5-8350U with 16 GB memory. For each run, a basket of purchasable or saleable bonds is created to match the given adjustment. We then report five metrics:

1. the duration-times-spread risk $|\Delta \mathrm{DTS}|$ expressed in bps;
2. the modified duration risk value $|\Delta \mathrm{MD}|$ expressed in bps;
3. the number of bonds $N_{\mathcal{B}asket}$ in the basket;
4. the axis matching ratio $R_{\mathcal{A}xis}$ defined as the number of bonds belonging to axis $\mathcal{A}$ to the number of bonds in the basket and expressed in %;
5. the ratio $R_{\mathcal{U}ninvested}$ of the uninvested amount to the adjustment amount expressed in %.

Among the different solutions, we will prefer those that lower the number of transactions, maximize the number of bonds on the axis and minimize the uninvested amount.

In Tables 6 and 7, we display the statistics of the five metrics listed above for the subscription/purchases process. We report nine statistics for each of them: the mean, the standard deviation, the skewness, the excess kurtosis, the lowest and the highest values, the median, and the 1st and 99th percentiles. The third and fourth statistics measure the asymmetry about the mean and the "*tailedness*" of the empirical distribution. The uninvested amount ratio $R_{\mathcal{U}ninvested}$ of purchases is the only metric whose skewness is not close to zero[19]. Its skewness is equal to 1.00, indicating that the mass of the distribution is concentrated to the left of the mean and that most of the outliers are present on the right side of the distribution. The excess kurtosis of 0.77 suggests heavy tails compared to the normal distribution. Figures 20–23 and 24 on page 46 illustrate the probability density functions of the five metrics. We note that DTS and MD risks evolve in the desired range[20]. Having 100% of the bonds belonging to axis $\mathcal{A}$ is only possible in purchases but maybe at the cost of high uninvested amounts. On average, we find that N_{Basket} (respectively $R_{\mathcal{A}xis}$ and $R_{\mathcal{U}ninvested}$) of purchases are lower (respectively higher) than those of sales and that the standard deviations are tighter for sales.

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
$\lvert\Delta\mathrm{DTS}\rvert$ (in bps)	1.19	0.27	-0.22	0.03	0.01	0.52	1.20	1.77	2.39
$\lvert\Delta\mathrm{MD}\rvert$ (in bps)	1.42	0.23	0.20	0.18	0.49	0.91	1.42	2.01	2.54
$N_{\mathcal{A}xis}$	35.87	2.27	-0.07	-0.14	26.00	31.00	36.00	41.00	45.00
$R_{\mathcal{A}xis}$ (in %)	91.10	3.83	0.12	-0.19	75.00	82.35	91.18	100.00	100.00
$R_{\mathcal{U}ninvested}$ (in %)	0.97	0.95	1.00	0.77	0.00	0.00	0.54	4.12	6.40

Table 6: Statistics of the subscription/purchases process

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
$\lvert\Delta\mathrm{DTS}\rvert$ (in bps)	1.39	0.21	-0.14	0.32	0.15	0.86	1.39	1.87	2.32
$\lvert\Delta\mathrm{MD}\rvert$ (in bps)	1.33	0.18	-0.14	-0.05	0.36	0.90	1.34	1.72	2.12
$N_{\mathcal{A}xis}$	38.36	1.42	-0.14	-0.00	32.00	35.00	38.00	41.00	44.00
$R_{\mathcal{A}xis}$ (in %)	82.38	2.22	0.08	0.46	72.22	76.92	82.05	87.50	94.59
$R_{\mathcal{U}ninvested}$ (in %)	0.75	0.59	0.19	-1.33	0.00	0.00	0.68	1.79	2.01

Table 7: Statistics of the redemption/sales process

4.3 Standalone analysis

In the section, we study the outcome of different definitions of crossover, mutation, and selection genetic operators on the metrics N_{Basket}, $R_{\mathcal{A}xis}$ and $R_{\mathcal{U}ninvested}$. We only present the subscrip-

---

[19] A value close to zero means that the tails on both sides of the mean balance out overall.

[20] Only two observations have their $\lvert\Delta\mathrm{MD}\rvert$ higher than 2.5 bps.

tion/purchases process results, but the redemption/sales process statistics are given in Appendix B on page 43.

4.3.1 Crossover operator

In Tables 8, 9 and 10, we investigated the impact of the crossover method. We notice that we obtain on average the lowest N_{Basket} and the highest $R_{\mathcal{A}xis}$ with a fixed crossover rate. The DW$_{\mathcal{C}rossover}$ method trails behind for these two metrics, with however the lowest standard deviation and the highest skewness. Regarding the ratio of uninvested amount $R_{\mathcal{U}ninvested}$, the ranking is reversed, and the DW$_{\mathcal{C}rossover}$ method has the lowest mean. However, all profiles exhibit a high positive skewness. As far as sales are concerned (see Tables 17, 18 and 19 on page 43), we observe that (1) apart from the DW$_{\mathcal{C}rossover}$ profile, all the other types of crossover exhibit almost the same mean and median, (2) DW$_{\mathcal{C}rossover}$ exhibits the lowest standard deviation and (3) the skewness is very close to 0.

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	35.87	2.27	-0.07	-0.14	26.00	31.00	36.00	41.00	45.00
Fixed	35.77	2.28	-0.06	-0.14	26.00	30.00	36.00	41.00	44.00
Long put spread	35.80	2.26	-0.07	-0.13	27.00	31.00	36.00	41.00	44.00
Short stock	35.92	2.25	-0.07	-0.13	28.00	31.00	36.00	41.00	44.00
Long put	35.96	2.24	-0.08	-0.15	27.00	31.00	36.00	41.00	45.00
DW$_{\mathcal{C}rossover}$	36.89	2.25	-0.28	0.01	29.00	31.00	37.00	42.00	44.00

Table 8: Impact of the crossover method on N_{Basket} (subscription/purchases process)

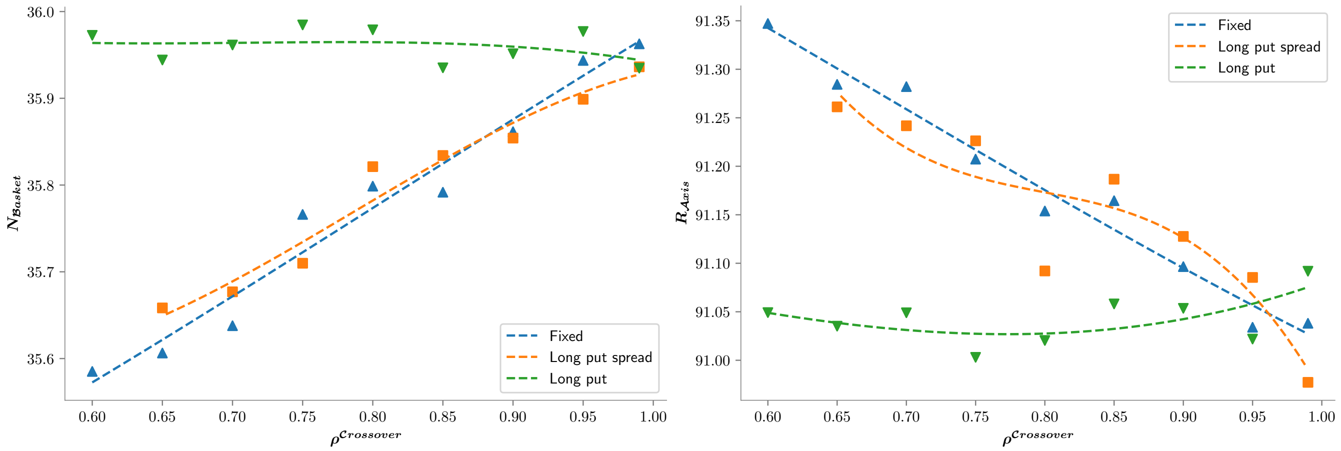
	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	91.10	3.83	0.12	-0.19	75.00	82.35	91.18	100.00	100.00
Fixed	91.18	3.88	0.10	-0.23	75.00	82.35	91.18	100.00	100.00
Long put spread	91.15	3.86	0.11	-0.20	75.68	82.35	91.18	100.00	100.00
Long put	91.04	3.78	0.13	-0.16	75.68	82.50	91.18	100.00	100.00
Short stock	91.04	3.78	0.11	-0.10	76.92	82.35	91.18	100.00	100.00
DW$_{\mathcal{C}rossover}$	90.19	3.48	0.36	0.07	78.95	82.50	89.47	97.37	100.00

Table 9: Impact of the crossover method on $R_{\mathcal{A}xis}$ (subscription/purchases process)

Hereafter, we focus on the behavior of profiles for which a fixed value is set for the crossover and we then discard the short stock and DW$_{\mathcal{C}rossover}$ profiles. In Figure 12, we report the averages of N_{Basket} and $R_{\mathcal{A}xis}$ per crossover rate. We note that the lowest values of fixed and long put spread profiles exhibit the best results. The lower the crossover rate, the lower N_{Basket} and the higher $R_{\mathcal{A}xis}$. This means that the GA has no need to systematically crossover the portfolios' population to obtain better results. The long put profile seems to be indifferent to the level of crossover rate.

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	0.97	0.95	1.00	0.77	0.00	0.00	0.54	4.12	6.40
$\text{DW}_{\mathcal{C}rossover}$	0.88	0.98	1.51	2.22	0.00	0.00	0.44	4.29	5.38
Short stock	0.95	0.93	1.00	0.78	0.00	0.00	0.53	4.08	5.65
Long put	0.96	0.94	1.03	0.84	0.00	0.00	0.52	4.12	6.40
Long put spread	0.98	0.95	0.97	0.70	0.00	0.00	0.55	4.11	6.14
Fixed	0.98	0.95	0.96	0.69	0.00	0.00	0.56	4.13	6.05

Table 10: Impact of the crossover method on $R_{\mathcal{U}ninvested}$ (subscription/purchases process)



(a) N_{Basket} (b) $R_{\mathcal{A}xis}$

Figure 12: Relationship between the crossover rate $\rho^{\mathcal{C}rossover}$, N_{Basket} and $R_{\mathcal{A}xis}$ (subscription/purchases process)

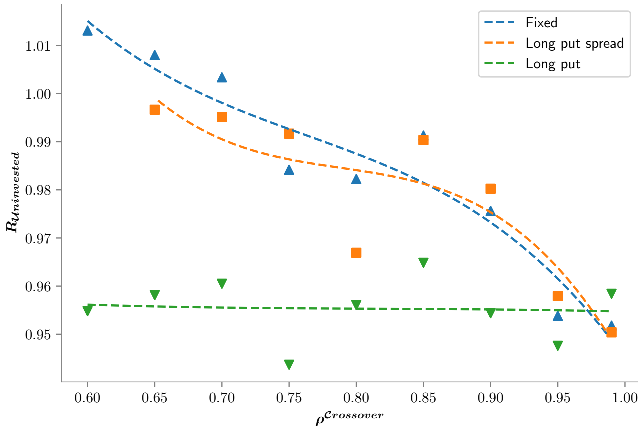


Figure 13: Relationship between the crossover rate $\rho^{\mathcal{C}rossover}$ and $R_{\mathcal{U}ninvested}$ (subscription/purchases process)

Figure 13 shows that the long put profile has on average the lowest $R_{\mathcal{U}ninvested}$ and again, this amount does not depend on the crossover rate. We also note that the higher the crossover rate, the lower the percentage of $R_{\mathcal{U}ninvested}$ for the other profiles. These results contrast with those obtained for N_{Basket} and $R_{\mathcal{A}xis}$ and can be understood because crossing over generates new portfolios and thereby helps to reduce the portion of the uninvested amount. If we look at sales in Figures 25 and 26 on page 48, we note that the range of average values is skinny and therefore we infer that the level of crossover is not critical.

4.3.2 Mutation operator

Regarding the type of mutation used in the subscription/purchases process, we note in Tables 11, 12 and 13 that the best profiles to be used for N_{Basket} and $R_{\mathcal{A}xis}$ are the short put and to a lesser extent the long stock. They have in common the fact that mutation increases linearly from 0 as a function of the first generations. The long call spread does better than the fixed profile and is preferred to reduce the uninvested amount. In Tables 20, 21 and 22 on page 44, we report the same metrics for the redemption/sales process. We note that previous findings hold for N_{Basket} and $R_{\mathcal{A}xis}$.

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	35.87	2.27	-0.07	-0.14	26.00	31.00	36.00	41.00	45.00
Short put	35.41	2.21	-0.04	-0.10	26.00	30.00	35.00	40.00	44.00
Long stock	35.52	2.19	-0.03	-0.05	26.00	30.00	36.00	41.00	43.00
Long call spread	35.86	2.17	-0.07	-0.11	27.00	31.00	36.00	41.00	44.00
Fixed	36.30	2.28	-0.14	-0.15	27.00	31.00	36.00	41.00	45.00
$\text{DW}_{Mutation}$	36.89	2.25	-0.28	0.01	29.00	31.00	37.00	42.00	44.00

Table 11: Impact of the mutation method on N_{Basket} (subscription/purchases process)

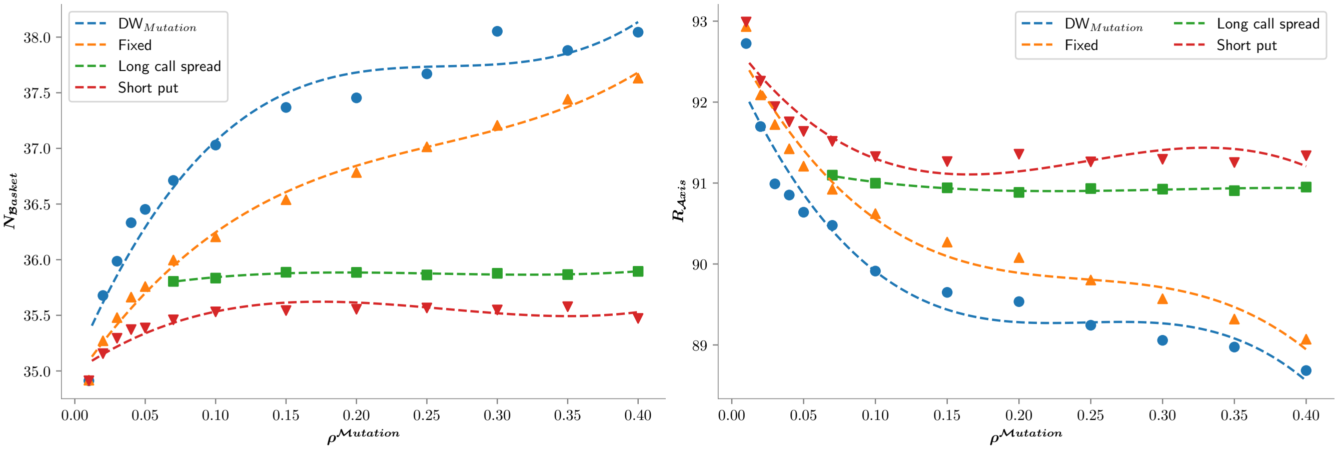
	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	91.10	3.83	0.12	-0.19	75.00	82.35	91.18	100.00	100.00
Short put	91.63	3.96	0.01	-0.34	75.68	82.50	91.43	100.00	100.00
Long stock	91.30	3.80	0.06	-0.08	77.14	82.35	91.43	100.00	100.00
Long call spread	90.95	3.70	0.10	-0.01	75.68	82.35	91.18	100.00	100.00
Fixed	90.69	3.74	0.20	-0.09	75.00	82.35	90.24	100.00	100.00
$\text{DW}_{Mutation}$	90.19	3.48	0.36	0.07	78.95	82.50	89.47	97.37	100.00

Table 12: Impact of the mutation method on $R_{\mathcal{A}xis}$ (subscription/purchases process)

As we did for the crossover, we look at the behavior of profiles for which a fixed value is set for mutation. Figure 14 shows a clear hierarchy between profiles: (1) short put, (2) long call spread, (3) fixed-rate and finally (4) $\text{DW}_{\mathcal{M}utation}$. The short put in the first position is consistent with the results of Tables 11 and 12, where short put and long stock profiles present the best

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	0.97	0.95	1.00	0.77	0.00	0.00	0.54	4.12	6.40
Long call spread	0.86	0.82	0.66	-0.64	0.00	0.00	0.48	2.65	4.68
DW$_{Mutation}$	0.88	0.98	1.51	2.22	0.00	0.00	0.44	4.29	5.38
Long stock	0.89	0.83	0.59	-0.83	0.00	0.00	0.50	2.67	4.37
Fixed	0.93	0.96	1.18	1.25	0.00	0.00	0.48	4.19	6.05
Short put	1.09	1.00	0.87	0.49	0.00	0.00	0.71	4.20	6.40

Table 13: Impact of the mutation method on $R_{\mathcal{U}ninvested}$ (subscription/purchases process)



(a) N_{Basket}

(b) $R_{\mathcal{A}xis}$

Figure 14: Relationship between the mutation rate $\rho^{\mathcal{M}utation}$, N_{Basket} and $R_{\mathcal{A}xis}$ (subscription/purchases process)

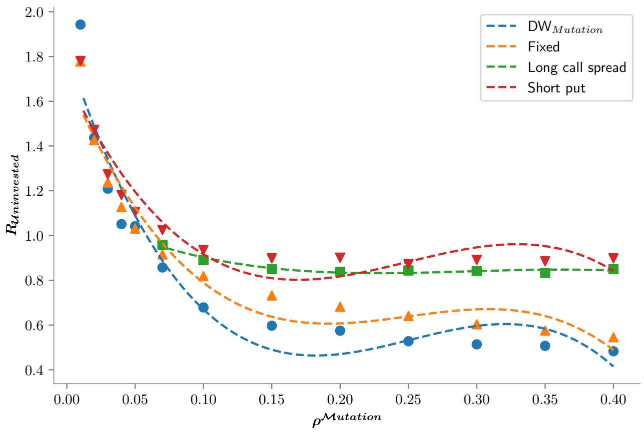


Figure 15: Relationship between the mutation rate $\rho^{\mathcal{M}utation}$ and $R_{\mathcal{U}ninvested}$ (subscription/purchases process)

results. We also observe that the lower the mutation rate, the better the results. For instance, the lowest average of N_{Basket} and highest average of $R_{\mathcal{A}xis}$ is found with a mutation rate of 1%. An interesting result concerns the curves of the short put and long call spread profiles. Indeed, we note that the outcome is a flat curve above 10%, meaning that it does not depend on the chosen mutation rate. Their GAs terminate on average after reaching only 10% of the maximum number of generations. These findings apply to sales as well, as shown in Figure 27 on page 49.

The flip side of the coin is with the lowest mutation rates since we experience high rates of uninvested amounts, as shown in Figure 15. With a mutation rate of 1%, 1.8% of the adjustment is not invested. Again, we observe a flat curve for short put and long call spread profiles above 10%. Figure 28 on page 49 shows that all sales profiles exhibit almost the same level of $R_{\mathcal{U}ninvested}$ if we exclude the first mutation rates of $\text{DW}_{\mathcal{M}utation}$.

4.3.3 Selection operator

Tables 14, 15 and 16 report the statistics related to N_{Basket}, $R_{\mathcal{A}xis}$ and $R_{\mathcal{U}ninvested}$. We note that both types of selection give almost the same statistics with a very slight preference for the traditional method, even though the expansive optimal sampling exhibits a lower standard deviation. This finding is also observable in Tables 23, 24 and 25 on page 45 when we focus on the redemption/sales process.

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	35.87	2.27	-0.07	-0.14	26.00	31.00	36.00	41.00	45.00
Traditional	35.83	2.28	-0.07	-0.15	26.00	31.00	36.00	41.00	44.00
Expansive OS	35.90	2.25	-0.07	-0.13	27.00	31.00	36.00	41.00	45.00

Table 14: Impact of the selection method on N_{Basket} (subscription/purchases process)

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	91.10	3.83	0.12	-0.19	75.00	82.35	91.18	100.00	100.00
Traditional	91.11	3.86	0.12	-0.22	76.32	82.35	91.18	100.00	100.00
Expansive OS	91.09	3.81	0.12	-0.17	75.00	82.35	91.18	100.00	100.00

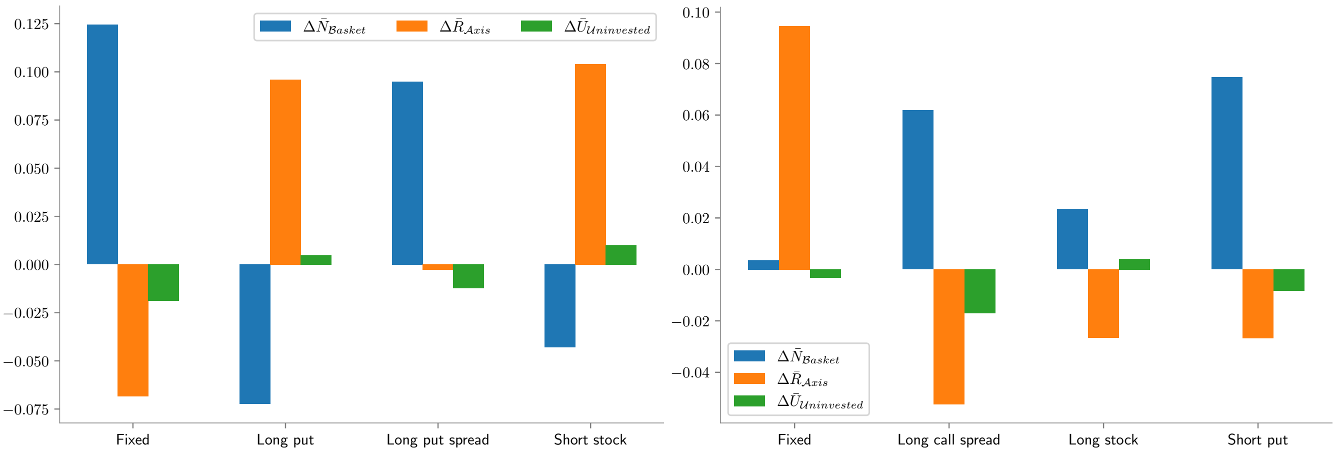
Table 15: Impact of the selection method on $R_{\mathcal{A}xis}$ (subscription/purchases process)

To assess the impact of the selection operator along with the crossover and mutation profiles, we define for each metric M the statistic $\Delta\bar{M} = \bar{M}_{\text{Traditional}} - \bar{M}_{\text{Expansive OS}}$, where $\bar{M}_{\text{Traditional}}$ and $\bar{M}_{\text{Expansive OS}}$ are the average values taken by the metric M when we choose the traditional or the expansive OS selection methods. In Figure 16, we report for each crossover and mutation profiles $\Delta\bar{N}_{\mathcal{A}xis}$, $\Delta\bar{R}_{\mathcal{A}xis}$ and $\Delta\bar{R}_{\mathcal{U}ninvested}$. We find better crossover results when we associate the long put and short stock profiles with the traditional selection scheme since $\Delta\bar{N}_{\mathcal{A}xis}$ is negative and $\Delta\bar{R}_{\mathcal{A}xis}$ is positive. We discard $\Delta\bar{R}_{\mathcal{U}ninvested}$ as it is too close to zero. The expansive OS scheme is suitable for fixed and long put spread profiles. This selection produces better results

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	0.97	0.95	1.00	0.77	0.00	0.00	0.54	4.12	6.40
Expansive OS	0.96	0.95	1.00	0.78	0.00	0.00	0.53	4.12	6.03
Traditional	0.98	0.95	0.99	0.76	0.00	0.00	0.54	4.12	6.40

Table 16: Impact of the selection method on $R_{\mathcal{U}ninvested}$ (subscription/purchases process)

for all mutation profiles except the fixed profile, where the traditional scheme is preferred since $\Delta\bar{R}_{\mathcal{A}xis}$ is positive and the two other metrics is close to zero. These findings are not relevant to the redemption/sales process because the range of metrics is very thin, as shown in Figure 29 on page 50.



(a) Crossover (b) Mutation

Figure 16: Impact of the selection method on the $\Delta\bar{M}$ metrics (subscription/purchases process)

4.4 Combined analysis

Ideally, we might be interested in low values of $N_{\mathcal{B}asket}$, high values of $R_{\mathcal{A}xis}$ and low values of $R_{\mathcal{U}ninvested}$ at the same time. For this purpose, we combine the z-scores of the three winsorized metrics. The winsorisation ensures that the average values used to standardize the variables are less affected by extreme values. For a given variable, the values for all observations are first ranked in ascending order. Then, for observations that lie outside the $1^{\text{st}} - 99^{\text{th}}$ percentile range, their value is set equal to the 1^{st} or 99^{th} percentile values. Then, we define the z-score associated with the i^{th} observation of the random variable w:

$$z_i(w) = \frac{w_i - \mu_w}{\sigma_w}$$

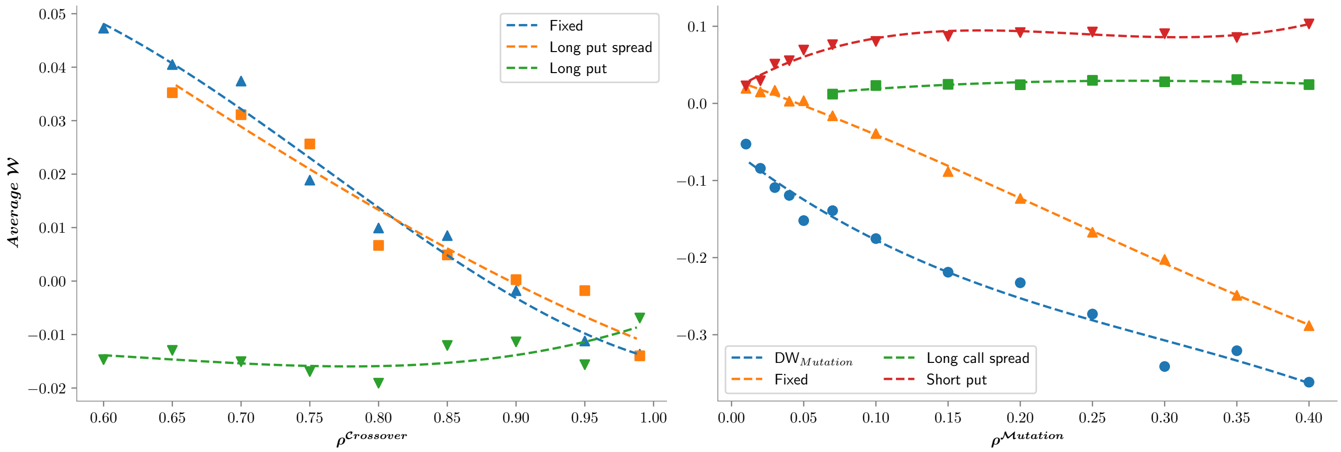
where μ_w and σ_w are respectively the mean and the standard deviation of the winsorized random variable w. The equally weighted z-score using the three winsorized metrics is equal to:

$$\mathcal{W}_i = \frac{-z_i(N_{\mathcal{B}asket}) + z_i(R_{\mathcal{A}xis}) - z_i(R_{\mathcal{U}ninvested})}{3}$$

The negative sign in the combination ensures that an observation having a high value of N_{Basket} or $R_{\mathcal{U}ninvested}$ gets a low z-score. We also define the z-score using only the first two metrics:

$$\mathcal{V}_i = \frac{-z_i\left(N_{Basket}\right) + z_i\left(R_{\mathcal{A}xis}\right)}{2}$$

Figure 17 shows the average z-score $\mathcal{W}$ per crossover or mutation rate. Contrary to mutation, the crossover does not seem to be too discriminating since the range is very thin. That said, we observe a decreasing trend for fixed and long put spread profiles as the rate increases and a flat curve for the long put spread as expected. If we look at mutation, the hierarchy is preserved but the highest z-scores for the short put profile are obtained for mutation rates above 10%. In Figure 18, we use the average z-score $\mathcal{V}$ instead. We find for crossover and mutation the same profiles as those we encountered above for $R_{\mathcal{A}xis}$ graphs.

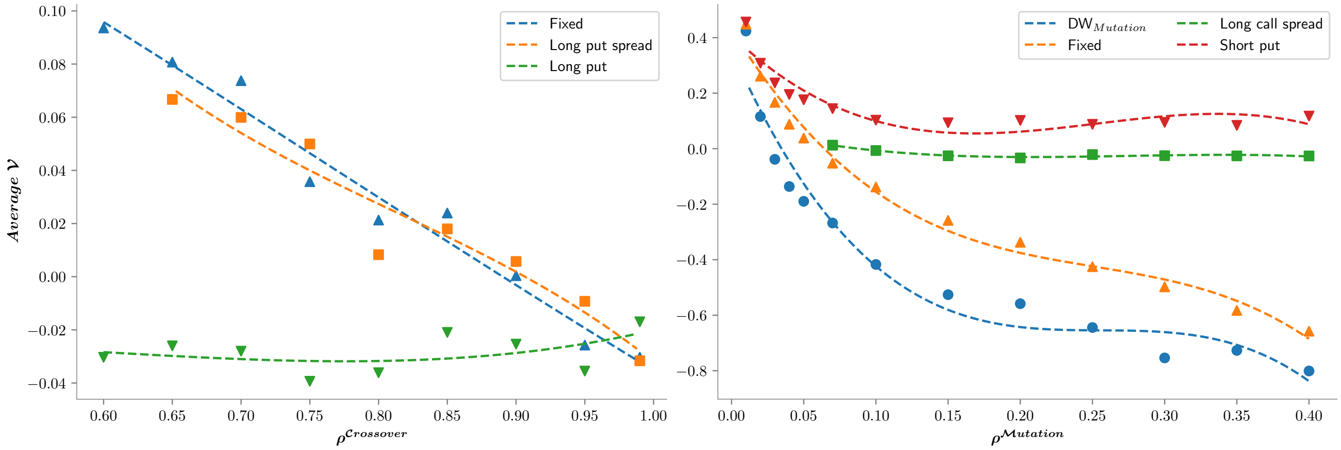


(a) Crossover

(b) Mutation

Figure 17: Relationship between the crossover/mutation rates and the average z-score $\mathcal{W}$ (subscription/purchases process)

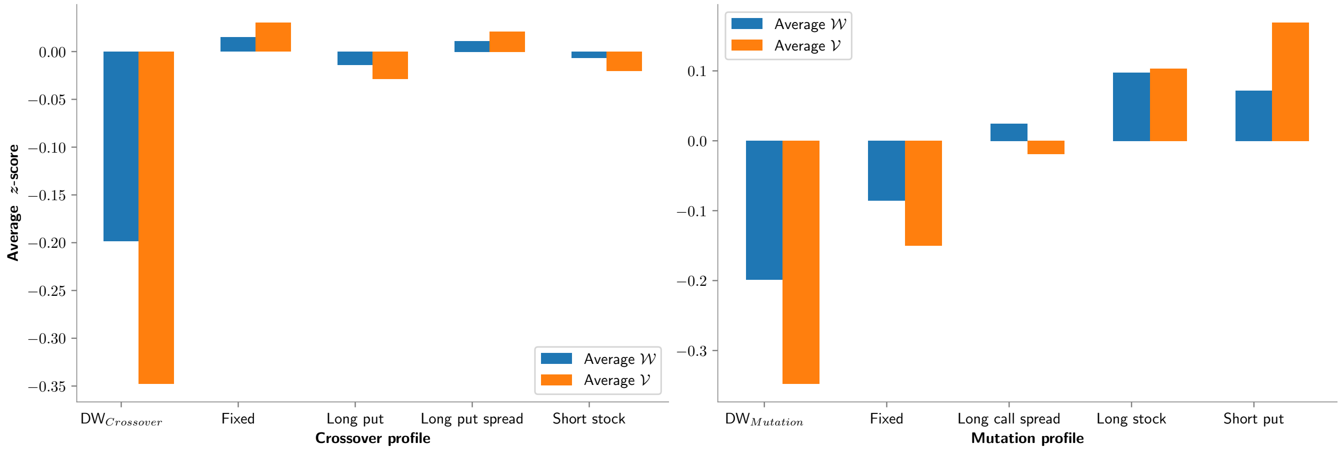
Figure 19 displays the average z-scores $\mathcal{W}$ and $\mathcal{V}$ per crossover and mutation profiles. By construction, the mean of the combined z-score is equal to zero when we consider all the observations. We notice that DW$_{\mathcal{C}rossover}$ and DW$_{\mathcal{M}utation}$ exhibit on average the worst z-scores. This method is definitely not suitable for this exercise of GA with its current settings. Dong and Wu (2009) point out that the results will not be satisfactory if the GA requires high speed (i.e. low number of generations). The results for other crossover profiles are however comparable. Therefore, we may prefer fixed crossover profiles as they show positive z-scores on average. If we look at mutation, both short put and long stock mutation profiles are to be selected since they are the only ones to show clearly positive z-scores.



(a) Crossover

(b) Mutation

Figure 18: Relationship between the crossover/mutation rates and the average z-score $\mathcal{V}$ (subscription/purchases process)



(a) Crossover

(b) Mutation

Figure 19: Average z-scores $\mathcal{W}$ and $\mathcal{V}$ per crossover and mutation profiles (subscription/purchases process)

5 Conclusion

This article uses genetic algorithms to match the reference portfolio's characteristics and handle liquidity and axis constraints when adjusting an amount of subscription or redemption. These algorithms' main contribution lies in their capacity to navigate into huge search spaces looking for optimal combinations of solutions, mainly when the objective function is discrete. Indeed, a candidate solution is a portfolio of bonds belonging to the reference portfolio where each bond has a discrete quantity set up as a multiple of the bond's minimum tradable.

We perform thousands of purchases and sales tests on a real portfolio consisting of 16 sectors and subject to axis constraints. We use different definitions of the genetic operators applying deterministic or self-adaptive rules to crossover, mutation, and selection. We find that the execution time of a few seconds and the outcome are very satisfactory given the considerable number of constraints to be satisfied overall and by sector. Using a population of 50 candidate solutions and a number of generations that does not exceed 500, we also find that GAs with fixed crossover and mutation rates increasing linearly along the number of generations are on average the most successful in minimizing the total number of transactions and maximizing the number of bonds belonging to a counterparty's book and the amount invested. These last findings may be enhanced if we amend the selection process to let parents and offspring struggle to access the next generation. Obviously, not all these results may be extended automatically to all real portfolios, especially when the axis is "*poor*", but they can be the first approach to solving similar automation issues.

Moreover, the genetic algorithm we have built is scalable, even though the chromosomes' definition may need to be amended to manage larger quantities. The extension of the genetic algorithm to more complex bond portfolio optimization problems is straightforward. It can particularly handle additional requirements, such as constraints on ESG scores and carbon metrics, that have now become mandatory when managing bond portfolios for tier-one institutional investors.

References

Ben Dor, A., Dynkin, L., Hyman, J., Houweling, P., van Leeuwen, E., and Penninga, O. (2007), Duration Times Spread, *Journal of Portfolio Management*, 33(2), pp. 77-100.

Ben Slimane, M., and de Jong, M. (2017), Bond Liquidity Scores, *Journal of Fixed Income*, 27(1), pp. 77-82.

Chehouri, A. ,Younes, R., Perron, J., and Ilinca, A. (2016), A Constraint-Handling Technique for Genetic Algorithms using a Violation Factor, *Journal of Computer Sciences*, 12, pp 350-362.

Darwin, C.R. (1859), *On the Origin of Species by means of Natural Selection and The Descent of Man and Selection in Relation to Sex*, John Murray, London.

Dong, M., and Wu, Y. (2009), Dynamic Crossover and Mutation Genetic Algorithm Based on Expansion Sampling, in Deng, H., Wang, L., Wang, F.L., and Lei, J. (eds), *Artificial Intelligence and Computational Intelligence 2009*, Lecture Notes in Computer Science, 5855, Springer, pp. 141-149.

Feng-Tse, L., Cheng-Yan, K., and, Ching-Chi, H. (1993), Applying the Genetic Approach to Simulated Annealing in Solving some NP-hard Problems, *IEEE Transactions on Systems, Man, and Cybernetics*, 23(6), pp. 1752-1767.

Goldberg, D. (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley.

Gray, F. (1953), *Pulse Code Communication*, U.S. Patent No. 2 632 058, Washington, DC.

Hassanat, A., Almohammadi, K., Alkafaween, E., Abunawas, E., Hammouri, A., and Prasath, V.B.S. (2019), Choosing Mutation and Crossover Ratios for Genetic Algorithms – A Review with a New Dynamic Approach, *Information*, 10(12), pp. 1-36.

Holland, J.H. (1975), *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, reprinted in 1992, MIT Press.

Neyman, J., (1992), On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection, in Kotz, S., and Johnson, N.L. (eds), *Breakthroughs in Statistics*, Springer Series in Statistics, Springer, pp 123-150.

Spall, J.C. (2003), *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons.

Wang, Z.G., Rahman, M., Wong, Y.S., and Sun, J. (2005), Optimization of Multi-pass Milling using Parallel Genetic Algorithm and Parallel Genetic Simulated Annealing, *International Journal of Machine Tools and Manufacture*, 45(15), pp 1726-1734.

A Appendix

A.1 Notations

We use the following notations:

- b is the reference (or benchmark) bond portfolio;
- x is a portfolio solution candidate;
- x^0 is the current bond portfolio;
- s is a given sector among the sectors of b;
- n_b is the number of bonds in Portfolio b;
- n_x is the number of bonds in Portfolio x;
- $N_{\mathcal{B}uckets}$ is the number of maturity buckets;
- DTS_i, MD_i, LTP_i and P_i are the duration-times-spread price, modified duration, liquidity-times-price, and the dirty price of Bond i;
- $\mathrm{MD}_i\left(\mathcal{B}ucket_j\right)$ is the contribution of the modified duration of Bond i to the j^{th} maturity bucket:

$$\mathrm{MD}_i = \sum_{j=1}^{N_{\mathcal{B}uckets}} \mathrm{MD}_i\left(\mathcal{B}ucket_j\right)$$

- $\Delta Q_i\left(x\right) = Q_i\left(x\right) - Q_i\left(x^0\right)$ is the change in quantity of Bond i in Portfolio x with respect to its quantity in the current portfolio x^0;
- $\mathcal{C}ash\left(x\right)$ is the amount of cash of Portfolio x;
- $\mathrm{NAV}\left(x\right)$ is the net asset value of Portfolio x:

$$\mathrm{NAV}\left(x\right) = \sum_{i=1}^{n_x} P_i \cdot Q_i\left(x\right) + \mathcal{C}ash\left(x\right)$$

- $\mathcal{A}djustment = \mathrm{NAV}\left(x\right) - \mathrm{NAV}\left(x^0\right)$ is the amount of adjustment;
- $\omega_i\left(x\right)$ is the weight of Bond i in Portfolio x:

$$\omega_i\left(x\right) = \frac{P_i \cdot Q_i\left(x\right)}{\mathrm{NAV}\left(x\right)}$$

- $\omega_{\mathrm{W}}\left(x, s\right)$ is the weight of Sector s in Portfolio x:

$$\begin{aligned}\omega_{\mathrm{W}}\left(x, s\right) &= \sum_{i \in \mathcal{S}ector(s)} \omega_i\left(x\right) \\ &= \frac{1}{\mathrm{NAV}\left(x\right)} \sum_{i \in \mathcal{S}ector(s)} P_i \cdot Q_i\left(x\right)\end{aligned}$$

- $\omega_{\mathrm{DTS}}\left(x,s\right)$ is the contribution to DTS of Sector s in Portfolio x:

$$\omega_{\mathrm{DTS}}\left(x,s\right)=\sum_{i\in\mathcal{S}ector(s)}\omega_{i}\left(x\right)\cdot\mathrm{DTS}_{i}$$

- $\omega_{\mathrm{MD}}\left(x,s\right)$ is the contribution to MD of Sector s in Portfolio x:

$$\omega_{\mathrm{MD}}\left(x,s\right)=\sum_{i\in\mathcal{S}ector(s)}\omega_{i}\left(x\right)\cdot\mathrm{MD}_{i}$$

- $\omega_{\mathrm{MD}(\mathcal{B}ucket_j)}\left(x,s\right)$ is the contribution of the modified duration of Sector s to the j^{th} maturity bucket in Portfolio x:

$$\omega_{\mathrm{MD}(\mathcal{B}ucket_j)}\left(x,s\right)=\sum_{i\in\mathcal{S}ector(s)}\omega_{i}\left(x\right)\cdot\mathrm{MD}_{i}\left(\mathcal{B}ucket_j\right)$$

  By construction, we have:

$$\begin{aligned}\omega_{\mathrm{MD}}\left(x,s\right) &= \sum_{j=1}^{N_{\mathcal{B}uckets}}\omega_{\mathrm{MD}(\mathcal{B}ucket_j)}\left(x,s\right)\\ &= \sum_{j=1}^{N_{\mathcal{B}uckets}}\sum_{i\in\mathcal{S}ector(s)}\omega_{i}\left(x\right)\cdot\mathrm{MD}_{i}\left(\mathcal{B}ucket_j\right)\\ &= \sum_{i\in\mathcal{S}ector(s)}\omega_{i}\left(x\right)\cdot\sum_{j=1}^{N_{\mathcal{B}uckets}}\mathrm{MD}_{i}\left(\mathcal{B}ucket_j\right)\\ &= \sum_{i\in\mathcal{S}ector(s)}\omega_{i}\left(x\right)\cdot\mathrm{MD}_{i}\end{aligned}$$

- $\Delta\omega_{i}\left(x\right)$ is the change in weight of Bond i in Portfolio x with respect to the current portfolio x^{0}:

$$\Delta\omega_{i}\left(x\right)=\frac{P_{i}\cdot\Delta Q_{i}\left(x\right)}{\mathrm{NAV}\left(x\right)}$$

- $\Delta\omega_{\mathrm{W}}\left(x,s\right)$ is the weight of the basket of bonds that belong to Sector s to be purchased or sold with respect to $\mathrm{NAV}\left(x\right)$:

$$\Delta\omega_{\mathrm{W}}\left(x,s\right)=\sum_{i\in\mathcal{S}ector(s)}\Delta\omega_{i}\left(x\right)$$

- $\Delta\omega_{\mathrm{DTS}}\left(x,s\right)$ is the contribution to DTS of the basket of bonds to be purchased or sold from Sector s in Portfolio x:

$$\Delta\omega_{\mathrm{DTS}}\left(x,s\right)=\sum_{i\in\mathcal{S}ector(s)}\Delta\omega_{i}\left(x\right)\cdot\mathrm{DTS}_{i}$$

- $\Delta\omega_{\mathrm{MD}(\mathcal{B}ucket_j)}\left(x,s\right)$ is the contribution of the modified duration of the basket of bonds to be purchased or sold from Sector s to the j^{th} maturity bucket in Portfolio x:

$$\Delta\omega_{\mathrm{MD}(\mathcal{B}ucket_j)}\left(x,s\right) = \sum_{i\in\mathcal{S}ector(s)} \Delta\omega_i\left(x\right)\cdot \mathrm{MD}_i\left(\mathcal{B}ucket_j\right)$$

- $\mathcal{R}_{\mathrm{W}}\left(x \mid b,s\right)$ is the weight risk measure:

$$\mathcal{R}_{\mathrm{W}}\left(x \mid b,s\right) = \left|\sum_{i\in\mathcal{S}ector(s)} \left(x_i - b_i\right)\right|$$

- $\mathcal{R}_{\mathrm{MD}}\left(x \mid b,s\right)$ is the modified duration risk measure:

$$\mathcal{R}_{\mathrm{MD}}\left(x \mid b,s\right) = \sum_{j=1}^{N_{\mathcal{B}uckets}} \left|\sum_{i\in\mathcal{S}ector(s)} \left(x_i - b_i\right)\cdot \mathrm{MD}_i\left(\mathcal{B}ucket_j\right)\right|$$

- $\mathcal{R}_{\mathrm{DTS}}\left(x \mid b,s\right)$ is the duration-times-spread risk measure:

$$\mathcal{R}_{\mathrm{DTS}}\left(x \mid b,s\right) = \left|\sum_{i\in\mathcal{S}ector(s)} \left(x_i - b_i\right)\cdot \mathrm{DTS}_i\right|$$

- $\mathcal{L}iquidity(x,s)$ is the liquidity penalty function of Sector s in Portfolio x:

$$\mathcal{L}iquidity(x,s) = \sum_{i\in\mathcal{S}ector(s)} x_i\cdot \mathrm{LTP}_i$$

- $\Delta\mathcal{L}iquidity\left(x,s\right)$ is the liquidity cost of the basket of bonds to be purchased or sold from Sector s in Portfolio x:

$$\Delta\mathcal{L}iquidity\left(x,s\right) = \sum_{i\in\mathcal{S}ector(s)} \frac{P_i\cdot\Delta Q_i\left(x\right)}{\mathrm{NAV}\left(x\right)}\mathrm{LTP}_i$$

- $\mathcal{A}xis(x,s)$ is the axis penalty function of Sector s in Portfolio x:

$$\mathcal{A}xis(x,s) = \#\left\{i\in\mathcal{S}ector\left(s\right) : \left(1-\delta_i\right)\cdot\left|\Delta Q_i\left(x\right)\right| > 0\right\}$$

  where $\delta_i = 1$ if Bond i is on the axis $\mathcal{A}$ and $\delta_i = 0$ otherwise;

- The fitness function is defined as:

$$\begin{aligned}\mathcal{F}itness\left(x,s\right) \quad &= \quad \mathcal{R}_{\mathrm{W}}\left(x \mid b,s\right) + \mathcal{R}_{\mathrm{MD}}\left(x \mid b,s\right) + \mathcal{R}_{\mathrm{DTS}}\left(x \mid b,s\right) + \\ &\qquad \lambda_L\cdot\mathcal{L}iquidity\left(x,s\right) + \lambda_A\cdot\mathcal{A}xis\left(x,s\right)\end{aligned}$$

- $N_{\mathcal{B}asket}(x)$ is the number of bonds purchased or sold in Portfolio x;

$$N_{\mathcal{B}asket}(x) = \#\left\{i \in x : \Delta Q_i\left(x\right) \neq 0\right\}$$

- $R_{\mathcal{A}xis}(x)$ is the percentage of bonds purchased or sold in Portfolio x that belong to the axis:

$$R_{\mathcal{A}xis}(x) = \frac{1}{N_{\mathcal{B}asket}(x)} \cdot \#\left\{i \in x : \delta_i \cdot \Delta Q_i\left(x\right) \neq 0\right\}$$

  where $\delta_i = 1$ if Bond i is on the axis $\mathcal{A}$ and $\delta_i = 0$ otherwise;

- $R_{\mathcal{U}ninvested}(x)$ is the percentage of the adjustment that remains uninvested:

$$R_{\mathcal{U}ninvested}(x) = 1 - \frac{1}{\mathcal{A}djustment} \cdot \sum_{i \in x} P_i \cdot \Delta Q_i\left(x\right)$$

A.2 Method of additional quantities

A.2.1 In-kind process

In this process, the reference portfolio is the current portfolio x^0. The weight risk measure can be written as follows:

$$\begin{aligned}
\mathcal{R}_{\mathrm{W}}\left(x \mid x^0, s\right) &= \left| \sum_{i \in \mathcal{S}ector(s)} \left(x_i - x_i^0\right) \right| \\
&= \left| \sum_{i \in \mathcal{S}ector(s)} \left(\frac{P_i \cdot Q_i\left(x\right)}{\mathrm{NAV}\left(x\right)} - \frac{P_i \cdot Q_i\left(x^0\right)}{\mathrm{NAV}\left(x^0\right)} \right) \right| \\
&= \left| \sum_{i \in \mathcal{S}ector(s)} \left(\frac{P_i \cdot \left(Q_i\left(x^0\right) + \Delta Q_i\left(x\right)\right)}{\mathrm{NAV}\left(x\right)} - \frac{P_i \cdot Q_i\left(x^0\right)}{\mathrm{NAV}\left(x^0\right)} \right) \right| \\
&= \left| \sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot \Delta Q_i\left(x\right)}{\mathrm{NAV}\left(x\right)} + \left(\frac{1}{\mathrm{NAV}\left(x\right)} - \frac{1}{\mathrm{NAV}\left(x^0\right)} \right) \sum_{i \in \mathcal{S}ector(s)} P_i \cdot Q_i\left(x^0\right) \right| \\
&= \left| \sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot \Delta Q_i\left(x\right)}{\mathrm{NAV}\left(x\right)} - \frac{\mathcal{A}djustment}{\mathrm{NAV}\left(x\right)} \left(\frac{1}{\mathrm{NAV}\left(x^0\right)} \sum_{i \in \mathcal{S}ector(s)} P_i \cdot Q_i\left(x^0\right) \right) \right| \\
&= \left| \sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot \Delta Q_i\left(x\right)}{\mathrm{NAV}\left(x\right)} - \frac{\mathcal{A}djustment}{\mathrm{NAV}\left(x\right)} \left(\frac{1}{\mathrm{NAV}\left(x^0\right)} \sum_{i \in \mathcal{S}ector(s)} P_i \cdot Q_i\left(x^0\right) \right) \right| \\
&= \left| \Delta\omega_{\mathrm{W}}\left(x, s\right) - \frac{\mathcal{A}djustment}{\mathrm{NAV}\left(x\right)} \omega_{\mathrm{W}}\left(x^0, s\right) \right|
\end{aligned}$$

where $\Delta\omega_{\mathrm{W}}\left(x, s\right)$ is the weight of the basket of bonds that belong to Sector s to be purchased with respect to $\mathrm{NAV}\left(x\right)$. We obtain similar formulas for the DTS risk measure:

$$\mathcal{R}_{\mathrm{DTS}}\left(x \mid x^0, s\right) = \left| \Delta\omega_{\mathrm{DTS}}\left(x, s\right) - \frac{\mathcal{A}djustment}{\mathrm{NAV}\left(x\right)} \omega_{\mathrm{DTS}}\left(x^0, s\right) \right|$$

and the MD risk measure:

$$\mathcal{R}_{\mathrm{MD}}\left(x \mid x^0, s\right) = \sum_{j=1}^{N_{\mathcal{B}uckets}} \left| \Delta\omega_{\mathrm{MD}(\mathcal{B}ucket_j)}\left(x, s\right) - \frac{\mathcal{A}djustment}{\mathrm{NAV}\left(x\right)} \omega_{\mathrm{MD}(\mathcal{B}ucket_j)}\left(x^0, s\right) \right|$$

where $\Delta\omega_{\mathrm{DTS}}\left(x, s\right)$ and $\Delta\omega_{\mathrm{MD}(\mathcal{B}ucket_j)}\left(x, s\right)$ are respectively the contribution to DTS of Sector s and the contribution of bucket j to the modified duration of Sector s by the basket of bonds to be purchased or sold.

A.2.2 Standard process

In the standard process, the reference portfolio is the benchmark b. The weight risk measure is equal to:

$$\begin{aligned}
\mathcal{R}_{\mathrm{W}}\left(x \mid b, s\right) &= \left|\sum_{i \in \mathcal{S}ector(s)} \left(x_i - b_i\right)\right| \\
&= \left|\omega_{\mathrm{W}}\left(x, s\right) - \omega_{\mathrm{W}}\left(b, s\right)\right| \\
&= \left|\sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot \left(Q_i\left(x^0\right) + \Delta Q_i\left(x\right)\right)}{\mathrm{NAV}\left(x\right)} - \omega_{\mathrm{W}}\left(b, s\right)\right| \\
&= \left|\sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot \Delta Q_i\left(x\right)}{\mathrm{NAV}\left(x\right)} - \left(\omega_{\mathrm{W}}\left(b, s\right) - \sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot Q_i\left(x^0\right)}{\mathrm{NAV}\left(x\right)}\right)\right| \\
&= \left|\sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot \Delta Q_i\left(x\right)}{\mathrm{NAV}\left(x\right)} - \left(\omega_{\mathrm{W}}\left(b, s\right) - \frac{\mathrm{NAV}\left(x^0\right)}{\mathrm{NAV}\left(x\right)} \sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot Q_i\left(x^0\right)}{\mathrm{NAV}\left(x^0\right)}\right)\right| \\
&= \left|\Delta\omega_{\mathrm{W}}\left(x, s\right) - \left(\omega_{\mathrm{W}}\left(b, s\right) - \frac{\mathrm{NAV}\left(x^0\right)}{\mathrm{NAV}\left(x\right)} \omega_{\mathrm{W}}\left(x^0, s\right)\right)\right|
\end{aligned}$$

The adjustment in the standard process is often performed at a constant NAV:

$$\mathcal{R}_{\mathrm{W}}\left(x \mid b, s\right) = \left|\Delta\omega_{\mathrm{W}}\left(x, s\right) - \left(\omega_{\mathrm{W}}\left(b, s\right) - \omega_{\mathrm{W}}\left(x^0, s\right)\right)\right|$$

For the DTS and modified duration risk measures, we obtain similar formulas. We have:

$$\mathcal{R}_{\mathrm{DTS}}\left(x \mid b, s\right) = \left|\Delta\omega_{\mathrm{DTS}}\left(x, s\right) - \left(\omega_{\mathrm{DTS}}\left(b, s\right) - \omega_{\mathrm{DTS}}\left(x^0, s\right)\right)\right|$$

and:

$$\mathcal{R}_{\mathrm{MD}}\left(x \mid x^0, s\right) = \sum_{j=1}^{N_{\mathcal{B}uckets}} \left|\Delta\omega_{\mathrm{MD}(\mathcal{B}ucket_j)}\left(x, s\right) - \left(\omega_{\mathrm{MD}(\mathcal{B}ucket_j)}\left(b, s\right) - \omega_{\mathrm{MD}(\mathcal{B}ucket_j)}\left(x^0, s\right)\right)\right|$$

A.2.3 Liquidity penalty function

As a reminder, the liquidity penalty function is equal to:

$$\begin{aligned}
\mathcal{L}iquidity\left(x, s\right) &= \sum_{i \in \mathcal{S}ector(s)} x_i \cdot \mathrm{LTP}_i \\
&= \sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot \left(Q_i\left(x^0\right) + \Delta Q_i\left(x\right)\right)}{\mathrm{NAV}\left(x\right)} \mathrm{LTP}_i \\
&= \sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot \Delta Q_i\left(x\right)}{\mathrm{NAV}\left(x\right)} \mathrm{LTP}_i + \sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot Q_i\left(x^0\right)}{\mathrm{NAV}\left(x\right)} \mathrm{LTP}_i \\
&= \sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot \Delta Q_i\left(x\right)}{\mathrm{NAV}\left(x\right)} \mathrm{LTP}_i + \frac{\mathrm{NAV}\left(x^0\right)}{\mathrm{NAV}\left(x\right)} \sum_{i \in \mathcal{S}ector(s)} \frac{P_i \cdot Q_i\left(x^0\right)}{\mathrm{NAV}\left(x^0\right)} \mathrm{LTP}_i
\end{aligned}$$

We notice that:

$$\mathcal{L}iquidity\left(x^{0},s\right)=\sum_{i\in\mathcal{S}ector(s)}\frac{P_{i}\cdot Q_{i}\left(x^{0}\right)}{\mathrm{NAV}\left(x^{0}\right)}\mathrm{LTP}_{i}$$

We deduce that:

$$\mathcal{L}iquidity\left(x,s\right)=\Delta\mathcal{L}iquidity\left(x,s\right)+\frac{\mathrm{NAV}\left(x^{0}\right)}{\mathrm{NAV}\left(x^{0}\right)+\mathcal{A}djustment}\mathcal{L}iquidity\left(x^{0},s\right)$$

where $\Delta\mathcal{L}iquidity\left(x,s\right)$ is the liquidity cost of the basket of bonds to be purchased or sold. For a given $\mathcal{A}djustment$, minimizing $\mathcal{L}iquidity\left(x,s\right)$ is equivalent to minimize $\Delta\mathcal{L}iquidity\left(x,s\right)$.

A.3 Lower bound of the modified duration risk

Using the triangle inequality, we have:

$$\sum_{j=1}^{N_{\mathcal{B}uckets}}\left|\sum_{i\in\mathcal{S}ector(s)}\left(x_{i}-b_{i}\right)\cdot\mathrm{MD}_{i}\left(\mathcal{B}ucket_{j}\right)\right|\geq\left|\sum_{j=1}^{N_{\mathcal{B}uckets}}\left(\sum_{i\in\mathcal{S}ector(s)}\left(x_{i}-b_{i}\right)\cdot\mathrm{MD}_{i}\left(\mathcal{B}ucket_{j}\right)\right)\right|$$

We notice that:

$$\begin{aligned}\sum_{j=1}^{N_{\mathcal{B}uckets}}\left(\sum_{i\in\mathcal{S}ector(s)}\left(x_{i}-b_{i}\right)\cdot\mathrm{MD}_{i}\left(\mathcal{B}ucket_{j}\right)\right) &= \sum_{i\in\mathcal{S}ector(s)}\left(x_{i}-b_{i}\right)\left(\sum_{j=1}^{N_{\mathcal{B}uckets}}\mathrm{MD}_{i}\left(\mathcal{B}ucket_{j}\right)\right)\\ &= \sum_{i\in\mathcal{S}ector(s)}\left(x_{i}-b_{i}\right)\mathrm{MD}_{i}\end{aligned}$$

because:

$$\mathrm{MD}_{i}=\sum_{j=1}^{N_{\mathcal{B}uckets}}\mathrm{MD}_{i}\left(\mathcal{B}ucket_{j}\right)$$

We conclude that:

$$\sum_{j=1}^{N_{\mathcal{B}uckets}}\left|\sum_{i\in\mathcal{S}ector(s)}\left(x_{i}-b_{i}\right)\cdot\mathrm{MD}_{i}\left(\mathcal{B}ucket_{j}\right)\right|\geq\left|\sum_{i\in\mathcal{S}ector(s)}\left(x_{i}-b_{i}\right)\mathrm{MD}_{i}\right|$$

A.4 Knapsack problem

In the knapsack problem (KP), we have a set of items $\{1,\ldots,n\}$ and a knapsack of limited capacity $\mathcal{K}$. To each item, we associate a positive profit p_j and a positive weight w_j. The optimization problem consists of selecting the set of items with the maximum profit that does not exceed the knapsack capacity. We can formulate it as an integer linear programming problem:

$$\begin{aligned}\left\{x_{1},\ldots,x_{n}\right\} &= \arg\max\sum_{j=1}^{n}p_{j}x_{j}\\ &\text{s.t.}\ \left\{\begin{array}{l}\sum_{j=1}^{n}w_{j}x_{j}\leq\mathcal{K}\\ x_{j}\in\{0,1\}\end{array}\right.\end{aligned}$$

If $\sum_{j=1}^{n} w_j \leq \mathcal{K}$, the solution is trivial since we select all the items. If $\sum_{j=1}^{n} w_j > \mathcal{K}$, the items' total weight exceeds the knapsack capacity $\mathcal{K}$ and we obtain an NP-hard problem.

B Additional results

B.1 Tables

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	38.36	1.42	-0.14	-0.00	32.00	35.00	38.00	41.00	44.00
Fixed	38.34	1.43	-0.13	0.00	32.00	35.00	38.00	41.00	44.00
Long put spread	38.35	1.42	-0.14	0.01	32.00	35.00	38.00	41.00	44.00
Short stock	38.36	1.39	-0.12	-0.08	33.00	35.00	38.00	41.00	44.00
Long put	38.36	1.40	-0.14	-0.02	32.00	35.00	38.00	41.00	44.00
$\text{DW}_{Crossover}$	38.75	1.32	-0.22	0.06	34.00	35.99	39.00	41.00	44.00

Table 17: Impact of the crossover method on N_{Basket} (redemption/sales process)

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	82.38	2.22	0.08	0.46	72.22	76.92	82.05	87.50	94.59
Short stock	82.41	2.21	0.08	0.49	73.68	76.92	82.05	87.50	91.89
Long put	82.41	2.19	0.10	0.50	72.97	76.92	82.05	87.50	94.44
Fixed	82.38	2.25	0.06	0.44	72.22	76.92	82.05	87.50	94.59
Long put spread	82.37	2.24	0.07	0.43	72.22	76.92	82.05	87.50	94.44
$\text{DW}_{Crossover}$	81.92	2.00	0.08	0.66	73.68	76.92	82.05	86.84	91.43

Table 18: Impact of the crossover method on $R_{\mathcal{A}xis}$ (redemption/sales process)

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	0.75	0.59	0.19	-1.33	0.00	0.00	0.68	1.79	2.01
Short stock	0.74	0.59	0.20	-1.32	0.00	0.00	0.68	1.79	1.99
Long put	0.75	0.59	0.20	-1.32	0.00	0.00	0.68	1.78	2.00
Long put spread	0.75	0.59	0.18	-1.34	0.00	0.00	0.68	1.79	2.01
Fixed	0.75	0.59	0.18	-1.34	0.00	0.00	0.68	1.79	1.98
$\text{DW}_{Crossover}$	0.77	0.60	0.15	-1.37	0.00	0.00	0.71	1.81	1.93

Table 19: Impact of the crossover method on $R_{\mathcal{U}ninvested}$ (redemption/sales process)

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	38.36	1.42	-0.14	-0.00	32.00	35.00	38.00	41.00	44.00
Short put	38.24	1.45	-0.13	-0.01	32.00	35.00	38.00	41.00	44.00
Long stock	38.26	1.43	-0.09	0.04	32.00	35.00	38.00	41.00	43.00
Long call spread	38.35	1.41	-0.12	-0.02	32.00	35.00	38.00	41.00	44.00
Fixed	38.46	1.38	-0.14	0.00	32.00	35.00	38.00	41.00	44.00
$\mathrm{DW}_{Mutation}$	38.75	1.32	-0.22	0.06	34.00	35.99	39.00	41.00	44.00

Table 20: Impact of the mutation method on N_{Basket} (redemption/sales process)

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	82.38	2.22	0.08	0.46	72.22	76.92	82.05	87.50	94.59
Short put	82.51	2.30	0.06	0.40	72.97	76.92	82.05	88.57	94.59
Long stock	82.48	2.30	0.05	0.50	72.97	76.92	82.05	87.50	94.44
Long call spread	82.40	2.22	0.04	0.43	72.22	76.92	82.05	87.50	94.44
Fixed	82.26	2.14	0.10	0.52	72.97	76.92	82.05	87.18	92.31
$\mathrm{DW}_{Mutation}$	81.92	2.00	0.08	0.66	73.68	76.92	82.05	86.84	91.43

Table 21: Impact of the mutation method on $R_{\mathcal{A}xis}$ (redemption/sales process)

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	0.75	0.59	0.19	-1.33	0.00	0.00	0.68	1.79	2.01
Long stock	0.73	0.58	0.21	-1.33	0.00	0.00	0.66	1.77	1.97
Short put	0.73	0.58	0.21	-1.33	0.00	0.00	0.66	1.78	2.01
Long call spread	0.75	0.59	0.17	-1.35	0.00	0.00	0.69	1.79	2.00
Fixed	0.76	0.59	0.17	-1.33	0.00	0.00	0.70	1.79	1.98
$\mathrm{DW}_{Mutation}$	0.77	0.60	0.15	-1.37	0.00	0.00	0.71	1.81	1.93

Table 22: Impact of the mutation method on $R_{\mathcal{U}ninvested}$ (redemption/sales process)

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	38.36	1.42	-0.14	-0.00	32.00	35.00	38.00	41.00	44.00
Traditional	38.34	1.42	-0.13	-0.01	32.00	35.00	38.00	41.00	44.00
Expansive OS	38.37	1.41	-0.15	0.00	32.00	35.00	38.00	41.00	44.00

Table 23: Impact of the selection method on N_{Basket} (redemption/sales process)

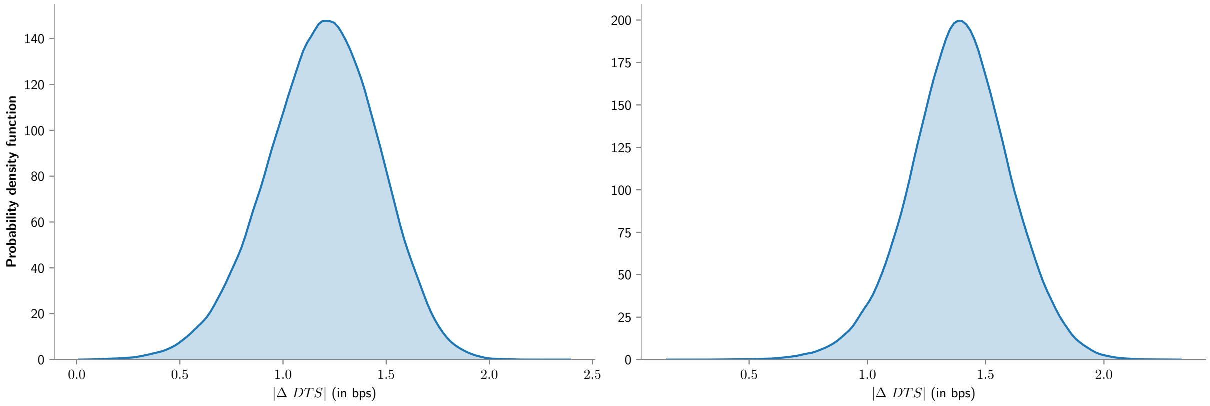
	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	82.38	2.22	0.08	0.46	72.22	76.92	82.05	87.50	94.59
Expansive OS	82.38	2.21	0.08	0.45	72.22	76.92	82.05	87.50	94.59
Traditional	82.38	2.24	0.07	0.47	72.22	76.92	82.05	87.50	94.44

Table 24: Impact of the selection method on $R_{\mathcal{A}xis}$ (redemption/sales process)

	Mean	Std. dev.	Skew.	Kurt.	Min.	Q_1	Median	Q_{99}	Max.
All	0.75	0.59	0.19	-1.33	0.00	0.00	0.68	1.79	2.01
Traditional	0.75	0.59	0.19	-1.34	0.00	0.00	0.68	1.79	2.01
Expansive OS	0.75	0.59	0.19	-1.33	0.00	0.00	0.68	1.79	2.00

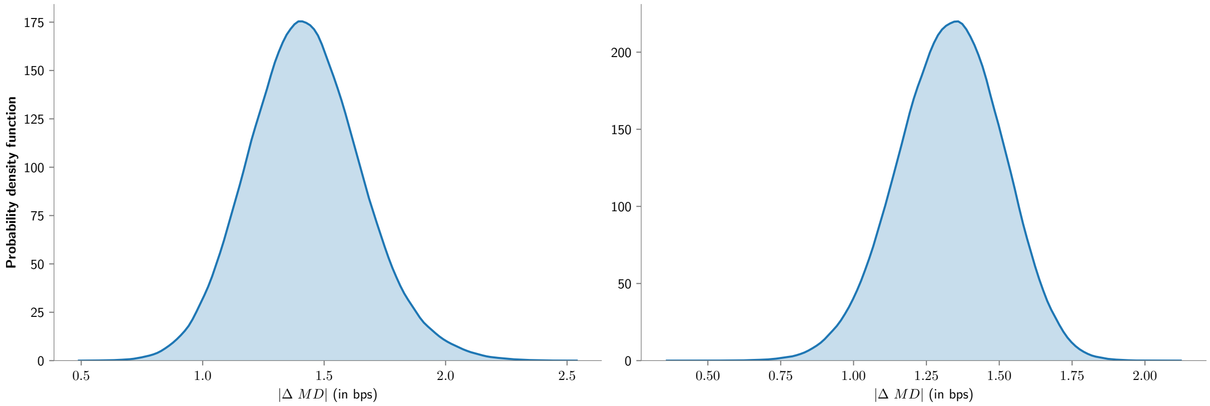
Table 25: Impact of the selection method on $R_{\mathcal{U}ninvested}$ (redemption/sales process)

B.2 Figures



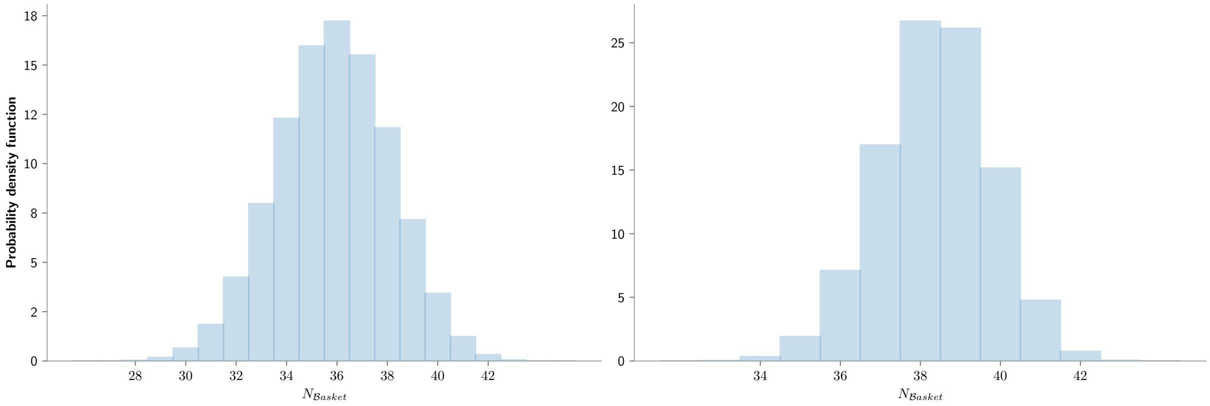
(a) Purchases (b) Sales

Figure 20: Probability distribution of |ΔDTS|



(a) Purchases (b) Sales

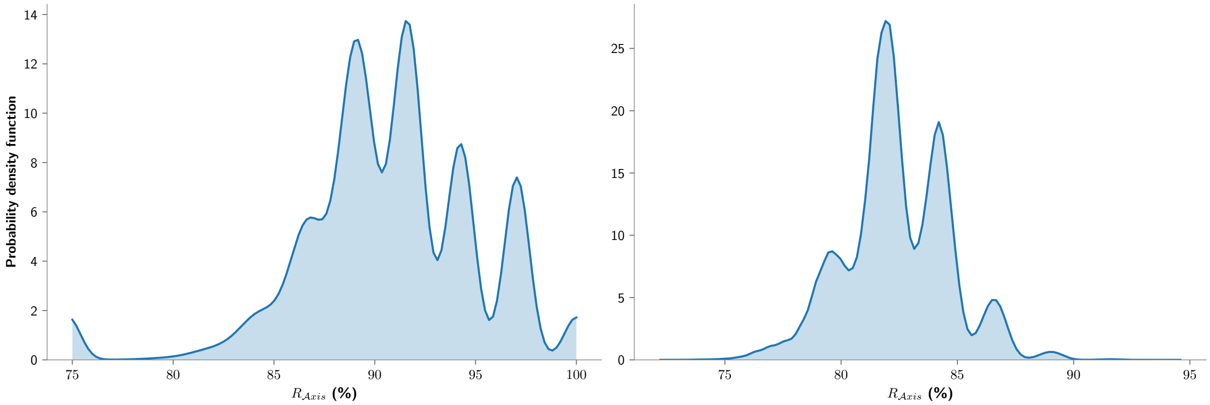
Figure 21: Probability distribution of |ΔMD|



(a) Purchases

(b) Sales

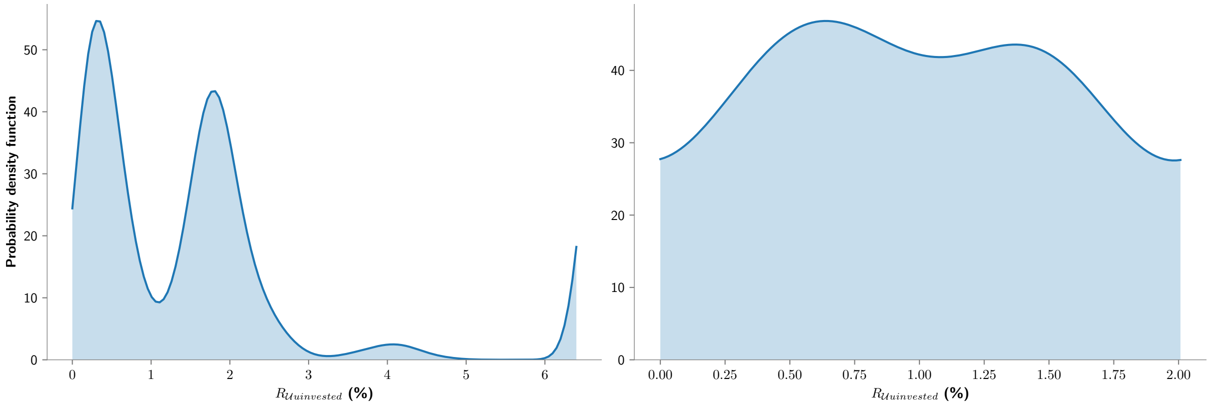
Figure 22: Probability distribution of N_{Basket}



(a) Purchases

(b) Sales

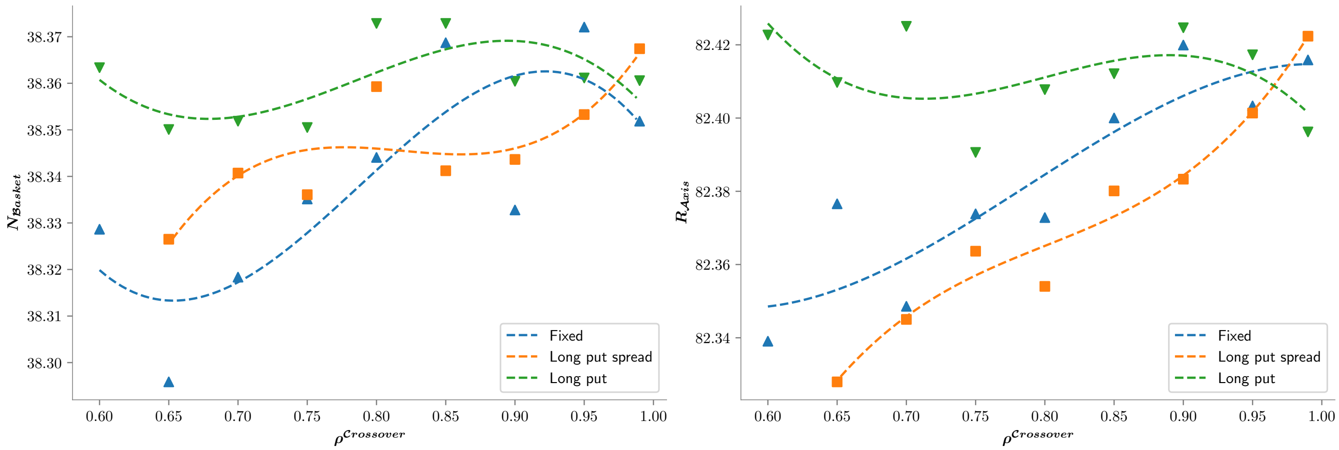
Figure 23: Probability distribution of $R_{\mathcal{A}xis}$



(a) Purchases

(b) Sales

Figure 24: Probability distribution of $R_{\mathcal{U}ninvested}$



(a) N_{Basket}

(b) $R_{\mathcal{A}xis}$

Figure 25: Relationship between the crossover rate $\rho^{Crossover}$, N_{Basket} and $R_{\mathcal{A}xis}$ (redemption/sales process)

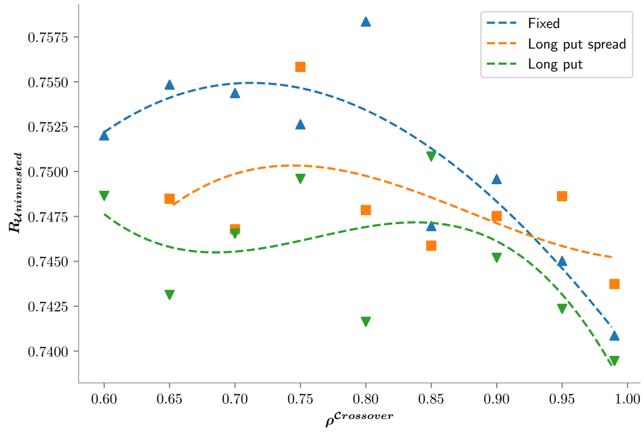
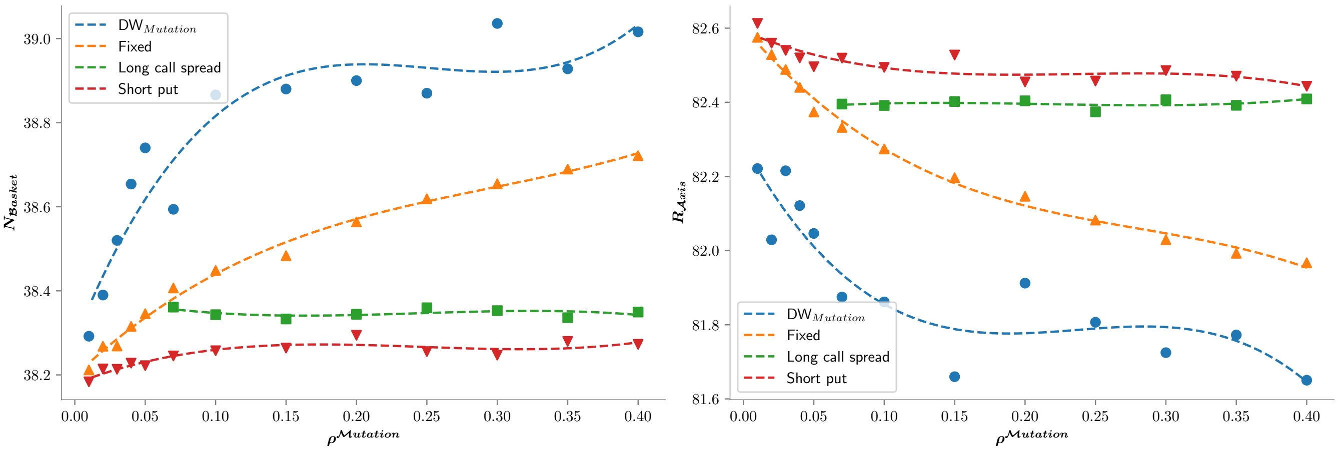


Figure 26: Relationship between the crossover rate $\rho^{Crossover}$ and $R_{\mathcal{U}ninvested}$ (redemption/sales process)



(a) N_{Basket}

(b) $R_{\mathcal{A}xis}$

Figure 27: Relationship between the mutation rate $\rho^{\mathcal{M}utation}$, N_{Basket} and $R_{\mathcal{A}xis}$ (redemption/sales process)

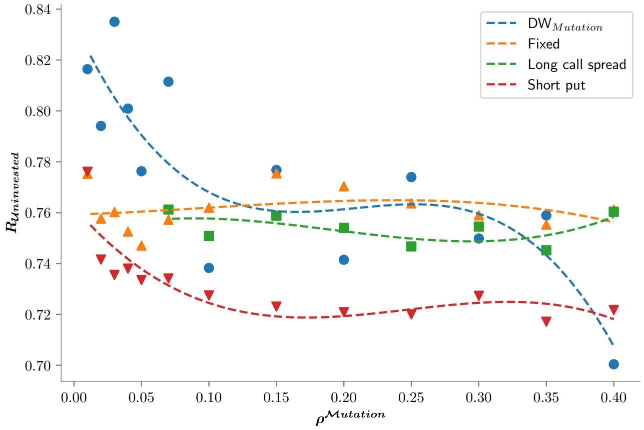
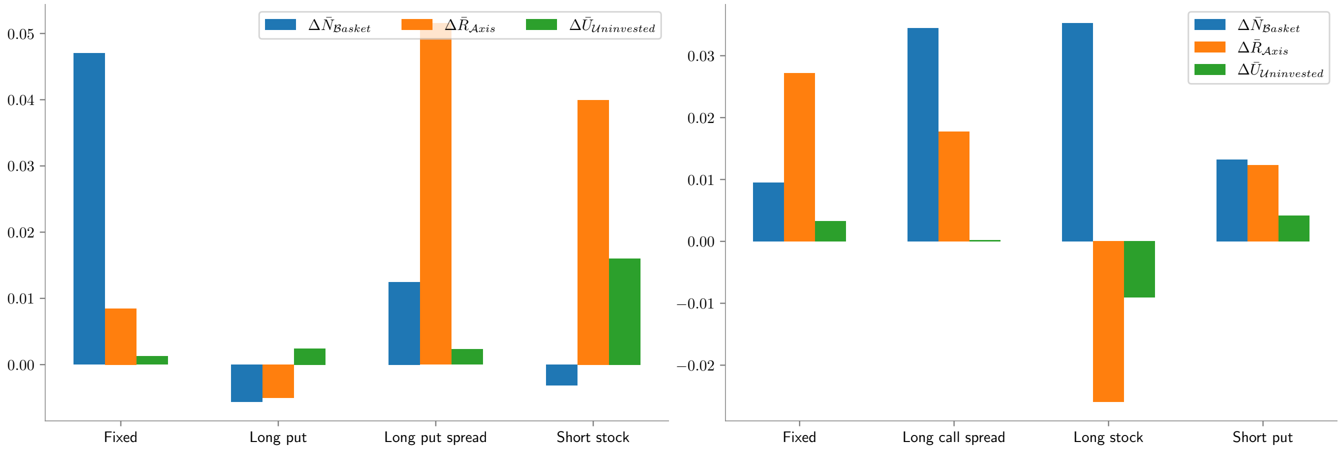
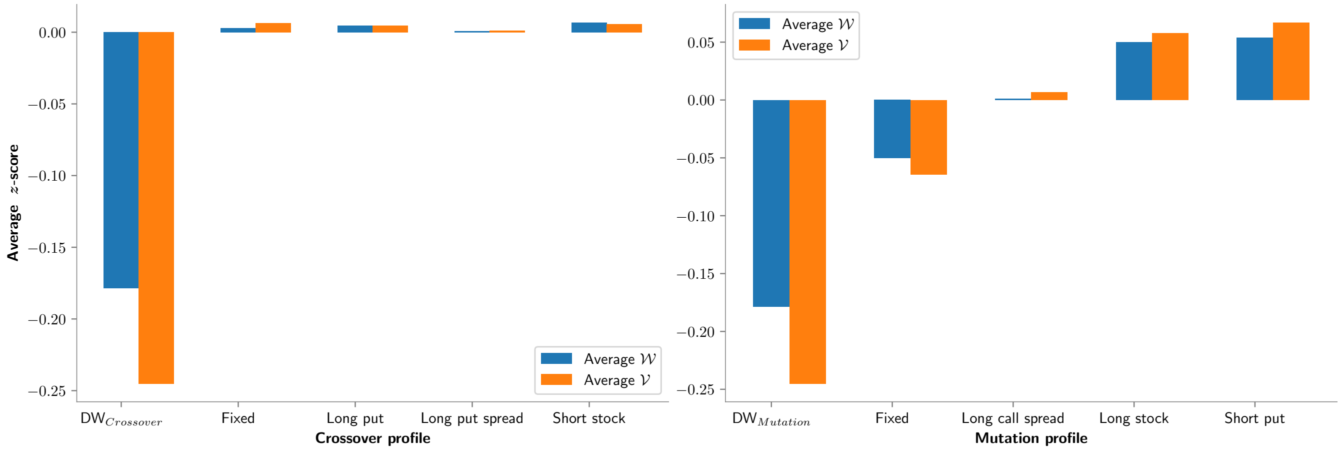


Figure 28: Relationship between the mutation rate $\rho^{\mathcal{M}utation}$ and $R_{\mathcal{U}ninvested}$ (redemption/sales process)



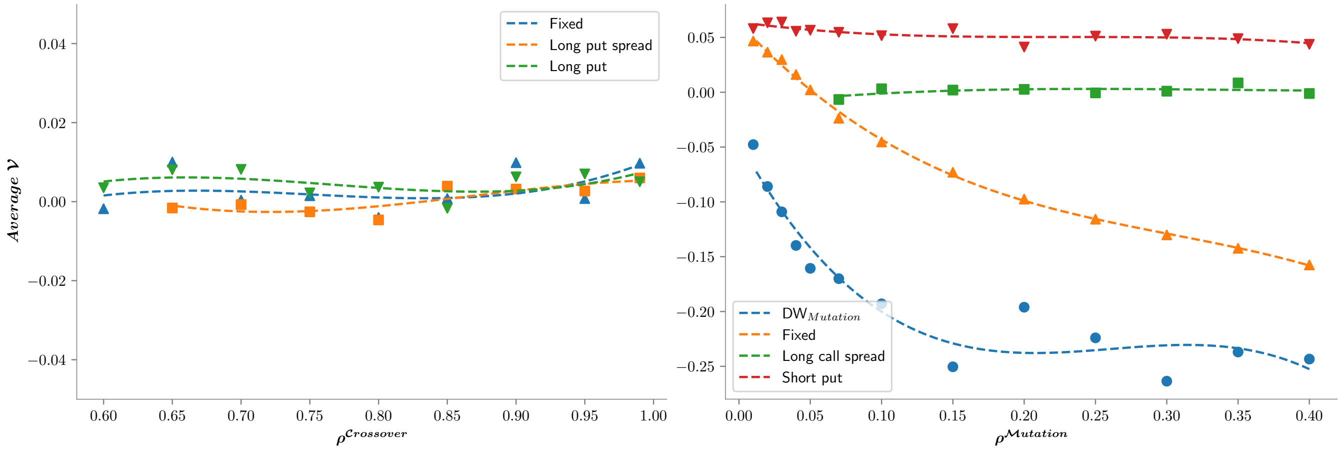
(a) Crossover (b) Mutation

Figure 29: Impact of the selection method on the $\Delta \bar{M}$ metrics (redemption/sales process)



(a) Crossover (b) Mutation

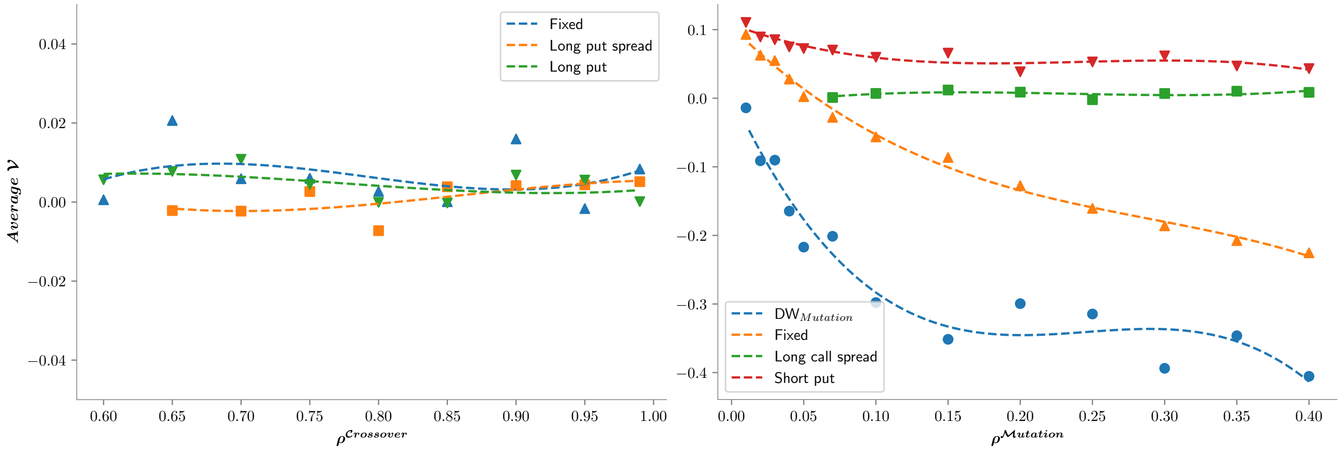
Figure 30: Average z-scores $\mathcal{W}$ and $\mathcal{V}$ per crossover and mutation profiles (redemption/sales process)



(a) Crossover

(b) Mutation

Figure 31: Relationship between the crossover/mutation rates and the average z-score $\mathcal{W}$ (redemption/sales process)



(a) Crossover

(b) Mutation

Figure 32: Relationship between the crossover/mutation rates and the average z-score $\mathcal{V}$ (redemption/sales process)

Chief Editors

Pascal BLANQUÉ
Chief Investment Officer

Philippe ITHURBIDE
Senior Economic Advisor

Amundi
ASSET MANAGEMENT

Working Paper
March 2021

DISCLAIMER

In the European Union, this document is only for the attention of "Professional" investors as defined in Directive 2004/39/EC dated 21 April 2004 on markets in financial instruments ("MIFID"), to investment services providers and any other professional of the financial industry, and as the case may be in each local regulations and, as far as the offering in Switzerland is concerned, a "Qualified Investor" within the meaning of the provisions of the Swiss Collective Investment Schemes Act of 23 June 2006 (CISA), the Swiss Collective Investment Schemes Ordinance of 22 November 2006 (CISO) and the FINMA's Circular 08/8 on Public Advertising under the Collective Investment Schemes legislation of 20 November 2008. In no event may this material be distributed in the European Union to non "Professional" investors as defined in the MIFID or in each local regulation, or in Switzerland to investors who do not comply with the definition of "qualified investors" as defined in the applicable legislation and regulation. This document is not intended for citizens or residents of the United States of America or to any «U.S. Person» , as this term is defined in SEC Regulation S under the U.S. Securities Act of 1933.

This document neither constitutes an offer to buy nor a solicitation to sell a product, and shall not be considered as an unlawful solicitation or an investment advice.

Amundi accepts no liability whatsoever, whether direct or indirect, that may arise from the use of information contained in this material. Amundi can in no way be held responsible for any decision or investment made on the basis of information contained in this material. The information contained in this document is disclosed to you on a confidential basis and shall not be copied, reproduced, modified, translated or distributed without the prior written approval of Amundi, to any third person or entity in any country or jurisdiction which would subject Amundi or any of "the Funds", to any registration requirements within these jurisdictions or where it might be considered as unlawful. Accordingly, this material is for distribution solely in jurisdictions where permitted and to persons who may receive it without breaching applicable legal or regulatory requirements.

The information contained in this document is deemed accurate as at the date of publication set out on the first page of this document. Data, opinions and estimates may be changed without notice. These data have been analyzed anonymously for scientific, statistical or historical research purposes.

You have the right to receive information about the personal information we hold on you. You can obtain a copy of the information we hold on you by sending an email to info@amundi.com. If you are concerned that any of the information we hold on you is incorrect, please contact us at info@amundi.com

Document issued by Amundi Asset Management, "société par actions simplifiée"- SAS with a capital of €1,086,262,605 - Portfolio manager regulated by the AMF under number GP04000036 – Head office: 90 boulevard Pasteur – 75015 Paris – France – 437 574 452 RCS Paris - www.amundi.com

Photo credit: iStock by Getty Images - monsitj/Sam Edwards

Find out more about
Amundi publications

research-center.amundi.com