Discovering mutual funds' holdings between disclosure dates

J. Bellina[1], I. Girerd-Potin[2], and D. Georges[3]

[1,2]Univ. Grenoble Alpes, Grenoble INP, CERAG, 38000 Grenoble France
[3]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 38000 Grenoble France

Abstract

Mutual funds are subject to regulations that require them to periodically disclose their portfolio holdings. Periods without information can be detrimental to investors as the exact risk and performance cannot be assessed and some internal manipulations can be performed in order to beautify the disclosed holdings. In this paper, we aim to create a method that can retrieve the exact portfolio holdings of a mutual fund on a daily basis using only publicly available data. We have developed two efficient methods, the rolling window with Sequential Oscillating Selection and the rolling window with Genetic Algorithm. In a limited investment universe, the two methods are able to retrieve the exact holdings of a portfolio and the voluntary changes in the portfolio composition, under the condition of a maximum number of voluntary changes in a given period. In addition, these methods allow us to find the exact dates on which the manager changed the portfolio.

Keywords: Genetic Algorithm, Portfolio holdings estimation, Mutual funds, Inverse problem

JEL classification codes: C61, C63, G23

1 Introduction

Mutual funds are a way for investors to increase their wealth by pooling their money with other investors to allow a professional fund manager to purchase securities or assets that will generate returns. Open-end funds are a particular type of funds that can issue an unlimited number of shares, so they are easily accessible to individual or institutional investors. Open-end funds worldwide managed over $63 trillion in 2020 mainly in the United States of America and Europe. They can be of different types which will impact how they invest. Mutual funds will try to create a portfolio that generates the best possible return following investment rules [1]. In the following, we will focus on open-end mutual funds.

Mutual funds have many stakeholders who can impact the fund's performance and operations. As described in the Figure 1, mutual funds primary stakeholders are the investors who pool their money in the fund. Indeed, since their goal in investing in a mutual fund is to increase their wealth, they can put pressure on the manager and the fund manager by threatening to withdraw their funds or add new ones. The relationship between the fund manager and the investor can raise agency issues when the objectives of both the investor and the manager are different. Regulators will have an impact on the fund by creating or removing rules to regulate funds' activity. In addition, companies are an important stakeholder because many of their stocks or bonds are held by mutual funds and therefore determine their performance. Stock ownership gives mutual funds the power to vote at shareholder meetings or to exert some pressure on the company through its stock market valuation. Finally, other mutual funds can impact a mutual fund through manipulations such as front-running or copycatting[2].

Mutual funds must regularly publish their portfolio holdings (i.e. the assets in which they are invested). Between publication dates, no information about their holdings is publicly available. This missing information can therefore cause problems, such as investors misjudging the actual risk and performance or the possibility for fund managers to engage in portfolio manipulation. In this paper, focused on equity mutual funds, some classical automation methods and feature selection algorithms have been used in order to recover the missing information. Indeed, the idea is to use only publicly available data, i.e. daily asset returns and daily fund returns in order to assess the exact composition of a mutual fund portfolio and to detect voluntary changes that have occurred during the period under study.

[1] Mutual funds must publish a prospectus that states their limitations in terms of investment choice. For example, a mutual fund may be limited in terms of proportion of stocks from a specific country or industry or may be required to invest only in stocks of sustainable companies.

[2] Front-running consists in selling or buying stocks for which it is known that the price will move due to important orders placed by investors. Copycat funds use public holdings disclosure in order to construct exactly the same portfolio as the disclosing fund, therefore removing any need for internal research

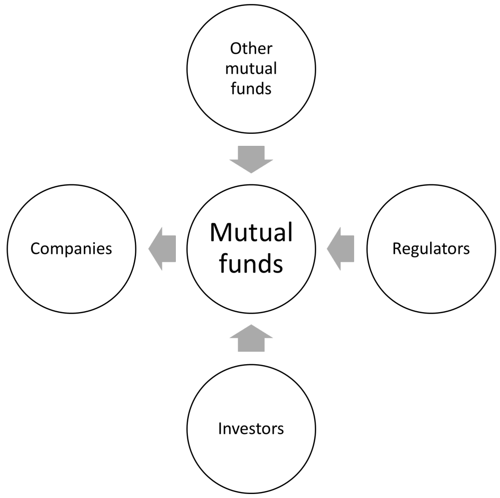


Figure 1: Mutual funds stakeholders

The paper shows that, in some cases, all the missing information can be recovered by accurately estimating the holdings of each stock in the portfolio for each day of the period. Some methods presented are also able to detect changes and the dates on which they occur. The study also compares the performance of the different methods for retrieving holdings and changes.

The holdings assessing problem has already been adressed in the literature. Georges and Girerd-Potin (2017), use an optimal state observer to do so. They show that given a sufficient time window, the holdings of a small portfolio of five indices in a investment universe of ten indices can be perfectly recovered even with a change during the period under study. However, the model is a first step with a simple simulated example and can be improved to better fit larger portfolios with more changes. In addition, Byrd, Bajaj, and Balch (2019) use a feature selection algorithm named Sequential Oscillating Selection method in order to try to evaluate the funds' holdings. They show that their SOS method outperforms a more traditional method: the Extended Linear Clones method, which is an extension of methods commonly used in index tracking. However, this paper doesn't address at all the problem of manager's changes during the estimation period.

Section 2 presents the debate about how often mutual funds should disclose their portfolio holdings. It provides the arguments for and against giving stakeholders the opportunity to know the holdings of funds on a daily basis. The following sections will deal with the inverse problem statement, the description of the models used, the data used and finally the results will be discussed before a concluding part.

2 Holdings disclosure frequency impact

First, the regulation of mutual funds in terms of holdings disclosure will be explained before stating the impact of increasing the disclosure frequency on the mutual funds stakeholders mentioned above. If the literature shows that the arguments in favor of increasing the disclosure frequency outweigh the arguments against, we have an incentive to develop tools that allow to know the composition of mutual fund portfolios more often than is required by regulation.

2.1 Holdings disclosure regulation

Because mutual funds manage money of investors, they have to be regulated and follow rules established to protect investors and help them compare performance and risk, for example. When talking about mutual funds and their regulation, portfolio holdings disclosure is a central topic. Portfolio holdings disclosures are official publications of the entire content of a mutual fund's portfolio at a given date. This means that funds must, for example, publish what is inside their portfolio the 1st of January and this disclosure must be published within a time frame set by the regulation under the name of lag. Funds can also try to increase this lag by making a request to the competent authority (SEC [3] in the USA or AMF [4] in France for example). The rules for these disclosures vary widely from country to country and raise many questions in the literature about their advantages and disadvantages.

According to the Investment Company Institute 2021 Factbook[5], the largest mutual funds industries are located in the United States of America, Australia, Japan and Europe. These different countries have different portfolio holdings disclosure rules. Some require mutual funds to disclose holdings very frequently, while in other countries there is no mandatory disclosure. For example, in the USA, one of the most regulated countries in terms of holdings disclosure, the Securities Act of 1933, the Investment Company Act of 1940 and its amendments[6] require mutual funds to disclose their portfolio holdings on a quarterly basis with a maximum lag of 60 days since 2004. Prior to that, the disclosure was done on a semi-annual basis. India is also a highly regulated country with a monthly holdings disclosure with an average lag of 10 open days. In contrast, in Australia, there is no mandatory disclosure and therefore, most funds that publicly disclose their holdings do so voluntarily. Some labels make an exception to this rule by imposing some disclosure such as the "KiwiSaver" funds. For Japan, the disclosure frequency depends on the settlement cycle of the fund, and can be annual or semi-annual with a maximum lag of 90 days. The European directive 2009/65/CE[7] makes it mandatory for funds in European Union coun-

[3] Securities Exchange Commission
[4] Autorité des Marchés Financiers
[5] https://www.ici.org/system/files/2021-05/2021_factbook.pdf
[6] https://www.sec.gov/rules/final/33-8393.htm
[7] https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32009L0065

tries to disclose their holdings at least twice a year, with 4 and 2 months of lag for the annual and semi-annual reports respectively. In France especially, this European directive is still the one followed by the AMF. Conversely, Spain has implemented a tightened regulation where funds have to disclose their holdings on a quarterly basis. China requires equity funds to disclose their entire holdings in annual and semiannual reports with 90 and 60 days lags respectively while the top-10 equity holdings and top-5 bond holdings must be published quarterly with of 15 open days lag.

However, many funds are already disclosing more often than required. Gimeno, Ortiz, and Sarto (2021) explain that over the past 20 years, voluntary disclosure has increased, showing the growing value placed on investor protection and awareness of true performance and risk. For example, the Morningstar Global Investor Experience Study on disclosure[8] shows that, despite local regulations, nearly 80% of Spain funds, 60% of U.S. funds and 50% of French funds already disclose their holdings on a monthly basis. Li et al. (2022) also explains that funds that decide to increase their voluntary portfolio disclosure frequency experience an increase in inflows, which may explain part of the increase in voluntary portfolio holdings disclosure.

2.2 Impact on investors

First, the frequency of disclosure will have an impact on the fund's investors. More frequent disclosure of holdings may have an impact on the fund's performance and also on the way investors correctly assess that performance and risk. The frequency is also important for verifying compliance with the fund's stated investment policy.

2.2.1 Disclosure and performance

A central question on the topic of portfolio disclosure frequency is its impact on the performance of disclosing funds. As discussed above, we could argue that portfolio manipulations such as front-running or copycat are detrimental to the performance of disclosing funds. However, this statement may not be true. Gregory-Allen, Balli, and Thompson (2019) study the impact of mandatory holdings disclosure on fund returns in Australia and New Zealand where no holdings disclosure is mandatory except for certain labeled funds. This label, created in 2007, made this type of disclosure mandatory in 2013. They used a difference-in-difference method to assess the impact of this policy change on the performance of these funds and no significant impact was observed. In addition, voluntary disclosure of holdings may also not be correlated with fund performance. Li et al. (2022), found that voluntary portfolio disclosure is not linearly correlated with performance and, more importantly, funds that change their voluntary holdings disclosure policy do not experience a significant decline in performance. In fact,they show that top-performing and worst-performing funds

[8] https://www.morningstar.com/lp/global-investor-experience-disclosure

tend to disclose less while middle-performing funds tend to disclose more often. They conclude that the SEC's 2016 decision to keep only quarterly disclosure mandatory helps top performing funds to avoid copycat and front-running and middle-performing funds to have more inflows without a drop in performance decrease.

2.2.2 Investor monitoring

More frequent holdings disclosure helps investors monitor the fund. Investors can then assess the fund's risk and performance, but also its strategy and the companies in which the fund is invested. For example, tracking portfolio companies could be particularly useful for Socially responsible funds.

First, Li et al. (2022) finds that the disclosure frequency is linearly correlated with institutional ownership, i.e. the more institutional investors a fund has, the more often it may disclose. This may be due to the ability of institutional investors to engage and monitor the funds in which they invest and thus their desire to have as much information as possible.

When funds don't disclose very often, many trades are missed and cannot be monitored by investors. Specifically, 18.5% (34.2%) of trades are missed when using quarterly (semiannual) instead of monthly disclosure according to Elton, Gruber, Blake, Krasny, and Ozelge (2010). Furthermore, the turnover computed with monthly holdings and the Morningstar one are very close, showing that most of the trades are captured in the monthly holdings. These missed trades can be misleading to some investors who are trying to evaluate the strategy of the fund in which they have invested. The quarterly data, for example, shows that most funds use a momentum strategy[9] which is not the case when using monthly data.

These missed trades may reduce investors' ability to assess the fund's true performance and risk. Chen, Gallagher, and Lee (2017) show that disclosed holdings underestimate fund performance, particularly for the top-performing ones, and that this effect is larger for less frequent disclosure dates. Thus, the more frequent the disclosures, the more they can help the investor to estimate the fund's true performance. Choi et al. (2022) shows that increasing the portfolio disclosure frequencies may decrease the information asymetry between fund investors and managers and thus increased the knowledge of investors about managerial skill. In order to measure the performance of a fund, some measures based on holdings have been proposed in the literature. Elton and Gruber (2020), in their review, argue that these measures are more robust to fund style changing and are better able to assess the manager's industry and asset selection skills. For example, Grinblatt and Titman (1989) created a measure

[9] Momentum strategy is an investment strategy which consists in buying (short-selling) securities based on the upward (downward) trend they are following. This strategy is based on the hypothesis that a well-established trend is likely to continue.

that compares the return of a portfolio with the actual weights and one with prior weights. Holdings can also be used to compute the portfolio *beta* as a weighted average of the individual stock *betas*. Unfortunately these measures cannot be used with low frequency disclosure policies. They would be unable to capture intra-quarter trading in mutual funds that creates significant value (Puckett and Yan (2011), Kacperczyk, Sialm, and Zheng (2008)).

The assessment of risk also depends on the frequency with which holdings are disclosed. Low frequency can mislead investors about the risk and risk shifts of funds. Elton et al. (2010), show, using monthly data, that high-return (low-return) funds tend to increase (decrease) risk in the second part of the reporting year. Thus, more frequent disclosure can help estimate risk shifts in the last five months of the year. They were able to properly estimate the risk of the portfolio using holdings to calculate more accurately the different risk measures like *beta* or standard deviation.

2.3 Impact on other mutual funds

The frequency of disclosure of a mutual fund's holdings may also have an impact on how other funds will be able to produce external manipulations. In particular, copycat and front-running are well studied in the literature.

Copycat consists in using the portfolio holdings of mutual funds in order to create a buy-and-hold strategy with these assets until the next disclosure, removing all need for internal research. Frank, Shackelford, Poterba, and Shoven (2004) show that the returns after expenses for copycat funds are statistically indistinguishable from the ones of high expense funds in the USA in the 1990's. They constructed a fund that uses a copycat strategy as follows: after each disclosure of a high expense fund, the copycat fund buys the exact same portfolio and holds it until the next disclosure date. To meet the allowed lag between the snapshot date and the publication one, they decided to perform this construction exactly 60 days after the snapshot. To compare the expense-adjusted returns, they assumed that the costs of copycat funds are similar to those of index funds because they don't have to do any research to perform their assets selection. Thus, even though their gross returns are lower than those of high expense funds, when costs are taken into account by comparing returns after expenses, the performance is similar and indistinguishable. In addition, they also found that reducing the frequency of disclosure improves the performance of copycat funds because it reduces the cumulative expenses. This last statement is somewhat surprising because Verbeek and Wang (2013) explain that the 2004 regulation change in the USA from semi-annual to quarterly reports increased the positive difference between the expense-adjusted returns of copycat funds and their targets (copycats outperform more the copied funds) and that this reduced the volatility of this difference. Thus, it should be easier for a copycat fund to free-ride an active fund with an increased frequency of holdings disclosure.

Front-running is considered as another drawback of holdings disclosure. It refers to the buying (selling) by other traders or market makers in anticipation of the fund buying (selling). The other traders or market makers are then able to raise (lower) the price at which the fund will buy (sell) their shares and thereby reduce their gross return. Dyakov and Verbeek (2013) managed to create a successful front-running strategy using public data. Since investors are eager to find best performing funds, they will therefore move from low-performing funds to high-performing funds, creating money flows. The authors first used the past performance of a fund to estimate inflows (outflows). As an inflow (outflow) is likely to make the fund increase (reduce) the quantity of some assets in its possession, they used the disclosed holdings to anticipate upward (downward) price pressure, i.e. they anticipate whether the fund will buy or sell many of assets. Using this information, they can now abuse distressed funds that have had recent outflows and are performing "fire-sales" to return money to their investors. One strategy to do this is to short sell the assets that will be sold by the distressed fund. Since the fund is likely to sell a large amount of that asset, the price will fall and generate return for the short seller. However, this front-running strategy is not necessarily harming the distressed fund as in some cases, it may have no impact on it.

2.4 Impact on the companies mutual funds invest in

The holdings' disclosure frequency can have an impact on the way the fund manager will invest.

Portfolio disclosure frequency raises many questions, including its relationship to performance or risk-taking for example. Li et al. (2022), study U.S. active equity funds to highlight the correlations between funds' portfolio disclosure policies and other of their characteristics. He points out that such disclosure is negatively correlated with risk-taking and liquidity. Indeed, funds that disclose more often tend to invest in less risky and liquid assets.

In addition,the increased frequency of holdings disclosure may raise some agency issues. In fact, as managers have to disclose more frequently, they seek to disclose good returns over the disclosure period and tend to buy stocks that offer good short-term returns. Agarwal, Vashishtha, and Venkatachalam (2018) show that managers have increased their portfolio turnover and reduced their average position duration to disclose portfolios with recent high returns stocks. This change in management strategy could be explained by the fact that asset managers, mainly young ones, are more concerned with their careers than with the final interest of investors. Moreover, the authors focus on the 2004 regulation change in the USA. They try to estimate the impact of this change on the innovation of companies through their patent production. They show that the increase in the frequency of holdings disclosure has caused the portfolio manager to focus more on short-term returns and thus on less innovative firms which can be considered as a negative externality.

2.5 Impact on regulators' way to control funds

Another important effect of frequent holdings disclosure is the reduction of portfolio manipulations aimed at improving the reported performance of funds, such as window dressing or portfolio pumping. These practices can be referenced as internal manipulations, and the monitoring of such manipulations could be useful to regulators.

Window dressing is the process of improving the appearance of a portfolio and its performance before showing it to clients or shareholders. For example, it can be done by buying (selling) recent winners (losers) in order to disclose a portfolio with mainly assets with good past performance. This manipulation exists in many markets such as USA (Agarwal, Gay, and Ling (2014)), Ortiz, Ramírez, and Sarto (2013)) or Taiwan (Hung, Lien, and Kuo (2020)), for example. Although this phenomenon is not a common practice in Spain, it does exist, especially during bear market periods, probably in response to past poor performance as explained by Ortiz et al. (2013). Hung et al. (2020) show that funds involved in window dressing tend to buy winners (sell losers) at high (low) prices, which hurts investors' wealth. They also explain that funds that engage in window dressing are more likely to do so with the portfolio they will present to clients or potential investors, i.e. in the annual report.

Portfolio pumping consists of artificially improving the performance of the assets held by the fund by buying them massively before the disclosure and selling them afterwards. This manipulation exists in different markets. In the USA, Duong and Meschke (2020) show evidence of portfolio pumping prior to annual reports for all funds and prior to quarterly reports especially for small cap funds. However, this phenomenon has diminished with the increased attention of regulators on this topic. In addition, Ouyang and Cao (2020) study portfolio pumping in China and show that one-third of the top holdings of the 2,000 funds studied experienced return reversals[10] which is characteristic of portfolio pumping. They also highlight evidence of "Sedan chair carrying" practice, consisting in using multiple funds of an asset management company in order to pump one "star fund" and thus increase its performance.

Thus, a change in regulation and especially an increase in the frequency of disclosure can reduce those types of portfolio manipulations. Verbeek and Wang (2013) show that the SEC regulation change of 2004 in the USA have increased the representativeness [11] of disclosed holdings and thus reduced internal manip-

[10] Return reversals consists in a switch in the trend of an asset's price. For example, we can talk of return reversal for a stock which price was increasing rapidly and suddenly starts decreasing.

[11] The representativeness measures how much the disclosed holdings represent the fund true strategy. It is computed as the tracking error between the reported fund returns and the

ulations such as window dressing or portfolio pumping.

3 Inverse problem statement

As shown in section 2, the lack of information between the holdings disclosure dates can lead to manipulations and misunderstanding of the performance and risk of the fund. This article then aims to create a method that can assess the portfolio composition of a fund for each day between the holdings regulatory disclosures. Indeed, punctual buy and sell actions of the fund manager lead to new initial weights of the portfolio which need to be assessed using the only available information, i.e. the asset returns, the portfolio return and some periodical publications of the portfolio composition.

3.1 Dynamics of holdings in a portfolio

Let T be the temporal horizon of the study, N_U be the size of the universe in which the portfolio can be constructed and N the size of the portfolio, which is unknown a priori.

The idea is to assess the weights $x(t) = (W_1(t), W_2(t), ..., W_{N_U}(t))^T$ of the portfolio for each day $t \in [0, T]$. The weights have to respect the following constraint:

$$\sum_{i=0}^{N_U} W_i(t) = 1, \forall t \in [0, T] \tag{1}$$

If the portfolio doesn't change due to a buy or sell action of the fund manager and no dividend is detached and reinvested by the fund manager, the weights of the portfolio will evolve according to the returns $(R_i(t))_{i \in [0, N_U]}$ of the assets at time t as follows:

$$W_i(t+1) = \frac{(1 + R_i(t))W_i(t)}{1 + \sum_{j=1}^{N_U} R_j(t)W_j(t)} \tag{2}$$

In order to keep the fraction definite, let's assume that the portfolio and the assets didn't lost all it's value overnight, i.e. $\forall t \in [0, T], \sum_{j=1}^{N_U} R_j(t)W_j(t) = R_P(t) \neq -1$.

However, the computed portfolio returns $\sum_{j=1}^{N_U} R_j(t)W_j(t)$ can be replaced by the observed portfolio returns $R_{P,obs}(t)$ as it is noiseless, which is a useful simplifi-

returns on a characteristics-based benchmark.

cation in terms of solving the inverse problem. The previous weights update is then linear in $W_i(t)$.

$$W_i(t+1) = \frac{(1+R_i(t))}{1+R_{p,obs}(t)} W_i(t) = A_i(t, R_{p,obs}(t)) W_i(t) \tag{3}$$

The following state-space representation can then be introduced:

$$x(t+1) = A(t, y_{obs}(t)) x(t) \tag{4}$$

$$y(t) = C(t) x(t) \tag{5}$$

where $A(t, y_{obs}(t)) = diag(A_1(t, y_{obs}(t)), ..., A_{N_U}(t, y_{obs}(t)))$, $y(t) = R_{p,obs}(t)$ and $C(t) = (R_1(t), ..., R_{N_U}(t))$ for each time step $t \in [0, T]$.

3.2 Observability analysis

The problem can be reduced to determining the state $x(t)$ given the observations $Y_{obs}(t, t+T) = (y_{obs}(t), ..., y_{obs}(t+T))$, the returns of the N_U assets of the universe $(C(j))_{j \in [t, t+T]}$ and the portfolio weights at the publication dates in the period $[t, t+T]$ $(x(t_{pub}))$.

Firstly, after a publication, as the exact portfolio composition is known, the weights update formula (3) can be used to determine the portfolio weights for each time step until the next change. A change in the composition can then be detected by comparing the predicted return of the portfolio with the measured one.

Let's assume for the next part that no portfolio holdings have been published and no changes occur during the period between t and $t+T$. If $T+1$ portfolio return observations are available in order to build a vector of size N_U and using the state-space representation (3), the estimation problem can be written under the following linear system:

$$\begin{pmatrix} y_{obs}(t) \\ y_{obs}(t+1) \\ \vdots \\ y_{obs}(t+T) \end{pmatrix} = \begin{pmatrix} C(t) I_N \\ C(t+1) A(t, y_{obs}(t)) \\ \vdots \\ C(t+T) A(t+T-1, y_{obs}(t+T-1)) ... A(t, y_{obs}(t)) \end{pmatrix} x(t) \tag{6}$$

$$Y_{obs}(t, t+T) = X(t, t+T) x(t) \tag{7}$$

where $x(t)$ is the composition of the portfolio to be assessed at time step t. The matrix $X(t, t+T)$ is squared if $T = N_U - 1$. If the matrix $X(t, t+T)$ is invertible, $x(t)$ is given by $X(t, t+T)^{-1} Y_{obs}(t, t+T)$. In order for $X(t, t+T)$ to be invertible, it needs to be full rank i.e. all columns need to be independent. In Georges and Girerd-Potin (2017), the authors have shown that full rank is guaranteed under mild conditions, mainly, the asset returns $R_i(t)$ are not equal to zero, $\forall t$, and the returns are distinguishable. If $T > N_U - 1$, a least-square technique can be used to compute $x(t)$, which also requires that the full rank condition is fulfilled.

4 Models and algorithms

In order to determine the composition of the portfolio at each time step in the studied period using only the publicly available data (asset and portfolio returns and some periodical compositions of the portfolio), several methods have been tested or proposed: the direct least-square approach, and, based on a reformulation of the estimation problem, two heuristic approaches, a Sequential Oscillating method (SOS) and a Genetic Algorithm (GA), associated to a rolling window method.

4.1 Direct least-square approach

The matrix inversion procedure described in section 3.2 can be generalized as a least-square problem, when $T > N_U - 1$. Indeed, the problem can be stated as finding the vector $x(t)$ that minimizes the least-square error:

$$\min_{x(t)} \|Y(t, t+T) - Y_{obs}(t, t+T)\|^2 \tag{8}$$

where $Y(t, t+T) = X(t, t+T)x(t)$ and $\|.\|$ denotes the L_2 norm.
Following the analysis provided in section 3.2, the solution of the problem (8) is given by a classical pseudo-inverse formulation:

$$x(t) = (X(t, t+T)^T X(t, t+T))^{-1} X(t, t+T)^T Y_{obs}(t, t+T). \tag{9}$$

if the problem is not well-posed in the sense that $T < N_U - 1$ (there are less portfolio returns than the size of the universe N_U), a regularization technique is needed to overcome the rank loss. A classical technique will consist in adding a regularization term to the formulation (8):

$$\min_{x(t)} \|Y(t, t+T) - Y_{obs}(t, t+T)\|^2 + \lambda_R \|x(t)\|^2 \tag{10}$$

where $\lambda_R > 0$ is the regularization coefficient to be tuned in order to get the best estimate. Then, the solution of (10) is given by:

$$x(t) = (X(t, t+T)^T X(t, t+T) + \lambda_R I_{N_U})^{-1} X(t, t+T)^T Y_{obs}(t, t+T) \tag{11}$$

where I_{N_U} denotes the $(N_U \times N_U)$ identity matrix.

In practice, T should be close enough to $N_U - 1$ to allow a meaningful estimate of the vector $x(t)$.

4.2 A new formulation to overcome practicability issue

In practice, if the size of the asset universe N_U is very large, the direct use of both the direct least-square approach is impracticable due to the observability condition that $T \geq N_U - 1$. For example, for a universe of 1800 assets, a total of 1800 time steps, at least, is needed, i.e. approximately 5 years of daily returns data containing no manager's changes for the methods to converge. For that reason, it is mandatory to consider reformulating the portfolio composition estimation problem as a large-scale combinatorial optimization problem in order to reduce the observability horizon constraint $T \geq N_U - 1$ to $T \geq N - 1$: Under the assumption that the maximum number of assets in the portfolio is known a priori, the estimation of the composition of a N-asset portfolio in a large universe of $N_U (N_U >> N)$ can be reformulated as a NP-hard mixed-integer optimal estimation problem of the form:

$$\min_{W_l(t), l \in I_p, \alpha_i \in \{0,1\}, i=1,\dots,N_U} \sum_{l=t}^{t+T} (y(l) - y_{obs}(l))^2 \tag{12}$$
$$+\lambda_R \|W\|^2, \ \lambda_R > 0, \ T \geq N - 1$$

subject to

$$W_i(l+1) = \frac{(1 + R_i(l))W_i(l)}{(1 + y_{obs}(l))}, \forall i \in I_p = \{i/\alpha_i = 1\} \tag{13}$$

$$y(l) = \sum_{j=1}^{N_U} \alpha_j R_j(l) W_j(l) \tag{14}$$

$$\sum_{j=1}^{N_U} \alpha_j \leq N \tag{15}$$

with α_i = integer affiliation variable (if the asset i belongs to the portfolio, $\alpha_i = 1$, otherwise $\alpha_i = 0$) and λ_R is the regularization coefficient. Constraint (15) is introduced to impose that the optimal estimated portfolio contains at most N holdings.

Of course, a direct enumeration technique (a brute force approach), which needs to make use of the approach described in 4.1, cannot be used to solve this problem since the number of combinations of possible smaller portfolios can be hugely large. For example, for a portfolio of 30 assets in a universe of only 100 assets, there are already 29×10^{24} possible portfolios.

In this paper, we propose to use heuristic approaches based on an optimal selection algorithm and a genetic algorithm in order to try to reduce the time steps needed to converge. These heuristic approaches will rely on the direct least-square approach (4.1) and will be performed on the sub-universe of N assets with $\alpha_j = 1$.

4.3 Heuristic methods

The idea behind using heuristic methods is to use less assets at one time. This can help to reduce the size of the universe of possible assets. Both methods are based on the direct least-square method and are aimed to find the portfolio that reduces the most the Root Mean Square Error (RMSE): $\sqrt{\frac{1}{T}\sum_{j=1}^{T}(y(t+j) - y_{obs}(t+j))^2}$.

4.3.1 Sequential Oscillating Selection method (SOS)

The Sequential Oscillating Selection method has been used first by Byrd et al. (2019).To perform the selection of assets inside the portfolio, the returns of the portfolio, the returns of the assets and the previous direct least-square approach will be needed.

The idea is to build the portfolio which is the closest in terms of returns to the ones of the portfolio we want to assess. To do so, the portfolio that reduces the most the RMSE between the created portfolio returns and the assessed ones will be chosen.

The steps of this algorithm are the following:

- Initially, the portfolio is assumed to be empty
- Add to the portfolio the asset of the universe that reduces the most the RMSE between the returns of the asset and the ones of the assessed portfolio.
- Add assets to the portfolio while it reduces the RMSE. Each asset will be added to the portfolio following the next steps:
  - For each asset left in the universe, add it temporarily to the portfolio
  - Use the direct least-square approach to determine the optimal initial state following the current portfolio
  - Use the asset returns to generate the constructed portfolio returns
  - If the RMSE between the constructed portfolio returns and the assessed ones is the lowest of all the possible assets in the universe and that it reduces the RMSE from the previous state, this asset is chosen and added permanently to the portfolio
  - Repeat those steps as long as adding another asset reduces more the RMSE
- Remove assets from the constructed portfolio to try to reduce more the RMSE. The asset to remove will be chosen as follows:
  - For each asset in the portfolio, remove it temporarily from the portfolio

  - Use the portfolio model to determine the optimal initial state following the current portfolio
  - Use the asset returns to generate the constructed portfolio returns
  - If the RMSE between the constructed portfolio returns and the assessed ones is the lowest of all the possible assets in the portfolio and that it reduces the RMSE from the previous state, this asset is chosen and removed permanently from the portfolio
  - Repeat those steps as long as removing another asset reduces more the RMSE

- Repeat again the adding and removing part until it stops improving the RMSE

At each step of the algorithm a stopping criteria is tested. Indeed, if the RMSE of the current portfolio is lower than 1.10^{-14}[12] and the current portfolio weights sum to 1, the algorithm is stopped and the current portfolio is considered as the final one.

4.3.2 Genetic Algorithm (GA)

This method is an adapted version of the classical genetic algorithms used to perform feature selection as introduced in Siedlecki and Sklansky (1989).

Genetic algorithms base idea is to create a population of individuals (in our case, an indiidual represents a sub-universe of assets represented as a binary vector in which 1 means the asset is in the portfolio and 0 means it is not) that will be tested and will evolve through generation in order to get close to the optimal solution. Each of the individual will be tested and scored using the direct least-square method and the following scoring method:

$$score = \frac{||R_P - R_{P,obs}||^2}{||R_{P,obs}||} + Kmax(0, \sum_{i=1}^{N_U} \alpha_i - N) \tag{16}$$

with R_P the returns of the assessed portfolio, $R_{P,obs}$ the returns of the true portfolio and K a coefficient set to 0.005.

After testing each individual of the current population, a crossover method and some mutations will be applied to the genes of the population in order to create a new generation that will be tested and scored. This crossover will be done until the maximum generations number is achieved or if the RMSE score between the assessed portfolio and the tested one is smaller than 1.10^{-14} and the portfolio weights sum to 1.

After each individual of the population is evaluated, some of them are chosen according to their score with the roulette-wheel selection(Goldberg (2013))

[12] Represents the machine zero

method to be parents for the new generation. A crossover will be performed on pairs of parents to generate a pair of children. This crossover method is constructed to try to remove stocks that are probably not in the true portfolio. The i-th element of the first child equals 1 only if the i-th element of both the first and second parent equals 1 while the i-th element of the second child equals 1 if one of the i-th element of the two parents equals 1. Some mutations of single elements of each child is then performed to create the new generation.

4.4 Rolling window method

The proposed heuristics methods are based on the direct least-square approach, thus changes in the composition of the portfolio by the portfolio manager can't be determined over a period. A method that can deal with these changes needs to be used.

The first idea to detect a change is to compare at each day the returns of the last known portfolio to the ones of the observed portfolio. Indeed, if the assessed portfolio is exact, then the returns for the t-th day of period T of both the assessed and the true portfolio should be the same. If it's not the case, this means that the portfolio has changed, then a new portfolio should be determined.

Mutual funds have to publicly publish regularly their portfolio holdings. Let's assume that the studied period starts from a publication date (t_{pub}) and has no other publication date. Indeed, the period between two publication dates is the interesting one to study as it is the period without information.

This publicly available data will be used in the rolling window method as the initial known holdings. The weights update model will then be used to update those holdings and compute, for each timestep the return of the portfolio. As long as the return of the updated portfolio match the one of the observed portfolio, the holdings are still updated for the next timestep. If the return doesn't match, then the hypothesis that a change in the portfolio occured is made, thus, a new portfolio needs to be determined, the Genetic algorithm or Sequential oscillating selection is then used.

The following steps are then reproduced for each timestep t of the period T:

- Update the last known portfolio with the asset returns and the weights evolution model
- Compute the return of the portfolio at t
- If the return of the current portfolio is the same as the observed one for t
  - Set the last known portfolio to the updated one
  - Go to timestep $t+1$

- If the return of the current portfolio is not the same as the observed one for t
  - Perform the GA or the SOS in order to retrieve a portfolio
  - Set the last known portfolio to the one given by the method
  - Go to timestep $t + 1$

5 Case study

In order to test the methods, a case study is proposed based on market data, i.e. asset returns, the returns of a fund and disclosed holdings.

5.1 Market data

The needed asset prices have been fetched from the Refinitiv Eikon Datastream [13] database. The idea is to build a universe in which portfolios will be constructed and will evolve. To do so, CAC40 asset prices have been fetched between 1st January 2015 and 31st December 2021. A total of 48 assets that have been or are still in the CAC40 and have no missing data over the period have been selected to build the universe.

As returns were needed to work with the weights update of the portfolios of the funds, the returns of those asset prices were computed.

Assets have been selected from the CAC40 to match with the 6-month periods without disclosure used to test the previously explained methods.

5.2 Preliminary tests

The first step to use the rolling window method is to determine a minimal size of the window needed for the methods to converge depending on the size N of the portfolio.

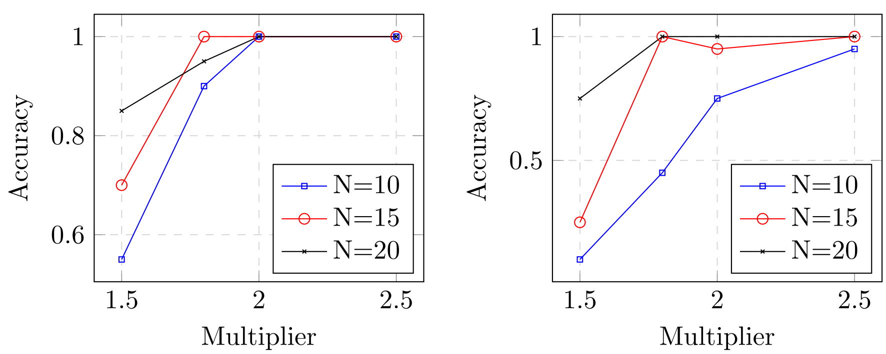
Twenty random portfolios of size $N = 10$ in a universe of size $N_U = 48$ with no change have been created over a 1 month period in January 2021. The heuristic methods (Genetic Algorithm and Sequential Oscillating Selection) have been tested on those portfolios in order to determine the minimal size T_{min} of the window needed for the methods to converge to a result. Theoretically, T_{min} depends on the size N of the portfolio, thus the studied window sizes are constructed as $\beta \times N$ with β, a multiplier, $\in \{1.5, 1.8, 2, 2.5\}$.

The result of a studied method is considered as true if the difference in terms of weights between the assessed initial portfolio and the true initial portfolio is

[13] https://www.refinitiv.com/en/products/datastream-macroeconomic-analysis

lower than $1e-14$[14]. The computed accuracy is then the proportion of tests that ran successfully over the 20 tested portfolios.

The Genetic Algorithm used to perform those tests is based on 1000 generations and a population of size 50 with 5 mutations per crossover child. The population used is reduced in order to limit the computation time. For both the SOS and GA methods, the λ_R regularization coefficient of the direct least-square approach has been set to $1e-5$ to avoid singular matrices problems.



(a) Accuracy of the GA method

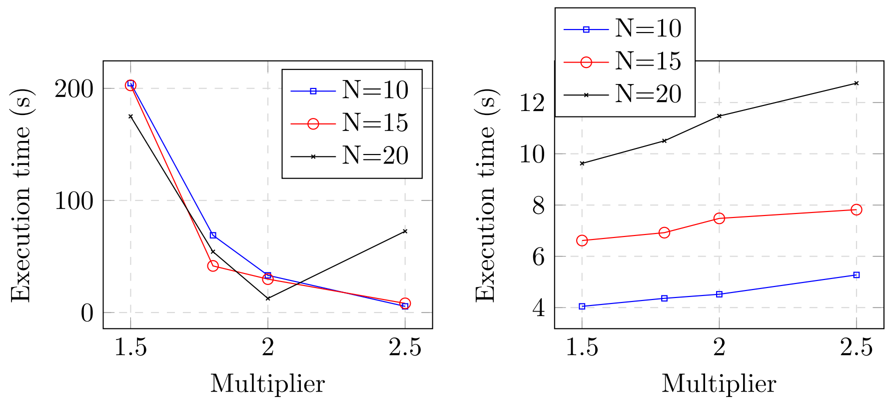
(b) Accuracy of the SOS method

Figure 2: Accuracy comparison of the Genetic Algorithm and Sequential Oscillating Selection methods over randomly created portfolios with no changes

As shown in fig. 2, the Genetic Algorithm has a better accuracy for $N = 10$ over smaller time windows. The sequential oscillating selection method seems to have the same results as the Genetic Algorithm for bigger portfolios. The minimal size of the study window for the SOS method seems not to be linear with the size of the portfolio whereas the Genetic Algorithm seems to converge for a multiplier $\beta = 2$ whatever the size N of the portfolio. Thus, the Genetic Algorithm seems to have better accuracies for smaller portfolios with small studied time window.

On the other hand, the fig. 3 depicts the mean execution time of both the methods for the same portfolios. It can clearly be seen that the SOS method is much faster to execute for small and large portfolios. However, the execution time of the Genetic Algorithm seems to decrease greatly when the multiplier increases. Indeed, the Genetic Algorithm execution time depends a lot on the number of generations needed to converge. If the data given is not sufficient, it will take more time to converge and then run for more generations. In addition, the execution time of the Genetic Algorithm seems to stay the same, or even decrease when the size of the portfolio increases. Indeed, when N increases,

[14] Representing the machine zero



(a) Execution time of the GA method (b) Execution time of the SOS method

Figure 3: Comparison of the execution time of the Genetic Algorithm and Sequential Oscillating Selection methods over randomly created portfolios with no changes

with N_U fixed, the number of possible α vectors is reduced, making the Genetic Algorithm more efficient.

Finally, the Genetic Algorithm seems to have better results in terms of accuracy with sufficient number of generations and population even if it takes more time to run. Indeed, a window with a multiplier of 2 seems to be sufficient for the Genetic Algorithm whereas one of 2.5 seems to be necessary for the sequential oscillating selection method. Those previous multipliers of time window sizes will be used for the next examples.

5.3 Artificial funds' portfolio construction

The rolling window method will be tested on artificial funds' portfolios that were created in order to know their exact composition at each date of the studied period and simplify the disclosure data fetching.

Simulated portfolios have been created over 6 month periods from July 2015 to July 2021. Indeed, those portfolios have been constructed over 6 month periods in order to mimic the periods without information in the official European regulation.

The simulated portfolios have 1, 2 or 5 evenly spread manager's changes [15] over the period. For example, in the portfolio with 5 changes, the composition of the fund portfolio will change at the beginning of each month, i.e. the change

[15] A manager's change occurs when the fund manager changes the composition of the fund portfolio by buying or selling stocks.

frequency is one month, while in the portfolio with 2 changes, a change will occur only at the beginning of the third month and the fifth one (the change frequency is then 2 months). Finally, for the portfolio with only 1 change, it will occur at the beginning of the fourth month (the change frequency is then 3 months). They've been constructed following the momentum strategy i.e. at each change date the portfolio is constructed using the ten assets with the best return over the last month. The size of those portfolio is set to ten.

The constructed portfolios are summarized in table 1:

Name	Size	Changes freq.(months)	Construction method
Momentum1	10	1	Momentum
Momentum2	10	2	Momentum
Momentum3	10	3	Momentum

Table 1: Constructed scenarios

For each of the methods, the weights of the first day will be used as the disclosed holdings for the period.

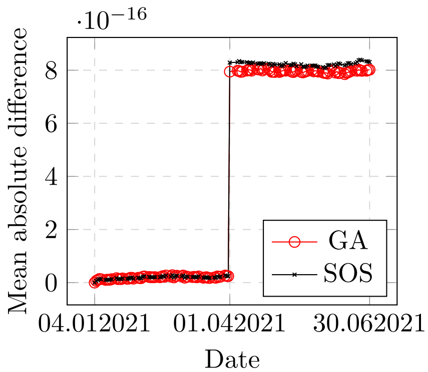
5.4 Results

The comparison of the two methods have been performed using the mean absolute difference between the assessed portfolio weights and true portfolio weights given by $\left(\frac{\sum_{t=1}^{M}\sum_{i=1}^{N_U} abs(W_{assessed,i,t} - W_{true,i,t})}{MN_U}\right)$.

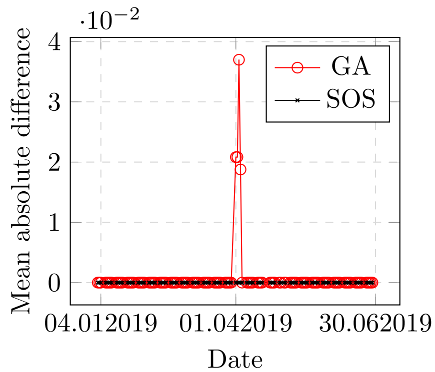
The comparison have been done on 6-month periods from the 1st of July of 2015 to the 30th of June of 2021. For clarity purposes, graphs represent the last period, from the 4th of January of 2021 to the 30th of June of 2021, which is representative of the rest of the studied period, and some interesting periods to comment. The rolling window methods will be denoted as SOS and GA in the graphs and the comparative table, depending on the feature selection method used to assess the portfolio when it changes.

The multipliers used for the size of the time window are the same as mentioned before: 2.5 for the SOS and 2 for the GA and the parameters of the GA are the same as the ones used for the preliminary tests (1000 generation with a population of 50 individuals).

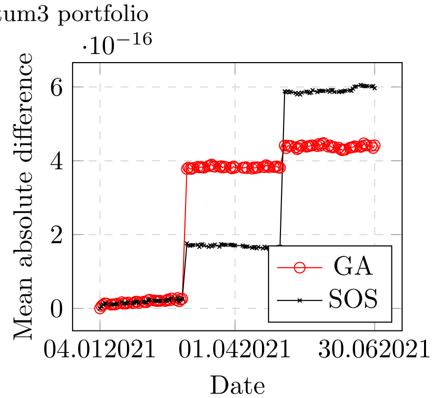
As shown in fig. 4 both the rolling window with SOS and GA converges to the true weights and determine the exact dates of portfolio changes. The performance of both methods are similar. Indeed, the assessed portfolio is exactly the same as the true one when the machine precision is taken into account except



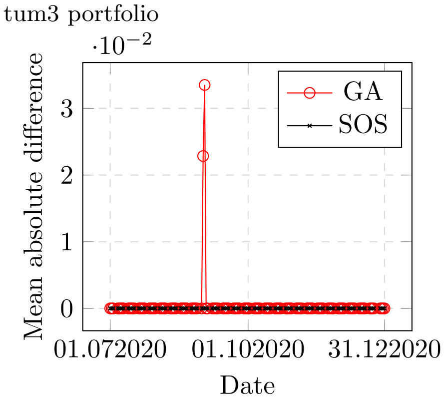
(a) Comparison of the models on the January-June 2021 period for the Momentum3 portfolio



(b) Comparison of the models on the January-June 2019 period for the Momentum3 portfolio



(c) Comparison of the models on the January-June 2021 period for the Momentum2 portfolio

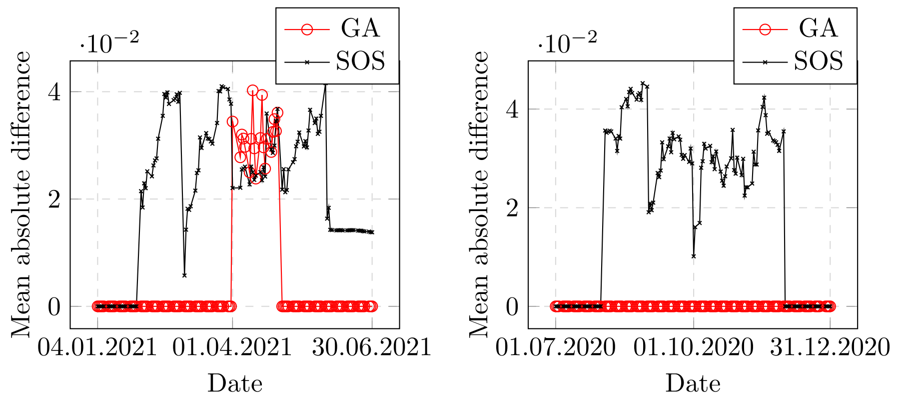


(d) Comparison of the models on the July-December 2020 period for the Momentum2 portfolio

Figure 4: Comparison of the Rolling window with Genetic Algorithm (GA) and Rolling window with Sequential oscillating selection (SOS) methods for Momentum3 and Momentum2 portfolios over 6-month periods

for some specific periods with the rolling window with GA. In addition, fig. 4b and fig. 4d shows that sometime the GA algorithm doesn't converge on the exact change date but some time steps later. In those examples, no problem is encountered as the period without change T is much bigger than the size of the window used for GA $\beta \times N_U = 2 \times 10$ and the method converges later. Those non converging dates can be counted as false positive changes detected by the method.

As shown on fig. 5, too many changes over the periods can result in not exact solutions. The rolling window with SOS method has a time window of



(a) Comparison of the models on the January-June 2021 period

(b) Comparison of the models on the July-December 2020

Figure 5: Comparison of the Rolling window with Genetic Algorithm (GA) and Rolling window with Sequential oscillating selection (SOS) methods for Momentum1 portfolios over 6-month periods

$T_{min,SOS} = 25$ days (2.5×10) which is greater than the period without changes (one month i.e. approximately 20 days). Thus, it cannot converge and find the true portfolio over the period without changes[16]. On the other hand, the rolling window with GA manages to converge most of the time to the exact portfolio. This last has not been able to converge on some periods, as shown in the fig. 5a which could be tackled by increasing the size of the population and/or the number of generations of the GA. However, if the size $T_{min,GA} = 20$ is greater than the period without changes, the method won't be able to converge either.

In addition, the rolling window methods can retrieve the date at which the manager changed the portfolio in the period. The dates have been correctly detected for the portfolio with only 1 change. Indeed, the two rolling window methods detected the 1st of April which is exact. For the portfolios with two changes, they occurred on the 1st of March and the 3rd of May, which have been detected by the models. As mentioned above, the rolling window with GA also detected some false positive change dates in addition to the true change dates. Finally, for the portfolios with 5 changes, as can be seen in the fig. 5, the rolling window with SOS couldn't converge nor detect the exact change dates most of the time. Indeed, apart from the first change which occurred on the 1st of February and sometimes the last one (the 1st of December, 2020 in fig. 5b), no other true change date has been detected. The rolling window with GA method came up with all the true change dates and also false positive change dates over some periods (as shown with the period in fig. 5a where the method detects the true

[16]The first period rely on the disclosed portfolio, thus the mean absolute difference is close to 0 as no SOS method have been used yet.

change date, the 1st of April, 2021 and then some false positive ones until the next true one, the 3rd of May 2021).

Nb. of changes per 6-month period	Method	Mean	Median	Accuracy
1	SOS	2.27963e-16	6.60641e-17	1
1	GA	6.34203e-05	6.60641e-17	1
2	SOS	4.87619e-16	1.90241e-16	1
2	GA	3.66943e-05	3.87711e-16	1
5	SOS	1.96723e-2	2.41021e-2	0.17
5	GA	7.81033e-4	5.81711e-16	0.97

Table 2: Comparative table of the precision of the methods according to the change frequency over all the periods from 07-2015 to 06-2021

Finally, the table 2 summarizes statistics about the precision (computed as before by the mean absolute difference between true and assessed weights of the different methods) according to the change frequency over the whole studied period (July 2015 to June 2021) and the accuracy. The accuracy has been computed as the number of periods for which the true portfolio has been successfully retrieved (the formerly mentioned false positive change dates are counted as a successful retrieving period if the method converges to an exact portfolio over the period without changes). The 6-month periods for Momentum1 portfolio are then divided in 5 periods with no changes to assess, 2 periods for Momentum2 and 1 for Momentum3. It can be seen that the accuracy for the rolling window with GA is much more accurate for the Momentum1 portfolio assessment than the rolling window with SOS. The mean value of the rolling window with GA is also higher than the one with SOS due to the previously mentioned falsely detected change dates for which the GA hasn't converged. However, the median is still very small showing that those errors are pretty rare.

6 Conclusion

The literature shows that the disclosure frequency of the composition of mutual funds remains debated with arguments for and against increasing this frequency. We claim that it is in the interest of investors and regulators to know the composition of mutual funds on a daily basis in order to better control the actions of fund managers and to avoid manipulations on their part. A better assessment of the performance, the risk and the respect of the funds' objectives is expected.

In this paper, we have shown that some feature selection methods can be used in order to discover the exact holdings of mutual funds for periods without disclosure and compared two of them, Genetic Algorithm and Sequential Oscillating

Selection, in terms of performance and computation time. We've also found that the Genetic Algorithm requires a smaller time window in order to converge than the Sequential Oscillating Selection. Moreover, we've developed two methods: the rolling window methods (with GA and SOS) that succeed in discovering those holdings, even with manager's voluntary changes, and detecting the dates of those changes, provided that the periodicity of changes is compatible with the constraint on the size of the observation period.

This paper is a proof of feasibility on simulated examples with small universes, small portfolios and some low frequency changes in the portfolio composition. The use of the Rolling window methods on real mutual funds and bigger universes will be a great test in order to discover the limits of the proposed methods.

Some further researches could focus on only determining the changes dates and the assets involved in those changes in order to reduce the number of observations needed. In addition, a better knowledge of the asset managers' strategies could help to improve and adapt the previous methods.

Finally, those methods could be used in order to detect some internal manipulations performed by mutual funds managers as window dressing or portfolio pumping.

References

Agarwal, V., Gay, G. D., & Ling, L. (2014, 11). Window dressing in mutual funds. *Review of Financial Studies*, *27*, 3133-3170. doi: 10.1093/rfs/hhu045

Agarwal, V., Vashishtha, R., & Venkatachalam, M. (2018, 5). Mutual fund transparency and corporate myopia. *Review of Financial Studies*, *31*, 1966-2003. doi: 10.1093/rfs/hhx125

Byrd, D., Bajaj, S., & Balch, T. H. (2019). Fund asset inference using machine learning methods: What's in that portfolio? *The Journal of Financial Data Science*.

Chen, Z., Gallagher, D. R., & Lee, A. D. (2017, 3). Testing the effect of portfolio holdings disclosure in an environment absent of mandatory disclosure. *Accounting and Finance*, *57*, 113-129. doi: 10.1111/acfi.12129

Choi, K.-S., Anderson, S., Baik, B., Berfeld, N., Core, J., Silva, T. D., ... Yoon, R. (2022). *The role of portfolio disclosures in mutual funds *.*

Duong, T. X., & Meschke, F. (2020, 2). The rise and fall of portfolio pumping among u.s. mutual funds. *Journal of Corporate Finance*, *60*. doi: 10.1016/j.jcorpfin.2019.101530

Dyakov, T., & Verbeek, M. (2013, 12). Front-running of mutual fund fire-sales. *Journal of Banking and Finance*, *37*, 4931-4942. doi: 10.1016/j.jbankfin.2013.08.013

Elton, E. J., & Gruber, M. J. (2020, 7). A review of the performance measurement of long-term mutual funds. *Financial Analysts Journal*, *76*, 22-37. doi: 10.1080/0015198X.2020.1738126

Elton, E. J., Gruber, M. J., Blake, C. R., Krasny, Y., & Ozelge, S. O. (2010, 5). The effect of holdings data frequency on conclusions about mutual fund behavior. *Journal of Banking & Finance*, *34*, 912-922. doi: 10.1016/J.JBANKFIN.2009.10.002

Frank, M. M., Shackelford, D. A., Poterba, J. M., & Shoven, J. B. (2004, 10). Copycat funds: Information disclosure regulation and the returns to active management in the mutual fund industry. *Journal of Law and Economics*, *47*, 515-541. doi: 10.1086/422982

Georges, D., & Girerd-Potin, I. (2017, 7). A discrete-time state observer approach to discovering portfolio holdings. *IFAC-PapersOnLine*, *50*, 946-951. doi: 10.1016/j.ifacol.2017.08.138

Gimeno, R., Ortiz, C., & Sarto, J. L. (2021). Mutual fund voluntary portfolio disclosure. *Emerging Markets Finance and Trade*, *57*, 1474-1488. doi: 10.1080/1540496X.2019.1629284

Goldberg, D. E. (2013). *Genetic algorithms.* Pearson Education India.

Gregory-Allen, R., Balli, H. O., & Thompson, K. (2019, 10). The impact of portfolio holdings disclosure on fund returns. *Pacific Basin Finance Journal*, *57*. doi: 10.1016/j.pacfin.2019.101172

Grinblatt, M., & Titman, S. (1989). Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of Business*. doi: 10.1086/296468

Hung, P. H., Lien, D., & Kuo, M. S. (2020, 11). Window dressing in equity mutual funds. *Quarterly Review of Economics and Finance*, *78*, 338-354. doi: 10.1016/j.qref.2020.05.003

Kacperczyk, M., Sialm, C., & Zheng, L. (2008, 11). Unobserved actions of mutual funds. *Review of Financial Studies*, *21*, 2379-2416. doi: 10.1093/rfs/hhl041

Li, J., Armstrong, C., Bushee, B., Evans, R., Goldstein, I., Guay, W., … Wermers, R. (2022). Mutual funds' discretionary portfolio disclosure policies the author sincerely thanks. *SSRN*.

Ortiz, C., Ramírez, G., & Sarto, J. L. (2013, 7). Assessment of window dressing using fund returns and portfolio holdings. *Spanish Review of Financial Economics*, *11*, 85-93. doi: 10.1016/j.srfe.2013.07.002

Ouyang, L., & Cao, B. (2020, 9). Selective pump-and-dump: The manipulation of their top holdings by chinese mutual funds around quarter-ends. *Emerging Markets Review*, *44*. doi: 10.1016/j.ememar.2020.100697

Puckett, A., & Yan, X. S. (2011). The interim trading skills of institutional investors. *Journal of Finance (John Wiley & Sons, Inc.)*, *66*(2), 601 - 633. Retrieved from `https://search-ebscohost-com.sid2nomade-1.grenet.fr/login.aspx?direct=true&db=bth&AN=59527266&lang=fr&site=ehost-live`

Siedlecki, W., & Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*, *10*.

Verbeek, M., & Wang, Y. (2013, 9). Better than the original? the relative success of copycat funds. *Journal of Banking and Finance*, *37*, 3454-3471. doi: 10.1016/j.jbankfin.2013.04.024