Application of Portfolio Optimization to Achieve Persistent Time Series

Adam Zlatniczki[1,2] · Andras Telcs[1,3,4]

Abstract

The greater the persistence in a financial time series, the more predictable it becomes, allowing for the development of more effective investment strategies. Desirable attributes for financial portfolios include persistence, smoothness, long memory, and higher auto-correlation. We argue that these properties can be achieved by adjusting the composition weights of the portfolio. Considering the fractal nature of typical financial time series, the fractal dimension emerges as a natural metric to gauge the smoothness of the portfolio trajectory. Specifically, the Hurst exponent is designed for measuring the persistence of time series. In this paper, we introduce an optimization method inspired by the Hurst exponent and signal processing to mitigate the irregularities in the portfolio trajectory. We illustrate the effectiveness of this approach using real data from an S&P100 dataset.

✉ Adam Zlatniczki
adam.zlatniczki@cs.bme.hu

Andras Telcs
telcs.andras@wigner.hu

1   Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Budapest, Hungary

2   Ericsson Hungary, Budapest, Hungary

3   Department for Computational Sciences, Wigner Research Centre for Physics, Budapest, Hungary

4   Department of Quantitative Methods, University of Pannonia, Veszprem, Hungary

⌆ Springer

1 Introduction

Financial technology (FinTech) is the emerging collection of technological innovations that focus on reducing costs and increasing efficiency to help firms have a competitive advantage through improved financial operations [9]. This naturally includes investments, hence constructing portfolios in automated ways.

There can be several financial assets available to investors in a market, each asset having a price that is continuously changing - these prices are thought of as signals that are usually modeled by continuous stochastic processes (geometric Brownian motion being one of the most well-accepted [17]). When an investor holds certain amounts of these assets, it is called a portfolio—the value of the portfolio is simply the superposition (linear combination) of the underlying signals, being a stochastic process itself.

Portfolio choice is practically a control problem, where we intend to change the composite of the portfolio in order to maximize profit or some other related objective function. Such an optimization problem can be formulated and solved in the framework of signal processing by tuning the gain of each signal (weights in the linear combination). In finance, the most basic idea is to control the "risk" of the investment, but this term is quite vague. Here we propose another control goal: to maximize the predictability of the portfolio.

Portfolio choice in its modern form was first proposed by Harry Markowitz [12]. He argued that to minimize risk, one should minimize the variance of the superposed signal's increments (in financial terms, the returns of the portfolio), and formulated this as a quadratic programming (QP) problem. This foundational work led to optimization techniques aiming for other risk measures that work better under stressed market conditions like value-at-risk, conditional value-at-risk, average and maximum drawdown, and so on [7, 15]. Markowitz's mean-variance framework is however the most well-studied, with many proposed improvements. The optimization problem has been improved for example by incorporating machine learning for estimating the covariance matrix via neural networks [20], or to predict returns and apply optimization on top of that [5, 16].

Other authors proposed to tackle the problem on different grounds, for example by incorporating the occasional presence of rare events [11] or phrasing it as a sorting problem [4]. Some even tried incorporating volatility as an additional asset [10]. Of course, properties that are not visible from the time series themselves are also important and can help optimization significantly if made available, for example, retail investor attention [18].

Intuitively, if the superposed signal is predictable, has positive trend-reinforcing capabilities, and shows long memory (i.e. its graph is smooth, its auto-correlation function is positive and vanishes slowly), then it seems to be a good choice for investment. Typical financial time series/signals exhibit self-similarity, and as such the above smoothness properties can be captured with the help of fractal metrics like the Hurst exponent or the fractal dimension [1]. Hurst devised in particular the exponent named after him to capture the persistence (trend-preserving) property of the historical records of the Nile's high water. Indeed, there are examples of trading strategies relying on the Hurst exponent itself [3].

The Hurst exponent (H) takes on values in the range $0 \leq H \leq 1$—the higher the exponent the longer the memory, as its calculation methods are related either directly or indirectly to the "largeness" of auto-correlation (for a detailed comparison of different approaches, see [6]). In the case of time series, fractal dimension (D) measures the smoothness of the path of the underlying stochastic process, hence $1 \leq D \leq 2$. For a smooth curve, $D = 1$, as is expected from a line. However, as the curve gets more zigzagged and loses its smoothness, the corresponding fractal dimension increases as well, up to the extreme where the curve practically fills a 2D space. The fractal dimension and the Hurst exponent are related as $D = 2 - H$, which also shows that longer memory processes with higher H have lower fractal dimensions. It follows that a higher H (or equivalently, a lower D) also indicates better predictability and positive trend-reinforcing, as the corresponding curve is smoother.

Minimizing the variance of signal increments sounds like a legitimate way to achieve a more deterministic signal that satisfies the smoothness properties we aim for, although no work has been done to verify this—later on, we take on this task as well and show that there is a better approach for improving predictability.

Fractal dimension captures the "raggedness" of a signal, and as such, it can be used to control risk [8]—smooth, strongly positively auto-correlated signals with long memory should lead to smaller fractal dimensions [1], therefore controlling for minimal fractal dimension should achieve these nice properties.

Bianchi, Pantanella and Pianese [2] proposed a scheme where the Hurst exponent is optimized, but they proposed to minimize it. This is quite different from our goals because minimizing the Hurst exponent achieves the opposite: a ragged, negative trend-reinforcing, short-memory process (more closely related to mean reversion). In another work, Pantanella and Pianese [14] argue that the Hurst exponent of the superposed signal should be maximized and the standard deviation of the signal changes minimized in a multi-objective problem. However, exact details on the solution are not available, but due to the non-convexity of Hurst exponent estimators, it is highly likely that at best local optima are found. This is a general problem with both fractal dimension and Hurst exponent, though: their estimators are not convex functions, therefore they can't be framed in terms of convex optimization problems, so practically only locally optimal solutions can be expected.

Based on Taguchi's Quality Engineering principles, Nedeltcheva and Ragsdell [13] show that Signal-to-noise (S/N) ratio can also be used as a measure of portfolio stability.

The S/N ratio originates from signal processing, where it measures the power of a signal compared to the power of background noise. The higher the S/N ratio is, the stronger the signal is compared to the noise, indicating a more clear signal that's less corrupted or obscured by said noise. The S/N ratio is used in time series analysis as well, in which case it is defined as the ratio of the mean and the standard deviation of the process, $S/N = \frac{\mu}{\sigma}$. A high S/N ratio means that $\frac{\sigma}{\mu}$ is low, which, interpreted from a financial point of view, indicates that the risk of the portfolio is negligible relative to its expected return. For this reason, the authors in [13] propose to maximize the S/N ratio for portfolio choice, although they do not solve the resulting optimization problem—instead, they rely on the generalized reduced gradient method to find some possibly local solution.

As pointed out in existing literature, financial assets/signals that achieve smoothness, positive auto-correlation of increments, and long memory are good candidates for investment. In this paper, we ask whether different financial assets can be combined linearly into a portfolio in a way that the portfolio itself exhibits these desirable properties. Such a method, given it exists, could be used as part of a multi-stage trading strategy: firstly, the portfolio could be constructed to obtain an artificial financial asset that has better predictability, then secondly, traders could implement their (compatible) trading strategies on top of this artificial asset with improved performance. This way, contrary to usual applications of portfolio optimization, the technique would be used as a kind of pre-processing step instead of as an investment strategy by itself. As such, measuring its effectiveness should be done differently as well, because while strategies are compared based on their effectiveness at generating returns and reducing risk, the effectiveness of this portfolio should be measured by its ability to achieve the aforementioned persistence. All these properties can be captured together by the Hurst exponent on its own, making it a good candidate for statistical comparison.

To answer our previous question, we proceed as follows. By using stochastic calculus, we derive an analytic model of the stochastic dynamics of the superposed signal. To make process increments more deterministic, we apply an L_2 dominance argument and arrive at an objective function for our control problem that turns out to be the maximization of the Signal-to-noise (S/N) ratio of the process increments. The assumptions behind our analytical model are somewhat simplistic, therefore we discard the model itself, but we keep its main intuitive insight: by maximizing the S/N ratio we make the process increments more deterministic, likely leading to increased Hurst exponents (more predictable time series). To show this empirically, we first identify an optimization problem with the necessary constraints for constructing portfolios. Starting from different principles, other authors arrived at similar optimization problems as well, but to the best of our knowledge, we are the first to derive corresponding numerical methods that solve for global optima, not just local ones. For our analysis, we consider the stocks of companies that are listed in the S&P100 index, as this provides plenty of options to optimize upon and also covers roughly 54% of US market capitalization (as of December 27, 2023). We use their adjusted daily close prices, available from *Yahoo! Finance*, between 2005-01-03 and 2022-12-30. We verify our hypothesis that maximizing the S/N ratio is expected to yield an increased Hurst exponent, resulting in a smoother signal, exhibiting stronger positive auto-correlation, and having longer memory than an arbitrary choice would.

The paper is organized in the following way. Section 2 describes the details of the control problem: the derivation of the stochastic dynamics in Sect. 2.1, the portfolio optimization problem in Sect. 2.2, and the means for its solution in Sect. 2.3. Section 3 presents the results, starting with the experiment design in Sect. 3.1, followed by statistical analysis in Sect. 3.2. Implications and further research directions are discussed in Sect. 4, followed by the conclusions in Sect. 5.

2 The Control Problem

2.1 Stochastic Dynamics of Superposed Signals

For illustrative purposes, let us focus on a financial market with two assets. Let the price of each asset be represented by signals X_t and Y_t, and let V_t denote the superposition of these signals. In financial terms, V_t is the value of a portfolio. Let a and b be the gain of the signals, which we would like to control (in finance, these indicate how much an investor should buy from the corresponding assets). The superposed signal is

$$V_t = aX_t + bY_t. \tag{1}$$

Even if we keep the control variables constant, the relative weights of the assets in the portfolio change as the signals change. The relative weights are

$$\begin{aligned} \alpha_t &= aX_t/V_t, \\ \beta_t &= bY_t/V_t. \end{aligned} \tag{2}$$

If we take the usual assumption that the signals follow geometric Brownian motions, we get the signal dynamics

$$\begin{aligned} dX_t &= \mu_X \, X_t \, dt + \sigma_x \, X_t \, dB_t, \\ dY_t &= \mu_Y \, Y_t \, dt + \sigma_Y \, Y_t \, dZ_t, \end{aligned} \tag{3}$$

where B_t and Z_t are two Brownian motions. In practice, the signals are likely not independent, so let's further assume that $Corr(B_t, Z_t) = \rho$. If W_t is another Brownian motion that is independent of B_t, then Z_t can be decomposed as follows

$$dZ_t = \rho dB_t + \sqrt{1 - \rho^2} dW_t \tag{4}$$

and also write

$$dY_t = \mu_Y \, Y_t \, dt + \sigma_Y \, Y_t \left(\rho dB_t + \sqrt{1 - \rho^2} dW_t \right). \tag{5}$$

Taking the assumption that V_t changes only as X_t and Y_t changes (called self-financing in the financial literature), we get

$$dV_t = a \, dX_t + b \, dY_t = \frac{\alpha_t V_t}{X_t} dX_t + \frac{\beta_t V_t}{Y_t} dY_t. \tag{6}$$

Finally, the infinitesimal percentage change in the superposed signal is

$$\begin{aligned} \frac{dV_t}{V_t} &= \alpha_t \left(\mu_X dt + \sigma_X dB_t \right) + \\ & \quad \beta_t \left(\mu_Y dt + \sigma_Y \left(\rho dB_t + \sqrt{1 - \rho^2} dW_t \right) \right) \\ &= \left(\alpha_t \mu_X + \beta_t \mu_Y \right) dt + \end{aligned}$$

$$\left[\left(\alpha_t \sigma_X + \beta_t \rho \sigma_Y \right) d B_t + \beta_t \sigma_Y \sqrt{1 - \rho^2} d W_t \right]. \tag{7}$$

Observe that the first term in the last stochastic differential equation (SDE) represents the deterministic, and the second term is the stochastic part of the superposed dynamics. We consider the relative change of signals instead of the absolute change because that's the common practice in financial engineering (it follows the idea of compound interests better).

2.2 Formal Statement of the Control Problem

If we take a look at SDE 7, we can see that the fractal dimension of V_t (which is in the range $[1, 2]$) can only be 1 if the term in the brackets vanishes and only the deterministic part remains. As long as the stochastic part remains, the raggedness of Brownian motion will be inherited to some extent. In some sense, the lower the fractal dimension is, the smoother V_t becomes, resulting in a more predictable portfolio. We could also take a look at SDE 7 from a Hurst-exponent point of view. The more dominant the deterministic part is, the higher the auto-correlation of process increments becomes, which again implies a more predictable portfolio.

Processes that have Hurst-exponents that are larger than 0.5 are also called persistent, or trend-reinforcing, which is a nice feature to have in a portfolio. Unfortunately, the stochastic term cannot be made to vanish completely. When we apply some form of practical control, we calculate α_0 and β_0, which directly translate to a specific value for a and b. While a and b are kept constant (at least for some time before control is applied again), α_t and β_t aren't—as the signals change, so do the relative weights, as is evident from Eq. 2. The reason we are not looking for a continuous control signal is that in practice, it is not possible to follow one - either due to technical limitations or transaction costs piling up. Instead, control should be reapplied at specific times, but that doesn't affect the way we phrase the control problem.

Nevertheless, we would like to achieve increased Hurst exponents (or decreased fractal dimensions). How do we do that? We practically want to make the relative process increments (or, in financial terminology, investment returns) constant. One immediately might jump to the conclusion that in order to achieve this we should minimize the stochastic term, for example by minimizing its L_2 norm, giving

$$
\begin{aligned}
\arg \min_{\alpha, \beta} L_2(stoch) &= \arg \min_{\alpha, \beta} L_2^2(stoch) \\
&= \arg \min_{\alpha, \beta} E\left[\left((\alpha \sigma_X + \beta \rho \sigma_Y) \, dB_t + \beta \sigma_Y \sqrt{1 - \rho^2} dW_t \right)^2 \right] \\
&= \arg \min_{\alpha, \beta} \left(\alpha^2 \sigma_X^2 + \beta^2 \sigma_Y^2 + \alpha \beta \sigma_X \sigma_Y \rho \right),
\end{aligned} \tag{8}
$$

where we used that $E[dB_t^2] = E[dW_t^2] = dt$, that B_t and W_t are independent, therefore $E[dB_t \, dW_t] = 0$, and omitted a dt multiplier (it does not affect the solution). We can see that this is equivalent to finding the Minimum Variance Portfolio.

However, there's an important observation that has to be made: since we generally cannot decrease the variance to 0, we don't necessarily make the process increments

automatically "more constant" by simply following this strategy. Let's consider two distributions, $N(10, 1)$ and $N(1000, 2)$. We get the intuition that the latter is "more constant-like", even though it has a higher variance. Taking this idea back, we can say that to have more deterministic increments, the optimization shouldn't try to minimize the stochastic part, but instead aim for a solution where the deterministic part dominates the stochastic part as much as possible. One way to formulate this is to maximize the L_2 norm of the deterministic part compared to the L_2 norm of the stochastic part:

$$\arg\max_{\alpha,\beta} \frac{L_2(det)}{L_2(stoch)} = \arg\max_{\alpha,\beta} \frac{|\alpha\mu_X + \beta\mu_Y|}{\sqrt{\alpha^2\sigma_X^2 + \beta^2\sigma_Y^2 + \alpha\beta\sigma_X\sigma_Y\rho}}, \tag{9}$$

where we omitted a $\sqrt{dt}$ multiplier (it does not affect the solution). We can see that this is conceptually equivalent to maximizing the Signal-to-noise (S/N) ratio of the superposed signal's increments.

Equation 9 gives us an objective function, but we still need to consider some constraints. The way α and β are defined (Eq. 2), they have to sum to unity. If investors are allowed to borrow assets (in financial literature, this is called taking short positions in the assets), then there are no constraints left. However, if investors can only spend their own money (in financial literature, this is called taking long positions in the assets), then α and β have to be non-negative.

At this point, we should generalize the control problem from two to an arbitrary number of signals (assets in the financial market). If we let w denote the vector of relative weights of the signals in the superposition, μ denote the vector of the drift of each signal's increments and let Σ denote the covariance matrix of the diffusion part of the signals' increments, then the optimization task can be written as

$$w^* = \arg\max_{w} \frac{|\mu^T w|}{\sqrt{w^T \Sigma w}},$$
$$1^T w = 1, \tag{10}$$
$$w \geq 0,$$

where 1^T is the transpose of a vector of ones. The last inequality is called the long-only constraint, which can be neglected if shorting is allowed.

2.3 Solution of the Control Problem

The optimization problem in (10) can not be tackled in its current form, but after the proper transformations, it becomes solvable. Let us focus on the long-short and the long-only cases one by one.

2.3.1 Solving the Long-Short Case

Let us first consider the long-short optimization problem, where the $w \geq 0$ constraint can be removed from (10). We can find the following slightly different problems that have the same solution:

$$\arg\max_w \frac{\left|\mu^T w\right|}{\sqrt{w^T \Sigma w}} = \arg\max_w \frac{\left(\mu^T w\right)^2}{w^T \Sigma w} = \arg\max_w \frac{w^T \left(\mu\mu^T\right) w}{w^T \Sigma w}, \qquad (11)$$

where the last form is simply a generalized Rayleigh quotient (with two positive semi-definite matrices). As such, its solution is the largest eigenvector of the generalized eigenvalue problem

$$\left(\mu\mu^T\right) w = \lambda \Sigma w. \qquad (12)$$

If v is the largest eigenvector, then the optimal solution w^* must be a scalar multiple of v that satisfies $1^T w^* = 1$, which is simply

$$w^* = \frac{v}{1^T v}. \qquad (13)$$

2.3.2 Solving the Long-Only Case

Dealing with the long-only constraint requires some extra effort because an analytic solution to (10) does not exist. However, it is possible to transform it into a standard convex optimization problem that can be solved numerically.

First, let $\Sigma^{\frac{1}{2}}$ denote the matrix square root of Σ. We can find another optimization task with the same solution as

$$\arg\max_w \frac{\left|\mu^T w\right|}{\sqrt{w^T \Sigma w}} = \arg\min_w \frac{\sqrt{w^T \Sigma w}}{\left|\mu^T w\right|} = \arg\min_w \frac{\left\|\Sigma^{\frac{1}{2}} w\right\|_2}{\left|\mu^T w\right|}, \qquad (14)$$

where $\|\cdot\|_2$ is the L_2 norm of a vector. Note that we can move $\left|\mu^T w\right|$ inside the L_2 norm just like a scalar multiplier, so with a change of variables we get

$$\arg\min_w \frac{\left\|\Sigma^{\frac{1}{2}} w\right\|_2}{\left|\mu^T w\right|} = \arg\min_w \left\|\Sigma^{\frac{1}{2}} y\right\|_2, \qquad (15)$$
$$y = \frac{w}{\left|\mu^T w\right|}.$$

Having an absolute value in an equality constraint is not something we can handle, but if we knew the sign of $\mu^T w$, then we could transform it into a simple linear equality constraint. This is fairly easy to achieve, we just have to split the optimization process into two phases: one where we are looking for a solution in the subspace $\mu^T w \geq 0$, and another where we are looking for a solution in the subspace $\mu^T w \leq 0$. Finally, we simply keep the better solution.

The Case of $\mu^T w \geq 0$

If we introduce $\mu^T w \geq 0$ as an inequality constraint, then $y = \frac{1}{\left|\mu^T w\right|} w = \frac{1}{\mu^T w} w$. After this change of variables, we have to make sure that if we find the solution y^* in this new space, the corresponding w^* satisfies $1^T w^* = 1$ and $w^* \geq 0$. The problem

can be transformed both into a quadratic program (QP) or a second-order cone program (SOCP) - depending on which one we choose, the objective function must be adapted appropriately. Because we already have an L_2 norm in our formulation, we go forward with SOCP for the sake of continuity. It should be noted though that SOCP-s are computationally more demanding, therefore for large problems the QP formulation (provided in Appendix A) might be preferred.

Let us focus on the constraints that must be posed in the y-space to enforce the original constraints in the w-space. First, we must make sure that the change of variables $y = \frac{1}{\mu^T w} w$ is defined. This is easily enforced by the constraint

$$\mu^T y = 1. \tag{16}$$

We also have to force the search into the subspace $\mu^T w \geq 0$. Since $w = (\mu^T w) y$ and $1^T w = 1$, we have $1^T y (\mu^T w) = 1$, or equivalently $1^T y = \frac{1}{\mu^T w}$. Therefore, to enforce $\mu^T w \geq 0$ we need to have

$$1^T y \geq 0. \tag{17}$$

Finally, we need to make sure that $w \geq 0$, which simply follows from

$$y \geq 0, \tag{18}$$

because we already have enforced $\mu^T w \geq 0$ via the previous constraints. SOCP-s have L_2 norms only in their constraints, so we need to introduce an artificial variable s as well, resulting in the final formulation

$$s^*, y^* = \arg \min_{s,y} s,$$
$$\left\| \Sigma^{\frac{1}{2}} y \right\|_2 \leq s,$$
$$\mu^T y = 1, \tag{19}$$
$$1^T y \geq 0,$$
$$y \geq 0.$$

When the optimization problem is solved, $1/s^*$ gives the maximal S/N ratio. We can reconstruct w^* by dividing y^* by $\frac{1}{\mu^T w^*}$, which happens to be equivalent to $1^T y^*$, leading to

$$w^* = \frac{y^*}{1^T y^*}. \tag{20}$$

The Case of $\mu^T w \leq 0$

As discussed before, $y = \frac{1}{|\mu^T w|} w$. In the $\mu^T w \leq 0$ subspace this is equivalent to $y = \frac{1}{-\mu^T w} w$. Following the ideas of the previous section, we have to make sure that this change of variables is defined, which is achieved by changing Eq. 16 to

$$-\mu^T y = 1. \tag{21}$$

Incorporating that $1^T w = 1$, we have $1^T y = \frac{1}{-\mu^T w}$, so keeping Inequality 17 intact achieves exactly this. Since now $-\mu^T w \geq 0$ is ensured, to enforce $w \geq 0$ we don't need to alter Inequality 18 at all either. The final formulation therefore becomes

$$s*, \, y* = \arg\min_{s,y} s,$$

$$\begin{aligned}
\left\| \Sigma^{\frac{1}{2}} y \right\|_2 &\leq s, \\
-\mu^T y &= 1, \\
1^T y &\geq 0, \\
y &\geq 0.
\end{aligned} \tag{22}$$

We can reconstruct $w*$ the same way, as given by Eq. 20.

Now that we have $1/s*$ available in both subspaces, we simply check which one is higher and select the corresponding $w*$ as the solution to the optimization problem (10).

3 Empirical Analysis

By using stochastic calculus, we derived an analytic model of the stochastic dynamics of the superposed signal/portfolio. To make process increments more deterministic, we applied an L_2-dominance argument and arrived at an objective function for our control problem that turned out to be the maximization of the Signal-to-noise (S/N) ratio of the process increments. The assumptions behind our analytical model are somewhat simplistic though, therefore at this point, we discard the model itself, but keep its main intuitive insight: by maximizing the S/N ratio we make the process increments more deterministic, likely leading to more predictable time series.

To show this empirically, we need to solve the optimization problem and construct portfolios on real data, then test whether results, where the S/N ratio is maximized, are significantly better than for other arbitrary portfolios. However, there are a few difficulties that have to be addressed. Before we dive into the details in Sect. 3.1, we give a high-level overview of the methodology.

Let's recall the original research question: is there a portfolio strategy that is expected to yield a smoother, more predictable time series? If so, what sort of evaluation is necessary? As we pointed out in the introduction, such a strategy is not meant to be used to generate returns directly. Instead, it should be used as a pre-processing step that provides a more predictable artificial asset that serves as an input for some other trading strategy that is going to generate the returns. Comparing the max-S/N to other portfolios through the usual means like expected return, value at risk, expected shortfall, and so on, is meaningless: these are meant to compare strategies that generate returns. Instead, the effectiveness of this portfolio should be measured by its ability to achieve the aforementioned fractal properties. Since the Hurst exponent captures them, it provides a good basis for comparison.

Formally, we want to show that the control we derived has a higher expected Hurst exponent than arbitrary portfolios do. If we let W_R denote weights obtained randomly and $W_{S/N}$ denote weights obtained by maximizing the S/N ratio (both being random vectors at this point), and $H(\cdot)$ denote the Hurst exponent corresponding to the weights in its argument, then we have to show that

$$E[H(W_R)] < E\left[H\left(W_{S/N}\right)\right]. \tag{23}$$

As we are going to prove later in Eq. 26, this is equivalent to

$$0 < E\left[H\left(w_{S/N,\mathcal{F}_t}\right) - E[H(W_R) \mid \mathcal{F}_t]\right], \tag{24}$$

where $w_{S/N,\mathcal{F}_t}$ is the optimal control vector that maximizes the S/N ratio if the information available up to time t is $\mathcal{F}_t$ (denoting the filtration of the processes). Moving from theory to practice, an expectation becomes an average, and a filtration becomes a sample. Observe that in our case this means that we take the difference of dependent measurements as both $H\left(w_{S/N,\mathcal{F}_t}\right)$ and $E[H(W_R) \mid \mathcal{F}_t]$ are calculated from the same sub-sample $\mathcal{F}_t$, indicating the necessity of a paired test for comparing averages across many $\mathcal{F}_t$. Because we don't know their distribution, we should rely on a non-parametric test: Wilcoxon's signed-rank test. However, not just $H\left(w_{S/N,\mathcal{F}_t}\right)$ and $E[H(W_R) \mid \mathcal{F}_t]$ are dependent, but the sub-samples of our dataset are too: the auto-correlation of time series is not zero, therefore any pair of neighboring, non-overlapping time windows will still share some common information. To have quasi-independent sub-samples, whenever we select a sub-sample (in the form of a time window), we need to skip a few observations so that at least the auto-correlation vanishes before the start of the next time window. There is another practical problem: estimating the Hurst-exponent. There are several estimators, each of them being sensitive to different characteristics of the data, depending on what their underlying statistical estimators (like re-scaled ranges, variances at different time lags, Fourier or wavelet spectrum,... etc.) are sensitive to. We refer the reader to [6] for a detailed description of such estimators and empirical evidence that shows that they can have quite different outcomes due to relatively high bias and variance, so relying on a single estimator is dangerous. To this end, we use 4 different Hurst-exponent estimators and do statistical testing on the results of each estimator separately. However, this way it becomes a multiple testing problem with dependent tests. In order to deal with it, we use the harmonic mean of the p-values to get a final p-value while controlling the strong-sense family-wise error rate [19]

For our analysis, we consider the stocks of companies that are listed in the S&P100 index, as this provides plenty of options to optimize upon and also covers roughly 54% of US market capitalization (as of December 27, 2023). We use their adjusted daily close prices, available from *Yahoo! Finance*, between 2005-01-03 and 2022-12-30. We drop 14 stocks ('ABBV', 'AVGO', 'BRK.B', 'CHTR', 'DOW', 'GM', 'KHC', 'MA', 'META', 'PM', 'PYPL', 'TMUS', 'TSLA', 'V') because they have missing values in the given time range. This leaves us with 87 time series, each having 4531 observations. In accordance with Eq. 7, we change from the actual prices to their relative increments when estimating parameters of the optimization problem,

namely the covariance matrix Σ and expected value vector μ. To have a normalized representation of the prices, we transform the time series to their cumulative returns (which practically means that we divide each element in a time series by its first observation). Data, Python source codes, and the Jupyter Notebook we used are made available on GitHub at https://github.com/adam-zlatniczki/max_snr_portfolio.

3.1 Experiment Design

Now that a high-level view of the analysis is given, we provide a more in-depth description of our methodology. As already stated before, we aim to show that

$$E\left[H\left(W_R\right)\right] < E\left[H\left(W_{S/N}\right)\right]. \tag{25}$$

In theory, if $\mathcal{F}_t$ is the filtration of the stochastic processes (signals), then this is equivalent to

$$
\begin{aligned}
0 &< E\left[H\left(W_{S/N}\right)\right] - E\left[H\left(W_R\right)\right] \\
&= E\left[H\left(W_{S/N}\right) - H\left(W_R\right)\right] \\
&= E\left[E\left[H\left(W_{S/N}\right) - H\left(W_R\right) \mid \mathcal{F}_t\right]\right] \\
&= E\left[E\left[H\left(W_{S/N}\right) \mid \mathcal{F}_t\right] - E\left[H\left(W_R\right) \mid \mathcal{F}_t\right]\right] \\
&= E\left[H\left(w_{S/N,\,\mathcal{F}_t}\right) - E\left[H\left(W_R\right) \mid \mathcal{F}_t\right]\right],
\end{aligned}
\tag{26}
$$

where $w_{S/N,\,\mathcal{F}_t}$ is the optimal control vector that maximizes the S/N ratio if the information available up to time t is $\mathcal{F}_t$. The conditional expectation disappears because $W_{S/N}$ is $\mathcal{F}_t$ measurable (i.e. the optimal weights are a deterministic function of the available information).

This derivation outlines how hypothesis testing should be done on data. Moving from theory to practice, an expectation becomes an average, and a filtration becomes a sample. Let us interpret Inequality 26:

– Focusing on a specific realization of $\mathcal{F}_t$ is practically equivalent to considering a specific sample.
– Given the sample, we can calculate $\hat{\Sigma}$ and $\hat{\mu}$ of the relative increments (in accordance with Eq. 7), solve the portfolio optimization problem that maximizes the S/N ratio, thus obtain $w_{S/N,\,\mathcal{F}_t}$. Given these weights, we can construct the portfolio's time series and calculate its Hurst exponent, $H\left(w_{S/N,\,\mathcal{F}_t}\right)$.
– $E\left[H\left(W_R\right) \mid \mathcal{F}_t\right]$ is the expected Hurst exponent of arbitrary strategies given the same sample. We can get this by generating many strategies (weights), calculating the corresponding Hurst exponents, and taking their average.
– We ask whether the difference between these two quantities is expected to be higher than zero, meaning that we take the average of the difference over many samples $\mathcal{F}_t$ and test whether it's significantly higher than 0.

While this process seems adequate, there are some further aspects we have to consider when we wish to apply statistical hypothesis testing, as we pointed out earlier:

– For each sample $\mathcal{F}_t$ we consider, $H\left(w_{S/N,\,\mathcal{F}_t}\right)$ and $E\left[H\left(W_R\right) \mid \mathcal{F}_t\right]$ are dependent, as they are calculated on the same sample. To this end, we need to use a paired test.

- We don't know the distribution of $H\left(w_{S/N}, \mathcal{F}_t\right) - E\left[H\left(W_R\right) \mid \mathcal{F}_t\right]$, therefore we need to use a non-parametric test.
- The previous two points narrow down our options for using Wilcoxon's signed-rank test for paired samples.
- We have only one sample at our disposal. When we split this up into several sub-samples (or time windows), the neighboring ones are not independent. This is because the processes are auto-correlated, and through this auto-correlation information seeps from one window to the next. To satisfy the requirements of the signed-rank test, we need to make the sub-samples at least quasi-independent, as illustrated below.
- Finally, Hurst exponent estimation is not trivial. There are many estimators, each being appropriate under somewhat different circumstances. We refer the reader to [6] for a detailed description of such estimators and empirical evidence that shows that they can have quite different outcomes due to relatively high bias and variance. Relying on a single estimator can easily lead to a misinformed decision, therefore multiple estimators should be used, and statistical testing needs to be done accordingly, as we will illustrate below.
- We also note that both data collection and testing must be done for long-only and long-short portfolios separately, these shouldn't be mixed together.

To make subsequent sub-samples quasi-independent, we apply the following scheme. First, we choose a sub-sample size n and select the first n observations as the first sub-sample. Then we calculate the auto-correlation of the relative increments of each time series in this sub-sample and select the largest significant lag l across them. Next, we ignore the observations in the $[n+1, n+l]$ range, and select the observations in the range $[n+l+1, n+l+n]$ as the second sub-sample. We re-calculate l and, again, skip this many sample points before selecting the third sub-sample. We keep iterating this approach as many times as possible before running out of sample points.

To deal with the Hurst exponent estimation problem, we use 4 different estimators: one based on rescaled range analysis (denote this by H_1), one based on the variance of lagged differences (denote this by H_2), one based on a robust estimator for the variance of lagged differences (denote this by H_3), and one based on fractal dimension calculation (denote this by H_4). We do the statistical testing on the results of each estimator separately. However, this way it becomes a multiple testing problem with dependent tests. In order to deal with it, we use the harmonic mean of the p-values to get a final p-value while controlling the strong-sense family-wise error rate [19].

Since we mentioned the minimum variance portfolio in Eq. 8, we extend our analysis with that as well. Data collection for the Wilcoxon signed-rank tests thus can be summarized as follows:

1. Identify a set of assets; let's focus on stocks of the S&P100 index, as proposed before.
2. Collect the adjusted close prices of said stocks over a large timespan as our signals. Drop those stocks that have missing values.
3. Choose a sub-sample size (or window size). We set it to 100, as it is neither too short nor too long.

4. Select the next 100 observations as a sub-sample.

   (a) Transform the time series to their relative changes.
   (b) Calculate the sample covariance matrix and sample means of relative changes.
   (c) Solve the optimization problem for $w_{S/N}$, form the cumulative returns of the portfolio, and calculate the four different Hurst exponent estimators.
   (d) Solve the optimization problem for w_{minvar}, form the cumulative returns of the portfolio, and calculate the four different Hurst exponent estimators.
   (e) Generate 10,000 random portfolios; for each, form the cumulative returns of the portfolio, calculate the four different Hurst exponent estimators; for each of the four types of Hurst estimators, average the 10,000 values.
   (f) Store the results: sub-sample index; H1, H2, H3, H4 of max-S/N portfolio; H1, H2, H3, H4 of min-var portfolio; average H1, average H2, average H3, average H4 of random portfolios.
   (g) Calculate the auto-correlation function of each (relative-change) time series, find the largest significant lag across them; and skip this many sample points.

5. Go back to Step 4; repeat this process as long as new sub-samples can be taken from the sample.

3.2 Results

We ran the experiment proposed in Sect. 3.1. Following the sub-sample selection algorithm, we obtained 37 quasi-independent sub-samples, each having 100 observations. As proposed, we collected the results for the long-short and long-only cases separately. The figures in this section were created with the Seaborn Python package.

Given shorting is allowed, Fig. 1 shows the distribution of the Hurst exponent, per type of estimator (H1-H4) and portfolio strategy (maximized S/N ratio, minimized variance, and random choice). We can see that the H1 and H3 estimators don't seem to be able to capture any differences. Even more, applying some kind of optimization rather increases their dispersion. On the other hand, H2 and H4 seem to be able to differentiate between the different strategies. Based on these two, maximizing the S/N ratio seems to yield higher Hurst exponents, outperforming the other strategies. Minimizing variance also shows an improvement compared to arbitrary portfolios, but many times it leads to even worse Hurst exponents than arbitrary strategies would.

Given only long positions are allowed, Fig. 2 shows the distribution of the Hurst exponent, per type of estimator and portfolio strategy. The results are harder to interpret in this case, the distributions have high overlaps. However, we can observe the ineffectiveness of H1 and H3 again, but unlike for long-short portfolios, the increase in dispersion due to the application of optimization doesn't seem to be present. Similarly to the long-short case, H2 and H4 indicate that maximizing the S/N ratio achieves higher Hurst exponents than the other strategies, but unlike in the long-short case, minimizing the variance doesn't seem to be able to achieve the same effect.

Based simply on the histograms, especially in the long-only case, determining whether optimization increases the Hurst exponent is not possible, as the distributions overlap too much, and different Hurst estimators have contradicting results. As pointed
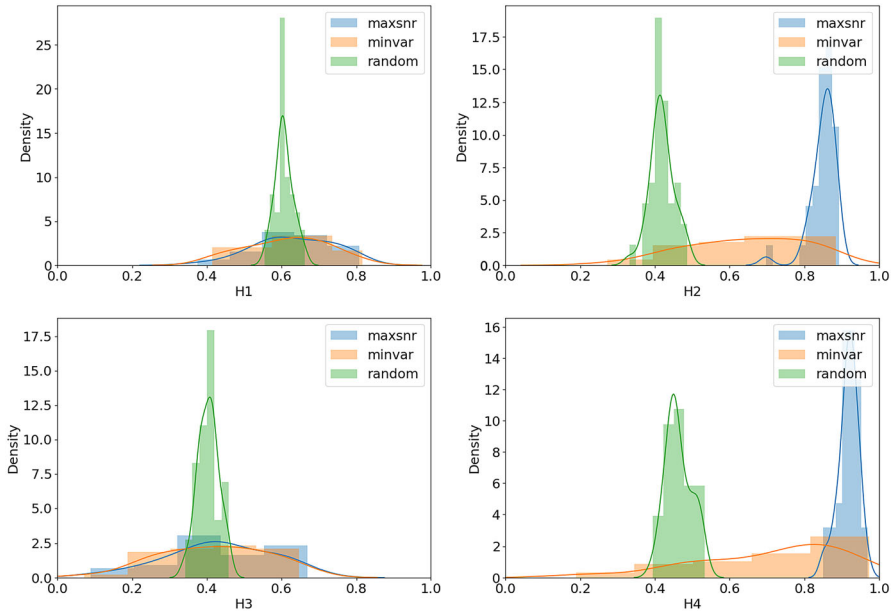
Fig. 1 Distribution of Hurst exponents of different estimators (H1–H4), across different portfolio strategies (maximized S/N ratio, minimized variance, and random weights), given shorting is allowed. H1 and H3, and H2 and H4 are quite similar in shape, but H1 also shows an upward bias of roughly 0.15 compared to the rest of the estimators. H1 and H3 don't seem to be able to differentiate between the different strategies. Even more, they indicate that applying optimization only increases dispersion, while H2 and H4 indicate that maximizing the S/N ratio of relative portfolio increments leads to much higher Hurst exponents—a contradicting result. H2 and H4 also indicate that minimizing variance many times leads to smoother functions, but unlike maximizing the S/N ratio, it can achieve even worse results than a random approach

out in the previous section, pairwise comparisons must be made to overcome this, combined with handling the multiple-testing problem. Non-normality of distributions is also evident. The latter points also support that Wilcoxon's signed-rank test is a good candidate for hypothesis testing and that the calculation of the harmonic mean p-value is necessary.

The results of statistical hypothesis testing are summarized in Table 1 for the long-short case, and in Table 2 for the long-only case. As it can be seen in the first two horizontal blocks of both tables, the null hypothesis is rejected (based on $hmp < 0.05$), meaning that maximizing the S/N ratio tends to yield a higher Hurst exponent than a minimum variance or a random approach. However, comparing the minimum variance and random strategies, the results are different: based on the third block, the null hypothesis is rejected (based on $hmp < 0.05$) in the long-short case, but accepted (based on $hmp > 0.05$) in the long-only case. This means that when shorting is allowed, minimizing the variance tends to provide a portfolio with a higher Hurst exponent than a random approach, but the same cannot be said when shorting isn't allowed. However, as we saw in Fig. 1, even if minimizing the variance can lead to significantly higher Hurst exponents, it still has a high chance of providing worse-
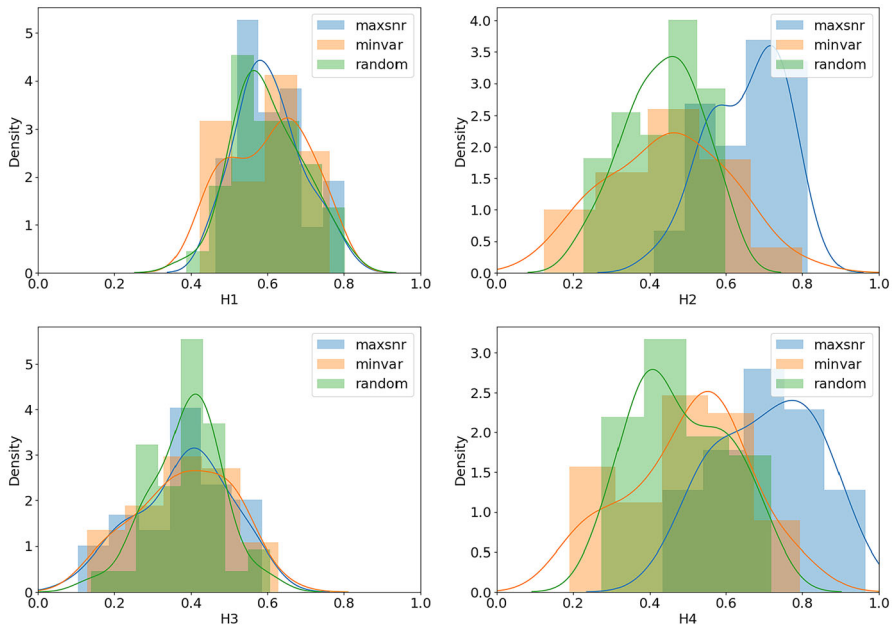
Fig. 2 Distribution of Hurst exponents of different estimators (H1–H4), across different portfolio strategies (maximized S/N ratio, minimized variance, and random weights), given shorting only long positions are allowed. H1 and H3 are quite similar in shape, but H1 also shows an upward bias of roughly 0.15 compared to the rest of the estimators. As in the long-short case (Fig. 1), H1 and H3 don't seem to be able to differentiate between the different strategies, but at least applying optimization doesn't seem to increase their dispersion. H2 and H4 indicate that maximizing the S/N ratio of relative portfolio increments leads to higher Hurst exponents, but minimizing variance doesn't. However, the increase in Hurst exponent seems less than in the long-short case

than-random performance. Overall, we can state that maximizing the S/N ratio is the better strategy if we want a portfolio with an increased Hurst exponent.

Looking at Figs. 1 and 2, we also get the intuition that when shorting is allowed, we can achieve higher Hurst exponents. Table 3 confirms this: the long-short option leads to a significantly higher exponent, except for the case of random portfolios. These results are also very intuitive, as in the long-short case we have a lower number of constraints, leaving a larger space for finding better solutions.

4 Discussion

First, let us summarize our findings. Maximizing the S/N ratio of process increments yields a portfolio with increased Hurst exponent, thus better predictability, hence proves to be a good pre-processing step before the application of some trading strategy. As a direct consequence, we conclude that the well-known maximum Sharpe-ratio portfolio (coming from the classical mean-variance portfolio optimization framework) also exhibits such beneficial properties, and so do portfolios obtained

Table 1 Hypothesis test results in the **long-short case**

		H1	H2	H3	H4	hmp
maxsnr	Mean	0.6336	0.8502	0.4311	0.9175	
minvar	Mean	0.6136	0.6436	0.4180	0.7075	
Wilcoxon	T	422	703	387	686	
	p	0.1473	7.275×10^{-12}	0.3008	1.506×10^{-9}	0.0
maxsnr	Mean	0.597	0.551	0.408	0.597	
Random	Mean	0.6072	0.4184	0.4039	0.4632	
Wilcoxon	T	460	703	448	703	
	p	0.0519	7.276×10^{-12}	0.0745	7.2760×10^{-12}	0.0
minvar	Mean	0.6136	0.6436	0.4180	0.7075	
Random	Mean	0.6072	0.4184	0.4039	0.4632	
Wilcoxon	T	382	688	398	660	
	p	0.3273	9.9681×10^{-10}	0.2461	9.4333×10^{-8}	0.0

Each horizontal block represents the results of a paired one-tailed Wilcoxon signed rank test between the two portfolio strategies indicated, comparing them along the different Hurst exponent estimators (H1–H4). The last row in each block shows the p-values for each test and the **hmp** column shows the combined *p-value*. In all three blocks, the null hypothesis is rejected (based on $hmp < 0.05$). The first block implies that maximizing the S/N ratio tends to yield a higher Hurst exponent than if we were minimizing variance, and the third block implies that minimizing the variance tends to yield a higher Hurst exponent than arbitrary portfolios. The second block simply shows that there's no contradiction. Observe that the results are different from the long-only case (Table 2), as here minimizing the variance performs much better in terms of increasing predictability

based on Taguchi's Quality Engineering principles, as these are closely related to the S/N ratio.

Minimizing variance can have a similar effect, but its effectiveness is significantly lower, and can even become worse than a random choice's. Shorting also proved to significantly increase the effectiveness of achieving higher Hurst exponents.

Starting from different principles, other authors arrived at similar optimization problems as well, but to the best of our knowledge, we are the first to derive corresponding numerical methods that solve for global optima, not just local ones.

Our findings might have implications for trader policies as well. The max-S/N technique proved to be a good pre-processing step before the application of some trading strategy, but it must match the existing policy. For example, if it's based on anti-persistency (the opposite of predictability), then applying the max-S/N technique as a pre-processing step should be avoided. Also, as our results indicate, the use of shorting significantly helps the effectiveness of the max-S/N technique—if possible, it should be incorporated.

For time increments, we used a whole day, as this is the smallest frequency at which stock prices were available openly. We could go into higher frequency data, but then we would face another problem: the current formulation considers only a 1-step increment. However, the high auto-correlation we wish to achieve implies that multiple steps need to be considered. By choosing daily increments, we indirectly do this, as a day

Table 2 Hypothesis test results in the **long-only case**

		H1	H2	H3	H4	hmp
maxsnr	Mean	0.6044	0.6528	0.3798	0.7066	
minvar	Mean	0.6035	0.4463	0.3833	0.4989	
Wilcoxon	T	336	690	326	669	
	p	0.5943	6.4028×10^{-10}	0.6509	2.7103×10^{-8}	0.0
maxsnr	Mean	0.6044	0.6528	0.3798	0.7066	
Random	Mean	0.6018	0.4340	0.3872	0.4821	
Wilcoxon	T	320	703	322	697	
	p	0.6835	7.2760×10^{-12}	0.6727	1.0186×10^{-10}	0.0
minvar	Mean	0.6035	0.4463	0.3833	0.4989	
Random	Mean	0.6018	0.4340	0.3872	0.4821	
Wilcoxon	T	335	391	340	420	
	p	0.6000	0.2803	0.5710	0.1544	0.2971

Each horizontal block represents the results of a paired one-tailed Wilcoxon signed rank test between the two portfolio strategies indicated, comparing them along the different Hurst exponent estimators (H1–H4). The last row in each block shows the p-values for each test and the **hmp** column shows the combined *p-value*. In the first two blocks, the null hypothesis is rejected (based on $hmp < 0.05$), meaning that maximizing the S/N ratio tends to yield a higher Hurst exponent than a minimum variance or a random approach. In the third block, the null hypothesis is accepted (based on $hmp > 0.05$), meaning that neither the minimum variance nor the random approach tends to yield higher Hurst exponents than the other

aggregates many "infinitesimal" increments, but as we increase data frequency, we start losing this property. However, moving to lower frequencies has a similar problem, but vice versa: too large increments might exceed the memory of the process, introducing irrelevant data, thereby decreasing the observable correlation between stocks, thus pushing Σ towards being a diagonal matrix where interdependencies of time series no longer can be utilized as much.

We also touched upon the somewhat higher computational complexity of solving SOCP-s, for which reason we present a QP formulation in Appendix A, as QP-s have a wider range of high-performance, concurrent solvers available either commercially (like CPLEX, GUROBI, MOSEK) or open-source (like HiGHS). With some modification, even real-time application of the max-S/N technique is possible: using exponentially weighted moving averages, both μ and Σ can be updated in an online fashion whenever new observations arrive while still controlling the effect of past observations. The existing weights can be used to warm-start the new optimization problem and find a new solution much faster, as small perturbations to the problem (especially in the more constrained long-only case) shouldn't result in a very different solution. It should be emphasized though that if the smoothing factor is low (giving high weight to new observations), then the distance between two consecutive estimates of the parameters might be large, possibly leading to solutions that are far from each other as well, in which case warm starting has reduced benefits. This must be validated, especially in the presence of high volatility. Due to the efficiency of QP

Table 3 Hypothesis test results comparing the long-short and long-only cases across different portfolio strategies

			H1	H2	H3	H4	hmp
maxsnr	Long-short	Mean	0.6336	0.8502	0.4311	0.9175	
	Long-only	Mean	0.6044	0.6528	0.3798	0.7066	
	Wilcoxon	T	433	703	479	700	
		p	0.1123	7.2759×10^{-12}	0.0273	3.6379×10^{-11}	0.0
minvar	Long-short	Mean	0.6136	0.6436	0.4180	0.7075	
	Long-only	Mean	0.6035	0.4463	0.3833	0.4989	
	Wilcoxon	T	359	655	421	637	
		p	0.4585	1.7825×10^{-7}	0.1509	1.3682×10^{-6}	0.0
Random	Long-short	mean	0.6072	0.4184	0.4039	0.4632	
	Long-only	Mean	0.6018	0.4340	0.3872	0.4821	
	Wilcoxon	T	385	274	416	292	
		p	0.3112	0.8788	0.1692	0.8152	0.3482

Each horizontal block represents the results of a paired one-tailed Wilcoxon signed rank test between the two portfolio strategies indicated, comparing them along the different Hurst exponent estimators (H1–H4). The last row in each block shows the p-values for each test, and the **hmp** column shows the combined *p-value*. In the first two blocks, the null hypothesis is rejected (based on $hmp < 0.05$), indicating that the option of shorting significantly increases the effectiveness of the portfolio. This is very intuitive, as the objective function is the same in both cases, but there are fewer constraints when shorting is allowed, making the feasible region much larger, likely encompassing a better solution. In the third block, the null hypothesis is accepted (based on $hmp > 0.05$), implying that regardless of the size of the feasible region, choosing an arbitrary strategy isn't likely to have any additional effect on the Hurst exponent

solvers (see the performance of *alglib* [1] for example), this is only of interest in the case of high-frequency or near-real-time trading though.

To test the sensitivity and robustness of our approach, we applied the method throughout a large time frame where many market conditions were present and collected the distribution of Hurst exponents in Figs. 1 and 2. We don't delve deeper into sensitivity analysis, as it is evident from the optimization problem: the condition number of the covariance matrix is the most important factor, as it determines how well-conditioned the problems become, hence how sensitive they are to small perturbations. To mitigate problems of robustness, one could use robust estimators for the optimization model's parameters: median, trimmed or winsorized mean for μ, and minimum covariance determinant for Σ. Quite naturally, the effectiveness of parameter estimation has some effect on the outcome, but this heavily depends on what trading strategy would be implemented on top of the artificial asset formed by the portfolio, the market conditions, and possibly other elements of the scenario it is going to be applied in. While our method likely has limitations in some of these scenarios, it is not possible to give a detailed analysis as a practically infinite space of problems would have to be spanned. Practitioners in the field routinely validate the applicability of any method they wish to use - following their well-established best practices should more than suffice as guidance to validate the applicability of our method as well.

When we derived the optimization problem, we used stationary parameters in the stochastic differential equations. While this leads to a limited model, it does not necessarily limit the applicability of the optimization technique itself that much (as indicated by the empirical analysis). Stationarity is usually assumed over some time window anyway, otherwise, statistical approaches have no sound basis. Also, as we indicated before, the theoretical model was only used to derive the intuition for the optimization problem—in the end, it was dropped, and the technique's effectiveness was demonstrated empirically. However, it should be noted that over large time frames, things tend to change, and assuming stationarity is no longer acceptable. The size of the time windows and the frequency of portfolio recalibration should be chosen accordingly. Since these are heavily influenced by the exact context in which our method is considered to be applied, they should be evaluated with care—just as for any other technique, practitioners should refer to existing best practices. Another approach would be to introduce non-stationary parameters and jumps into the stochastic differential equations, and to derive a possibly more refined optimization problem, but this goes beyond the scope of this paper, we leave it as an interesting direction for future research.

5 Conclusions

In this paper, we set out to find a control mechanism that can find a linear superposition of financial signals (a portfolio) that is smooth, has positive auto-correlation, and has long memory. Such a technique could be used as a sort of pre-processing step that generates a predictable portfolio that could be used as an artificial asset in another trading strategy. We found that maximizing the Signal-to-noise ratio of relative

---

[1] https://www.alglib.net/optimization/quadraticprogramming.php.

portfolio increments achieves this goal. We also found that minimizing the variance instead can have a similar effect, but its effectiveness is significantly lower, and can even become worse than random choice. As a direct consequence, we concluded that the well-known maximum Sharpe-ratio portfolio (coming from the classical mean-variance portfolio optimization framework) also exhibits such beneficial properties, and so do portfolios obtained based on Taguchi's Quality Engineering principles, as these are closely related to the S/N ratio. As expected, shorting also proved to significantly increase the effectiveness of achieving more predictable portfolios. Starting from different principles, other authors arrived at similar optimization problems as well, but to the best of our knowledge, we are the first to derive corresponding numerical methods that solve for global optima, not just local ones. When we derived the optimization problem, we assumed stationary processes. While this is not necessarily a limiting factor, it should be noted that over large time frames, things tend to change, assuming stationarity is no longer acceptable. The size of the time windows and the frequency of portfolio recalibration should be chosen accordingly with care.

Introducing non-stationary parameters, as well as jumping into the stochastic differential equations and deriving a more refined optimization problem is an interesting direction for future research.

Data Availability Data and code for reproducing the results is available at https://github.com/adam-zlatniczki/max_snr_portfolio.

Declarations

Conflict of interest The authors have no conflict of interest to declare.

A Quadratic Programming Formulation

In Sect. 2.3.2 we transformed the original problem to a SOCP form, which guarantees that global optima can be found via standard solvers for SOCP-s. However, SOCP solvers are computationally more demanding, therefore we show how the original

problem can be formulated as a quadratic program, which is easier to solve and has a wider range of high-performance solvers available.

First, we find the following alternative objective functions with the same solution:

$$\arg\max_w \frac{|\mu^T w|}{\sqrt{w^T \Sigma w}} = \arg\min_w \frac{\sqrt{w^T \Sigma w}}{|\mu^T w|} = \arg\min_w \frac{w^T \Sigma w}{(\mu^T w)^2}. \tag{27}$$

Introducing the change of variable $y = \frac{w}{\mu^T w}$, we can rewrite the problem as

$$\arg\min_y y^T \Sigma y,$$
$$y = \frac{w}{\mu^T w}. \tag{28}$$

Similarly, as before, we need to pose constraints in the y space that make sure that the corresponding w satisfy $1^T w = 1$ and $w \geq 0$. To make the change of variables well-defined, we have to introduce $\mu^T y = 1$. Unfortunately, $y \geq 0$ doesn't necessarily imply $w \geq 0$, unless we enforce $\mu^T w \geq 0$. Incorporating $1^T w = 1$ indirectly, we get $1^T y = \frac{1}{\mu^T w}$, hence we can enforce $\mu^T w \geq 0$ by introducing the constraint $1^T y \geq 0$. Now we can safely say that $y \geq 0$ will also imply $w \geq 0$. Summarized, the optimization problem can be written as

$$y^* = \arg\min_y y^T \Sigma y,$$
$$\mu^T y = 1,$$
$$1^T y \geq 0,$$
$$y \geq 0, \tag{29}$$

where we obtain $w^* = \frac{y^*}{1^T y^*}$.

If we are looking for a solution in the $\mu^T w \leq 0$ half-space, then we need to introduce $1^T y \leq 0$ and $y \leq 0$ instead, resulting in

$$y^* = \arg\min_y y^T \Sigma y,$$
$$\mu^T y = 1,$$
$$1^T y \leq 0,$$
$$y \leq 0, \tag{30}$$

where we similarly obtain $w^* = \frac{y^*}{1^T y^*}$.

References

1. Beran, J.: Statistics for Long-Memory Processes. Chapman and Hall/CRC, New York (2017). https://doi.org/10.1201/9780203738481
2. Bianchi, S., Pantanella, A., Pianese, A.: Financial portfolio selection in a nonstationary Gaussian framework. In: The Role of the University in the Analysis of Current Economic Crisis, vol. 1, pp. 619–627. România de Mâine Publishing House, Spiru Haret University, Bucharest (2009)

3. Bui, Q., Slepaczuk, R.: Applying hurst exponent in pair trading strategies on nasdaq 100 index. Physica A Stat. Mech. Appl. **592**, 126784 (2022). https://doi.org/10.1016/j.physa.2021.126784

4. Emamat, M., Mota, C., Mehregan, M., Moghadam, M., Nemery, P.: Using electre-tri and flowsort methods in a stock portfolio selection context. Financial Innov. **8**(11) (2022)

5. Grudniewicz, J., Slepaczuk, R.: Application of machine learning in algorithmic investment strategies on global stock markets. Res. Int. Bus. Finance **66**, 102052 (2023). https://doi.org/10.1016/j.ribaf.2023.102052

6. Hamza, A.H., Hmood, M.Y.: Comparison of hurst exponent estimation methods. J. Econ. Admin. Sci. **27**(128), 167-183 (2021). https://doi.org/10.33095/jeas.v27i128.2162

7. Han, C., Wang, K.: Stressed portfolio optimization with semiparametric method. Financial Innov. **8**(27) (2022)

8. Kantardzic, M., Sadeghian, P., Shen, C.: The time diversification monitoring of a stock portfolio: an approach based on the fractal dimension. In: Proceedings of the 2004 ACM Symposium on Applied Computing, SAC '04, pp. 637–641. ACM, New York, NY, USA (2004). https://doi.org/10.1145/967900.968034

9. Kou, G., Akdeniz, O.O., Dincer, H., Yüksel, S.: Fintech investments in european banks: a hybrid it2 fuzzy multidimensional decision-making approach. Financial Innov. **7**(39) (2021)

10. Latoszek, M., Slepaczuk, R.: Does the inclusion of exposure to volatility into diversified portfolio improve the investment results? Portfolio construction from the perspective of a Polish investor. Econ. Bus. Rev. **6**(1), 46–81 (2020). https://doi.org/10.18559/ebr.2020.1.3

11. Lv, W., Pang, T., Xia, X., Yan, J.: Dynamic portfolio choice with uncertain rare-events risk in stock and cryptocurrency markets. Financial Innov. **9**(73) (2023)

12. Markowitz, H.M.: Portfolio selection. J. Finance **7**(1), 77–91 (1952)

13. Nedeltcheva, G.N., Ragsdell, K.: A new approach to optimal investment portfolio management. Int. J. Glob. Manag. Stud. **2**(2), 16–34 (2011)

14. Pantanella, A., Pianese, A.: Minimum risk portfolios using MMAR. In: MCBE'09 Proceedings of the 10th WSEAS International Conference on Mathematics and Computers in Business and Economics (2009)

15. Pfaff, B.: Risk-optimal portfolios, chap. 12, pp. 228–273. John Wiley & Sons, Ltd (2016). https://doi.org/10.1002/9781119119692.ch12

16. Slusarczyk, D., Slepaczuk, R.: Optimal Markowitz portfolio using returns forecasted with time series and machine learning models. Working Papers 2023-17, Faculty of Economic Sciences, University of Warsaw (2023). https://ideas.repec.org/p/war/wpaper/2023-17.html

17. Steele, J.M.: Stochastic Calculus and Financial Applications. Springer, New York (2010)

18. Wen, F., Xu, L., Ouyang, G., Kou, G.: Retail investor attention and stock price crash risk: evidence from China. Int. Rev. Financial Anal. **65**, 101376 (2019). https://doi.org/10.1016/j.irfa.2019.101376

19. Wilson, D.J.: The harmonic mean p-value for combining dependent tests. In: Proceedings of the National Academy of Sciences **116**(4), 1195–1200 (2019). https://doi.org/10.1073/pnas.1814092116

20. Wysocki, M., Sakowski, P.: Investment portfolio optimization based on modern portfolio theory and deep learning models. Working Papers 2022-12, Faculty of Economic Sciences, University of Warsaw (2022). https://ideas.repec.org/p/war/wpaper/2022-12.html

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.