Can hedge fund elites consistently beat the benchmark? A study of portfolio optimization

Stéphane Meng-Feng Yen a,1, Ying-Lin Hsu b,2, Yi-Long Hsiao c, * a
Department of Accounting and Graduate Institute of Finance, National Cheng Kung University, Taiwan
b Department of Applied Mathematics, National Chung Hsing University, Taiwan
c Department of Finance, National Dong Hwa University, Taiwan

Abstract

This study aims to explore whether a regularly updated portfolio of outperforming hedge funds can consistently beat the corresponding hedge fund dataset index. If yes, moreover, the second question concerns whether portfolio optimization approaches can lead to an even better performance than the naive equal-weighting method. The dataset spans the January-1994 to August-2008 period and is classified into four main categories - Macro, Equity Hedge, Relative Value and Event Driven. Based on a seven-factor model, this study applies the Step-SPA test to each category of funds and examines the statistical significance of the studentized fund alpha over the selection period of 3–7 years in length. A ‘winner’ portfolio of funds, namely, consisting of funds with statistically significant, positive studentized alpha, can be formed at the end of the selection period and held for 1 up to 3 years. We find that the winner portfolio tends to beat the dataset indexes during the holding period, irrespective of the time span for the selection and the holding periods investigated. Moreover, two of the three optimization approaches employed, the Probabilistic Global Search Lausanne and the Genetic Algorithm, prove to further enhance the performance of the equal-weighted winning portfolio.

Keywords:
Hedge funds
Step-SPA test
GA
PGSL
Performance persistence

Article info

Article history:
Received 15 October 2014
Accepted 9 April 2015
Available online 26 July 2015

1. Introduction

The industry of hedge funds has consistently performed well except in the year of 2008 when a financial tsunami occurred as a result of a substantial subprime mortgage default. Previous evidence in the literature has shown from a variety of perspectives that the hedge fund elites, namely, those beating the benchmarks, are capable of consistently remaining outperformers over time. Moreover, most previous efforts have ignored the data-snooping bias problem. Controlling for the data-snooping bias, Kosowski, Naik, and Teo (2007) apply the cross-sectional alpha bootstrap technique, introduced by Kosowski, Timmermann, Wermers and White (2006), to examine the statistical significance of each hedge fund’s alpha in the presence of a large quantity. Their test allows us to know whether a hedge fund whose alpha ranks with a certain place delivers a statistically significant alpha with that certain ranking. Based on the bootstrapped alpha distribution constructed by extracting and ranking from small to large the bootstrapped alpha under each of a large number of, say 1000, resamples, in particular, the observed sample alpha of a hedge fund can be compared to such a bootstrapped distribution which then leads to the p-value and statistical significance according to a given level of significance. Similarly, Romano and Wolf (2005) pioneer an alternative approach, known as the stepwise reality check (SRC), to mitigating the adverse effect of data-snooping bias in the context of large-scale multiple hypothesis testing. Instead of testing the null hypothesis for each single hedge fund using its corresponding bootstrapped distribution, Romano and Wolf’s method focuses on the bootstrapped distribution for the best-performing fund’s alpha and screens all resulting outperformers. Due to the different characteristics of these two alternatives, conducting the former approach is more time-consuming than the latter although they tend to generate different statistical inferences. However, both studies above select only a fixed sample period for testing the

http://dx.doi.org/10.1016/j.apmrv.2015.04.001

Copyright © 2015 College of Management, National Cheng Kung University. Production and hosting by Elsevier Taiwan LLC. All rights reserved.
statistical significance of the fund alphas without considering the effect of a structural change of the hedge fund dataset on their results. Also, they both fail to test the statistical significance of the alpha of the identified hedge fund elites as a whole over the holding period.

Another main issue that has been ignored in the literature of hedge fund research involves the portfolio optimization. One possible cause for the dearth of such efforts might be the size of the amount for initial investment. Given that it is not extraordinary for the typical hedge fund to require an initial investment of 200,000 USD up to 1,000,000 USD, the issue of constructing and managing a portfolio for hedge funds only concerns fund of hedge funds, sizable endowment funds, pension funds or other large financial institutions, rather than wealthy individuals. However, if the union of accessible hedge funds are reduced to only a few, e.g., less than ten, elites, a portfolio with optimized weights on component funds will be attractive to and much more affordable for both large financial institutions and individual wealthy investors.

It is the main thrust of this study to, in the first place, investigate whether the performance of outperforming hedge funds sustains over the holding period of a variety of length in time while using a recently proposed test, the stepwise test for superior predictive ability (or Step-SPA test introduced by Hsu, Hsu, & Kuan, 2010; hereafter HHK), to control for the data-snooping bias. The Step-SPA test improves on the test power of Romano and Wolf’s SRC. Second, this study endeavors to extend the effort above to test whether portfolio optimization techniques would enhance the performance of an equal-weighted portfolio of hedge funds outperformers. Having surveyed the recent development in the literature of optimization techniques for portfolio, this study applies the recently-developed probability global search Lausanne (hereafter PGS), the commonly-adopted genetic algorithm (hereafter GA) as well as the outdated Newton–Raphson algorithm (hereafter NR) as a benchmark. In this study, empirical results suggest that both equally-weighted and weights-optimized portfolios of good hedge funds screened by the Step-SPA test are generally able to outperform the corresponding HFR database index.

The rest of the article is organized as follows. Section 2 reviews the literature on performance persistence of hedge funds and on portfolio optimization. Section 3 briefly documents the dataset, the three optimization methods and the Step-SPA test while Section 4 discusses the empirical results and brings about some implications of them. Section 5 concludes.

2. Literature review

2.1. Performance persistence of hedge funds

Two main approaches to testing the performance persistence of mutual funds or hedge funds have been developed in the literature. The first approach examines the entire universe of funds using either the contingency table-based (nonparametric) method or the regression-based (parametric) method. In particular, a fund is referred to as a winner (loser) if its risk-adjusted return (alpha value or information ratio) is larger than (less than) the mean or median of the entire sample of funds. Following this categorization, Goetzmann and Ibbotson (1994), Brown and Goetzmann (2003) apply the Cross Product Ratio (CPR) test (also known as the Odds Ratio test) to test whether past winners (losers) tend to be next winners (losers). The CPR is given by

$$CPR = \frac{(\text{no. of winners at } t) \times (\text{no. of winners at } t - 1)}{(\text{no. of losers at } t) \times (\text{no. of losers at } t - 1)}$$

Accordingly, the CPR equals 1 under the null hypothesis of no persistence. The test statistic of the null hypothesis is given by

$$z = \frac{\log(CPR)}{\text{Std}(\log(CPR))}$$

where $\text{Std}(\log(CPR)) = \sqrt{\frac{1}{\text{no. of winners at } t - 1 \times \text{no. of losers at } t - 1} + \frac{1}{\text{no. of winners at } t - 1 \times \text{no. of losers at } t - 1} + \frac{1}{\text{no. of winners at } t - 1 \times \text{no. of losers at } t - 1} + \frac{1}{\text{no. of losers at } t - 1 \times \text{no. of winners at } t - 1}}$

Under the assumption of independent observations, this $z$-statistic will follow a standard normal distribution asymptotically. Agarwal and Naik (2000) adapt the method for multi-periods. Given a series of nine consecutive periods, for instance, the probability for a fund to be winner for nine straight periods is $(0.5)^9$ under the null hypothesis of no persistence. The probability for a fund to be winner for five straight periods and then loser for the remaining four periods is $(0.5)^3$. In this context, the Kolmogrov–Smirnov ($K–S$) test is employed to test whether the observed probability for a fund to be winner in all scenarios is significantly different from its theoretical counterpart as illustrated above. Statistical significance of the $K–S$ test would then indicate performance persistence.

In contrast with the non-parametric approach above, Edwards and Caglayan (2001) and Agarwal and Naik (2000) explore this issue using a regression-based (parametric) method. In particular, they calculated the risk-adjusted return (e.g., the alpha value by a factor model) for a series of consecutive periods. The alpha values of all funds of period $t-1$ are regressed upon those of period $t$. For the case of 10 consecutive sample periods, such regression will be performed nine times and one can apply the Fama–Macbeth test to check if the mean of the nine slope estimates is significantly positive. If yes, the universe of funds tested demonstrate performance persistence.

The second main approach, in contrast, sorts all funds into different equally-weighted portfolios according to the funds’ performance (e.g., return in excess of the risk-free rate or risk-adjusted return) during the portfolio-formation period, and tests whether each portfolio’s performance, if statistically significant, persists into the portfolio-holding period. Parametric (standard) $p$-value or KTWW's cross-sectionally bootstrapped $p$-value is used to test whether the portfolio’s alpha value is statistically significant. Their relative performance, i.e., the spread between the returns of the best portfolio and the poorest portfolio, can also be tested in this way to see if it persists across time. (See Carhart (1997)).

The factor models such as CAPM, Sharpe’s (1966) single-factor model, three-factor model built by Fama and French (1993), and Carhart’s (1997) four-factor model are commonly used to evaluate the abnormal returns of mutual funds. Mr. Lynch really owned talent and skill in selecting underpriced stocks. Chen, Jegadeesh, and Wermers (2000) study the stock positions and active trades
Many extensions to the basic Newton's method have appeared in approach the function's solution, which is analytically insoluble. It uses a tangent line on the objective function to this seminal numerical approach to optimization algorithms for adjusting the portfolio weights. Obradovic, Tang, and Thapar (2009) survey the following global the portfolio's conditional value-at-risk. For hedge funds, Minsky, and Parpas (2009) apply a global search algorithm to optimize to optimizing a portfolio's weights on individual assets. Maringer and Parpas (2009) apply a global search algorithm to optimize the higher order moments in portfolio selection. Krokhmal, Palmquist, and Uryasev (2002) devise a global search approach to optimizing the expected returns of a portfolio given constraints on the portfolio's conditional value-at-risk. For hedge funds, Minsky, Obradovic, Tang, and Thapar (2009) survey the following global optimization algorithms for adjusting the portfolio weights.

2.2.1. Newton’s method

In the book, entitled ‘Method of Fluxions,’ Newton introduces this seminal numerical approach to finding the optimal solution of a function. It uses a tangent line on the objective function to approach the function's solution, which is analytically insoluble. Many extensions to the basic Newton’s method have appeared in the literature such as Broyden, Dennis, and Moré’s (1973) quasi-Newton method.

2.2.2. Genetic Algorithm (GA)

Based on the principle of natural selection in the course of living being’s evolution, Holland (1975) introduces the genetic algorithm. Goldberg and Lingle (1985) applies it to the problem of optimization. In the context of portfolio construction, the GA selects weights on individual asset at random from the current population weight range for each asset to be parents. They are then used to generate the children for the next generation via crossing over the parents. During the process of the population’s evolution toward an optimal solution, the GA allows a small probability for mutation.

2.2.3. Simulated annealing

Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) documents the simulated annealing algorithm. It concerns using a probability search algorithm model that simulates the physical process of heating a material and then gradually cooling down its temperature to avoid defects, thus minimizing the system energy. At each step of the simulated annealing algorithm, the current solution is replaced by another solution which is selected depending on the difference between the functional values at the two points and the temperature variable, which is systematically lowered during the process. However, this method does not ensure global optimum. Kirkpatrick (1983) improves on Metropolis et al.’s (1953) work and provides a similar but better algorithm.

2.2.4. Probabilistic Global Search Lausanne (PGSL)

Raphael and Smith (2003) propose a global search algorithm based on the simulated annealing method mentioned above. Following an initial global search based on a uniform distribution assumed for all variables depending on the value of the target function, the PGSL intensifies the probability density on regions corresponding to a good value of the target function. By iteratively increasing the probability density for regions of a good solution and reducing the search space along these regions, the PGSL approaches the optimal solution.

2.2.5. Pattern search

The Pattern Search randomly selects a point (value) for each of all driving variables along the initial search space. Each of these initial values is multiplied by a vector (of a constant or variable length) to form a region for further search. If any new point within each generated region corresponds to a better solution for the target function, the initial point is replaced by the new point. Multiplying the new point by the vector above constructs a new search region which substitutes for the previous search region. Iteration of such a procedure will lead to the optimal solution. If the vector length is constant, such a method is called the Generalized Pattern Search (GPS). If the vector is of a variable length, in parallel, this method is called the Mesh Adaptive Search (MADS).

Minsky et al. (2009) suggest that the PGSL performs best among the methods introduced above in 21 out of 24 scenarios of nonlinear optimization problems. As a result, we apply this method in the following analysis. Due to its popularity in portfolio optimization, however, the study also includes the GA as a benchmark approach in addition to the Newton Raphson method.

3. Data and methodology

3.1. Data

The dataset was bought from the Hedge Fund Research, Inc and spans the September 1994 to August 2008 period. During the sample period, a total of 13098 hedge funds are covered by the dataset, including 5023 funds still existent by the end of our sample period, August 2008, and 8075 defunct funds. Including the universe of defunct funds help purge our results of the survivorship bias. As our focus is on the benefits of using a portfolio optimization technique on its performance, we also categorize the dataset into five sub-samples according to the main strategy defined by the dataset vendor, HFR. In particular, the five strategies are Marco, Event-Driven, Relative Value, Equity Hedge, and Fund of Funds. The monthly returns of each hedge fund, net of all fees, are entered into our workhorse, the Step-SPA test, to screen significant outperforming funds.

3.2. Step-SPA test

The literature substantially implements the bootstrap method to control for the data-snooping bias. White (2000) pioneers the use of bootstrapping by introducing the Bootstrap Reality Check (hereafter BRC). White’s BRC tests the null hypothesis under which the best model (strategy) in a large group of competitors cannot beat the benchmark model (strategy) in terms of a specific loss function (e.g., average daily returns for funds). However, the BRC method is less effective when the sample contains too many underperforming models (strategies). In light of this drawback,3 Funds might be moved from the live funds list to the defunct funds list for reasons such as liquidation due to bad performance, having been closed to new investors due to unmanageable size, being merged by other funds and so on.
Hansen (2005) devised a superior predictive ability (SPA) test, which increases the rejection rate of the null hypothesis by reducing the number of poor models in the sample. A drawback of both the BRC and the SPA test is that they only test the best model of the entire union. In the context of portfolio management, however, a fund-of-funds manager is inclined to pick as many good individual funds as possible. Thus, this study incorporates the Step-SPA test to screen as many good hedge funds as possible while controlling for the data-snooping bias.

For a typical investor who allocates assets among hedge funds or for a fund-of-funds manager, the challenge is how to identify the number of models for some variable and let \( d_{kj} \) denote the mean of the \( k \)-th element of the \( j \)-th bootstrap of Politis and Romano (1994) stationary bootstrap approach. The null in Eq. (1) would be rejected if

\[
\sqrt{T} \bar{d} - \mu \xrightarrow{D} N(0, \Omega),
\]

where \( \bar{d} = T^{-1} \sum_{t=1}^T d_{it} \) and \( \Omega \equiv \lim_{T \to \infty} \text{var}(\sqrt{T} (\bar{d} - \mu)) \), and the symbol \( \xrightarrow{D} \) denotes convergence in distribution. The assumption above allows \( d_{it} \) to be weakly dependent over time, which justifies Politis and Romano’s (1994) stationary bootstrap and ensures the heteroskedasticity and autocorrelation consistency of their kernel estimator for \( \Omega \).

Naive individual hypothesis testing such as in Eq. (1) will allow fortunate hedge funds to sneak in, which shapes the data-snooping problem given a good chunk of funds being tested collectively. A typical example is White (2000), which renders his test too conservative to reject false null hypothesis. In particular, let \( \Omega \) refer to an consistent estimator for \( \Omega \) with the \( (i, j) \)-element \( \hat{\sigma}_{ij} \). Let \( \hat{\sigma}_k^2 \equiv \hat{\sigma}_{kk} \) and \( A_{rt} = -\hat{\sigma}_k^2 (2 \log \log T)^1/2 \). Under the SPA test, the test statistic under the null is given by \( \text{SPA}_T = \max_{1 \leq k \leq m} \sqrt{T} \bar{d}_k \) which is similar to the BRC test. Furthermore, we define \( \bar{\mu} \) as a vector with the \( k \)-th element \( \bar{\mu}_k = \hat{\mu}_k \cdot 1(\bar{d}_k \in [\hat{\mu}_k - \hat{\sigma}_k, \hat{\mu}_k + \hat{\sigma}_k]) \), where \( 1(\cdot) \) denotes the indicator function. Since \( \sqrt{T} \bar{d}_k \) can be rewritten as \( \sqrt{T} (\bar{d}_k - \mu_k) + \sqrt{T} \mu_k \), adding \( \sqrt{T} \mu_k \) to the bootstrapped distribution of max \( (\sqrt{T} (\bar{d}_k - \mu_k)) \) re-centered the distribution leftward and enhances the rejection rate of the null under theSPA test relative to under the BRC test.

Another drawback of the BRC test is that it does not identify all models that significantly deviate from the null hypothesis. Rejecting the null hypothesis by the BRC test only suggests that there exists at least one model with \( \mu_k > 0 \). Based on the BRC test, Romano and Wolf (2005) propose a stepwise procedure that can identify as many models with \( \mu_k > 0 \) as possible, while asymptotically controlling the family-wise error (FWE) rate,\(^\dagger\) the probability of rejecting at least one of the correct hypotheses. This test, also known as the Step-RC test, is practically more useful than the BRC test. For example, a fund-of-funds manager ought to be more interested in finding out the funds that can beat the benchmark, rather than just knowing the best performing fund. To implement the stepwise procedure, HHK re-arrange \( \bar{d}_k \) in a descending order. A top model \( k \) would be rejected if \( \sqrt{T} \bar{d}_k \) is greater than the bootstrapped critical value, where bootstrapping is computed as in the BRC test. If none of the null hypotheses is rejected, the process stops; otherwise, we remove the rejected models from the data and bootstrap the critical value again using the remaining data. In the new sample, a top model \( i \) would be rejected if \( \sqrt{T} \bar{d}_i \) is greater than the newly bootstrapped critical value. The procedure continues until no more model can be rejected. Hansen (2005) and Romano and Wolf (2005) show that using studentized statistics \( \sqrt{T} \bar{d}_k / \hat{\sigma}_k \) would render the test more powerful.

Apart from studentization, HHK’s Step-SPA test re-centers the bootstrapped distribution of the null in Eq. (1) in the manner of Hansen (2005), and as a result, the Step-SPA test ought to be more powerful than the Step-RC test. In particular, the Step-SPA test is based on the following statistics: \( \sqrt{T} \bar{d}_k \) (for \( k = 1, 2, \ldots, n \)) and a stepwise procedure analogous to that of the Step-RC test. Using the stationary bootstrap of Politis and Romano (1994), let \( \bar{d}_k = \frac{\sum_{t=1}^T d_{it}^*}{T} \) be the sample mean of the \( b \)-th bootstrap resample, \( d_{it}^* \). Repeating this procedure \( B \) times yields an empirical distribution of \( \sqrt{T} \bar{d}_k / \hat{\sigma}_k \) with \( \hat{\sigma}_k \) re-centered. Given the pre-specified level \( \alpha_0 \) (equivalent to the FWE rate in the SRC framework), the bootstrapped SPA critical value is determined as

\[
\hat{q}^{*}_{\alpha_0} = \max_{1 \leq k \leq m} \left( \hat{q}_{\alpha_0}, 0 \right),
\]

with \( \hat{q}_{\alpha_0} = \inf \{ q : \text{P}[\sqrt{T} \max_{1 \leq k \leq m} \left( \bar{d}_k - \hat{\mu}_k + \hat{\sigma}_k \right) \leq 0] \geq 1 - \alpha_0 \} \), the \((1-\alpha_0)\)-th percentile of the re-centered empirical distribution, and \( \text{P} \) is the bootstrapped probability measure. The Step-SPA test with the pre-specified level \( \alpha_0 \) then proceeds as follows.

\(^\dagger\) The possibility of falsely rejecting at least one correct null hypothesis.
1. Re-arrange $\mathcal{D}_k$ in a descending order.
2. Reject the top model $k$ if $\sqrt{\mathcal{D}_k}$ is greater than $q_{\alpha}(\text{all})$, the critical value bootstrapped as in $\mathcal{D}_\text{sub}$, the critical value bootstrapped as in Eq. (3) using the complete sample. If no model can be rejected, the procedure stops; otherwise, go to next step.
3. Remove $\mathcal{D}_k$ of the rejected models from the data. Reject the top model $i$ in the sub-sample of remaining observations if $\sqrt{\mathcal{D}_i}$ is greater than $q_{\alpha}(\text{sub})$, the critical value bootstrapped as in Eq. (3) from the sub-sample. If no model can be rejected, the procedure stops; otherwise, go to next step.
4. Repeat the third step till no model can be rejected.

Based on simulation results, HHK suggest that their Step-SPA test is more powerful than Romano and Wolf’s Step-RC test. Note that the discussion of empirical results throughout this article is based on a level of 10%. However, results relevant to a level of 1% and 5% are similar but not reported for brevity.

Moreover, the Mean-Variance Method from Markowitz (1952) is used to describe the return and volatility of the portfolios. The concept is extended from the efficient frontier to measure the relationship between the returns and risks of portfolios. When the expected return $i\mu$, we want to find a set of weights of assets to minimize the risk (volatility) of the portfolio. Therefore, we use the Lagrange Multiplier Method to solve the solution. The process is based on a set of constraints on weights. First, the weights of all component hedge funds should sum to one. Second, there is no short selling allowed for any individual fund. Lastly, if a portfolio is composed of the significant outperforming funds selected, respectively, from the four categories of Macro, Event-Driven, Relative Value and Equity Hedge, we impose a minimum requirement of 5% on the sum of weights on all screened funds for each category. The benchmark index for such a portfolio is the HFRI Fund of Funds Composite Index which comprises about 800 individual funds across the four aforementioned categories of the HFR hedge fund database. In contrast, if a portfolio is composed of all significant outperformers selected from the entire database without categorization, there is no minimum requirement for the portfolio’s position in each category of funds. The benchmark index in this case is the HFRI Fund Weighted Composite Index, which comprises more than 2000 individual hedge funds across all fund types including Fund of Funds according to the HFR database.

3.3. Portfolio formation approaches

After screening all elite (statistically outperforming) hedge funds out of the entire universe during a pre-specified period by the Step-SPA test, we proceed to the next stage of how to form them into a portfolio to test its performance persistence. A naïve practice is to construct an equal-weighted (EW) portfolio. Testing the statistical significance of the performance of this EW portfolio over the holding period amounts to examining the ‘hot-hand’ effect (also known as the momentum effect). Without any type of portfolio optimization, if an EW portfolio of significant outperforming hedge funds proves persistent in performance across time, the ‘hot-hand’ effect is confirmed once again but from a data-snooping-bias free perspective.

It is our second overarching objective to check if applying an optimization technique can further enhance the performance of the EW approach. To achieve this goal, we apply the Step-SPA test separately to the four main sets of hedge funds categorized by the HFR database, Macro, Event-Driven, Relative Value and Equity Hedge. Given that a certain number of outperforming funds can be selected from each category, the following step is to form an optimal portfolio, the performance of which is then compared to those of the EW portfolio and the Fund of Hedge Funds benchmark index. Regarding the portfolio optimization method, we choose three approaches based on the results of the literature review discussed in Section II. In particular, we apply the PGSL, the GA and the NR method to form the optimal portfolio and compare all three sets of results. All three approaches should be applied with the following constraints on weights. First, the weights of all

3.3.1. The Probability Global Search Lausanne (PGSL)

Based on the concept of direct search, Raphael and Smith (2003) propose the PGSL. In the definition of Trosset (1998), a direct search for a numerically optimal solution is any algorithm that depends on the objective function only through ranking a countable set of function values. A direct method does not involve computing or approximating values of derivatives. It uses the value of the objective function only to decide whether a solution ranks higher than other solutions. The PGSL assumes that optimal solutions can be found through focusing search around sets of good solutions. The search space is sampled using a probability distribution function for each axis of the multi-dimensional search space. In the beginning of the search process, a uniform distribution with intervals of constant width is assumed as the PDF for each dimension. A probability distribution function is then updated by increasing probability and decreasing the width of intervals of the regions with good values for the objective function. During the process, a focus algorithm is used to gradually shrink the search space by modifying the minimum and maximum of each dimension of the search space. The PGSL framework consists of the following four nested cycles: the sampling cycle, the probability updating cycle, the focusing cycle, and the subdomain cycle. For more detail and the parameter setting, we refer to Raphael and Smith (2003).

3.3.2. The Newton method

In Method of Fluxions, Newton (1736) introduces his numerical approach to solving equations without closed-form solutions. To give a simple example, the solution is given by finding the intersection point of the function’s tangent line and the horizontal axis as follows.

Given a function $f(x)$ and an initial value $x_0$, the following iterative procedure leads to the final solution upon convergence of the difference between the updated solution and its previous version.

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}, \quad i = 0, 1, \ldots, k.$$  

In the multi-variate case, let $x_k = [x_1, \ldots, x_k]^T$ and $Hf(x)$ be the Hessian matrix of function $f(x)$ with $\nabla f(x)$ as its gradient. The solution is given by

$$x_{k+1} = x_k - [Hf(x_k)]^{-1}\nabla f(x_k), \quad k \geq 0.$$  

5 The HFRI Fund of Funds Composite Index is an equal-weighted index.

6 In contrast to the HFRI Fund of Funds Composite Index, the HFRI Fund Weighted Composite Index is a value weighted index.
Based on Newton’s approach, Broyden et al. (1973) suggest a variant of Newton’s method, the quasi–Newton method, as follows.

Let $B_k$ be the estimator of the Hessian matrix above and be defined by

$$
\Delta x_k = -\alpha_k(B_k)^{-1}\nabla f(x_k),
$$

where $\alpha$ is chosen according to the Wolfe conditions, and $x_{k+1} = x_k + \Delta x_k$. The gradient is given by $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ and the Hessian matrix can be updated by

$$
B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T \Delta x_k} - \frac{B_k \Delta x_k (B_k \Delta x_k)^T}{\Delta x_k^T B_k \Delta x_k}.
$$

Defining

$$
H_{k+1} = (B_{k+1})^{-1} = \left( I - \frac{y_k \Delta x_k^T}{y_k^T \Delta x_k} \right)^T H_k \left( I - \frac{y_k \Delta x_k^T}{y_k^T \Delta x_k} \right) + \frac{\Delta x_k \Delta x_k^T}{y_k^T \Delta x_k},
$$

the quasi–Newton method concerns substituting the new Hessian matrix into the Newton method.

### 3.3.3. The Genetic Algorithm (GA)

Holland (1975) proposes the genetic algorithm (or GA) for searching the optimal solution based on Charles Darwin’s theory that living things evolve by natural selection. At each step, the GA selects individuals at random from the current population to be parents and uses them to produce the children for the next generation. During the evolution process, there is a small probability of mutation. In this way, the population evolves toward an optimal solution.

This artificial intelligence-type technique features a random multiple-point computation, which would discover an optimal solution with a less opportunity to fall within local optimum than the one by the conventional single-point computation. The GA also transforms the problem of interest into a continuous and differentiable fitness function with coded parameters, which provides more possibility of finding the optimal solution than the conventional algebra. For the GA to be applicable to an empirical problem, it involves steps including configuring the framework for the fitness function, coding, selection, crossover rate and mutation rate. Given the absence of a standard guideline for setting the parameters’ values of the GA framework, we refer to Lin’s (1997) empirical results and employ the following set of parameters: 100 for the size of population, 0.8 for the crossover rate, 0.03 for the mutation rate, 3000 for the number of generations.

### 3.3.4. Benchmark indexes

To avoid the hind-sight problem of assessing the performance of equal-weighted or artificially optimized portfolio of the elite hedge funds based on historical data, it is necessary to form the portfolio, hold it for a period of time, and test its performance during the holding period relative to some benchmark index. We use the HFR Fund of Hedge Funds Index to gauge the performance of portfolios consisting of significant outperforming funds selected, respectively, from the first four sub-samples, i.e., Macro, Event-Driven, Equity Hedge and Relative Value. We turn to the HFR Hedge Fund Composite Index to measure the performance of portfolios consisting of significant outperforming funds selected from all five types of funds pooled together.

### 3.3.5. Objective function (I): Sharpe ratio

Since the first stage of forming the optimal FoHF portfolio is to identify all the significant outperforming funds via the Step-SPA test, an equal-weight version of such a portfolio is supposed to perform best among its equal-weight competitors. In such a circumstance, the conventional Markowitz mean–variance approach to finding the optimal risky portfolio by maximizing the portfolio’s Sharpe ratio seems an intuitive method of improving on the equal-weight portfolio’s performance. The Sharpe ratio of a portfolio is given by $\frac{E(r_p - r_f)}{\sigma_p}$, where $r_p$ and $r_f$ denote, respectively, the portfolio’s return and the risk-free asset return, and $\sigma_p$ the standard deviation of the portfolio’s excess return, i.e., $r_p - r_f$. Nevertheless, the constraint on short-selling any individual hedge fund complicates the application of the mean–variance approach. To circumvent this problem, we adopt the Newton–Raphson algorithm, which is used by much optimizer software such as the Solver in the Microsoft Excel package, to come up with the optimal set of weights in the context of forbidden short-selling on any individual fund.

#### 3.3.6. Objective function (II): proximity to the benchmark index’s systematic risk

The literature on portfolio optimization suggests that the mean–variance technique is not a guaranteed performance booster if it is applied based on the historical data. Incorporating forecast views of the future might overcome the failure of the mean–variance approach. However, predicting the future outcome of a hedge fund is not a trivial task and is beyond the scope of this research. We cling to the historical data and, as a result, the constrained optimal portfolio under the mean–variance framework is not necessarily superior to its equal-weight counterpart according to the literature.

The second objective function we would like to use stems from the concept of mimicking the systematic risk of a benchmark hedge fund index but in the hope of constructing a portfolio with a return better than that of the index. The rationale behind this practice is that this study aims to devise a passive–active strategy by forming a portfolio of funds each with a significant, positive alpha (or studentized alpha) and allowing its risk coming from the set of systematic risk factors to track the portfolio’s relevant benchmark index. The former constitutes the active part of the strategy, while the latter shapes up the passive part. As a result, the second objective function is called the proximity to the benchmark index’s systematic risk. Mathematically formulated, it is given by

$$
\sum_{i=1}^{7} \beta_i \hat{\alpha}-\hat{\alpha},
$$

where $\beta_i$ and $\hat{\alpha}$ refer to the parameter estimate for the $i$-th risk factor of the FH seven factor model in Eq. (4) applied to the returns of the FoHF portfolio and of the benchmark index, respectively. Minimizing the objective function using any of the three optimization methods gives the optimal set of weights under all weight constraints detailed in the 3.3 section.\footnote{Another objective function can be $\sum_{i=1}^{7} (\beta_i - \hat{\beta}_i)^2$. However, it turns out that all three optimization approaches deliver the same set of optimal weights. This finding suggests the possibility of a closed-form solution for the optimal set of portfolio weights with respect to the minimal $\sum_{i=1}^{7} (\beta_i - \hat{\beta}_i)^2$.}

$$
r_t^t = \alpha + \sum_{k=1}^{7} \beta_k F_{k,t} + e_t,
$$

where $r_t$ is the month risk-premium return of the $i$-th Hedge Fund at time $t$. $\alpha$ is the external non-normal return of the $i$-th Hedge Fund. $\beta_i$ is the regression coefficient with respect to the $k$-th factor. $F_{k,t}$ is the month return of the $k$-th factor at time $t$, which the factors are as follows: S&P 500 return minus risk-free rate, Wilshire small cap minus large cap return, Change in the constant maturity yield of the 10-year Treasury, Change in the spread of Moody’s Baa minus the 10-year Treasury, Return of a portfolio of lookback straddles on bond futures, Return of a portfolio of lookback straddles on foreign
It is therefore neither feasible nor efficient for an investor to hold a hedge fund for less than a year. On the other hand, 3 yrs for the holding period should be long enough for the average hedge fund to sustain good performance. The literature on performance persistence of funds discussed above also involves holding periods equal to or less than 3 yrs.

The results contained in Table 2 suggest that the Step-SPA test tends to identify more outperformers under the studentized alpha statistic than under the basic alpha statistic, irrespective of fund categorization and holding period. This finding is in accord with the fact that the studentized alpha statistic measures the size of a fund’s abnormal return given a fixed level of idiosyncratic risk a fund has taken over the data period. On the contrary, the basic alpha statistic only measures the abnormal return itself. As a result, the variance of the alpha statistic across different funds is supposed to be larger than that of the studentized alpha statistic. Given that the bootstrapped null distribution under the Step-SPA test might well be dominated by those funds with a poor and volatile performance, the Step-SPA test power is increased by studentizing the test statistic (see Hansen (2005)).

Since the Step-SPA test is not able to identify any outperforming fund in several 7-year periods for individual fund categories and for the entire uncategorized set of funds when alpha is used as the criterion, the following discussion is based on results of the 7-year formation period only. The results of all other formation periods are similar and available upon request from the authors.

4. Empirical results

4.1. Portfolio formation periods

Taking into account the various market conditions across time, the study rolls over both the portfolio formation period and the holding period so that the hedge fund performance persistence could be tested based on a span of the hedge fund history as long as possible. Moreover, both the formation period and the holding period are switched forward by the length of the holding period to ensure that adjacent holding periods are non-overlapping. In particular, Table 1 summarizes all formation periods across time. Since hedge fund returns are observed every month at most, it would not make sense to use only one year of historic data, i.e., 12 monthly returns, for the Step-SPA test. The study thus begins with two years or 24 months for the length of the formation period. Given that both long-standing and short-lived (or young) funds had existed at the same time in the history of hedge funds, this study also examines 3, 4, 5, 6 and 7 years. For the sake of brevity, however, the following discussion is based on results of the 7-year formation period only. The results of all other formation periods are similar and available upon request from the authors.

4.2. Portfolio holding periods

As regards the span of the holding period, the study examines 1, 2 and 3 yrs. For a typical hedge fund, an advance notice from the investor is required at least six months before his/her redemption. It is therefore neither feasible nor efficient for an investor to hold a hedge fund for less than a year. On the other hand, 3 yrs for the holding period should be long enough for the average hedge fund to sustain good performance. The literature on performance persistence of funds discussed above also involves holding periods equal to or less than 3 yrs.

The results contained in Table 2 suggest that the Step-SPA test tends to identify more outperformers under the studentized alpha statistic than under the basic alpha statistic, irrespective of fund categorization and holding period. This finding is in accord with the fact that the studentized alpha statistic measures the size of a fund’s abnormal return given a fixed level of idiosyncratic risk a fund has taken over the data period. On the contrary, the basic alpha statistic only measures the abnormal return itself. As a result, the variance of the alpha statistic across different funds is supposed to be larger than that of the studentized alpha statistic. Given that the bootstrapped null distribution under the Step-SPA test might well be dominated by those funds with a poor and volatile performance, the Step-SPA test power is increased by studentizing the test statistic (see Hansen (2005)).

Since the Step-SPA test is not able to identify any outperforming fund in several 7-year periods for individual fund categories and for the entire uncategorized set of funds when alpha is used as the criterion, the following discussion is based on results of the 7-year formation period only. The results of all other formation periods are similar and available upon request from the authors.

4.3. Performance persistence test

The study tests whether the good performance of hedge fund elites sustains over a holding period of 1 through 3 years using a regression method.\textsuperscript{8} In case of disappearance of any selected fund
in the portfolio from the database during the holding period, the fund’s weight is proportionately allocated to all surviving members of the portfolio. If none of the portfolio’s component funds is able to sustain over the holding period according to the database, this study assumes 100% of recovery of the assets value in the month before the portfolio’s breakdown and uses the 3-month US T-bill rate as the capital’s return for the remaining months before the next portfolio reshuffle based on the Step-SPA test.

The results of Tables 3 and 4 show that the alpha of the EW portfolio is statistically significant and positive irrespective of whether the elite portfolio is made up of funds selected independently from the four categories or from all uncategorized funds. This finding suggests the hot-hand effect among the hedge funds reported by the HFR hedge fund dataset; namely, the past winner fundamentally from the four categories or from all uncategorized funds.

Notes. 1. The equal-weighted portfolio is formed by equally weighting those outperforming funds selected independently from each of all four fund categories. 2. The test statistic of the Step-SPA test used is the t-statistic of fund alpha. The Step-SPA test is employed to identify statistically outperforming funds relative to the Fung and Hsieh 7-factor model during a series of continuously switching period of 7 years. 3. The HFR Fund of Funds Index’s performance is identical for holding period of both 2 and 3 years because the two concatenated time series of consecutive and non-overlapping holding periods correspond to the same Sep. 2002-Aug. 2008 period.

Table 3

<table>
<thead>
<tr>
<th>Formation period: 7Y</th>
<th>Holding period</th>
<th>Mean return (%)</th>
<th>Cumulative return (%)</th>
<th>Sharpe ratio</th>
<th>Alpha (%)</th>
<th>Studentized alpha</th>
<th>Information ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>EW</td>
<td>1y</td>
<td>0.5315</td>
<td>55.7350</td>
<td>1.5432</td>
<td>0.3052</td>
<td>(0.001)</td>
<td>3.6320</td>
</tr>
<tr>
<td></td>
<td>2y</td>
<td>0.4928</td>
<td>42.2838</td>
<td>1.5891</td>
<td>0.1958</td>
<td>(0.006)</td>
<td>2.8236</td>
</tr>
<tr>
<td></td>
<td>3y</td>
<td>0.6127</td>
<td>55.1313</td>
<td>3.1889</td>
<td>0.3466</td>
<td>(0.001)</td>
<td>7.1280</td>
</tr>
<tr>
<td>HFR Fund of Funds Index</td>
<td>1y</td>
<td>0.4818</td>
<td>48.7423</td>
<td>0.7605</td>
<td>0.1700</td>
<td>(0.137)</td>
<td>1.5020</td>
</tr>
<tr>
<td></td>
<td>2y</td>
<td>0.5488</td>
<td>47.3929</td>
<td>0.8915</td>
<td>0.1643</td>
<td>(0.243)</td>
<td>1.1795</td>
</tr>
<tr>
<td></td>
<td>3y</td>
<td>0.5488</td>
<td>47.3929</td>
<td>0.8915</td>
<td>0.1643</td>
<td>(0.243)</td>
<td>1.1795</td>
</tr>
</tbody>
</table>

Notes. 1. The equal-weighted portfolio is formed by equally weighting those outperforming funds selected independently from each of all four fund categories. 2. The test statistic of the Step-SPA test used is the t-statistic of fund alpha. The Step-SPA test is employed to identify statistically outperforming funds relative to the Fung and Hsieh 7-factor model during a series of continuously switching period of 7 years. 3. The HFR Fund of Funds Index’s performance is identical for holding period of both 2 and 3 years because the two concatenated time series of consecutive and non-overlapping holding periods correspond to the same Sep. 2002-Aug. 2008 period.

Table 4

<table>
<thead>
<tr>
<th>Formation period: 7Y</th>
<th>Holding period</th>
<th>Mean return (%)</th>
<th>Cumulative return (%)</th>
<th>Sharpe ratio</th>
<th>Alpha (%)</th>
<th>Studentized alpha</th>
<th>Information ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>EW</td>
<td>1y</td>
<td>0.5756</td>
<td>61.7347</td>
<td>2.3212</td>
<td>0.3410</td>
<td>(&lt;0.001)</td>
<td>4.7636</td>
</tr>
<tr>
<td></td>
<td>2y</td>
<td>0.5448</td>
<td>47.7117</td>
<td>2.1159</td>
<td>0.2652</td>
<td>(0.002)</td>
<td>3.2975</td>
</tr>
<tr>
<td></td>
<td>3y</td>
<td>0.6174</td>
<td>55.7039</td>
<td>4.8552</td>
<td>0.3705</td>
<td>(&lt;0.001)</td>
<td>10.7515</td>
</tr>
<tr>
<td>Composite HFR Hedge Funds Index</td>
<td>1y</td>
<td>0.6499</td>
<td>70.6666</td>
<td>1.0405</td>
<td>0.3219</td>
<td>(&lt;0.002)</td>
<td>3.2241</td>
</tr>
<tr>
<td></td>
<td>2y</td>
<td>0.7489</td>
<td>69.8331</td>
<td>1.3103</td>
<td>0.3163</td>
<td>(0.013)</td>
<td>2.5510</td>
</tr>
<tr>
<td></td>
<td>3y</td>
<td>0.7489</td>
<td>69.8331</td>
<td>1.3103</td>
<td>0.3163</td>
<td>(0.013)</td>
<td>2.5510</td>
</tr>
</tbody>
</table>

Notes. 1. The equal-weighted portfolio is formed by equally weighting those outperforming funds selected together from all pooled, uncategorized funds. 2. The test statistic of the Step-SPA test used is the t-statistic of fund alpha. The Step-SPA test is employed to identify statistically outperforming funds relative to the Fung and Hsieh 7-factor model during a series of continuously switching period of 7 years. 3. The Composite HFR Hedge Funds Index’s performance is identical for holding period of both 2 and 3 years because the two concatenated time series of consecutive and non-overlapping holding periods correspond to the same Sep. 2002-Aug. 2008 period.

Table 5

<table>
<thead>
<tr>
<th>Formation period: 7Y</th>
<th>Holding period</th>
<th>Mean return (%)</th>
<th>Cumulative return (%)</th>
<th>Sharpe ratio</th>
<th>Alpha (%)</th>
<th>Studentized alpha</th>
<th>Information ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>EW</td>
<td>1y</td>
<td>0.5315</td>
<td>55.7350</td>
<td>1.5432</td>
<td>0.3052</td>
<td>(0.001)</td>
<td>3.6320</td>
</tr>
<tr>
<td></td>
<td>2y</td>
<td>0.4928</td>
<td>42.2838</td>
<td>1.5891</td>
<td>0.1958</td>
<td>(0.006)</td>
<td>2.8236</td>
</tr>
<tr>
<td></td>
<td>3y</td>
<td>0.6127</td>
<td>55.1313</td>
<td>3.1889</td>
<td>0.3466</td>
<td>(&lt;0.001)</td>
<td>7.1280</td>
</tr>
<tr>
<td>NR</td>
<td>1y</td>
<td>0.5797</td>
<td>60.2157</td>
<td>0.7284</td>
<td>0.3804</td>
<td>(0.074)</td>
<td>1.8147</td>
</tr>
<tr>
<td></td>
<td>2y</td>
<td>0.5561</td>
<td>48.4861</td>
<td>1.1327</td>
<td>0.2887</td>
<td>(0.019)</td>
<td>2.4126</td>
</tr>
<tr>
<td></td>
<td>3y</td>
<td>0.8567</td>
<td>83.4863</td>
<td>1.6205</td>
<td>0.4800</td>
<td>(0.002)</td>
<td>3.1890</td>
</tr>
<tr>
<td>PGSL</td>
<td>1y</td>
<td>0.6033</td>
<td>64.2563</td>
<td>0.9655</td>
<td>0.3849</td>
<td>(0.023)</td>
<td>2.3149</td>
</tr>
<tr>
<td></td>
<td>2y</td>
<td>0.4876</td>
<td>41.4592</td>
<td>0.9702</td>
<td>0.2302</td>
<td>(0.044)</td>
<td>2.0501</td>
</tr>
<tr>
<td></td>
<td>3y</td>
<td>0.8337</td>
<td>80.6240</td>
<td>1.6396</td>
<td>0.4577</td>
<td>(0.002)</td>
<td>3.2421</td>
</tr>
<tr>
<td>GA</td>
<td>1y</td>
<td>0.5947</td>
<td>63.3432</td>
<td>1.0397</td>
<td>0.3719</td>
<td>(0.016)</td>
<td>2.4578</td>
</tr>
<tr>
<td></td>
<td>2y</td>
<td>0.4268</td>
<td>35.4866</td>
<td>0.8032</td>
<td>0.1758</td>
<td>(0.06)</td>
<td>1.6388</td>
</tr>
<tr>
<td></td>
<td>3y</td>
<td>0.7737</td>
<td>73.3284</td>
<td>1.6998</td>
<td>0.4236</td>
<td>(0.001)</td>
<td>3.3832</td>
</tr>
</tbody>
</table>

Notes. 1. The equal-weighted portfolio is formed by equally weighting those outperforming funds selected independently from each of all four fund categories. 2. To form a portfolio using any of the three portfolio optimization approaches, a 5% minimum weight requirement is imposed on each fund category, of which at least one fund is selected by the Step-SPA test. 3. The Composite HFR Hedge Funds Index’s performance is identical for holding period of both 2 and 3 years because the two concatenated time series of consecutive and non-overlapping holding periods correspond to the same Sep. 2002-Aug. 2008 period.
4.4. Performance of portfolio optimization

Having confirmed the hot-hand effect in the previous subsection, this study proceeds to discuss whether using any of the three portfolio optimization techniques is able to enhance the performance generated by the naive equal-weight method. The relevant results are reported in Tables 5 and 6. The results of Table 5 suggest a mixed finding in that the mean return, cumulative return and alpha of all three optimization approaches tend to be superior to the naive EW method, while, on the contrary, the Sharpe ratio, studentized alpha and information ratio of the three approaches tend to be tarnished by the EW method. The results of Table 6 show that the three optimization approaches generally outperform the naive EW method in terms of both pre-risk-adjustment and risk-adjusted measures. A pronounced exception occurs in the case of 3-year holding period, in which the Sharpe ratio, studentized alpha and information ratio of the portfolio constructed by the three optimization approaches are less than the EW portfolio.

In addition, the results of Table 5 do not suggest a clear-cut ranking among the relative performance of the three optimization approaches. It all depends on the length of holding period and which measure is used. However, we observe that the PGSL tends to generate the most stable performance across the three choices of holding period if evaluated by any of the four risk-adjusted measures. Taking the Sharpe ratio as reported by Table 5 as an example, PGSL gives the worst 0.9655 for the 1-year holding period and the best 1.6998 for the 3-year holding period. The range between the worst case and the best case is 0.7441. In comparison with the Sharpe ratio’s min—max range of the PGSL, the min—max ranges of the NR and the GA are, respectively, 1.6205—0.7284 (¼0.8921) and 1.6998—0.8032 (¼0.8966).

In a nutshell, the results of this subsection suggest that the three optimization approaches are able to provide a better performance than what the naive EW method does under the circumstance of selecting elite funds from the four fund categories independently using the Step-SPA test. Moreover, the PGSL tends to generate the most stable risk-adjusted performance across the three different lengths of holding period among the three optimization approaches.

5. Concluding remarks

Our empirical results tend to suggest that both equally-weighted and weights-optimized portfolios of good hedge funds screened by the Step-SPA test are generally able to outperform the corresponding HFR database benchmark index, regardless of the performance measure used. Moreover, the difference among the equally-weighted portfolio and the weights-optimized portfolio of elite hedge funds is not significantly different in terms of their alpha estimate. However, the standardized alpha of the weights-optimized portfolio tends to be obviously larger than that of the equally-weighted portfolio. Of all three optimization techniques, the PGSL tends to outperform the Newton–Raphson and Genetic algorithms.

As regards the optimal length of portfolio formation period, the results are not able to suggest a consistent choice. However, we do find it best to hold the portfolio for one year relative to two and three years.

References


