UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Master 1 Econometrics, statistics

# Evaluation of the outperformance of hedge funds


| *Authors* | *Id* |
|---|---|
| Chaimaa Bentalha | 11 706 034 |
| Awa Keita | 11 700 102 |
| Zhanike Yevniyeva | 12 011 178 |


Presented to:
Rania Hentati-Kaffel, Philippe De Peretti

January 24, 2021

# Abstract

In this paper, we conduct various tests to analyze the absolute and relative performance of returns of hedge funds indices following different strategies in the BarclayHedge database for monthly data from 1997 to 2020. The main objective of our tests is to study the randomness theory for different strategies using Wald-Wolfowitz runs tests, autocorrelation tests based on the ARIMA model. Based on the two-tailed runs tests with the null hypothesis of randomness and one-tailed runs tests allowing to check for clustering we consider clustering for the Equity Market Neutral Index, the European Equities Index, the Event-driven Index, the Fixed Income Arbitrage Index and the Multi-strategy Index. We use Ljung-Box statistics, SCAN method, as well as Dickey-Fuller tests to investigate autocorrelation orders and conclude the stationarity of strategies indices datasets. Confirmed by many existing pieces of research, we can validate the non-normality of the hedge funds strategies indices performance. Based on obtained performance measures, such as Sortino ratio, the CAGR and the ECDF we assume that nine strategies outperformed the selected S&P500 and DJIA benchmarks, with Equity Market Neutral, Global Macro and European Equities strategies with a higher level of outperformance certainty. Nevertheless, we highlight the influence of biases because we cannot perceive our conclusions isolatedly from the biases that are pertinent to hedge funds and might affect the performance persistence. Finally, despite the results of the tests and crisis outperformance of the hedge funds strategies indices, there is a perpetual issue emblematic for hedge funds of high managerial fees that impend the overall performance of selected indices. Thus, despite plausible pertinence the fees might enforce a substantial drag on the returns that investors earn on their investments.

# Contents

# Introduction

The first semblance of hedge fund reportedly appeared in 1949 by Alfred Winslow Jones. Nonetheless, it was only in 1966 when hedge funds could draw the financial analysts attention when a Fortune article reported that Alfred Winslow Jones' hedge funds had significant returns compared to mutual funds (*The Amazing Story Behind The World's First Hedge Fund.* 2016). Since then, hedge funds have been attracting investors interest and have been growing exponentially.

Thus, the hedge fund industry is still generally perceived as being able to outperform the market's capacity to generate positive returns, but financial crises might question this capacity (Hentati-Kaffel and De Peretti, 2015). Even today it remains difficult to define a hedge fund, but here we propose to define it as a common investment vehicle that is actively managed by a limited number of investors. The performance of a hedge fund is measured in units of absolute return. The performance of hedge funds is much discussed in the literature, and more precisely its persistence. It refers to the ability of a hedge fund to have regular performances that can be positive or negative (Connor and Woo, 2004). Studies agree on finding persistence in the performance of hedge funds in the short term (between one and six months). In the best-case scenario, some hedge funds record a persistence of performance over a full year. Moreover, performance differs according to strategy (Stafylas et al., 2016).

The purpose of this report is to investigate the capacity of hedge funds to outperform more traditional financial management approaches. In this paper, we present the results of the runs tests on different strategies to test the hypothesis of randomness (via two-sided tests) and clustering (via one-sided tests). We also look at the results of autocorrelation tests conducted with ARIMA procedure, and the results of standard statistical tests, such as normality or goodness of fit. Based on the computations of the Sortino ratio and the Compound Annual Growth Rate, we compare the performance of our strategies to benchmarks (S&P500 and Dow Jones Average Index). Finally, based on our tests, we are able to conclude the non-randomness and persistence of selected strategies, and their possible outperformance comparing with the benchmarks. For example, based on the performance measures assessment out of nine selected strategies Equity Market Neutral, Global Macro and European Equities strategies indices demonstrate the most reliable performance. The paper is structured as follows: In Section 1 we present the database selection and present our methodological approaches; in Section 2 we present the results of runs tests, autocorrelation tests, normality tests and performance ratios calculation; in Section 3 we present our findings considering the performance comparison, hedge funds specific biases, the impact of hedge funds managerial fees, as well as comparison of indices performance during crisis periods.

# 1 Data and approach

In this section, we justify our choice of database and provide an overview of our strategies as well as particular basic statistical measures. Besides, we present the summary of the methodology implemented to answer our problem.

## 1.1 Database selection

In order to carry out our examination paper on hedge fund performance, we had to find sufficient and reliable data. We decided to retrieve our databases from the BarclayHedge database provider. Indeed, the site allows us to access 31 distinct indices grouping data from various funds whenever possible. In addition, for each index, we have the quantity of funds included for the calculations. Another important feature and advantage of the BarclayHedge indices database is its accessibility comparing with existing hedge funds databases. The indices are updated as and when the monthly returns are reported by the hedge funds. For each index, the database provides us with monthly performance data from January 1997 to December 2020. For November and December 2020 performances, these are estimated performances.

The choice of database is of paramount importance in the analysis of hedge fund performance because hedge funds are not managed in the same way. Thus, databases have biases and these biases differ from one database to another. Moreover, the number of hedge funds differs from one database to another (there are 10,520 funds in total for BarclayHedge). Compared to other databases such as EurekaHedge, the BarclayHedge database (like those of HFR and TASS) yields a lower survival bias implying lower average returns but a higher persistence of (short-term) performance. Moreover, the BarclayHedge database is characterized by its low proportion of missing observations for Asset Under Management, indeed, a 12% missing observation is recorded against 36% for EurekaHedge (it is 19% for the HFR database) (Joenväärä et al., 2012).

Therefore, a more simplified access to the BarclayHedge database and a lower bias justifies our choice to conduct our analysis using data from this provider.

## 1.2 Description of strategies

For our study we have selected nine different indices for different strategies: Equity Long Bias Index, Equity Long / Short Index, Equity Market Neutral Index, European Equities Index, Event-Driven Index, Fixed Income Index, Global Macro Index, Multi-Strategy Index and Technology Index (Table 1).

- The **Equity Long Bias** strategy of long bias funds is a combination of a neutral fund and a long term fund. A long bias fund will generally have a ratio of long positions of at least 40% compared to short positions (a long / short fund can become a long bias fund and vice versa depending on the asset allocation). This strategy allows, by maintaining a long bias, to take advantage of market rises that would not necessarily have taken place if the assets were continuously redistributed in the different positions.

| Strategy Index | December ROR | Number of funds reporting | YTD through December |
|---|---|---|---|
| Equity Long Bias Index | 4.75% | 364 | 16.02% |
| Equity Long / Short Index | 3.53% | 254 | 9.26% |
| Equity Market Neutral Index | 0.70% | 76 | -1.14% |
| European Equities Index | 2.65% | 56 | 4.51% |
| Event Driven Index | 4.06% | 64 | 11.16% |
| Fixed Income Arbitrage Index | 0.73% | 19 | 10.43% |
| Global Macro Index | 3.54% | 98 | 9.53% |
| Multi Strategy Index | 2.64% | 112 | 4.63% |
| Technology Index | 5.09% | 33 | 34.14% |

Table 1: Overview of used Strategy Indices from The BarclayHedge Indices database

- The objective of the **Equity Long Short** strategy is regional or sector-specific oriented. The objective is to invest shares in the long and short sides of the market. The size of the capitalization is of insignificant importance to managers who can move from a net long position to a net short position and they are able to hedge their risk by using futures and options.

- The **Equity Market Neutral** strategy consists of simultaneously creating portfolios of the same size equities in the same country, which are matched in long and short positions. Neutral portfolios that will endeavour to be beta or currency neutral can be created, while other well-constructed portfolios will constrain certain risks such as exposure to diverse existing sectors. The manager has a leverage effect to improve his returns.

- The index for the **European Equitie**s strategy is an arithmetic average of the net returns of the funds in this category. These funds have a minimum allocation of 90% of the portfolio to Western European countries and the United Kingdom. The strategy consists of investing in equities on the long and short sides of the market.

- In the **Event-Driven** strategy, the emphasis is on events considered special. The objective is to invest in the occurrence of these events in order to capture the price movements generated. These events can be the bankruptcy, restructuring or liquidation of a company.

- The objective of the **Fixed Income Arbitrage** strategy is to take advantage of possible price anomalies between financial securities with related interest rates. Global arbitrage allows for significant and consistent returns.

- The **Global Macro** strategy is a hedge fund strategy that takes into account general trends and economic events in the market. Investments are made worldwide and the portfolios of these funds may include equities, derivatives or commodities.

- The **Multi-Strategy** index is a concentrate of funds that allocate their capital between strategies belonging to several hedge fund disciplines, making it difficult to allocate them to a particular category.

- The **Technology** strategy focuses on one sector in particular, which is the technology sector. The index represents an average of the net returns of the funds in the category for which Barclays has data for each month.

## 1.3 Strategies Indices Overview

Table 2 presents the basic statistical measures for the selected strategies. Preliminarily we might assume that the strategies with the most important mean value (Equity Long Bias and Technology indices) are also the most volatile.

| Strategy | N | Mean | Median | Mode | Std Deviation | Variance |
|---|---|---|---|---|---|---|
| Equity Long Bias Index | 287 | 0.00812187 | 0.0095 | 0.008 | 0.03355896 | 0.0011262 |
| Equity Long / Short Index | 287 | 0.00677192 | 0.0062 | -0.0049 | 0.02002019 | 0.00040081 |
| Equity Market Neutral Index | 287 | 0.00362718 | 0.0036 | 0.004 | 0.00859829 | 0.00007393 |
| European Equities Index | 287 | 0.00728491 | 0.0066 | 0.0038 | 0.02261634 | 0.0005115 |
| Event-Driven Index | 287 | 0.00680488 | 0.0082 | 0.0093 | 0.02052626 | 0.00042133 |
| Fixed Income Arbitrage Index | 287 | 0.00462439 | 0.00462 | 0.0058 | 0.01329479 | 0.00017675 |
| Global Macro Index | 287 | 0.00556491 | 0.0042 | -0.0059 | 0.01616934 | 0.00026145 |
| Multi Strategy Index | 287 | 0.00609375 | 0.0072 | 0.0043 | 0.01335173 | 0.00017827 |
| Technology Index | 287 | 0.00975087 | 0.0091 | 0.0028 | 0.03694471 | 0.00136491 |

Table 2: Strategies indices basic statistical measures

## 1.4 Methodology

In the literature, different models have been implemented to test the ability of hedge funds to outperform more traditional funds (benchmarks). In our study, we were led to conduct various specific tests to address our problem. We began by conducting runs tests to test the randomness of our data. The Wald-Wolfowitz test is relevant because it allows us to account for the differences that exist between two data distributions and in our case we compared different strategies with benchmarks. Classic Wald-Wolfowitz's test supposes as follows: the null hypothesis of two populations of data having the same distribution and the alternative hypothesis of different distribution. Wald-Wolfowitz's method consists of counting the number of runs to study the similarities between the two populations of data (our strategies and our benchmarks). Compared to other tests, the Wald-Wolfowitz test is potentially the best performing non-parametric test on nonuniform samples of the same size (Magel and Wibowo, 1997). In addition, the test can be conducted as a two-sided and one-sided test as it was inquired in recent related research (Hentati-Kaffel and De Peretti, 2015).

To test the auto-correlation in our data we used the ARIMA procedure. This procedure is applicable considering our data being time series (with the regular time interval) and continuous and allows us to report a possible auto-correlation (Mahan et al., 2015).

To test the normality in our data, we used the univariate procedure under SAS which outputs the descriptive statistics for the dataset. The univariate procedure is simple to implement on SAS and is quite complete because it allows us to assess normality by doing normality tests such as the Shapiro-Wilk test, the Kolmogorov-Smirnov test, the Anderson-Darling test and the Cramér-von Mises test.

As of the final phase of our analyses, we conducted performance tests. To do so, we looked at the Sortino ratio and the CAGR. The Sortino ratio is recognised as a more preferable measurement tool to the Sharpe ratio because in the calculation of the Sortino ratio the risk is considered as an underperformance compared to the benchmark (here S&P 500 or DJIA). Moreover, the greater the number of observations, the less the Sortino ratio is biased in relation to the Sharpe ratio. The two measurement instruments are roughly equal when the observations are normally distributed, but this is not always the case for hedge funds (Chaudhry and Johnson, 2008). Finally, we calculate the CAGR to smooth the returns of our strategies in order to compare them more easily with our benchmarks.

# 2 Tests

## 2.1 Wald-Wolfowitz (Runs) test for randomness

Run is defined as an increasing or decreasing values series. The foundation of the general runs test is that within a random data set the probability of the following value in a sequence is larger or smaller than the previous one is following a binomial distribution (Bradley, 1968). In order to investigate randomness in hedge funds indices performance over time we use runs test statistical procedure also known as Wald-Wolfowitz test (Wald and Wolfowitz, 1940). This type of test is often used to address observation of randomness in time series. Wald-Wolfowitz was also used to test the presence of contagion during crisis periods along with the Z-test (Kim and Lee, 2017).

### 2.1.1 Wald-Wolfowitz Test Description

Generally, the Wald-Wolfowitz runs test is employed to verify if two independent samples issue from the same population. In our case, we conduct the test in several stages firstly normalizing the model using the median value of the suite. Normalized performance data is used to construct runs two-tailed and one-tailed tests. The null hypothesis in a two-sided test is a randomness hypothesis the rejection of which leads to the conclusion of non-randomness. The rejection of one-sided null hypothesis is in favour of clustering.

The total number of sample size is $n + m$ with n, the number of positive filtred values, and m, the number of negative filtred values. Hence, the number of the possible arrangements will be binomial of the form:

$$\binom{n+m}{n} = \binom{n+m}{m}$$

When the value of R, number of runs, is even, the probability of obtaining an observed or a smaller value of R will be:

$$P(R \geq R') = \frac{1}{\binom{n+m}{n}} \sum_{R=2}^{R'} (2) \binom{n-1}{\frac{R}{2}-1} \binom{m-1}{\frac{R}{2}-1}$$

When R is odd, the probability will be defined as:

$$P(R \geq R') = \frac{1}{\binom{n+m}{n}} \sum_{R=2}^{R'} \left[ \binom{n-1}{k-1}\binom{m-1}{k-2} + \binom{n-1}{k-2}\binom{m-1}{k-1} \right] with\ R = 2k-1$$

When the sample size is sufficient, the sampling distribution of runs under the null hypothesis might be assessed as approximately normal. The expected value and variance of R will be:

$$E(R) = \frac{2nm}{n+m} + 1 \qquad V(R) = \frac{2nm \times (2nm - n - m)}{(n+m)^2 \times (n+m-1)}$$

The Wald-Wolfowitz test statistic is computed as follows:

$$Z = \frac{R - E(R)}{\sqrt{V(R)}}$$

For relative performances runs tests we used relative performances for each strategy computed using the S&P 500 Index and the Dow Jones Industrial Average Index respectively.

### 2.1.2 Runs tests implementation and discussion

Table 3 synthesises the results of the runs test computed for the absolute performance of each strategy and its relative performance using benchmarks: number of runs, Wald-Wolfowitz test statistic, two-tailed and one-tailed p-values. We separated our runs tests in two principal segments: tests for absolute indices performance and their relative performance employing the S&P 500 and Dow Jones Indices as benchmarks.

Based on the p-values of two-tailed tests, we might reject with an acceptable degree of certainty the null hypothesis of randomness for Equity Market Neutral, European Equities, Event-Driven, Fixed Income Arbitrage, Multi-Strategy indices. One-tailed tests p-values of these strategies permit us to reject the null hypothesis in favour of clustering. Therefore, the absolute performances of five of the nine selected strategies tend to cluster, by way of explanation, supposedly exhibit persistence based on the runs tests. An unusual observation has arisen for Equity Long / Short Strategy index: forasmuch as we cannot reject the null hypothesis of randomness, we reject the null hypothesis for a one-tailed test. The persistence trend of explicit strategies indices might be due to the specificity of market behavioural tactics.

Diversely, we conclude in non-rejection of randomness for the entirety of relative performance for all the selected strategies indices, consequently failing to prove clustering trend. This observation is not unexpected if we adhere to the theory of market random-walk when we assume that fluctuations in stocks and indices prices follow the same distribution and are independent (Chen, 1996). In 2013 the Sveriges Riksbank Nobel Prize in Economic Sciences was awarded jointly to Eugene F. Fama, Lars Peter Hansen and Robert J. Shiller "for their empirical analysis of asset prices." (*The Sveriges Riksbank Prize In Economic Sciences In Memory Of Alfred Nobel 2013* n.d.). Stated in their works, Fama, 1970, Hansen, 1982 and Shiller, 1981 argued the predictability of future stocks prices in short terms but supposed diverse market efficiency models and standardized methods that would enable the predictability

| Strategy | Runs | Walf-Wolfowitz Z | Pr>\|Z\| | Pr>Z |
|---|---|---|---|---|
| Equity Long Bias Index | 137 | -0.94368 | 0.3453 | 0.1727 |
| Equity Long Bias Index (S&P relative) | 134 | -1.29863 | 0.1941 | 0.097 |
| Equity Long Bias Index (DJIA relative) | 140 | -0.41517 | 0.678 | 0.339 |
| Equity Long / Short Index | 131 | -1.4833 | 0.138 | 0.069 |
| Equity Long / Short Index (S&P relative) | 131 | -1.54012 | 0.1235 | 0.0618 |
| Equity Long / Short Index (DJIA relative) | 143 | -0.11847 | 0.9057 | 0.4528 |
| Equity Market Neutral Index | 126 | -2.02037 | 0.0433 | 0.0217 |
| Equity Market Neutral Index (S&P relative) | 135 | -1.06623 | 0.2863 | 0.1432 |
| Equity Market Neutral Index (DJIA relative) | 149 | 0.59235 | 0.5536 | 0.2768 |
| European Equities Index | 122 | -2.49594 | 0.0126 | 0.0063 |
| European Equities Index (S&P relative) | 132 | -1.42165 | 0.1551 | 0.0776 |
| European Equities Index (DJIA relative) | 140 | -0.47388 | 0.6356 | 0.3178 |
| Event-Driven Index | 123 | -2.48788 | 0.0129 | 0.0064 |
| Event-Driven Index (S&P relative) | 139 | -0.59235 | 0.5536 | 0.2768 |
| Event-Driven Index (DJIA relative) | 149 | 0.59235 | 0.5536 | 0.2768 |
| Fixed Income Arbitrage Index | 93 | 5.79342 | <.0001 | <.0001 |
| Fixed Income Arbitrage Index (S&P relative) | 139 | -0.59235 | 0.5536 | 0.2768 |
| Fixed Income Arbitrage Index (DJIA relative) | 149 | 0.59235 | 0.5536 | 0.2768 |
| Global Macro Index | 146 | 0.2986 | 0.7652 | 0.3826 |
| Global Macro Index (S&P relative) | 135 | -1.06623 | 0.2863 | 0.1432 |
| Global Macro Index (DJIA relative) | 143 | -0.11847 | 0.9057 | 0.4528 |
| Multi Strategy Index | 99 | -5.43062 | <0.0001 | <0.0001 |
| Multi Strategy Index (S&P relative) | 134 | -1.29863 | 0.1941 | 0.097 |
| Multi Strategy Index (DJIA relative) | 142 | -0.35417 | 0.7232 | 0.3616 |
| Technology Index | 131 | -1.54012 | 0.1235 | 0.0618 |
| Technology Index (S&P relative) | 145 | 0.11847 | 0.9057 | 0.4528 |
| Technology Index (DJIA relative) | 143 | -0.11847 | 0.9057 | 0.4528 |

Table 3: Two-sided and one-sided tests for randomness of indices returns against clustering or mixing.

over more prolonged periods. Consequently, we might assume the randomness of the relative performance of hedge funds strategies indices can be imposed by the volatility of benchmark indices.

## 2.2 Autocorrelation Tests

Another widely-applied method for detecting non-randomness in data is autocorrelation measurement. Introduced by Box and Jenkins, autocorrelation plots allow checking randomness in data ascertained by computing autocorrelation values throughout the time lags (G. Box and Jenkins, 1976). Besides, in the model identification stage autocorrelation plots are used for autoregressive, moving average time series models. With the $Y_1, Y_2, Y_3, ..., Y_N$ measurements through the $X_1, X_2, X_3, ..., X_N$ period, the lag $k$ autocorrelation function will be defined as:

$$r_k = \frac{\sum_{i=1}^{N-k}(Y_i - \overline{Y})(Y_{i+k} - \overline{Y})}{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}$$

In our analysis, we will assess the lags in order to detect non-randomness. Another important step in our analysis is a determination on the stationarity of the model and detecting seasonality.

### 2.2.1 Tests of Residuals

The null hypothesis of residuals being white noise is tested using the chi-square statistics applying Ljung and G. E. P. Box, 1978 computations:

$$\chi_m^2 = n(n+2)\Sigma_{k=1}^{m}\frac{r_k^2}{(n-k)} \ with \ r_k = \frac{\Sigma_{t=1}^{n-k}a_t a_{t+k}}{\Sigma_{t=1}^{n}a_t^2} \ with \ a_t, \ residual \ series$$

### 2.2.2 The SCAN Method Description

Given a stationary or nonstationary time series $\{z_t : 1 \leq t \leq n\}$ using the corrected mean form $\tilde{z}_t = z_t - \mu_z$, using a true autoregressive order of $p + d$ and with a true moving-average order of $q$ we apply the SCAN method to analyze eigenvalues of the correlation matrix of the autoregressive moving average (hereinafter referred to as ARMA) process (Brockwell and Davis, 1987). For autoregressive test order $m = p_{min}, \ldots, p_{max}$ and for moving-average test order $j = q_{min}, \ldots, q_{max}$ we perform the stages as follows:

1. With $Y_{m,t} = (\tilde{z}_t, \tilde{z}_{t-1}, \ldots, \tilde{z}_{t-m})'$ we compute the $(m+1) \times (m+1)$ matrix with:

$$\begin{aligned}
\hat{\beta}(m, j+1) &= \left(\sum_t Y_{m,t-j-1}Y'_{m,t-j-1}\right)^{-1}\left(\sum_t Y_{m,t-j-1}Y'_{m,t}\right)\\
\hat{\beta}^*(m, j+1) &= \left(\sum_t Y_{m,t}Y'_{m,t}\right)^{-1}\left(\sum_t Y_{m,t}Y'_{m,t-j-1}\right)\\
\hat{A}^*(m,j) &= \hat{\beta}^*(m, j+1)\hat{\beta}(m, j+1)
\end{aligned}$$

   with $t$ ranging from $j + m + 2$ to $n$.

2. We find the smallest eigenvalue, $\hat{\lambda}^*(m,j)$, of $\hat{A}^*(m,j)$ and its corresponding normalized eigenvector, $\Phi_{m,j} = (1, -\phi_1^{(m,j)}, -\phi_2^{(m,j)}, \ldots, -\phi_m^{(m,j)})$. We obtain the squared canonical correlation estimate of $\hat{\lambda}^*(m,j)$.

3. With $\Phi_{m,j}$ as $AR(m)$ coefficients, we obtain the residuals for $t = j + m + 1$ to $n$ applying the following formula: $w_t^{(m,j)} = \tilde{z}_t - \phi_1^{(m,j)}\tilde{z}_{t-1} - \phi_2^{(m,j)}\tilde{z}_{t-2} - \cdots - \phi_m^{(m,j)}\tilde{z}_{t-m}$.

4. From the sample autocorrelations of the residuals, $r_k(w)$, we approximate the standard error of the squared canonical correlation estimate by

$$var(\hat{\lambda}^*(m,j)^{1/2}) \approx d(m,j)/(n-m-j)$$

   with $d(m,j) = (1 + 2\sum_{i=1}^{j-1} r_k(w^{(m,j)}))$.

The test statistic that is used as an identification criterion is

$$c(m,j) = -(n-m-j)ln(1-\hat{\lambda}^*(m,j)/d(m,j))$$

which asymptotically follows $\chi^2_1$ if $m = p + d$ and $j \geq q$ or if $m \geq p + d$ and $j = q$. SCAN table that we obtain is constructed using $c(m,j)$ to determine $\hat{\lambda}^*(m,j)$ significantly different from zero. The ARMA orders are identified by finding a (maximal) rectangular pattern in which the $\hat{\lambda}^*(m,j)$ are insignificant for all test orders $m = p + d$ and $j \geq q$.

### 2.2.3 Dickey-Fuller Tests

In the case where the series is non-stationary, the ordinary least squares are not normally distributed. This is a matter of time series with a simple unit root (Dickey and Fuller, 1979). The augmented Dickey-Fuller test is widely used to inquire about the stationarity of the model. This test allows assessing zero mean stationary, single mean stationary, and time trend stationary models.

We assume that the autoregressive process is zero mean when it is of a form:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_{p+1} y_{t-p-1} + \epsilon_t$$

A single mean autoregressive process is of a form:

$$y_t - \mu = \alpha_1(y_{t-1} - \mu) + \alpha_2(y_{t-2} - \mu) + \cdots + \alpha_{p+1}(y_{t-p-1} - \mu) + \epsilon_t$$

A linear time trend autoregressive process is of a form:

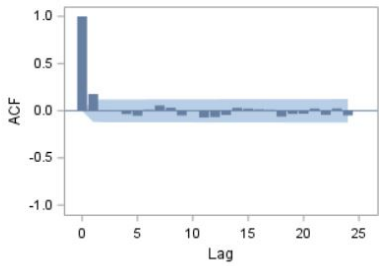
$$y_t - \mu - \beta t = \alpha_1(y_{t-1} - \mu - \beta(t-1)) + \alpha_2(y_{t-2} - \mu - \beta(t-2)) + \cdots + \alpha_{p+1}(y_{t-p-1} - \mu - \beta(t-p-1)) + \epsilon_t$$

The tau test statistic allows us to test the null hypothesis of non-stationarity.

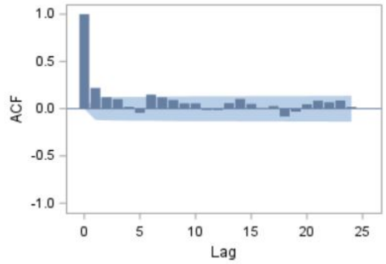
### 2.2.4 Autocorrelation Tests Implementation

In order to execute autocorrelation and trend analysis, we have produced autocorrelation function plots based on strategies indices performances. Based on the obtained plots (Figure 1) we might preliminarily conclude stationarity of the strategies indices. There are four strategies for which we can observe higher autocorrelation graphically: Equity Market Neutral, European Equities, Fixed Income Arbitrage, and Multi-Strategy. Whereas it is precipitously to deduce the autoregressive character of the model at this stage, further we might want to apply AR or autoregressive integrated moving average (hereinafter referred to as ARIMA) model process for modelling this type of data. Another considerable observation based on the autocorrelation function plots is that there is no negative autocorrelation recognised at lag one for any of the strategies, which is significant for further ARIMA application.

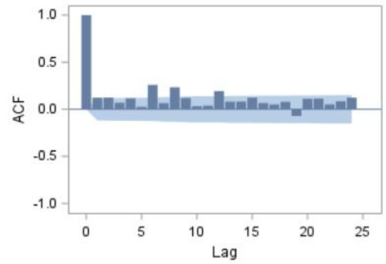
The Chi-Square values obtained for different lags allow us to evaluate white noise hypothesis for each strategy (Table 4). We successfully reject the null hypothesis of residuals following white noise distribution for seven out of nine selected strategies: the hypothesis cannot be rejected for Equity Long Bias and Global Macro strategies. This finding is in line with the conclusions obtained on the basis of our runs test: we could not reject the



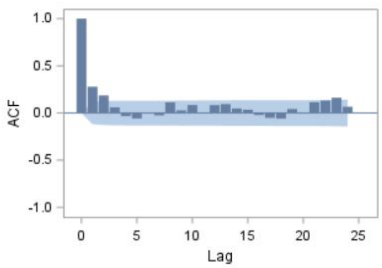

(a) Equity Long Bias



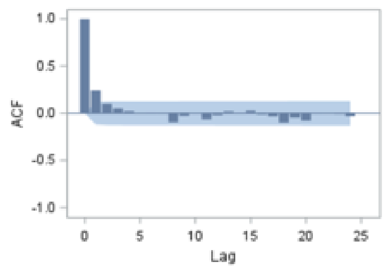

(b) Equity Long / Short



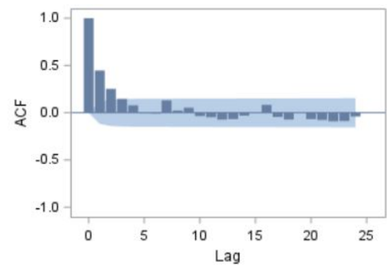

(c) Equity Market Neutral



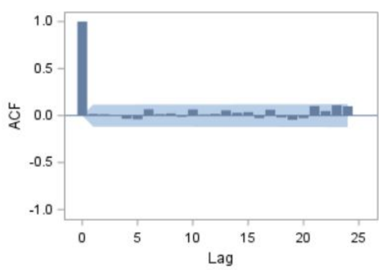

(d) European Equities



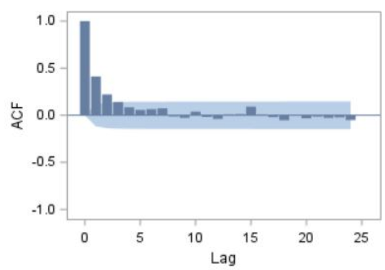

(e) Event-Driven



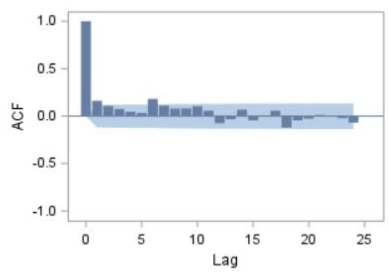

(f) Fixed Income Arbitrage




(g) Global Macro




(h) Multi-Strategy




(i) Technology

Figure 1: Strategies Indices Autocorrelation Function Plots

| Strategy | $\chi^2 Lag$ 6 | $\chi^2 Lag$ 12 | $\chi^2 Lag$ 18 | $\chi^2 Lag$ 24 | Zero Mean $\tau$ (1) | Single Mean $\tau$ (1) | Trend $\tau$ (1) |
|---|---|---|---|---|---|---|---|
| Equity Long Bias Index | 10.21 | 14.74 | 16.81 | 18.84 | -10.96 | -10.99 | -11.12 |
| Equity Long / Short Index | 28.57 | 37.43 | 44.84 | 51.95 | -9.64 | -9.63 | -10.11 |
| Equity Market Neutral Index | 34.29 | 68.03 | 80.76 | 98.12 | -9.83 | -9.81 | -11.32 |
| European Equities Index | 34.87 | 43.35 | 49.14 | 68.81 | -8.91 | -8.89 | -9.42 |
| Event-Driven Index | 20.84 | 25.12 | 28.83 | 31.74 | -9.69 | -9.73 | -9.86 |
| Fixed Income Arbitrage Index | 84.34 | 93.12 | 99.1 | 107.88 | -8.23 | -8.27 | -8.31 |
| Global Macro Index | 2.25 | 3.92 | 7.06 | 18.75 | -11.58 | -11.65 | -12.3 |
| Multi Strategy Index | 73.53 | 76.44 | 80.03 | 81.61 | -8.41 | -8.45 | -8.84 |
| Technology Index | 23.28 | 36.68 | 44.54 | 47.02 | -9.87 | -9.86 | -9.89 |

Table 4: Autocorrelation Check for White Noise and Augmented Dickey-Fuller Unit Root Tests results

null hypothesis of randomness for Equity Long Bias, Equity Long Short, Global Macro, and Technology Strategies indices. Noteworthily, we observe that Global Macro strategy index possesses the largest p-value in runs test, as well as the most insignificant autocorrelation values: this strategy is the most likely of the observed to tend to randomness trend.

To confirm the observations that we could have made based on the autocorrelation function plots, we address the tau tests statistics values for zero mean, single mean and time trend forms indicated in the table as well (Table 4). As all of the constructed test statistics are significant, we might confidently reject the null hypothesis of non-stationarity and state that all of the strategies models are stationary.

## 2.3 Goodness of fit and Normality

Many authors have thoroughly studied and brought out their findings on non-normality of hedge funds returns, paying attention that the degree of non-normality might vary depending on various aspects (Mitchell and Pulvino, 2001, Fung and D. Hsieh, 2001, Kouwenberg and Ziemba, 2007).

### 2.3.1 Measures of Skewness and Kurtosis

**Skewness.** The Fisher-Pearson coefficient of skewness adjusted for sample size is calculated as follows:

$$Sk = \frac{\sqrt{N(N-1)}}{N-2} \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^3/N}{\sigma^3}$$

The coefficient of skewness allows studying the symmetry or asymmetry of a dataset. Gaussian distribution skewness has to be equal to zero and signifies the symmetricity. Negative skewness indicates data skewed left and positive skewness indicates data skewed right (Joanes and Gill, 1998).

**Kurtosis.** The weighted coefficient of kurtosis is calculated as follows:

$$\mathrm{Ku} = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^4/N}{\sigma^4} - 3$$

The measure of kurtosis allows us to test whether data are heavy-tailed or light-tailed relativistic to a normal distribution. Hence, data sets with high kurtosis tend to have outliers, whereas data sets with low kurtosis lack outliers (Pearson, 1905).

### 2.3.2 Tests for Normality Description

**Shapiro-Wilk Test.** The Shapiro-Wilk W statistic represents the ratio of the best estimator of the variance (Shapiro and Wilk, 1965). The rejection of the null hypothesis of normality is justified in the case of small values of W. The process for calculating the probability of obtaining a W statistic less than or similar to the detected value depends on n. The probability distribution of W is known for n = 3 and is used to determine the p-value. For $n > 4$ we have to normalize as follows:

$$Z_n = \begin{cases} (-\log(\gamma - \log(1 - W_n)) - \mu)/\sigma & \text{if } 4 \leq n \leq 11 \\ (\log(1 - W_n) - \mu)/\sigma & \text{if } 12 \leq n \leq 2000 \end{cases}$$

with $\sigma, \gamma, \mu$ calculated with from computation results.

**Empirical Distribution Function Goodness-of-Fit Tests.** A set of $n$ independent observations $X_1, \ldots, X_n$ with a common distribution function $F(x)$ is defined by the empirical distribution function (D'Agostino and Stephens, 1986). In its turn, the empirical distribution function, $F_n(x)$ is calculated as:

$$\begin{array}{ll} F_n(x) = 0, & x < X_{(1)} \\ F_n(x) = \frac{i}{n}, & X_{(i)} \leq x < X_{(i+1)} \quad i = 1, \ldots, n-1 \\ F_n(x) = 1, & X_{(n)} \leq x \end{array}$$

**Kolmogorov-Smirnov Test.** The Kolmogorov-Smirnov D statistic is defined as:

$$D = \sup_x |F_n(x) - F(x)|$$

**Anderson-Darling Test.** The Anderson-Darling Statistic is defined as (Anderson and Darling, 1952):

$$A^2 = n \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 \left[F(x)\left(1 - F(x)\right)\right]^{-1} dF(x)$$

**Cramér–von Mises Test.** The Cramér-von Mises Statistic is defined as (Anderson and Darling, 1962):

$$W^2 = n \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 dF(x)$$

and is computed with

$$W^2 = \sum_{i=1}^{n} \left(U_{(i)} - \frac{2i-1}{2n}\right)^2 + \frac{1}{12n}$$

### 2.3.3 Implementation of Tests for Normality

We gather from Table 5 and Figure 2 below that all of our hedge fund strategy indices deviate from normality: Table 5 reports that all of our indices are leptokurtic (because SAS returns the excess Kurtosis: $Ku - 3$) and right-skewed (except for the Equity Long Bias, Event-Driven, Fixed Income Arbitrage and Multi-Strategy indices), which is in line with our normality test results. We find in Table 5 that the normality hypothesis is rejected for each one of the tests mentioned earlier.

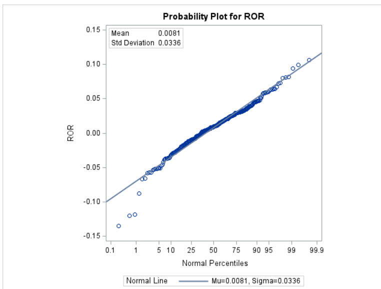
In that respect, the standard deviation alone fails to properly measure risk or performance, mainly because it does not distinguish returns that exceed the mean from returns that fall below the mean (Atilgan and Bali, 2013). Metrics such as downside deviation would be preferred here.

## 2.4 Performance tests

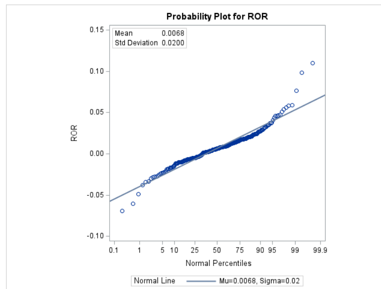
When it comes to evaluating the performance of our strategy-specific hedge fund indices, a few tools come to mind including, among other things, the Sharpe, information and Sortino

| **Strategy** | **Sk** | **Ku** | **W** | **Pr <W** | **D** | **Pr >D** | $W^2$ | **Pr >**$W^2$ | $A^2$ | **Pr >**$A^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Equity Long Bias Index | -0.55018 | 2.0953 | 0.9713 | <0.0001 | 0.06304 | <0.0100 | 0.25333 | <0.0050 | 1.53472 | <0.0050 |
| Equity Long / Short Index | 0.73126 | 4.7631 | 0.93305 | <0.0001 | 0.08933 | <0.0100 | 0.6361 | <0.0050 | 3.88017 | <0.0050 |
| Equity Market Neutral Index | 0.10116 | 1.5795 | 0.97261 | <0.0001 | 0.08273 | <0.0100 | 0.48088 | <0.0050 | 2.62454 | <0.0050 |
| European Equities Index | 1.56319 | 8.4841 | 0.88073 | <0.0001 | 0.10841 | <0.0100 | 0.8757 | <0.0050 | 5.61798 | <0.0050 |
| Event-Driven Index | -1.56639 | 8.9956 | 0.90416 | <0.0001 | 0.08579 | <0.0100 | 0.55806 | <0.0050 | 3.42494 | <0.0050 |
| Fixed Income Arbitrage Index | -4.90706 | 45.345 | 0.6612 | <0.0001 | 0.17784 | <0.0100 | 3.29124 | <0.0050 | 18.4078 | <0.0050 |
| Global Macro Index | 0.67023 | 1.5616 | 0.97272 | <0.0001 | 0.06317 | <0.0100 | 0.2644 | <0.0050 | 1.7122 | <0.0050 |
| Multi Strategy Index | -2.26845 | 12.224 | 0.84331 | <0.0001 | 0.10229 | <0.0100 | 0.98873 | <0.0050 | 6.13134 | <0.0050 |
| Technology Index | 0.69463 | 4.1389 | 0.94352 | <0.0001 | 0.08297 | <0.0100 | 0.49162 | <0.0050 | 2.97957 | <0.0050 |

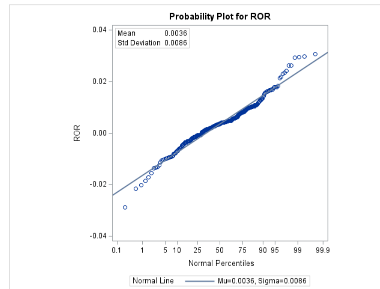
Table 5: Strategies indices Tests for Normality



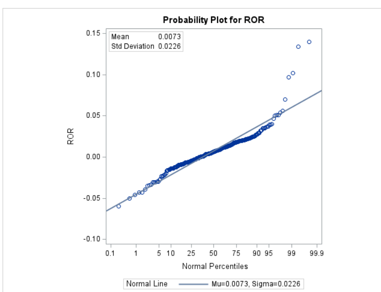

(a) Equity Long Bias



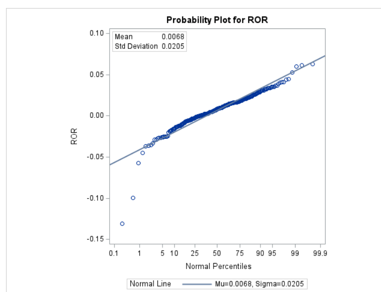

(b) Equity Long / Short



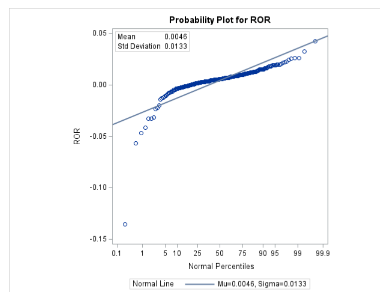

(c) Equity Market Neutral



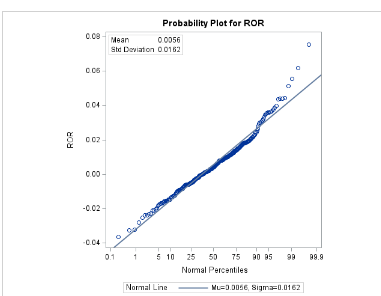

(d) European Equities



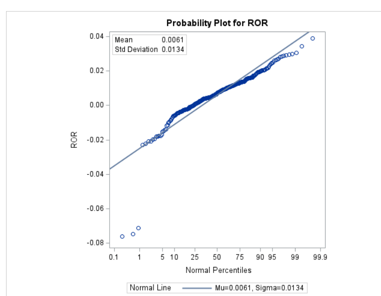

(e) Event-Driven



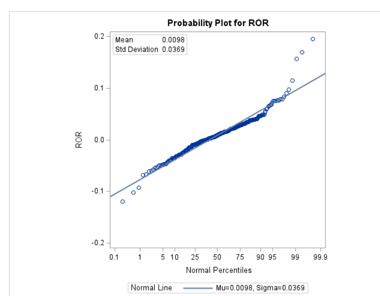

(f) Fixed Income Arbitrage




(g) Global Macro




(h) Multi-Strategy




(i) Technology

Figure 2: Strategies Indices Normal Q–Q Plots

ratios. Since hedge funds returns aren't normally distributed, using the Sharpe or the information ratio, *which imposes a normality assumption*, to evaluate their performance would lead us to inaccurate results (Sortino and Price, 1994). These indicators are both based on the standard deviation of portfolio returns, which, as mentioned earlier, is not the best tool for performance or risk evaluation when it comes to non-normally distributed returns. Thus, it would be more appropriate to focus on indicators that take into account the non-normality of hedge fund returns such as the Sortino ratio and the compound annual growth rate (CAGR).

### 2.4.1 Sortino ratio

When measuring and comparing hedge funds indices performance, the Sortino ratio is a better indicator than the Sharpe ratio since it takes into account the downside deviation *( i.e a downside risk measure that only focuses on returns falling below a specific target)* (Rollinger and Hoffman, 2013), rather than the regular standard deviation.

The Sharpe ratio is defined as :

$$Sharpe\ ratio = \frac{R_p - r_f}{\sigma_p}$$

with $R_p$ as the portfolio rate of return, $r_f$ as the risk free rate, $\sigma_p$ as the standard deviation of portfolio returns. Whereas the Sortino ratio, S is defined as :

$$S = \frac{R_p - r_f}{\sigma_d}$$

with $R_p$ as the portfolio rate of return, $r_f$ as the risk free rate, $\sigma_d$ as the Downside risk deviation.

Thus, it is fair to say that the main difference between the Sharpe and Sortino ratios lies in the fact that the Sharpe ratio *(or more specifically, standard deviation used in the parameter)* ignores the non-normality of hedge fund returns, whereas the Sortino ratio takes it into account using the downside risk deviation.

| **Strategy Index** | **Sortino** |
|---|---|
| Equity Long Bias Index | 0.429 |
| Equity Long / Short Index | 0.784 |
| Equity Market Neutral Index | 1.147 |
| European Equities Index | 0.801 |
| Event Driven Index | 0.551 |
| Fixed Income Arbitrage Index | 0.479 |
| Global Macro Index | 0.988 |
| Multi Strategy Index | 0.708 |
| Technology Index | 0.551 |
| S&P500 | 0.283 |
| DJIA | 0.278 |

Table 6: Sortino ratios for hedge fund indices and benchmarks using below-average returns

Table 6 reports Sortino ratios for the nine strategy-specific hedge fund indices we are focusing on as well as the benchmarks. Here, in order to determine the downside risk deviation, we have used the average return as a minimum threshold. So it would be equal to the standard deviation of returns that fall below the average return rate:

$$\sigma_d = StdDev\ (returns < average\ return\ rate)$$

We have also used negative returns in order to calculate the downside risk deviation. So it would be equal to the standard deviation of returns that fall below zero:

$$\sigma_d = StdDev(returns < 0)$$

The resulting Sortino ratios are reported in Table 7.

| Strategy Index | Sortino |
|---|---|
| Equity Long Bias Index | 0.434 |
| Equity Long / Short Index | 0.806 |
| Equity Market Neutral Index | 1.204 |
| European Equities Index | 0.829 |
| Event Driven Index | 0.506 |
| Fixed Income Arbitrage Index | 0.362 |
| Global Macro Index | 1.089 |
| Multi Strategy Index | 0.595 |
| Technology Index | 0.565 |
| S&P500 | 0.289 |
| DJIA | 0.282 |

Table 7: Sortino ratios for hedge fund indices and benchmarks using negative returns

For both approaches, we have used the Chicago Board Options Exchange interest rate 10 year Index (CBOE TNX) as the risk free rate. Since we have calculated our ratios for each index from monthly data, tables 6 and 7 report the arithmetic average of the monthly Sortino ratios for each strategy and benchmark index.

A higher Sortino ratio is preferred as it means that the hedge fund index has generated a higher excess return per unit of downside risk (Atilgan and Bali, 2013). In that respect, all nine of our strategy hedge fund indices generate a superior Sortino ratio in comparison with the benchmarks.

Specifically, with both methods, we find that the Equity Market Neutral index has the highest Sortino ratio which is equal to 1.2 approximately, followed by the Global Macro index whose Sortino ratio is equal to 1 approximately. On the other hand, the Fixed Income Arbitrage index as well as the Equity Long Bias index have the lowest Sortino ratios.

This implies that the Equity market neutral index has generated an excess return of about 1.2% per unit of downside risk deviation, which makes it the best performing index among those we have evaluated. Overall, our strategy hedge fund indices outperform the S&P500 and DJIA benchmarks, for both methods.

#### 2.4.2 Compound annual growth rate

As explained earlier, most existing performance indicators fail to measure accurately fund performance when returns are not normally distributed *(which is often the case in hedge* *funds)*, hence our choice of using the compound annual growth rate (hereinafter referred to as CAGR).

CAGR is a constant rate of return over a specific period, it measures the growth of an investment under the assumption that it grew at the same rate and was re-invested (compounded) each year. The CAGR is defined as :

$$CAGR(t_0, t_n) = \left(\frac{V(t_n)}{V(t_0)}\right)^{\frac{1}{t_n - t_0}} - 1$$

where $V(t_0)$ denotes the beginning value and $V(t_n)$ denotes the ending value.

Table 8 reports compound annual global rates for hedge fund strategy indices as well as for the S&P 500 and DJIA indices.

| **Strategy Index** | **CAGR** |
|---|---|
| Equity Long Bias Index | -1,35% |
| Equity Long / Short Index | -0,46% |
| Equity Market Neutral Index | 0,17% |
| European Equities Index | 0,19% |
| Event-Driven Index | -0,58% |
| Fixed Income Arbitrage Index | -0,63% |
| Global Macro Index | -0,19% |
| Multi-Strategy Index | -0,23% |
| Technology Index | -2,20% |
| S&P 500 Index | -2,37% |
| DJIA Index | -1,41% |

Table 8: Compound annual growth rate for hedge fund indices and benchmarks between 1997 and 2019

Here, it is quite clear that most of our hedge fund indices have a higher CAGR than the S&P 500 and DJIA indices. More specifically, we find that the European Equities index has a 0.19% CAGR, followed by the Equity Narket Neutral index with a 0.17% CAGR. The Global Macro index has a -0.19% CAGR, which makes it the third better performing index. We should note here that these three indices are also among the top three based on their Sortino ratios. On the other hand, the Technology index has a -2.2% CAGR which is the lowest one. The DJIA and the S&P 500 indices have a -1.41% and -2.37% CAGR respectively. Thus, the Technology index does not outperform the DJIA index, but it still outperforms the S&P 500 index. In conclusion, all of our hedge fund strategy indices outperform both our benchmarks, except for the Technology index which only outperforms the S&P 500 index.

### 2.4.3 Empirical cumulative distribution function analysis

The empirical cumulative distribution function (hereinafter referred to as ECDF or CDF) is defined as:

$$F_N(x) = \text{percent of nonmissing values} \leq x = \frac{\text{number of values} \leq x}{N} \times 100\%$$

The ECDF is an increasing step function with a vertical jump of $\frac{1}{N}$ for each $x$ value equal to the observed value. Based on the Glivenko and Cantelli, 1933 theorem the function converges to the underlying distribution with the probability of one.

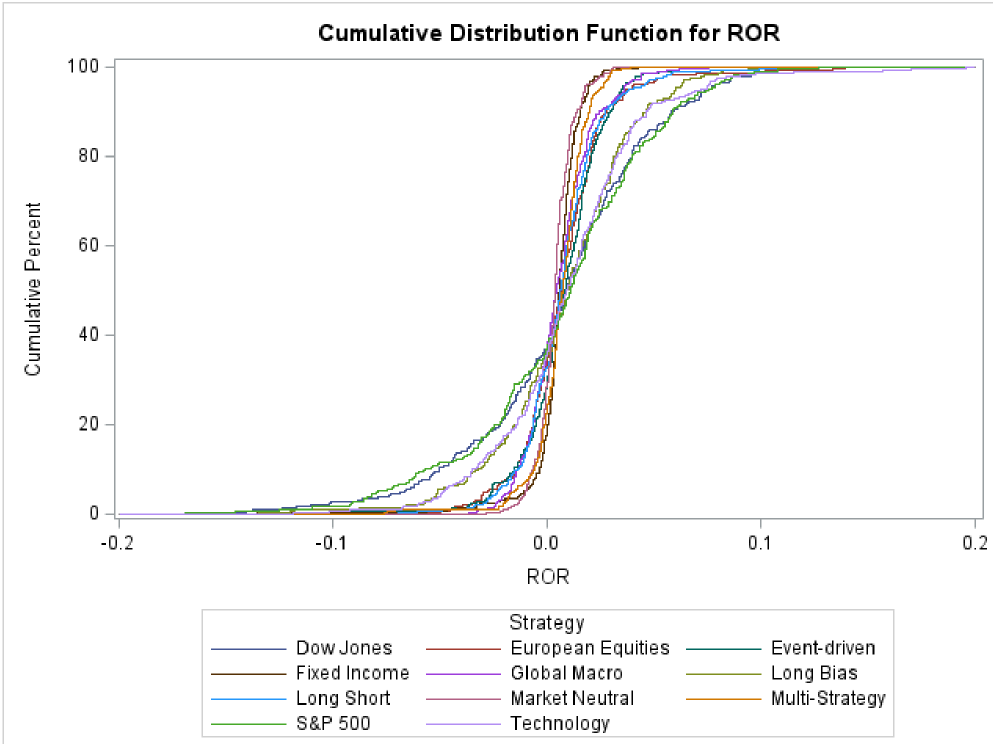



Figure 3: Strategies and Benchmarks Empirical cumulative distribution function plot

As we can see in Figure 3, the selected strategies indices have almost vertical CDF which is due to their high kurtosis and non-normality.

Performance wise, the figure indicates that, at first (for negative returns), our selected indices outperform the S&P 500 and DJIA benchmarks in the sense that the benchmarks have a higher amount of negative returns. When it comes to positive returns, the conclusion is the same: as a matter of fact, 85% of the benchmark returns are around 0.05 versus over 90% for the strategy indices.

The Table 9 presents the associated quantiles for any given level for strategies indices and benchmarks. The presentation of quantile definitions differs depending on the software and the table presented below is of the default quantile definition of 5 for SAS software (Hyndman and Fan, 1996).

# 3 Performance comparison

## 3.1 Results of performance tests

The performance tests conducted earlier, whether it is the Sortino ratio, the compound annual growth rate, or the ECDF lead us to the conclusion that in most cases, the hedge fund strategy indices outperform both the S&P 500 and DJIA benchmarks.

| **Strategy** | **100% Max** | **99%** | **95%** | **90%** | **75% Q3** | **50% Median** | **25% Q1** | **10%** | **5%** | **1%** | **0% Min** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Equity Long Bias Index | 0.0996 | 0.082 | 0.06 | 0.046 | 0.029 | 0.009 | -0.01 | -0.032 | -0.051 | -0.118 | -0.135 |
| Equity Long / Short Index | 0.1099 | 0.0762 | 0.038 | 0.0274 | 0.0164 | 0.0062 | -0.0049 | -0.0141 | -0.0231 | -0.0492 | -0.0692 |
| Equity Market Neutral Index | 0.0307 | 0.0296 | 0.0178 | 0.0141 | 0.0081 | 0.0036 | -0.0009 | -0.0069 | -0.01 | -0.0202 | -0.0287 |
| European Equities Index | 0.1395 | 0.1017 | 0.0367 | 0.0276 | 0.018 | 0.0066 | -0.005 | -0.0141 | -0.028 | -0.0457 | -0.0595 |
| Event-Driven Index | 0.063 | 0.0602 | 0.0348 | 0.0296 | 0.0178 | 0.0082 | -0.0026 | -0.0155 | -0.0257 | -0.0572 | -0.1309 |
| Fixed Income Arbitrage Index | 0.0423 | 0.0261 | 0.0191 | 0.0152 | 0.01 | 0.0058 | 0.0016 | -0.0035 | -0.0106 | -0.0467 | -0.1355 |
| Global Macro Index | 0.0753 | 0.0556 | 0.0361 | 0.0243 | 0.0141 | 0.0042 | -0.0051 | -0.0138 | -0.0177 | -0.0324 | -0.0363 |
| Multi Strategy Index | 0.0388 | 0.0304 | 0.0243 | 0.0196 | 0.013 | 0.007 | 0.0005 | -0.0058 | -0.0154 | -0.0711 | -0.0762 |
| Technology Index | 0.1958 | 0.1565 | 0.0718 | 0.0463 | 0.0284 | 0.0091 | -0.0086 | -0.0349 | -0.0482 | -0.0923 | -0.1188 |
| S&P 500 | 0.12684 | 0.10755 | 0.07518 | 0.05858 | 0.03561 | 0.01106 | -0.0181 | -0.0568 | -0.079 | -0.1251 | -0.1694 |
| DJIA | 0.11837 | 0.10605 | 0.07574 | 0.05786 | 0.034 | 0.00833 | -0.0161 | -0.0507 | -0.0669 | -0.1374 | -0.1513 |

Table 9: Quantiles for given level for Strategies Indices and Benchmarks

Nevertheless, we should highlight one very crucial fact: it is important to keep in mind that our data may be biased, which would affect all the tests carried out previously and lead us to inaccurate results and conclusions.

## 3.2 Biases in strategies performance data

Whereas it is challenging to quantify the impact of bias on the index performance it has to be taken into account when considering outperformance of hedge funds strategies indices in comparison with funds or indices that do not exhibit the same biasedness.

### 3.2.1 Reporting and self-selection bias

Within the available hedge fund databases, certain biases exist in terms of reporting endowment: it is up to hedge funds to disclose or not their performance information. The bias might be variously influenced by several aspects, such as hedge fund size, its age, manager ability (Horst and Verbeek, 2005). Horst and Verbeek have concluded that the self-selection bias is more likely to occur for hedge funds with low returns history. Agarwal et al., 2010 find that the hedge funds' inclination to self-report is consonant with the tradeoffs between the advantages of access to prospective investors and incremental expenses of revealing trading confidence. In their paper, they state that during the period of 1980-2008, a stronger incentive to self-report to databases is typical for young and medium-sized funds with more diversified and higher-frequency trading strategies, seemingly in order to advertise their funds and attract potential investors. Horst and Verbeek attempt to correct the self-selection bias by using the weighting for performance measures, they conclude that following the correction the positive persistence pattern might be further strengthened including the annual horizon.

### 3.2.2 Instant history bias

Instant history bias also called backfill bias might occur in case of inconsistent reporting that in their turn can disproportionately inflate the hedge fund performance. Thus, the average return in the funds' database is upwardly biased within backfilling the fund's performance by data vendor (Fung and D. A. Hsieh, 2004). Bontschev and Eling, 2013 propose to omit the returns of the first 12 months of the dataset during the construction of the sample in order to eliminate instant history bias. To oppose the fact of potential backfilling, some hedge fund databases might need to restrict the extent to which hedge fund managers can backfill their performance up to forbidding.

### 3.2.3 Survivorship bias

In the situations where the database provider removes failing funds from the index, the past index values have to be adjusted to remove the data of the dropped fund. Considering the logic that the hedge fund with a poor performance is more likely to fail we assume the upward bias existence in the index. Based on their study, Joenväärä et al., 2012 highlight multiple hypotheses on survivorship bias issuing from the differences in database properties. Based on the findings they confirm the hypothesis suggesting that databases that consist of larger funds with the higher survivor and backfilling biases might exhibit moderately lower performance persistence.

## 3.3 Hedge funds fees impact

Whereas we can judge that hedge funds outperform market benchmarks we have to take into consideration the standard hedge funds fee arrangements that might influence our perception of overall performance.

Unlike mutual fund managers, hedge fund managers charge incentive fees for their performance. As a result, mutual fund managers are more prudent in their management while hedge fund managers are more inclined to take risks to increase their returns and have more flexibility in their management. In doing so, they attract larger and more sophisticated investors who have the same interests as they do. In this regard, Ackermann et al., 1999 show that when incentive fees rise from 0% to a median value of 20%, the Sharpe ratio increases, on average, by 66%.

On the other hand, Ben-David et al., 2020 state in their paper that based on their long-run studies of hedge funds performance the aggregate incentive fee rate is 2.5 times the average contractual rate (i.e., approximately 50% instead of 20% as presumed). Based on the results of their analysis they found that due to fees asymmetric structure investors earned $228.2bn in aggregate gross profits on their hedge fund investments between 1995 and 2016, forasmuch as $113.3bn of incentive fees they paid. Therefore, considering the possible outperformance of hedge funds, the fees might enforce a substantial impediment on the returns that investors earn on their investments.

## 3.4 Performance during crisis

Apart from the conclusions based on the performance tests we have to take into consideration the persistence effect which might be significantly influenced during crisis periods. The market performance during the crises becomes unquestionably more volatile. From what we see graphically (Figure 4), benchmarks are significantly more sensitive to the market volatility comparing with the hedge funds indices, especially during the crisis periods.

If we compare different strategies performance during crisis times (Figure 5), we can conclude that certain strategies are more exposed to market volatility than others. If we compare different strategies performance during crisis times, we can conclude that certain strategies are more exposed to market volatility than others: Event-Driven, Equity Long Bias, Technology strategies indices exhibit unstable performance during the observed time period. In order to investigate the behaviour of hedge funds strategies indices and benchmarks, we divided our dataset into three periods: Financial crisis period from June 2007 to March 2009, COVID-19 crisis period starting from February 2020 until the latest available database date, and out of crisis period from April 2009 to January 2020. Table 10 integrates the mean-variance measures for four types of periods for every strategy and selected benchmarks.

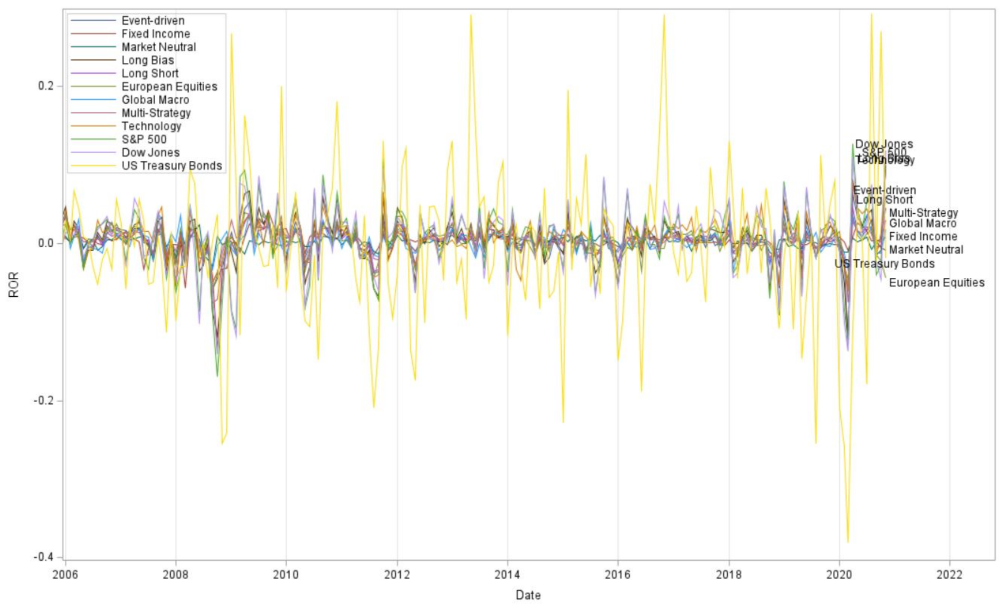



Figure 4: Strategies and Benchmarks Comparative Performance

| Strategy | Whole Period | | During Financial Crisis | | During COVID-19 Crisis | | Out of crisis | |
|---|---|---|---|---|---|---|---|---|
| | **Mean** | **Variance** | **Mean** | **Variance** | **Mean** | **Variance** | **Mean** | **Variance** |
| Equity Long Bias Index | 0.00812 | 0.0011262 | -0.0147 | 0.00156623 | 0.02155 | 0.00453938 | 0.00679 | 0.00064631 |
| Equity Long / Short Index | 0.00677 | 0.00040081 | -0.0046 | 0.00035321 | 0.00654 | 0.00101347 | 0.00399 | 0.00016271 |
| Equity Market Neutral Index | 0.00363 | 0.00007393 | 0.00035 | 0.00018994 | -0.0018 | 0.00005726 | 0.00187 | 0.00003197 |
| European Equities Index | 0.00728 | 0.0005115 | -0.0045 | 0.0004303 | -0.007 | 0.0007833 | 0.00443 | 0.00023162 |
| Event-Driven Index | 0.0068 | 0.00042133 | -0.0077 | 0.00044485 | 0.00796 | 0.00307986 | 0.00539 | 0.00025326 |
| Fixed Income Arbitrage Index | 0.00462 | 0.00017675 | -0.015 | 0.00112214 | 0.00842 | 0.00010288 | 0.00562 | 0.00005916 |
| Global Macro Index | 0.00556 | 0.00026145 | 0.00236 | 0.00035275 | 0.00552 | 0.00049578 | 0.00261 | 0.00014565 |
| Multi Strategy Index | 0.00609 | 0.00017827 | -0.0054 | 0.00070154 | 0.00226 | 0.00095447 | 0.00449 | 0.00009524 |
| Technology Index | 0.00975 | 0.00136491 | -0.0004 | 0.00072087 | 0.02417 | 0.00228024 | 0.0076 | 0.00051797 |
| S&P 500 | 0.00655 | 0.00199408 | -0.0274 | 0.00346554 | 0.01471 | 0.00676211 | 0.01146 | 0.00133047 |
| DJIA | 0.00631 | 0.00193613 | -0.0247 | 0.00286503 | 0.00812 | 0.00730217 | 0.01076 | 0.00123849 |

Table 10: Crisis and out of crisis periods comparative performance

We will try to have an idea of market behaviour based on Modern Portfolio Theory framework for assembling a portfolio that is enabling the maximized expected return for a given level of risk (variance) (Markowitz, 1952). Using the simplified Markowitz's mean-variance analysis we conclude that whereas during the whole period it is hard to determine the most efficient index, S&P 500 and Dow Jones indices are the least preferable according to

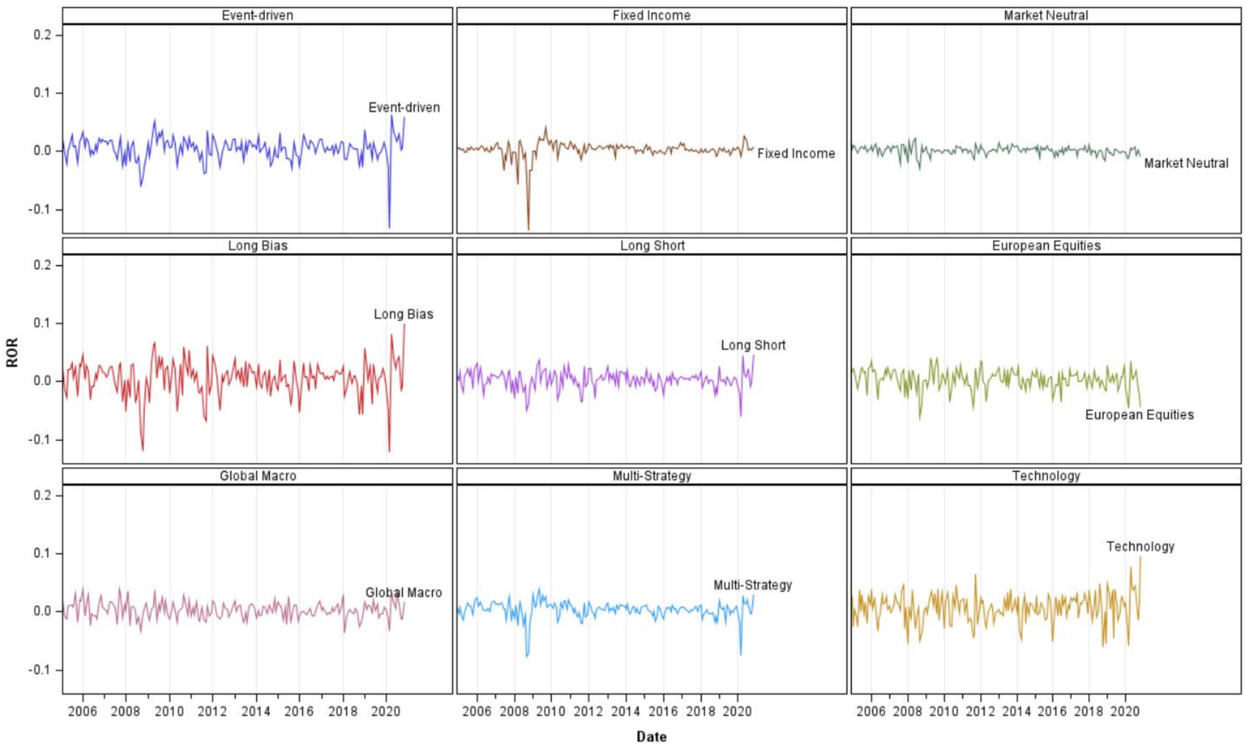



Figure 5: All Strategies Comparative Performance

the mean-variance criteria during the financial crisis period. We cannot decisively conclude on the indices performance persistence during COVID-19 crisis or out of crisis periods based on the before mentioned simplified measures. Nevertheless, we can presume that our hedge funds strategies indices are more persistent and demonstrate better performance during crisis times in contrast with benchmarks.

# Conclusion

This study provides an extensive analysis of how the hedge fund industry has performed over time. In our paper, we have adopted different methods to test the hypotheses of randomness and clustering considering both absolute and relative performances of selected strategies indices including runs tests that appear to be the inception of our outperformance analysis. In order to do that, we have conducted Wald-Wolfowitz runs tests to begin with, which has allowed us to analyze the randomness of our selected hedge fund strategy indices. After implementing the runs test for the absolute performance of each strategy and its relative performance using benchmarks, we have reached the following conclusions: the absolute performance of five out of nine strategy indices rejects the null hypothesis of randomness and do tend to cluster. However, we were unable to confirm the existence of a clustering trend for the relative performance of all nine of our indices.

Secondly, we implemented autocorrelation tests based on the ARIMA model and our main findings were the following: the ACF plots seem to show autocorrelation for eight out of nine indices (except for the Global Macro index), we reject the null hypothesis of residuals following white noise distribution for seven out of nine selected strategies, all of the strategy models are stationary.

Then, after implementing and analyzing tests for normality, we found that the normality hypothesis is rejected for all selected indices. Since the hedge fund strategy indices all deviate from normality, we chose to evaluate their performance using the Sortino ratio and the CAGR. Both these metrics led us to the same results: In most cases, the selected indices outperform both the S&P 500 and DJIA benchmarks.

However, it is important to keep in mind that our data may be biased, which would alter all of our previous test results. Therefore, it would be interesting to conduct robustness tests on our results (Eling and Schuhmacher, 2007).

# References

Ackermann, C., McEnally, R., & Ravenscraft, D. (1999). The performance of hedge funds: Risk, return, and incentives. *The journal of Finance*, *54*(3), 833–874.

Agarwal, V., Fos, V., & Jiang, W. (2010). Inferring reporting biases in hedge fund databses from hedge fund equity holdings. *Working Paper, Columbia Business School.*

*The amazing story behind the world's first hedge fund.* (2016). %3Chttps://www.businessinsider.com/alfred-winslow-jones-started-the-first-hedge-fund-2016-8?IR=T Accessed 23 January 2021

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics.*, *23*, 193–212.

Anderson, T. W., & Darling, D. A. (1962). On the Distribution of the Two-Sample Cramer–von Mises Criterion. *Annals of Mathematical Statistics. Institute of Mathematical Statistics.*, *33*(3), 1148–1159.

Atilgan, Y., & Bali, T. (2013). The Performance of Hedge Fund Indices. *Borsa Istanbul Review*, *13*(3), 30–52.

Ben-David, I., Birru, J., & Rossi, A. (2020). The Performance of Hedge Fund Performance Fees. *National Bureau of Economic Research Working Paper Series.*

Bontschev, G., & Eling, M. (2013). Factors that affect the performance of distressed securities hedge funds. *Journal of Derivatives Hedge Funds*, *19*, 159–180.

Box, G., & Jenkins, G. (1976). *Time Series Analysis Forescasting And Control.* Holden-Day.

Bradley, J. (1968). *Distribution-free statistical tests.* Prentice Hall.

Brockwell, P. J., & Davis, R. A. (1987). *Time Series: Theory and Methods.* Springer-Verlang.

Chaudhry, A., & Johnson, H. L. (2008). The efficacy of the Sortino ratio and other benchmarked performance measures under skewed return distributions. *Australian Journal of Management*, *32*(3), 485–502.

Chen, P. (1996). A Random Walk or Color Chaos on the Stock Market? Time-Frequency Analysis of SP Indexes. *Studies in Nonlinear Dynamics Econometrics.*, *1*(2).

Connor, G., & Woo, M. (2004). An introduction to hedge funds.

D'Agostino, R. B., & Stephens, M. (1986). Goodness-of-Fit Techniques. *Marcel Dekker.*

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association.*, *74*, 427–431.

Eling, M., & Schuhmacher, F. (2007). Does the Choice of Performance Measure Influence the Evaluation of Hedge Funds? *Journal of Banking Finance*, *31*(9), 2632–2647.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance.*, *25*(2), 383–417.

Fung, W., & Hsieh, D. (2001). The risk in hedge fund strategies: Theory and evidence from trend followers. *Review of Financial Studies.*, *14*, 313–341.

Fung, W., & Hsieh, D. A. (2004). Hedge fund benchmarks: A risk-based approach. *Financial Analysts Journal.*, *60*(5).

Glivenko, V., & Cantelli, F. P. (1933). Sulla determinazione empirica delle leggi di probabilità. *Giorn. Ist. Ital. Attuari.*, *4*.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica.*, *50*(4), 1029–1054.

Hentati-Kaffel, R., & De Peretti, P. (2015). Detecting performance persistence of hedge funds. *Economic Modelling.*, *47*, 185–192.

Horst, J., & Verbeek, M. (2005). Fund liquidation, self-selection, and look-ahead bias in the hedge fund industry. *Review of Finance*, *11*, 605–632. https://doi.org/10.1093/rof/rfm012

Hyndman, R. J., & Fan, Y. (1996). Sample quantiles in statistical packages. *American Statistician.*, *50*, 361–365.

Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society.*, *47*, 183–189.

Joenväärä, J., Kosowski, R., & Tolonen, P. (2012). New 'stylized facts' about hedge funds and database selection bias. *Imperial College Business School*.

Kim, T., & Lee, H. (2017). The contagion versus interdependence controversy between hedge funds and equity markets. *European Financial Management.*, *24*(3), 309–330.

Kouwenberg, R., & Ziemba, W. (2007). Incentives and risk taking in hedge funds. *Journal of Banking and Finance.*, *31*, 3291–3310.

Ljung, G. M., & Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika.*, *65*, 297–30.

Magel, R. C., & Wibowo, S. H. (1997). Comparing the powers of the Wald-Wolfowitz and Kolmogorov-Smirnov tests. *Biometrical Journal.*, *39*(6), 665–675.

Mahan, M. Y., Chorn, C. R., & Georgopoulos, A. P. (2015). White Noise Test: detecting autocorrelation and nonstationarities in long time series after ARIMA modeling. *Proceedings 14th Python in Science Conference (Scipy 2015)*.

Markowitz, H. M. (1952). Portfolio Selection". The Journal of Finance. *The Journal of Finance.*, *7*(1), 77–91.

Mitchell, M., & Pulvino, T. (2001). Characteristics of risk in risk arbitrage. *Journal of Finance.*, *56*, 2135–2175.

Pearson, K. (1905). The Error Law and its Generalizations by Fechner and Pearson. A Rejoinder. *Biometrika.*, *4*, 169–212.

Rollinger, T., & Hoffman, S. (2013). *Sortino : A 'Sharper' Ratio*. http://redrockcapital.homestead.com/Sortino___A___Sharper___Ratio_Red_Rock_Capital.pdf

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*, 591–611.

Shiller, R. J. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *The American Economic Review.*, *71*(3), 421–436.

Sortino, F., & Price, N. (1994). Performance Measurement in a Downside Risk Framework. *The Journal of Investing*, *3*(3), 59–64.

Stafylas, D., Anderson, K., & Uddin, M. (2016). Recent advances in hedge funds' performance attribution: Performance persistence and fundamental factors. *International Review of Financial Analysis*, *43*, 48–61.

*The sveriges riksbank prize in economic sciences in memory of alfred nobel 2013*. (n.d.). https://www.nobelprize.org/prizes/economic-sciences/2013/summary/ Accessed 20 January 2021

Wald, A., & Wolfowitz, J. (1940). On a Test Whether Two Samples are from the Same Population. *The Annals of Mathematical Statistics.*, *11*(2), 147–162.