BACHELOR THESIS

Hedge Fund Replication in Times of Change

submitted by:

Moritz Johannes Maidl

(student number: 100825)

submitted on: 15.01.2024

Examiner: Prof. Dr. Christian Schmitt

Supervisor: Prof. Dr. Christian Schmitt

Affidavit

I hereby declare that this thesis was written without the help of third parties and only with the specified sources and aids. All used passages have been marked. This thesis has not yet been submitted in the same or a similar form to any examination institute.

Moritz Maidl

Munich, 15.01.2024

Table of contents

List of Illustrations

Table Dictionary

List of Abbreviations

OLS	Ordinary Least Squares
GFC	Global Financial Crisis

1 Introduction

1.1 Background and Motivation

According to recent data published by data provider BarkleyHedge, assets under management in the hedge fund industry have reached more than five trillion U.S. Dollars in the second quarter of 2023. This represents a significant increase of about two-thirds compared to the three trillion invested before the COVID-19 pandemic.[1] Despite already having experienced strong inflows during the bull market in the last decade, professionals expect continued growth rates of around 3.5% on average until 2032 with AuM reaching levels of more than 7.1 trillion U.S. Dollars.[2] These developments run counter to the fact that hedge funds have on average, as shown later, underperformed market benchmarks like the S&P500 over the period mentioned. Apparently, superior performance does not always play a role for investors looking to invest their money. Instead, it is common knowledge that an essential attraction of hedge funds is their ability to add diversification from traditional assets to an investor's portfolio. As Global Markets Insights point out in their 2023 hedge fund markets summary, next to the growing interest of institutional investors and high-net-worth individuals as well as hopes for improved economic conditions, is a rising demand for diverse investment strategies in the hedge fund industry.[3]

Diversification is still a decisive argument for investments in hedge funds and while AuM continue to grow, investors may wonder where the funds' returns come from. Managers claim that these returns can be attributed to their superior skills in understanding the markets and subsequently making profitable investment decisions. Contradicting that way back in 2001, Asness, Krail and Liew show that hedge funds are indeed exposed to systematic risk factors[4], followed by numerous studies coming to similar conclusions. A recent study like additionally examines a systematic decline in returns attributable to managerial skills.[5] Common criticism of hedge funds also includes illiquidity and high fees. The method of hedge fund replication tries to offer a solution to these three problems. With the seminal research by Hasanhodzic and Lo in 2006, the replication of hedge fund returns, as well as the construction of hedge fund clones, became popular for the first time. Since then, many studies have covered this task, leading to numerous different results on whether it is possible to calculate investable products that lack the three named issues while generating similar returns as the hedge funds.

Looking at the state of the economy, market trends, and important events like the COVID-19 crash, as well as the substantial rise of interest rates in recent months, it seems appropriate to investigate how these changes have influenced developments in the field of hedge fund replication.

---

[1] Cf. BarkleyHedge.
[2] Cf. Wadhwani (2023).
[3] Cf. Wadhwani (2023).
[4] Cf. Asness, et al. (2001), p. 1.
[5] Cf. Cherian, et al. (2020), p. 21.

The task of replicating hedge fund returns includes more than just predicting the returns of a hedge fund. It is essential to understand how hedge funds are exposed to systematic risk factors to be able to conclude where the returns come from. Only with this knowledge, the replication of hedge fund returns can be carried out.

1.2 Research Objectives

The following three research questions arise from the current analysis:

1. How are the factor exposures of hedge funds structured across different investment strategies in the selected period, also considering market developments in recent years?

2. Can the assumption of a decline in returns attributable to manager skill with hedge funds being significantly exposed to systematic risk factors be supported?

3. Can factor-based replication generate out-sample clones with similar performance compared to their respective hedge fund index over an extended period?

The thesis is structured as follows: Section 2 provides a review of the literature covering both the developments in literature as well as the theory behind hedge funds, the attribution of hedge fund returns, and the replication of hedge fund returns. Section 3 describes the statistical methods and later performance tests, followed by section 4 with a presentation and evaluation of the chosen risk factors and hedge fund indices. In section 5, which is the linear regression analysis, the empirical results are summarized and interpreted. Section 6 builds on the gained knowledge by calculating out-of-sample returns and clones. The thesis ends with a conclusion in section 7.

2 Literature Review / Theory

This section provides an up-to-date literature review on the history of research on hedge funds, hedge fund returns attribution, and hedge fund replication combined with the theory behind hedge funds and their replication. Understanding how hedge funds work, especially where their returns come from, why investors choose to invest in them, and why replication is still worth researching, is essential. With this knowledge, later investigation results can be correctly classified and interpreted.

2.1 Hedge Funds

Going back to the very beginning, when Alfred Winslow Jones created one of the first hedge funds in 1949, he introduced a way of investing capital that was entirely new for the financial markets. As a journalist and sociologist, Jones was the first who 'hedged' his portfolio of stocks against market drops by using leverage and short-selling. He did not receive much attention for his actions until a few years later and the hedge fund industry did not take off until the 1990s.[6] He also implemented a fee structure still used today by many hedge funds. This typical fee structure consists of a fixed management fee of 1-2% combined with an incentive fee of around 20% of the net gains.[7] Hedge fund managers aim to generate exceptional performance and 'abnormal' returns to compensate for these fees.

Since Jones' first fund, hedge funds have changed, but in principle, they still gather money from investors and invest these resources to achieve positive returns. Other than mutual funds, hedge funds generally now use more adaptable investment strategies. They often aim to generate profits in many market conditions by using techniques like leverage, short-selling, and other high-risk investment tactics that mutual funds typically do not use.[8] They often serve as a source for diversification in the investors' portfolios.

To make the case for why the replication of hedge funds is still worth investigating, it is imperative to look at the benefits and downsides of hedge funds. The three main downsides are common knowledge: high fees, low liquidity and very little transparency. Hedge fund managers try to make this up, as described earlier, by delivering abnormal returns for their investors.

Comparing the fee structure of an average hedge fund to the fees in an Exchange-Traded Fund (ETF), it is evident that an ETF with similar performance would be a more attractive investment. A 20% performance fee can make the incentive fee look comparably low depending on how well the hedge fund performs. This assumption may be contrary to a recent analysis examining the decline of hedge fund alpha combined with a steady inflow of money into hedge funds. They suspect that as managers increasingly lack the ability to

6 Cf. Ubide (2006).
7 Cf. Getmansky et al. (2015), p. 3.
8 Cf. SEC Investor Bulletin Hedge Funds (2013).

generate superior performance, the incentive fee perhaps plays a more important role.[9]

Certain hedge funds may also impose an exit fee of up to 5% to refund the remaining investors for the expenses associated with adjusting the fund's portfolio after an investor withdraws.[10] It is therefore also important to keep in mind that many hedge funds implement a lock-up period, during which investors are bound to the fund for a specific duration. This is not the case with ETF’s for instance.

Last but not least is the problem of the lack of transparency. Although hedge funds have improved their transparency in the last years after the GFC as they are looking to attract more investors[11], the Securities and Exchange Commission (SEC) still warns of the significant discretion many hedge fund managers give themselves. Investors, therefore, have, other than with, for instance, ETFs, very little knowledge of what is happening inside the products they buy. If a manager changes his strategy by changing the assets and/or the leverage, they might create an unwanted correlation with other assets in the investor's portfolio.

Nowadays, the hedge fund industry offers various strategies to cover all kinds of demands for diversification and outperformance in certain areas. These strategies include Equity, Event Driven, Macro, Relative Value and Multi-Strategy. Of course, there are more, and all strategies used in this study will be described in section 4.1, ‘Hedge Fund Indices’.

2.2 Attribution of Hedge Fund Returns

As mentioned, hedge funds are often seen as a convenient source for diversifying an investor's portfolio. Their above-average returns are supposedly generated from exploiting incorrect pricing and unconventional sources of risk premia. Claiming to run products with low correlation to traditional assets while generating abnormal returns for their investors, managers are charging high performance fees. However, academic literature shows that while many managers claim that the returns they generate can be brought into context with their superior ability to understand the market and, therefore, make profitable investment decisions, in reality, much of the returns are attributable to exposure to systematic risk factors.[12] Many studies cover the attribution of hedge fund returns and the following are probably the most influential in the context of this study.

In a 2001 study, Asness, Krail and Liew empirically examined the described common claims of hedge fund managers in a time frame from 1994 to 2000. For hedge fund indices, they find that they have significantly more market exposure than initially thought and do not add overall value during the examined period. This means that while in simple regression, in their example, the S&P500, the claims of modest market exposure and

[9] Cf. Cherian et al. (2020), p. 21.
[10] Cf. Kat & Palaro (2005), p. 6.
[11] Cf. Soerensen & Hansen (2020), p. 10.
[12] Cf. Cherian et al. (2020), p. 2.

positive value are correct, in reality, the returns attributable to systematic risk factors are much more significant, plus the high returns in the time frame are due to, as they call it, “swimming with the tide”.[13]

Three years later, in 2004, Fung and Hsieh also researched the attribution of hedge fund returns, this time with a non-conventional approach using asset classes that are directly observable from the market. They find that for average hedge fund portfolios, their proposed factors can explain up to 90% of the returns, which can be seen as very contrary to the stated claims hedge fund managers often make. In addition to that, they analyze the time-varying behavior of the alternative alphas and betas, which make up the returns, to show how funds-of-hedge funds adjust their strategies over time.[14]

Extending this, in 2010, Bali, Brown and Caglayan conducted empirical tests on the exposures of different hedge funds to various financial and macroeconomic risk factors. With the results of their study, they support the assumption even more that hedge fund returns are indeed exposed to systematic risk factors. They show that there is a significant correlation between hedge fund returns and important economic indicators, not just in certain economic situations, which leads to the situation that the exposures can predict cross-sectional variations in hedge fund returns.[15]

The foundation of this empirical study is the linear regression method, which can use risk factors with linear and non-linear payoffs to predict the values of a dependent asset. In many hedge funds, non-linear payoffs are part of the strategy, so there is already a significant amount of literature covering non-linear payoffs in hedge funds.[16] This must be considered when choosing the risk factors later.

Cherian, Kon and Li also find in their 2020 study on hedge fund index attribution and replication that since the GFC, the manager-specific alpha has declined significantly, which they attribute to the still growing funds in North America, which, as a result, face diminishing opportunistic returns. They also empirically show how alternative betas make up a large part of the variance in the indices, with systematic risk factors explaining up to 81%.[17]

2.3 Hedge Fund Replication

With the knowledge that hedge fund returns may arise in large parts from exposure to systematic risk factors, it is no wonder that researchers have conducted studies on replicating their returns for more than 20 years. However, back to the disadvantages of hedge funds: their fees are high, they are illiquid and very opaque. A desirable investment

[13] Cf. Asness, et al. (2001), p. 4.
[14] Cf. Fung & Hsieh (2004), p. 1.
[15] Cf. Bali et al. (2013), p. 41.
[16] Cf. Cherian et al. (2020), p. 7.
[17] Cf. Cherian et al. (2020), p. 35 f.

would be a financial product that could offset all three points and deliver a similar return with a similar risk exposure compared to the actual hedge fund. That is, in short, the overall goal of hedge fund replication.

There are three main replication approaches, with the last two being the subject of most of the scientific works in this field: the method of replicating the strategies, the payoff distribution method, and the linear factor-based replication method. In the strategy replication method, a manager uses a rule-based approach designed to copy a hedge fund's trades to achieve comparable returns. A problem with this approach is that if it only copies the trades, it somehow becomes a hedge fund, which differs from the matter's meaning. The payoff distribution method focuses on replicating the statistical properties of hedge fund return distributions. However, this approach faces significant problems: it can often be more complex and expensive than directly executing the hedge fund strategy. Third, the linear factor-based replication method constructs a portfolio using linear regression with investable factors as the independent variables. The aim is to create similar betas and risk premiums compared with the hedge fund or hedge fund index that is being replicated.[18]

As previously stated, this study will focus on the linear factor-based replication method when creating full replicas and hedge fund index clones. Hasanhodzic and Lo laid the foundations for this strategy in 2006 when they published the first complete study on factor-based replication with fixed-weight windows and rolling window regressions. Most papers in the years after used their strategies and findings as starting points for their research. They found that using six different common, liquid, exchange-traded risk factors in linear regression can create index clones with similar returns to their respective counterparts. However, they also notice that most clones are inferior and certainly need further improvement like being created with non-linear methods.[19] The clones still at least partly fulfill their tasks as they are more transparent, cheaper and passive than the clones they aim to replicate.

In the years after, the factor-based method was continuously expanded by using more and different factors, different hedge funds and hedge fund indices, changing the time frames, adding factors with non-linear payoffs, or implementing more advanced techniques like the Kalman filter into the rolling window regression.

Two recent studies by Cherian, J.A. et al. in 2022 and Westphal Soerensen, M. et al. in 2020 use the linear regression factor-based approach with time frames covering the last two decades. Both struggle to create clones that perform similarly to the chosen hedge fund index strategies. They examine that it is still hard to build clones that can keep up with hedge funds during all market situations. However, they both write that it may be too early to give up on hedge fund replication strategies.

[18] Cf. Getmansky et al. (2015), p. 59.
[19] Cf. Hasanhodzic & Lo (2006), p. 43.

3 Method

This section introduces the theoretical foundations behind the investigations in the sections ‘Regression Analysis’ and ‘Replication Analysis’. As this is an empirical study, it is necessary to understand how the statistical methods work, why values are treated in a certain way, and how to interpret later results correctly. It starts by explaining the (multiple) linear regression, on which all analyses in this study are built. Then, it continues with a vital part of the regression, the estimation of the parameters in the model. This is done using the OLS method. After that, the knowledge gained on decomposing hedge fund returns can be extended by introducing a way to replicate the indices. This study uses the factor-based rolling windows replication method. First, out-of-sample replication returns will be calculated and evaluated, followed by the final part of the work, the creation and interpretation of hedge fund index clones.

3.1 Linear Regression

Linear regression is an extension of the standard statistical regression. In finance, simple linear regression is a statistical model that tries to explain the linear relationship between an independent assets returns and a dependent assets returns. Extending this, multiple linear regression uses two or more independent factors to predict the outcome of the dependent variable.[20] In the case of this work, the regression variant uses a given set of risk factors with either linear or non-linear payoffs to decompose the returns of a hedge fund index into the manager-specific alpha and the part of the returns, which can be attributed to systematic risk exposures.

The following formula defines the multiple linear regression:

$$\mathrm{R_t} = \alpha + \beta_1 \mathrm{F_{1t}} + \cdots + \beta_i \mathrm{F_{it}} + \varepsilon_t \tag{1}$$

with:

$\mathrm{R_t} =$ Hedge Fund Index Return at time t

$\alpha =$ Alpha (Manager − specific)

$\beta_i =$ Beta (Risk exposure for risk factor i)

$\mathrm{F}_{it} =$ Return of risk factor i, at time t

$\varepsilon_t =$ Estimated specific risk in the return of the hedge fund

[20] Cf. Blokhin (2023).

As shown in Soerensen and Hansen in 2020, the returns of an asset, the dependent variable in this case, can be explained with this linear formula. It is composed of the alpha (α) and the risk factors (F_{it}) with their associated sensitives, which are called betas (β_i).[21] Outlined in the introduction, the alpha describes the part of the risk exposure of a hedge fund that is attributable to superior managerial skill. However, this is only the basic formula for conducting a linear regression.

For this thesis, much of the replication efforts are built upon the research of Hasanhodzic and Lo from 2006, who, as described in the literature section, have laid the foundations of linear factor-based hedge fund replication. They were not the first to create a model like this but used it more complexly than anyone before.

This is an extended formula based on the linear regression formula:

$$\mathrm{R_{it}} = \alpha_\mathrm{i} + \beta_\mathrm{i1}\,\mathrm{Stock_t} + \beta_\mathrm{i2}\,\mathrm{Bond_t} + \beta_\mathrm{i3}\,\mathrm{Currency_t} + \beta_\mathrm{i4}\,\mathrm{Credit_t} + \beta_\mathrm{i5}\,\mathrm{Volatility_t} + \qquad (2)$$

$$+ \beta_\mathrm{i6}\,\mathrm{Commodities_t} + \beta_\mathrm{i7}\,\mathrm{Size_t} + \beta_\mathrm{i8}\,\mathrm{Value_t} + \beta_\mathrm{i9}\,\mathrm{Momentum_t} + \beta_\mathrm{i10}\,\mathrm{OTM\ Short\ Put_t}$$

$$+ \varepsilon_\mathrm{it}$$

with:

$\mathrm{i} = \text{Hedge Fund Index}$

$\mathrm{t} = \text{Defined point in time}$

As displayed here, the formula is very similar to the one used by Cherian, Kon and Li in their 2020 research on the possibility of hedge funds having resurrected as traditional beta.[22] It also just expands the multiple linear regression formula (1) by inserting the pre-defined risk factors into the equation.

Section 4, ‘Data Analysis’, defines and develops the risk factors shown in the formula.

Hasanhodzic and Lo also mention in their 2006 study that the manager-specific alpha is not risk-free.[23] It is not just added to the other risk factors and their associated sensitives as a fixed given return but instead can be influenced by risk factors that are different than the ones used in this study. It can change over time due to different reasons. This must be kept in mind for the later interpretation of regression results.

[21] Cf. Soerensen & Hansen (2020), p. 39.
[22] Cf. Cherian et al. (2020), p. 14.
[23] Cf. Hasanhodzic & Lo (2006), p. 11.

3.1.1 Ordinary Least Squares

The idea behind the linear regression model in the context of hedge fund replication is to find parameters that make the replicated returns / clones follow the returns of the hedge fund indices as closely as possible. One of the most popular ways of doing this is using the OLS method. This method tries to estimate the sensitives, later associated with their respective factors, so the disturbance is as close to zero as possible. Compared to the method of maximum likelihood, the mathematical implementation of OLS is much easier.[24]

Seven different assumptions can be listed that lay the groundwork for the best possible estimation of the parameters in the linear regression.[25] Expanded on the conditions and requirements of this study, these assumptions are as follows:

1. “The regression model is linear in the parameters, though it may or may not be linear in the variables.”[26]
2. Generally, the independent variables (risk factors) should be independent of the error term.
3. The mean value of the disturbance shall be zero.
4. The variables should have the same variance, meaning there is homoscedasticity present.
5. The returns should not be subject to autocorrelation.
6. It is essential that the number of observations is greater than the number of parameters to be estimated → the number of months in this work is 165.
7. The values of the different variables are not allowed to be the same.

However, the authors of the cited study add that these assumptions may not all be realistic, and in fact, the predictions made on these assumptions matter more.[27] Not all of these assumptions will be tested for in this study.

3.2 Factor-Based Replication

Multiple linear regression shows a way of decomposing the dependent variables returns. This helps to understand the factor exposures, risk profiles, and performance drivers of the chosen hedge fund indices. The factor-based replication method transfers this into the world of hedge fund replication. Models based on factor replication aim to reproduce the (monthly) returns of a chosen hedge fund index by constructing a clone that imitates the risk factors and their corresponding sensitivities. By identifying and understanding these risk factors and their associated sensitivities, it should be possible to create a portfolio that uses the chosen risk factors to yield returns comparable to those of the hedge fund

[24] Cf. Damodar & Porter (2009), p. 55.
[25] Cf. Damodar & Porter (2009), p. 62 ff.
[26] Cf. Damodar & Porter (2009), p. 62.
[27] Cf. Damodar & Porter (2009), p. 68 f.

index.[28] The risk factors can have linear and non-linear payoffs and the model is executed with the rolling-window method, which will be described later.

The formula for the factor-based replication with rolling windows is:

$$R_{it-k} = \alpha_i + \beta_{i1}\,\mathrm{Stock}_{t-k} + \beta_{i2}\,\mathrm{Bond}_{t-k} + \beta_{i3}\,\mathrm{Currency}_{t-k} + \beta_{i4}\,\mathrm{Credit}_{t-k} + \tag{3}$$

$$\beta_{i5}\,\mathrm{Volatility}_{t-k} + \beta_{i6}\,\mathrm{Commodities}_{t-k} + \beta_{i7}\,\mathrm{Size}_{t-k} + \beta_{i8}\,\mathrm{Value}_{t-k} +$$

$$+ \beta_{i9}\,\mathrm{Momentum}_{t-k} + \beta_{i10}\,\mathrm{OTM\ Short\ Put}_{t-k} + \varepsilon_{it}$$

with:

$$k = 1 \ldots 60$$

This formula combines the linear regression formula (2) and the rolling window method[29]. It is evident that hedge fund indices change their exposures over time[30], which is why their replicated returns and clones must be able to do the same. In the fixed-weight method, the “portfolio weights are fixed through time for each fund”. This leads to a certain “look-ahead” bias[31], which is not desirable. Rolling windows are, therefore, the method of choice.

The first step is to define the size of the window, which will be 60 months (e.g., five years) in this work. This is long enough to create a statistically acceptable regression but not too long to miss important changes in the index's development by accident or capture them only partially later. The window is set to start at month 1, which means it conducts a regression of months 1 to 60. Then it moves one window right and again regresses over a 60-month time frame, which spans now from month 2 to month 61. This continues until the right side of the window meets the last month (165) in the data frame. To describe it more technically, the rolling alphas and betas are calculated by a rolling window, which goes from (t – 60) to (t - 1) for each month t and each fund i, with t being the current month. To obtain the replicated return for a specific window, all the sensitives are multiplied with their respective factors from month t and then summed up. This means the replication-created returns span from month 61 to month 165, delivering 105 monthly returns. As already mentioned, the chosen time frame and the risk factors are defined in section 4 ‘Data Analysis’.

[28] Cf. Soerensen & Hansen (2020), p. 39.
[29] Cf. Cherian et al. (2020), p. 25.
[30] Cf. Hasanhodzic & Lo (2006), p. 22.
[31] Cf. Hasanhodzic & Lo (2006), p. 19.

3.2.1 Hedge Fund Clones

The final part of this work is the calculation of hedge fund clones.

The construction of the clones is based on a slightly shortened formula:

$$R_{it-k} = \beta_{i1}\,\mathrm{Stock}_{t-k} + \beta_{i2}\,\mathrm{Bond}_{t-k} + \beta_{i3}\,\mathrm{Currency}_{t-k} + \beta_{i4}\,\mathrm{Credit}_{t-k} + \beta_{i5}\,\mathrm{Volatility}_{t-k} + \beta_{i6}\,\mathrm{Commodities}_{t-k} + \beta_{i7}\,\mathrm{Size}_{t-k} + \beta_{i8}\,\mathrm{Value}_{t-k} + \beta_{i9}\,\mathrm{Momentum}_{t-k} + \beta_{i10}\,\mathrm{OTM\ Short\ Put}_{t-k} + \varepsilon_{it} \tag{4}$$

with:

$$k = 1 \dots 60$$

$$\beta_{i1} + \cdots + \beta_{i10} = 1$$

The most apparent change is omitting the intercept (alpha)[32] because there is no possibility to somehow invest into returns attributable to the manager's specific skills.

Next, for every window, the betas are set to sum to one to ensure the same amount of money is invested into the clone as the hedge fund index. This is done by summing the betas up, subtracting them from one and then taking either a long or short position in the risk-free asset according to the difference.[33] This will be later part of the interpretation of the clone returns as it, in the case of this study, has an unintentionally strong influence on the clones returns. Generally, a positive beta value means going long in an asset and a negative beta value means going short.

Finally, an additional normalization is implemented to ensure that the index and the clone returns have the same volatility. It is crucial when comparing the performance of the clone and the index. Hasanhodzic and Lo first did this with their replication efforts in 2006 to make a fair comparison possible.[34] The normalization works by dividing the hedge fund index's standard deviation by the clone's standard deviation. The result is the renormalization factor, which is then multiplied with the clone returns.

It must be noted that the renormalization factor changes the leverage of the clone. A factor >1 means positive leverage, and a factor <1 means negative leverage.

32 Cf. Hasanhodzic & Lo (2006), p. 20.
33 Cf. Soerensen & Hansen (2020), p. 40.
34 Cf. Hasanhodzic & Lo (2006), p. 21.

3.3 Performance Measurements

The performance and how close and consistent the forecasted values of the out-of-sample predicted returns and clones are compared to the hedge fund index values can only be evaluated with proper statistical measurements.

For the analysis of the predicted hedge fund returns only the mean returns will be compared after carrying out a paired t-test. For the analysis of the hedge fund clones, the mean returns and the Sharpe ratios will be compared. In addition to that, the following performance measurements will be conducted: RMSE, Tracking Error and Theil's Inequality Coefficient.

3.3.1 Mean Return

The mean return is a simple way of describing an investment's average return over a defined period. It is used to capture the central tendency of the return distribution so investors can estimate what they can earn from this investment in the future.

In the calculation, all the returns during the defined period are summed up and divided by the number of returns:

$$\text{Mean Return } = \frac{(\text{Return}_1 + \text{ Return}_2 + \ldots + \text{Return}_\text{n})}{n} \tag{5}$$

with:

$\text{n } = \text{number of all returns}$

This study uses the mean return when comparing the realized returns from hedge fund indices with the expected returns from the out-of-sample replications and clones.

3.3.2 Sharpe Ratio

William F. Sharpe developed the Sharpe ratio and first proposed it in 1966. It is "designed to measure the expected return per unit of risk for a zero investment strategy."[35] It tells investors how much excess return they receive for the extra volatility they endure for holding a riskier asset.

In the calculation, the risk-free asset is subtracted from the returns of the asset, therefore

[35] Cf. Sharpe (1993).

creating the excess return which is divided by the standard deviation of the portfolio's excess returns:

$$\text{Sharpe Ratio} = \frac{\mathrm{r_p} - \mathrm{r}_f}{\sigma_\mathrm{p}} \tag{6}$$

with:

$\mathrm{r_p}$ = Return of the portfolio (hedge fund index or clone)

$\mathrm{r_f}$ = Return of the risk − free rate

σ_p = Standard deviation

Therefore, a higher Sharpe ratio is desirable as it indicates a more attractive risk-adjusted return. Later, the returns for the clones and their respective hedge fund indices will be calculated in excess of the risk-free rate, so in principle, all the following tests will be conducted with returns that are in excess of the risk-free rate.

3.3.3 Root Mean Square Error

The RMSE measures the difference between predicted and actual values, in the case of this study, the expected returns from the clones and the realized returns from the indices. It effectively shows how close the expected returns are to the realized returns by checking the accuracy of the predictive model against the actual performance.

In the calculation, the realized returns are subtracted from the predicted returns, with the results being squared. Then, the mean is calculated from these results and in the last step, the square root is taken:[36]

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(predicted - realized)^2} \tag{7}$$

The result should be as low as possible, meaning that the predicted returns of the clones are close to the realized returns of the hedge fund indices.

[36] Cf. StatisticsHowTo.

3.3.4 Tracking Error

The tracking error measures the difference between the expected and realized returns. It quantifies how closely a portfolio follows the index to which it is benchmarked. In the case of this study, how closely the predicted returns of the clones follow the realized returns of the hedge fund indices.

In the calculation, the realized returns are subtracted from the predicted returns and the results are multiplied with the standard deviation:[37]

$$Tracking\ Error = \sigma * (Predicted\ Returns - Realized\ Returns) \tag{8}$$

The result should be as low as possible, meaning that the clones are closely following the realized returns of the hedge fund indices.

3.3.5 Theil's Inequality Coefficient

Theil's inequality coefficient measures how accurate the predictions made by the clone are compared to the realized returns of the hedge fund indices. The clone returns must be good predictions and not random guesses.

In the calculation, the RSME is divided by the square root of the square of the mean clone return minus the square root of the square of the mean hedge fund index return:[38]

$$Theil's\ Inequality\ Coefficient = \frac{RMSE}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(predicted)^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}(realized)^2}} \tag{9}$$

The result should be as low as possible, meaning that the clones are good at predicting the realized returns of the hedge fund indices.

With all the empirical methods, models and performance measurements set, the following section deals with selecting and processing the data later used in sections 5 and 6.

---

[37] Cf. Chen (2020).
[38] Cf. Soerensen & Hansen (2020), p. 55.

4 Data Analysis

For the research tasks ahead in this study, data from hedge fund indices and data from common available liquid risk factors are needed. Choosing appropriate data for the linear regression and the construction of hedge fund clones is crucial to laying a foundation for statistically valid and understandable results. Therefore, historical returns from trusted sources like Eurekahedge, Investing.com and the Federal Reserve Bank of St. Louis are being used. Despite the larger size, the data still needs to be chosen carefully, so before being used in the research, the obtained data will be subject to some research, transformations and statistical tests.

The overall time frame is from January 2010 through September 2023. This is after the GFC, followed by one of the longest bull markets in history (2009 – 2020). However, it also includes very recent market events like the drawdown due to the COVID-19 pandemic, the Russian war against Ukraine, the impacts of high inflation and an almost unprecedented rise of interest rates after a very long time of rates being near zero. To date, very few have explored these recent events in the context of hedge fund replication. As mentioned before, the number of studies available on this topic, especially around and in the years after the GFC, is quite large, so adding a more current perspective to the discussion will give new insights into the development of hedge fund replication.

Additionally, all data used is collected in US dollars for both the hedge fund indices and risk factors.

4.1 Hedge Fund Indices

Since this study aims to provide a broad view of the developments and the current situation in hedge fund replication, it is more beneficial to use hedge fund indices instead of individual funds. These are designed to give an investor an overview of the general performance managers archive with the strategies. The indices used in this study are downloaded from Singapore-located data provider Eurekahedge, which has an extensive and primarily up-to-date database of hedge funds and hedge fund indices. This is helpful because there is no need to replace missing data with zeros, and the indices are likely high quality. Eurekahedge follows the so-called “Eurekahedge Equal Weighted Hedge Fund Index methodology,” which means that its indices are neither asset-weighted nor with median returns. The monthly index values are the means of the monthly returns of all hedge fund constituents, and the returns are the gains and losses of the total portfolio values by performance (net of all fees), with AUM inflows/outflows being excluded from it. They aim to “simply give an overview of the average performance of hedge funds, without attempting to highlight monthly inflows and unjustly overweigh the performance of certain funds due to good marketing staff or location in investor hot spots.”[39] Being free to use

[39] Cf. Eurekahedge Equal Weighted Hedge Fund Index.

and providing reliable data, downloading data from Eurekahedge is, therefore, the best choice for this study.

As Westphal Soerensen, M. et al. pointed out in 2020, having as many constituting funds in an index as possible is essential to get a broad performance for research.[40]

The following is a list of all ten investment strategies used in this work with their respective amount of funds as well as the abbreviations later used:

- Eurekahedge Hedge Fund Index (3241 Funds) → EHIMain
- Arbitrage (90 Funds) → EHIArbitrage
- CTA/Managed Futures (405 Funds) → EHIManagedFutures
- Distressed Debt (19 Funds) → EHIDistressedDebt
- Event Driven (127 Funds) → EHIEventDriven
- Fixed Income (496 Funds) → EHIFixedIncome
- Long/Short Equities (1235 Funds) → EHILSEquity
- Macro (217 Funds) → EHIMacro
- Multi-Strategy (355 Funds) → EHIMultiStrategy
- Relative Value (62 Funds) → EHIRelativeValue

The strategies above are the main strategies Eurekahedge lists on its website, with the Eurekahedge Hedge Fund Index being the flagship index, which combines several strategies with the overall goal to “provide a broad measure of the performance all underlying hedge fund managers irrespective of regional mandate.”[41]

Below is an extended overview of the Eurekahedge-defined hedge fund strategy definitions[42] to familiarize with and to set the groundwork for later interpretation of results in the analysis of the regression and replication. Some of the strategies, of course, do overlap in certain parts of their implementation.

[40] Cf. Soerensen & Hansen (2020), p. 29.
[41] Cf. Eurekahedge Hedge Fund Index.
[42] Cf. Eurekahedge Equal Weighted Hedge Fund Index.

4.1.1 Arbitrage

The Eurekahedge arbitrage index comprises hedge funds whose managers conduct some form of arbitrage strategy. This strategy involves the simultaneous purchase and sale of an asset to profit from a difference in the price in two different markets, exploiting market inefficiencies. It works by buying an asset in a market where the price is “lower” and selling it in another where the price is “higher”. Depending on the strategy and market situation, it bears almost no risk, leading to more efficient markets by correcting these inefficiencies. Popular sub-strategies are “triangle arbitrage,” which is just the simple exploitation of price differences between three different currencies; “merger arbitrage,” which is “mostly the betting on failure or success of M&A’s”[43] and “capital structure arbitrage,” which looks for pricing imbalances within a company's debt and equity instruments to find profitable trades, often aiming for market-neutral positions. Arbitrage strategies, therefore, often claim to be market-neutral.

4.1.2 CTA/Managed Futures

Hedge fund managers who follow a CTA/Managed Futures strategy trade futures contracts in various markets such as commodities, currencies, interest rates, and equity indices. They use different strategies like trend following and global macro but also systematic strategies. According to Eurekahedge, they conduct their transactions directly or through a CFTC-registered Commodity Trading Advisor. “Most managed futures programs are systematic and quantitatively driven via mathematical models and computational power to help guide their trading decisions.”[44] This point sets this strategy apart from all the others.

4.1.3 Distressed Debt

Managers conducting this strategy buy companies' debt in financial distress or undergoing reorganization at lowered prices. Commonly, these securities are sold by investors seeking to avoid the risks associated with companies near or in bankruptcy. This means they can often be purchased at a significant discount, offering the potential for high returns if the company recovers. These discounts are due to the debts often being “illiquid and inefficiently priced because of forced or emotional selling by investors, low coverage by analysts, or a high degree of risk aversion by the market.”[45]

[43] Cf. Baker & Filbeck (2017), p. 208.
[44] Cf. Inglis (2017).
[45] Cf. Baker & Filbeck (2017), p. 216.

4.1.4 Event Driven

Like the arbitrage strategy, the event driven strategy takes advantage of pricing inefficiencies that may occur before or after corporate events, such as mergers, acquisitions, spin-offs, bankruptcies, or other significant corporate changes. The managers try to use detailed analysis to predict the outcome and timing of these events. Often, hedge funds move between different event driven strategies depending on the current market situation.[46] Risk-arbitrage and distressed debt are sub-strategies of the event driven strategy.

4.1.5 Fixed Income

This strategy invests in fixed income securities and tries to use anomalies in pricing them to its advantage. By using leverage, fixed income arbitrage attempts to profit from the spreads between related bonds or fixed income securities. It includes investing in high-yield bonds, distressed securities, or mortgage-backed securities.

4.1.6 Long/Short Equities

“A long/short equity strategy is a portfolio management approach that most resembles the one originally followed by the Alfred W. Jones , the first hedge fund manager.”[47] This strategy seeks to take advantage of market risks by going long in positions that may increase in value and going short in positions that may decrease in value. The net market exposure of the fund can, at best, be fully hedged through short-selling. The use of leverage, options, and futures can amplify absolute returns. Other than market-neutral strategies, long/short equity strategies often involve directional bets, where the manager predicts that certain stocks will perform better than others, irrespective of overall market trends.

4.1.7 Macro

Macro funds use a top-down investment approach by exploiting economic trends and changes in policy that can influence global markets (the strategy is therefore often called ‘global macro’). Managers have “mandates to invest with derivatives or leverage in a wide variety of markets, including currencies, commodities, interest rates, and similar assets.”[48] They often use leverage and derivatives to increase their exposure, significantly influencing the fund's performance.

---

[46] Cf. Baker & Filbeck (2017), p. 207 f.
[47] Cf. Baker & Filbeck (2017), p. 186.
[48] Cf. Baker & Filbeck (2017), p. 228.

4.1.8 Multi-Strategy

Multi-strategy fund managers achieve diversification through the adoption of multiple investment strategies. A hedge fund can, therefore, allocate capital flexibly to the most favorable opportunities at any given market event. There are both single multi-strategy hedge funds and multi-strategy funds of hedge funds.[49] The volatility and risk-return profile for multi-strategy funds can vary depending on the strategies chosen and how aggressively they are pursued.

4.1.9 Relative Value

"Relative value trading is an investment strategy in which fund managers attempt to identify and exploit pricing discrepancies among the same or related securities using long and short positions."[50] Sub-strategies include fixed income arbitrage, capital structure arbitrage, and long/short equities. Relative value strategies try to be market-neutral, meaning they profit regardless of the overall market's direction.

4.1.10 Hedge Fund Index Performance

With the time frame and strategies set, evaluating how each strategy has performed over the last almost 14 years, from January 2010 to September 2023, is necessary. As managers allow only limited transparency, evaluating the performance when comparing different strategies is even more critical. There are a few general findings in the literature regarding the performance and volatility of hedge funds. On the one hand, it has been proven that hedge funds do consistently beat mutual funds but not market indices as well as that they are more unstable than both.[51] On the other hand, research by Liang and Kat from 1999 disputes this as they observed data from 1990 to 1999 and discovered that hedge funds were less volatile than market benchmarks like the S&P500.[52] It seems to depend at least partly on what time frame is being analyzed and which market benchmarks are used.

In addition, it is said that hedge funds perform better with less volatility in bear markets but struggle to keep up with the market index's performance in bull markets. A study by Metzger and Shenai from 2019 shows that between June 2007 and January 2017, a time frame that covers the entire GFC, only three of ten strategies performed worse than the S&P500.[53] The market index also had the worst drawdown of all assets when the crisis hit. This means the managers can deliver more stable investments to their investors by

[49] Cf. Baker & Filbeck (2017), p. 259.
[50] Cf. Baker & Filbeck (2017), p. 242.
[51] Cf. Ackermann et al. (1999), p. 833.
[52] Cf. Liang (1999), p. 10.
[53] Cf. Metzger & Shenai (2019), p. 12.

exposing them to less volatility, which is especially valuable in bear markets. The conclusion is that a significant part of the performance must be due to superior managerial investment skills.

It indeed seems to be very different when looking at a more extended period covering a bull market in large parts. Figure 1 shows the performance of the ten hedge fund strategies and the S&P500 index over a total of 165 months. It is striking how much better the market index performs compared to all other funds. While the S&P500 has quadrupled, nine out of ten funds only doubled, with the distressed debt strategy standing slightly above. The risk-free rate hardly moves but shows some gains starting in the middle of 2020, probably due to the interest rate hikes.

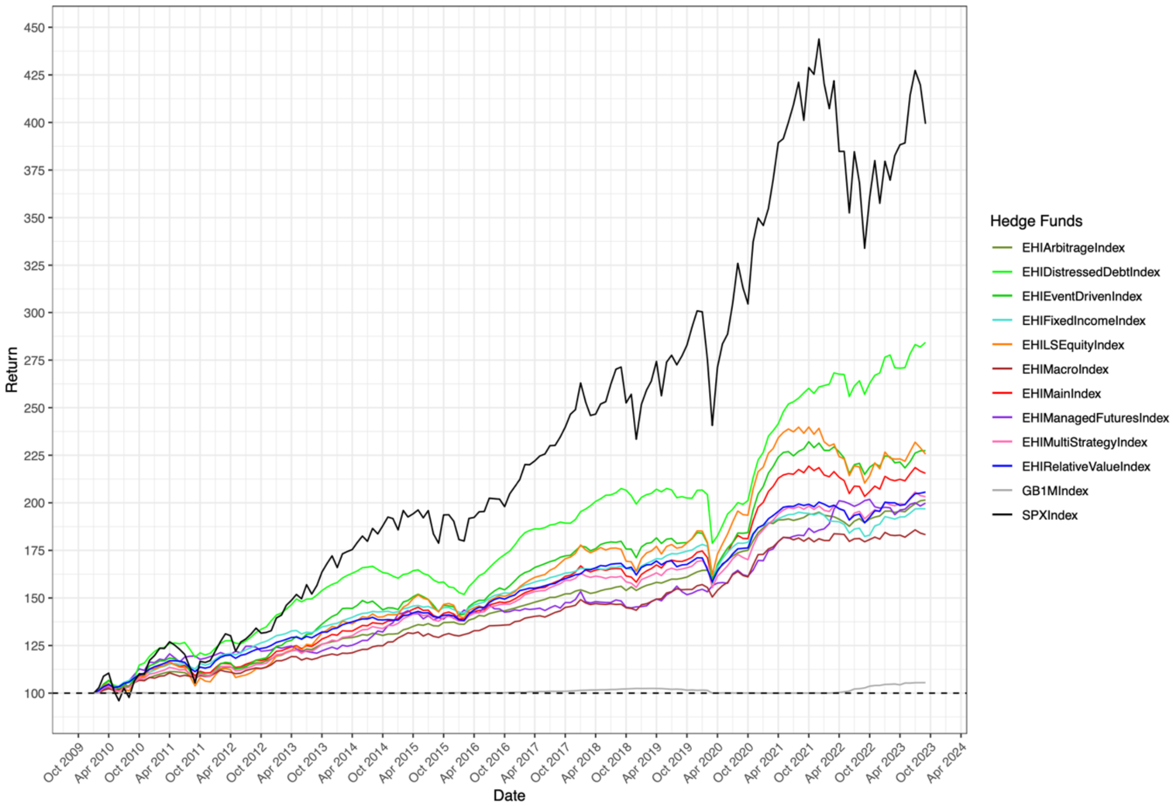


Figure 1: Performance of the Hedge Fund Indices, S&P500 & Risk-free rate

The visualization confirms the assumption that hedge funds underperform market indices in bull markets. While the hedge fund indices deliver positive returns, these may come from something other than a superior manager's skill. This is similar to Asness, Krail and Liew, who suggested in 2001 that the good performance of hedge fund returns between January 1994 and September 2000 came from "swimming with the tide".[54] It is also visible that the hedge fund indices are less volatile, especially during the COVID-19 crash and in the correction after the bull market following the rebound from COVID-19. If only these short periods were examined, the hedge fund indices would probably perform similar or better than the market index, as they did during the GFC.

[54] Cf. Asness, et al. (2001), p. 4.

4.1.11 Hedge Fund Index Autocorrelation

Illiquidity and serial correlation affect some hedge fund strategies, which can lead to seemingly 'smoothed' returns. This phenomenon means that the observed volatility is lower than the underlying volatility, leading to a Sharpe ratio that may not accurately reflect the actual risk profile.[55]

Therefore, a Ljung-Box test is conducted to check for overall randomness based on the chi-squared distribution, with a joint null hypothesis that the first five lags are not correlated. The results of this test, as shown in Table 1, are somewhat surprising, with none of the tests qualifying for autocorrelation.

HF Index Strategy	Funds	Lag 1 autocorrelation coefficient	Q* Ljung Box Statistic (Lag 1)	Q* Ljung Box Statistic (Lags 1 to 5)
EHI Main	3241	0,09	1,26	2,49
EHI Arbitrage	90	0,09	1,35	5,28
EHI CTA/Managed Futures	405	-0,05	0,36	4,98
EHI Distressed Debt	19	0,21	7,55*	8,98
EHI Event Driven	127	0,10	1,73	2,14
EHI Fixed Income	496	0,08	1,21	3,18
EHI Long/Short Equity	1235	0,06	1,25	0,61
EHI Macro	217	0,02	0,08	0,61
EHI Multi Strategy	355	0,06	0,68	2,87
EHI Relative Value	62	0,14	3,28	5,28

Chi-squared distribution at 5%-level: Lag 1 = 3.841459 ; Lag 5 = 11.0705

* means evidence of significance at the 5%-level

Table 1: Hedge Fund Index Autocorrelation Test

The threshold is only passed once, by the first lag of the Distressed Debt index. Figure 2 on the next page shows the visualized statistics for this test.

This means that no correction for autocorrelation is needed. As previously explained, funds with illiquid assets tend to use smoothing techniques to flatten out their returns.

[55] Cf. Cherian et al. (2020), p. 9.

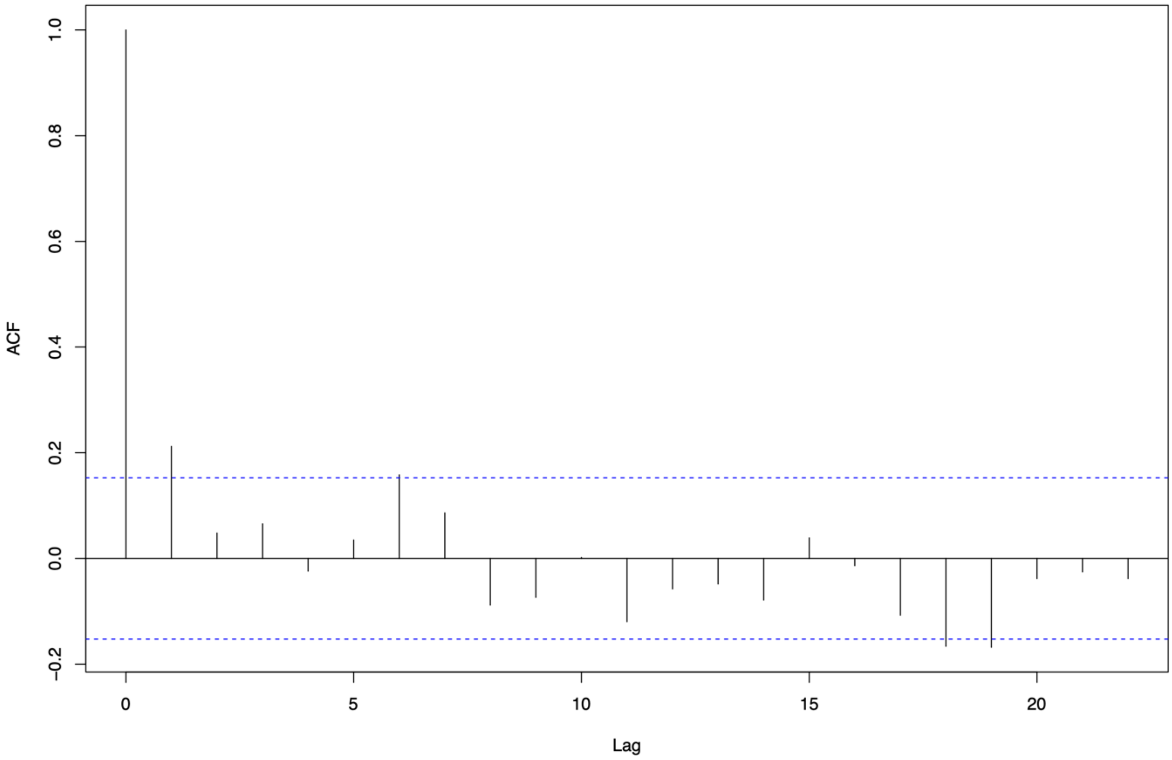


Figure 2: Autocorrelation in EHI Distressed Debt

4.2 Risk Factors

The risk factors used in this work are carefully chosen, based on existing research, and expanded on my own. They are similar to those used by Cherian, Kon and Li in 2020, except they are closer to reality.[56] For the success of the upcoming linear regression, the right choice of risk factors is not only important but decisive. All data has been collected from the data providers Investing.com and the Federal Reserve Bank of St. Louis. The first months of some data sets had to be filled with zeros because no data existed there.

The first six risk factors are similar to the ones used by Hasanhodzic and Lo in their replication efforts from 2006, which serves as the starting point for most research conducted on this topic in the years after.

They used currencies, bonds, credits, stocks, commodities and volatility. These are the U.S. Dollar Index, the Lehman Corporate AA Intermediate Bond Index, the spread between the Lehman BAA Corporate Bond Index and the Lehman Treasury Index, the S&P500, the Goldman Sachs Commodity Index and the first-difference of the end-of-month value of the CBOE Volatility Index.[57]

The Fama-French factors Size and Value have also been chosen as stock-level

[56] Cf. Cherian et al. (2020), p. 12.
[57] Cf. Hasanhodzic & Lo (2006), p. 10 f.

systematic risk factors. The Size factor shows the excess return of a portfolio of small-cap stocks over a portfolio of large-cap stocks, as research has found that small companies tend to outperform large companies.[58] The Value factor shows the excess returns of value stocks over growth stocks, because value stocks tend to outperform growth stocks.[59]

As section 2, ‘Literature Review / Theory’, describes, many hedge funds are exposed to factors with non-linear payoffs. To add non-linear factors to the regression, Cherian, J.A., et al. built in their 2020 study an OTM Short Put factor to capture tail risks and a momentum-type asset constructed from lookback straddles.[60] Here, I have settled with a Downside Hedged ETF and a Momentum Factor ETF as they seem liquid enough to fulfill the same task. This also makes carrying out the linear regression later more realistic. It must be noted that the risk factors they constructed and the risk factors used in this study were not tested for correlation. Therefore, it cannot be said with certainty that they behave in a similar manner.

These are all risk factors used in the regression later:

- Risk-free asset: U.S. 1-Month Treasury Bill
- Equities: S&P500 minus the risk-free asset
- Fixed Income: iShares 7-10 Year Treasury Bond ETF minus the risk-free asset
- USDX: U.S. Dollar Index
- Spread: Moody's Seasoned Baa Corporate Bond Yield vs. Yield on 10-Year Treasury Constant Maturity
- Volatility: ProShares VIX Short-Term Futures ETF
- Commodities: iShares S&P GSCI Commodity-Indexed Trust
- Size: Russell 2000 (small cap) - Russell 1000 (large cap)
- Value: Russell 1000 Value - Russell 1000 Growth
- Momentum: iShares MSCI USA Momentum Factor ETF
- OTM Short Put: Invesco S&P 500® Downside Hedged ETF

[58] Cf. Fama & French (1993), p. 14.
[59] Cf. Fama & French (1993), p. 14.
[60] Cf. Cherian et al. (2020), p. 11.

Following this order, the abbreviations for the risk factors used in this study are: GB1M, SPX, IEF, DXY, BAA10Y, VIXY, GSG, SIZE, VALUE, MTUM and PHDG.

The isolation of the risk-free asset has a relatively limited impact on the returns of both assets, as the returns of GB1M are very small. This can be seen in figure 4 in the following section.

4.2.1 Risk Factor Performance

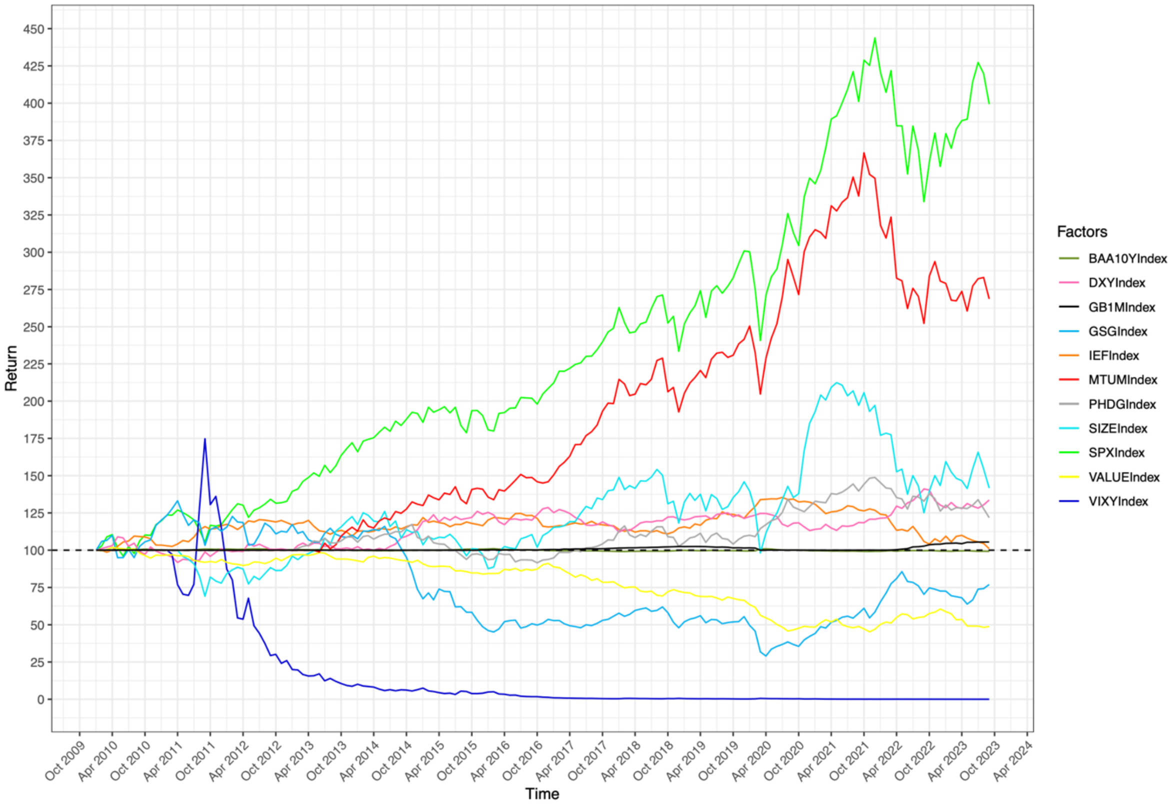


Figure 3: Performance of the risk factors and risk-free asset

Figure 3 visualizes the performance of all chosen risk factors, including the risk-free asset similar to the hedge fund indices in figure 1, again, for the entire period of 165 months. There are a few things that can be noticed in the graphs. Both the SPX and the MTUM factors outperform the others by a significant margin with most factors only adding a maximum of about 50% to their value. The spread (BAA10Y) has no visible movement, while all three risk factors, GSG, VALUE and VIXY, show negative performances. This is especially true for the VIXY factor, which has the goal profiting from increases in the expected volatility of the S&P500.

This line-up of risk factors should be able to capture most movements in the markets, making it well-suited for the linear regression.

4.2.2 Correlation Test

Before starting the regression analysis, the risk factors are tested for normality. This is done to avoid multicollinearity, which means that the independent variables are correlated.

	BAA10Y	SPX	IEF	DXY	VIXY	GSG	SIZE	VALUE	MTUM	PHDG
BAA10Y	1,00	-0,33	0,41	0,16	0,38	-0,51	-0,37	-0,21	-0,25	0,01
SPX	-0,33	1,00	-0,13	-0,47	-0,75	0,50	0,83	-0,11	0,77	0,57
IEFR	0,41	-0,13	1,00	-0,06	0,22	-0,38	-0,15	-0,37	0,06	0,02
DXY	0,16	-0,47	-0,06	1,00	0,28	-0,43	-0,52	0,04	-0,28	-0,18
VIXY	0,38	-0,75	0,22	0,28	1,00	-0,43	-0,74	-0,08	-0,60	-0,31
GSG	-0,51	0,50	-0,38	-0,43	-0,43	1,00	0,54	0,26	0,31	0,15
SIZE	-0,37	0,83	-0,15	-0,52	-0,74	0,54	1,00	0,04	0,72	0,46
VALUE	-0,21	-0,11	-0,37	0,04	-0,08	0,26	0,04	1,00	-0,26	-0,38
MTUM	-0,25	0,77	0,06	-0,28	-0,60	0,31	0,72	-0,26	1,00	0,62
PHDG	0,01	0,57	0,02	-0,18	-0,31	0,15	0,46	-0,38	0,62	1,00

Table 2: Correlation test of risk factors and the risk-free asset

It is visible that most of the factors show only very little correlation. The majority is below |0.50|, but some are high. The SPX has high positive correlations above 0.75 with the SIZE factor (0.83) and the MTUM factor (0.77). This is contradicts the general opinion that the factors used in a linear regression should not be highly correlated. Therefore, removing these two risk factors from the list is worth considering. However, only the SIZE factor will be omitted, as the MTUM factor represents one of the non-linear risk factors and its correlation is not much above 0.75. Tests have shown that omitting the SIZE factor only leads to a marginal decrease of the R-squared values in the linear regressions.

With these results, the theoretical foundations, as well as the data preparations, are complete. The following section will deal with the linear regression analysis of the hedge fund indices.

5 Regression Analysis

As described in section 3, 'Method' earlier, multiple linear regression uses two or more independent factors to predict the outcome of a dependent variable. In the following analysis, this method is used extensively. This regression variant uses a given set of risk factors with either linear or non-linear payoffs to decompose the returns of a hedge fund index into the manager-specific alpha and the part of the returns, which can be attributed to systematic risk exposures. Section 4, 'Data Analysis', gives insight into the risk factors as the independent variables and the hedge fund indices as the dependent variables.

The following regression analysis covers the entire time frame of this study, which is again from January 2010 until September 2023. This allows 165 monthly data points to be examined, with ten different hedge fund indices and nine different risk factors, which, according to assumption 6 in section 3.1.1, is more than enough for a statistically valid analysis.

The linear regression delivers a wide range of results. Figure 4 shows exemplary the regression results of the linear regression of the Eurekahedge Hedge Fund Index (EHIMain):

```
>   summary(regression_EHIMainReturn)

Call:
lm(formula = EHIMainReturn ~ BAA10YReturn + SPXReturn + IEFReturn +
    DXYReturn + VIXYReturn + GSGReturn + VALUEReturn + MTUMReturn +
    PHDGReturn, data = DataAll)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0119541 -0.0042675  0.0001479  0.0029116  0.0201801

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.0026478  0.0005224   5.069 1.13e-06 ***
BAA10YReturn -1.9107246  0.3735821  -5.115 9.17e-07 ***
SPXReturn     0.2408794  0.0256917   9.376  < 2e-16 ***
IEFReturn     0.0643248  0.0326875   1.968   0.0509 .
DXYReturn    -0.0032996  0.0285229  -0.116   0.9081
VIXYReturn    0.0007201  0.0045067   0.160   0.8733
GSGReturn     0.0275476  0.0106477   2.587   0.0106 *
VALUEReturn   0.0082265  0.0238854   0.344   0.7310
MTUMReturn    0.0420012  0.0223006   1.883   0.0615 .
PHDGReturn   -0.0417602  0.0280474  -1.489   0.1385
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.006134 on 155 degrees of freedom
Multiple R-squared:  0.8277,    Adjusted R-squared:  0.8177
F-statistic: 82.74 on 9 and 155 DF,  p-value: < 2.2e-16
```

Figure 4: Regression summary of the EHI Main Index

This summary can be created in RStudio with the lm-function, which conducts a linear regression using the risk factors and the hedge fund index. The line “Call:” shows which dependent and independent variables are used, with the following line, “Residuals,” examining the differences between the observed values and the values predicted by the model. Several different types of values are displayed in the “Coefficients”-section. For every risk factor, including the intercept, the “Estimate”, “Std. Error”, “t value” and “Pr(>|t|)” are calculated. The estimates are the most important values as they represent the expected change in the dependent variable for a one-unit change in the corresponding independent variable when all other variables are held constant. A positive value, therefore, means that when the risk factor increases its value, the hedge fund index is probably going to do so, too. It works the other way around the same way. The standard error shows how precise the estimate is, with a low value indicating a more precise result and vice versa. For the t statistic, the null hypothesis is that there is no association between the independent and dependent variables. It is calculated by dividing the estimate value by the standard error. If the t value is large, as with the Intercept, BAA10Y, SPX and GSG risk factors, there is a good chance that the association is statistically significant. Last in the “Coefficients”-section is the p-value, which tests the statistical significance of each risk factor. Usually, a p-value below 0.05 (e.g., 5%) is seen to be significant. The codes for significance are shown in the line below the “Coefficients”. The number of observations minus the number of all factors, including the intercept, are the degrees of freedom with the residual standard error showing the average size of the errors between the values predicted and the original values. Next, there are also both r-squared values. As the adjusted R-squared value also considers the number of independent variables used in the regression, it delivers a more accurate result regarding how much of the variation of the hedge fund indices is explained by the risk factors. Last is the F-statistic, which tests if the created model of estimates is significant as a whole, different from the t values, which only test the significance of the individual risk factors.

Not all of these results are evaluated in the linear regression analysis of all hedge fund indices. In the following sections, only the values of the R-squared, the estimates of the betas, and the intercept are being interpreted.

The whole list of the regression results, including the t-stats, are specified in the Appendix in table 3.

5.1 Regression of Hedge Fund Indices

Overall, the value of the R-squared is relatively high for almost all hedge fund strategies in the 165 months covered. The average value is 0.68, meaning that the nine risk factors can explain more than two-thirds of the variance in the hedge fund indices returns.

The highest R-squared values, all 0.80 and above, are found in the strategies Multi Strategy (0.80), the EHI Main Index (0.82), Event Driven (0.82) and Long/Short Equities (0.85). They all have very significant exposures in the factors BAA10Y and the SPX factor, with Event Driven being additionally exposed to the GSG.

The lowest R-squared values can be found with the strategies Arbitrage (0.67), Macro (0.58) and CTA/Managed Futures (0.04). The CTA/Managed Futures index is probably in parts constructed with hedge funds following a market-neutral strategy. So, the low R-squared value demonstrates that this strategy is indeed, on average, market neutral. However, as described in section 4.1.2, 'CTA/Managed Futures', it needs to be mentioned that the strategy is often systematic and quantitatively driven via mathematical models and computational power. This could mean that a "simple" multiple linear regression model like this is not able decompose the returns with the given nine risk factors in an acceptable way. Managers who run funds with Arbitrage strategies often claim that these are almost market-neutral. As Arbitrage shows an R-squared value of almost 0.70, it is clear that this claim is, on average, not valid, and a significant amount of the strategy variation can be explained by the risk factors.

5.2 Beta Exposures

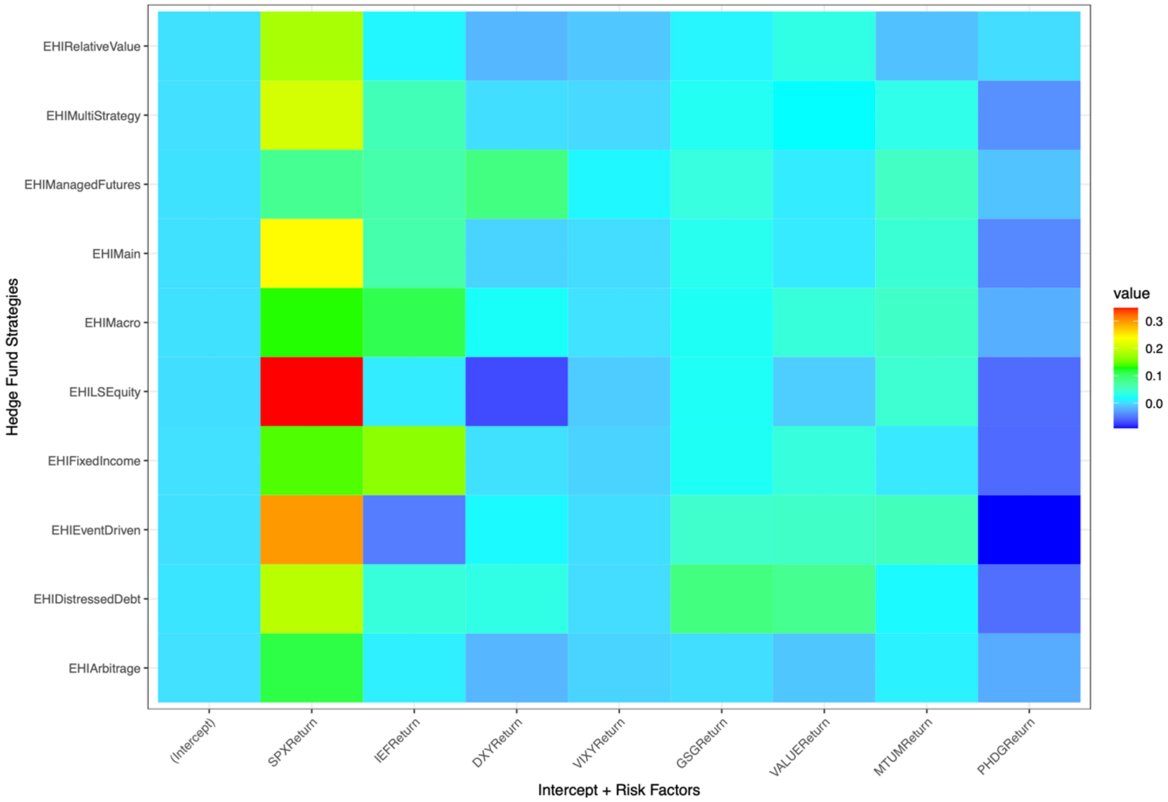


Figure 5: Heatmap of hedge fund index beta exposures without BAA10Y

Figure 5 displays a heatmap for the beta exposures / estimates in the linear regression using the coefficient estimates. For reasons of clarity, the exposures to BAA10Y, which is the yield spread between Moody's Seasoned Baa Corporate Bond Yield and the Yield on 10-Year Treasury Constant Maturity, have been omitted from the heatmap. A complete heatmap with all factors' exposures can be found in the Appendix as figure 16. The estimates of the spread are many times higher or lower compared to those of the other factors. When showing them together with the others, it is visually difficult to observe differences between values that are distributed close to zero.

The risk factor BAA10Y is very significant or highly significant in eight of ten strategies, all eight with values below -1. The betas are below -2 for the strategies Long/Short Equities (-2.45), Event Driven (-3.19), Fixed Income (-3.30) and Distressed Debt (-3.65). There are two ways to interpret these values; the first one was delivered by Cherian, Kon and Li in their 2020 regression analysis of the period between January 2000 and March 2020. They assess that because bonds and spreads are based on yields and not prices, their values in the model should be higher than those of other factors. In this study, the bond is an ETF and, therefore, does not have yields, but the monthly yield spread from Moody's is the same. For the strategies Fixed Income and Distressed Debt, they also generate beta values close to -3, which are both the most extreme, like in this study. Similar to their study, although not as high, the data in this one shows a significant positive above-average exposure of Fixed Income and Macro to the bond (IEF) risk factor. They conclude that "due to the inverse relationship between bond prices and yields, a more negative beta for Bond and Credit implies increased market exposure."[61] This could be a part of the explanation for the exposures. However, it is probably not very realistic that the market exposures of these strategies are as high as shown in this work. Due to rising interest rates, they may have changed. When looking at the spread's development since 2000, it is evident that there has not been much change, at least in the spread itself. Another reason for these high values may lie in the linear regression itself. It can be seen in figure 3 in section 4.2.1, 'Risk Factor Performance', that the spread has always had returns very close to zero and therefore, not visibly moving in almost 14 years. There is a chance that the OLS method figuratively realizes that the spread is relevant for decomposing the index returns but needs to leverage it high to make a difference in the model. This assumption is hard to prove and, therefore, not more than a guess, but it may be the reason for such extreme values together with the first interpretation. However, it will later create further problems in calculating the hedge fund clones in section 6.2, 'Out-Of-Sample Clones'.

Moving on to the other risk exposures, most are relatively low, as shown in figure 5. However, some hedge fund strategies sometimes have higher exposures to certain risk factors. The S&P500 (SPX) is especially important, as 9 out of 10 strategies show a highly significant beta exposure towards it. This is somewhat concerning as it implies that almost all strategies are exposed above-average to equities compared to the other risk factors. It means, as described with the results of the EHI Main Index that a change in the value of the SPX can significantly influence the value of the hedge fund indices. Investors wanting to build a diversified portfolio should take this into account. Only the CTA/Managed Futures strategy is neither exposed to the spread and the S&P500 nor to any other risk factor. This is in line with the interpretation of the strategy from the previous section. As

[61] Cf. Cherian et al. (2020), p. 16.

mentioned before, the Macro and the Fixed Income strategies are both strongly and highly exposed to the IEF, which is unsurprising. Macro uses economic trends and policy changes, which include hedging on bonds. Fixed Income, as the name says, is investing in the bond market, often with leverage. Interestingly, a few strategies also have exposures to commodities, with Event Driven and Distressed Debt being strongly and highly exposed.

5.3 Alpha Exposures

It is also imperative to analyze and interpret the exposure to the intercept, which represents, as described earlier, the hedge fund manager-specific alpha. The following column chart shows the average alpha for all ten strategies:

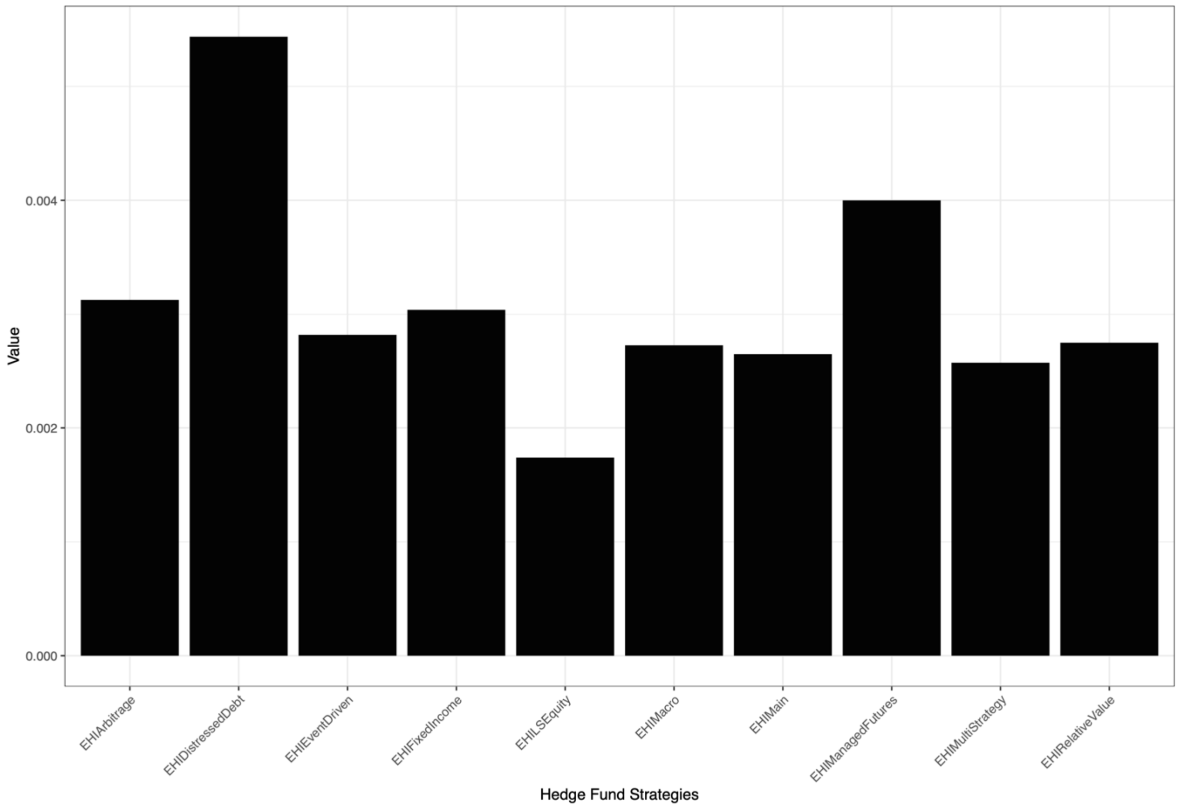


Figure 6: Average hedge fund index alphas

Since all the alphas are positive, managers can still generate returns attributable to their individual skills. Except for the strategies Distressed Debt (0.54), Managed Futures (0.40) and Long/Short Equities (0.17), the values are very similar, with an average of about 0.31%. Cherian, J.A., et al. found in their 2020 study that for 20 years (2000 - 2019), the average alpha was 0.47% in North America, which is noticeably higher than in this work.

Also, the alpha of the Distressed Debt strategy is much higher than their study[62], possibly due to good investment opportunities during and after the COVID-19 pandemic. Figure 1 in section 4.1.10, 'Hedge Fund Index Performance', shows that managers following this strategy have significantly outperformed all other strategies, especially during the bull market after the rebound from the COVID-19 crash.

The difference between the average alpha in their study and this thesis opens up again the question of whether, as in the introduction mentioned, the alpha value and therefore the returns attributable to the manager-specific skills are declining. They showed that alphas have declined by comparing the values of two different time spans, pre- and post-GFC, as well as with rolling windows over the entire time span. It sharply declined during the GFC, followed by a strong rebound. However, from 2014 until the end of their studies in March 2020, alphas have declined systematically across all strategies in North America.[63]

Figure 7 shows the development of the alphas with 60-month rolling windows across the entire period of this study, therefore delivering the first result for December 2014. All graphs shown below can be found in the appendix as figure 17.

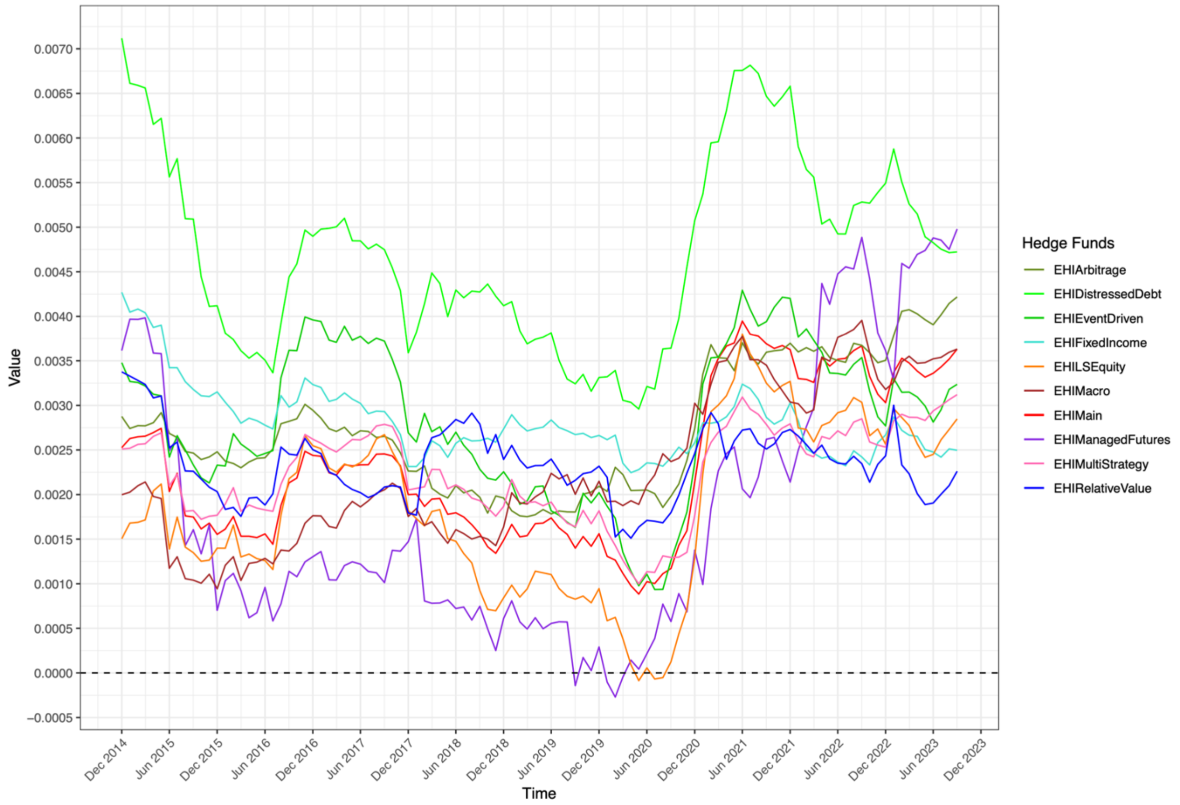


Figure 7: Rolling-window alphas for all hedge fund indices

[62] Cf. Cherian et al. (2020), p. 17.
[63] Cf. Cherian et al. (2020), p. 21.

The alphas of almost all strategies seem to follow a very similar trajectory. Generally, the trend is that they all decline from the start until the beginning of 2020. The assessment of a “steady and systematic decline” in manager-specific alphas since 2014[64] can be supported. Of course, the individual developments in the two studies are different, primarily due to different risk factors used in the linear regression, but the overall trend is the same. All alphas stay positive except for the strategies CTA/Managed Futures and Long/Short Equities. The development of the CTA/Managed Futures alpha is fraught with great uncertainty, as shown in the analysis of the linear regression, so it does not necessarily have to be negative in reality.

It is clearly visible that there is a significant turnaround of the alphas in all strategies in the middle of 2020. In the short period between June 2020 and June 2021, except for Fixed Income, all strategies at least doubled their alphas, with some more than tripling it. These lifted values mostly stay at their high levels, sometimes adding limited gains until the end of September 2023. As the COVID-19 pandemic hit at the beginning of 2020, the markets declined sharply. This can be seen in figure 1 in section 4.1.10, ‘Hedge Fund Index Performance’. Without an exemption, all alphas decline in the same period, some very strong, others only limited. During the rebound from the collapse and the subsequent strong bull market in 2021, hedge fund alphas are increasing significantly, as described above. In their paper from 2004, Fung and Hsieh developed the hypothesis that alphas increase during bull markets but decrease when markets deliver negative returns.[65] The results from this analysis support their assumption. As the opportunities for creating alpha rise again after many years, the performance fee for the managers may become more important again, especially considering persistent high inflows into hedge funds, as mentioned in the introduction.

This analysis concludes that, indeed, large parts of the hedge fund indices returns can be attributed to systematic risk factors. However, only these parts of the returns can be replicated in a clone, as the manager-specific alpha is not investable. Due to significant increases in the alphas in the last three years, it may have become more challenging to create clones that are sufficient alternatives to their counterparts again.

The following section will deal with efforts on predicting hedge fund index returns by first calculating out-of-sample expected returns and afterwards building hedge fund clones.

[64] Cf. Cherian et al. (2020), p. 21.
[65] Cf. Fung & Hsieh (2004), p. 23.

6 Replication Analysis

As summarized above, the results from section 5, ‘Regression Analysis’, deliver valuable insights on how the risk profiles of the hedge fund strategies are structured, how much of the returns can be attributed to either the managers-specific skills or systematic risk factors and by what extend the former one has changed over the period covered by the analysis. Especially considering developments in the markets in recent years, this knowledge is part of the interpretations for this section's results.

The model of linear factor-based hedge fund replication is described extensively in the methods section, covering the theory of the method itself and the procedure behind the use of the rolling-window method. In short, factor-based hedge fund replication aims to predict a hedge fund's monthly returns, in this study's case, a hedge fund index, by setting up a clone with common liquid risk factors and their corresponding sensitives, which are determined by linear regression.

Recent studies and those from many years ago have shown that, as described in section 2.3, ‘Hedge Fund Replication’, linear factor-based replication is a promising field to investigate. While clones created in many tests are indeed more liquid, transparent and probably cheaper than their counterparts, they cannot outperform them on average over time spans longer than a few years. For instance, clones performed better in the three years before the COVID-19 pandemic than hedge fund indices across the board, with some even delivering positive returns when the actual hedge fund indices had negative returns.[66] This is a clear proof that clones can outperform hedge funds using common systematic risk factors. It also shows how the decline in alpha, as shown in the regression analysis, contributes to increasingly successful replications.

With the collapse of the markets at the start of the COVID-19 pandemic, a very strong bull market in the year after, much higher interest rates and as found out a sudden and sustainable increase of alpha in almost all strategies, it is time to reevaluate how well linear factor-based replication is handling the task of creating hedge fund clones. Clones can only become a viable alternative to hedge funds if they can deliver comparable performance results over an extended period.

6.1 Out-Of-Sample Replications

Before building hedge fund clones, out-of-sample expected returns are calculated in R. This is done using rolling windows with the factor-based replication model. As described in section 3.2, ‘Factor-Based Replication’, other than the fixed-weight method, rolling windows can adapt to changes in the hedge fund indices exposures. This may be useful as figures 1 and 6 indicate that the hedge fund indices may be subject to significant exposure movements. The expected returns are then compared to the realized returns of

[66] Cf. Cherian et al. (2020), p. 29 f.

the hedge fund indices. First, the overall correlation is evaluated. Next, the differences between the performances are examined, and a paired t-test is conducted to see if there is any significance in the difference between both. The reason for calculating predicted returns of the hedge fund indices is to find out, in principle, if the regression coefficients calculated based on returns from the past can predict returns in the future at least roughly.

The full overview of the results for all investment strategies can be found in the appendix in table 4.

Most expected returns are highly correlated with the realized returns of the hedge fund indices. Nine out of ten show a correlation above 0.70, seven above 0.80. The expected returns of Long/Short Equities are even 0.90, which is undoubtedly very high. This means that “the high correlations between predicted returns from the rolling window regression model and realized returns demonstrate that there are obvious patterns of time-varying beta exposure to market risk factors for many hedge fund indices.”[67] Only the CTA/Managed Futures strategy shows a correlation below 0.20 with just 0.16. This is not surprising as results in the previous section have shown that the nine risk factors in the regression model cannot adequately capture the strategy's exposures.

To examine if the predicted returns and the realized returns are significantly different from each other, a paired t-test is conducted in R. The null hypothesis is that the mean difference between both is zero and the alternative hypothesis is that the mean difference is not equal to zero. Figure 8 shows exemplary the results of the paired t-test of the Eurekahedge Hedge Fund Index (EHI Main) from RStudio:

```
> t_test_result <- t.test(OutSampleData$EHIMainReturn, OutSampleData$EHIMainReplicationReturn, paired = TRUE)
> t_test_result

	Paired t-test

data:  OutSampleData$EHIMainReturn and OutSampleData$EHIMainReplicationReturn
t = 0.0081428, df = 104, p-value = 0.9935
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -0.001428002  0.001439777
sample estimates:
mean difference
    5.8879e-06
```

Figure 8: Paired t-test between EHIMain & EHIMainReplication

The overview displays a range of different information. The "data" line defines which columns from the data frames are subject to the test. The t-value, one of the most important values in the test, displays the mean difference between the paired samples relative to the variability of the differences. In the case of this test, the value is very low, indicating a small difference. With 105 months tested, the number of degrees of freedom (df) is 104. The high p-value of 0.9935 implies that both expected and realized returns are

[67] Cf. Cherian et al. (2020), p. 26.

not significantly different. The test is conducted at a confidence interval of 95%, which sets the mean difference between -0.001428002 and 0.001439777. The mean difference is very low (0.0000058879) and, therefore, in the confidence level range, so there is no indicator for rejecting the null hypothesis.

All other strategies show similar results, meaning that none of the expected returns is significantly different than the respective realized return of the hedge fund index. Every t-test has a p-value above the threshold of 0.05 (e.g., 5%) and a mean difference that lies in the boundaries set by the confidence interval.

In the last step, the differences between the average returns are examined visually and with numbers. The expected returns of the EHI Main, Multi Strategy and Relative Value calculated to two decimal places are similar to the realized returns. Only three strategies, Arbitrage, CTA/Managed Futures, and Macro, have lower average returns than their counterparts. Overall, none of the expected average returns is more different than |0.08| than their respective realized return.

The following two plots show the benchmarked expected and realized returns of the Long/Short Equities strategy, which has the highest correlation (0.90), and the CTA/Managed Futures strategy, which has the lowest correlation (0.16). The blue line represents the realized returns and the red line represents the expected returns from the replication model. All plots from the other eight strategies can be found in the appendix in figure 18.

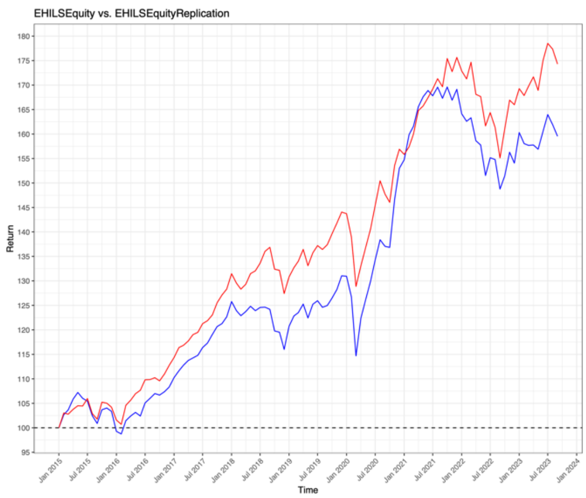


Figure 10: Performance of EHI LSEquity & EHI LSEquityReplication

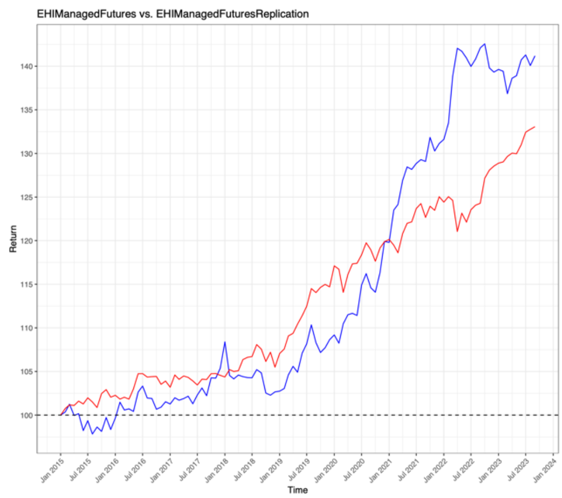


Figure 9: Performance of EHI Managed-Futures & EHI ManagedFuturesReplication

It is visible that the expected returns of the Long/Short Equities strategy have a very similar development compared to the realized returns. This is important for the following section: the predicted returns, which again include the intercept (alpha), can follow the realized returns precisely, including from June 2020 to June 2021 in that very strong bull market after the rebound from the COVID-19 crash. As described before, the expected returns of the CTA/Managed Futures strategy do not follow the expected returns similarly, with additionally underperforming them by a large margin.

In summary, besides the CTA/Managed Futures strategy, the expected returns of all investment strategies are very closely related to the realized returns of their respective hedge fund indices. This implies that using common liquid risk factors, the returns of hedge fund indices can be predicted on a qualitatively high level.

6.2 Out-Of-Sample Clones

Expanding on these favorable results, in the final part of this work, out-of-sample hedge fund clones are being constructed. The returns of the clones are calculated in the same way as the expected returns in the previous section, except for a few differences. This is done to answer the third research question from the introduction, asking whether it is possible to use the factor-based replication method to generate out-sample clones that show similar performance compared to their respective hedge fund index over an extended period. The “extended period” means the entire period of the thesis covering almost 14 years. A clone being able to outperform visibly or at least show an average performance similar to its respective hedge fund index during that period could be an attractive investment.

Section 3.2.1, ‘Hedge Fund Clones’, already covers the method of conducting factor-based linear hedge fund replication using rolling windows. In short, the differences between simply calculating expected returns and building clones are as follows: For the linear regression, the intercept representing the manager-specific alpha is omitted. This is necessary because investing in the manager's skills is not possible. Additionally, the calculated sensitives for each month are set to sum to one to ensure the same amount of money is invested into the clone as in the hedge fund index. This is done by putting the difference between one and the sum of all betas into the risk-free asset. As mentioned in the ‘Hedge Fund Clones’ section, this has an unintentionally high influence on the returns of the clone. Section 5.2, ‘Beta Exposures’ shows that the linear regression model calculates for the hedge fund indices a, compared to other risk factors, either very high or very low exposure to the risk factor spread (BAA10Y). Possible reasons for these results have already been discussed. It leads to a very high leverage in the risk-free asset. As described, the procedure is based on a study by Soerensen and Hansen from 2020. They do not provide any numbers regarding the exposure of their chosen investment strategies to different risk factors, including the calculated leverage in the risk-free asset. It can only be assumed that the differences between 1 and the sum of all betas in their study are probably a bit lower than in this work. Figure 1 shows that the risk-free asset has, most of the time, returns close to zero. However, in the last months of the chosen period, beginning in the middle of 2022, the returns of the risk-free asset have multiplied due to the sudden and strong interest rate hikes. As the leverage representing the calculated difference between 1 and the sum of all betas continues to be high, the risk-free asset becomes a very significant influence on the clones returns.

Figures 11 and 12 compare the hedge fund index returns and the clone returns of the Fixed Income strategy with and without the spread (BAA10Y).

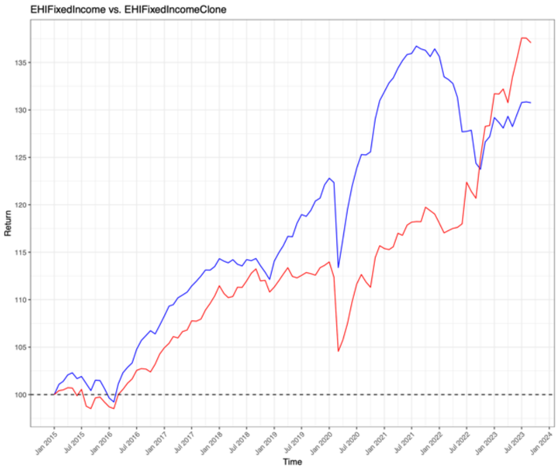


Figure 11: Performance of EHI FixedIncome vs. EHI FixedIncomeReplication with BAA10Y

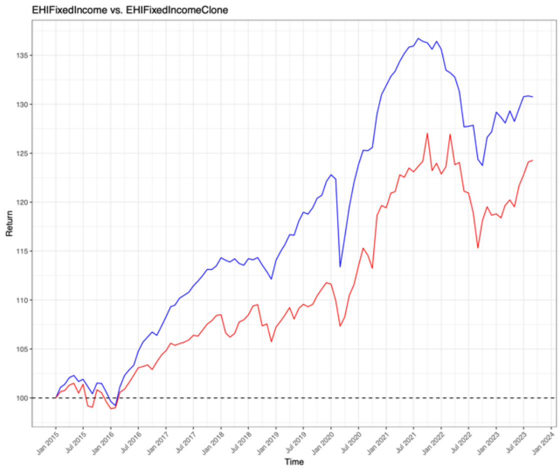


Figure 12: Performance of EHI FixedIncome vs. EHI FixedIncomeReplication without BAA10Y

Figure 11 shows the clone (red) with the spread and Figure 12 without the spread. Therefore, the clone in the left plot has significantly higher leverage in the risk-free asset. The difference between both calculations is visible as the clone in figure 11 performs worse during the COVID-19 crash but shows a sudden rise when the returns of the risk-free asset increase from the middle of 2022. The clone in figure 12, on the other hand, obviously underperforms the hedge fund index but can follow the realized returns similarly. The behavior displayed in figure 11 is not wanted; therefore, the decision has been made to omit the spread when calculating the sensitives for the clone returns. Separate tests have shown that the spread itself only modestly influences the clone returns.

After implementing these changes, the hedge fund index returns and the clone returns are adjusted to the risk-free rate. This is done by subtracting the risk-free rate from the returns.

Finally, as described in the 'Method' section, the clone returns are normalized to compare both fairly. For instance, if the clone has a lower mean return but also a lower volatility compared to the hedge fund index, after the renormalization, it would have the same standard deviation as the hedge fund index but a higher average return than before. It must be pointed out again that this changes the leverage of the clone.

One reason for building hedge fund clones is to generate lower fees than the actual hedge fund. The systematic risk factors chosen for this study, of course, are with gross-of-fee returns and, therefore, similar to Cherian, Kon and Li in their 2020 study, the fee structure is assumed to be low, having "a negligible effect on the overall results."[68] Because of this,

[68] Cf. Cherian et al. (2020), p. 29.

the possible fees behind each clone will not be subject in the interpretation of the results.

In this part of the analysis, three generated clones will be compared to their respective hedge fund indices. Besides, in section 6.2, this comparison is more extensive as it includes several performance tests. First, the correlations and the difference between the returns are examined again. Because the returns are normalized, there is no longer a need to check the standard deviation. Next, the Sharpe Ratios will be compared, followed by several statistical tests. These are as extensively described in the 'Methods' section: The RMSE, the tracking error and Theil's inequality coefficient. All four aim to measure how well the factor-based replication method handles the prediction of the clone returns regarding the performance and how close the clone returns are to the actual hedge fund index returns.

A complete overview of the results for all investment strategies can be found in the appendix in table 5. All other generated plots with the performance of the clones and the performance of the corresponding hedge fund indices can also be found in the appendix as figure 19.

First is the Eurekahedge Hedge Fund Index (EHI Main):

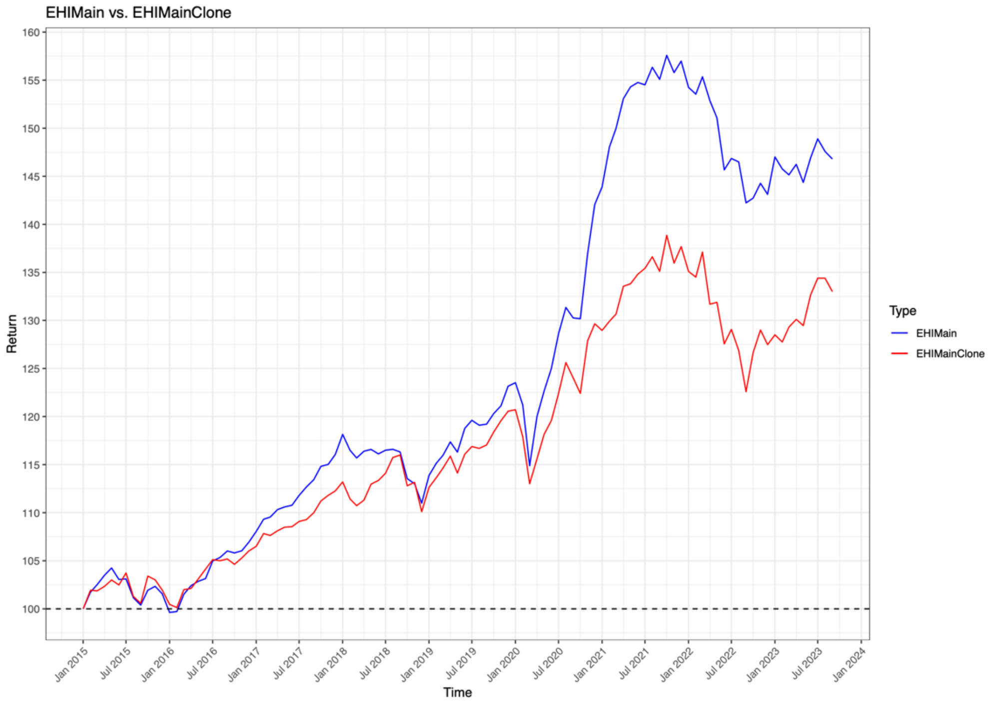


Figure 13: Performance of EHI Main vs. EHI MainClone

The clone is performing significantly worse than the actual hedge fund index. While both are benchmarked to start at 100, the hedge fund index ends at about 146.8, while the clone only reaches 133. Therefore, the Sharpe Ratio is not good, with just 0.18 compared to 0.25 for the index. Looking at how both graphs develop over time, the clone returns can follow the hedge fund index returns reasonably well until about the middle of 2020. This is when the markets have rebounded from the COVID-19 crash and are transferring into the subsequent bull market. The overall correlation between both returns is high, with 0.84, which is only 0.02 less than the correlation for the same strategy in the previous section. The problem is that the clone misses most parts of these significant increases in value in the year after the rebound. After this time, it continues to follow the index in a similar pattern. The fact that it cannot replicate this sudden increase in value could be attributed to various reasons. The first one may be that the windows in the rolling-window method used in the regression are too long. This would lead to a lack of reactivity, meaning the clone is unable to adapt quickly enough to those changes. However, this is rather unlikely due to two reasons. The first one is that separate tests, for instance, with 36-month windows, have shown no significant improvement to the clone's performance, and second, the clone is very much able to calculate the COVID-19 crash and the subsequent rebound in similar to the hedge fund index. The likelier reason for the failure to replicate the bull market is the sudden increase of alpha in the exposure of the hedge fund indices. As extensively described in section 5.3 ‘Alpha Exposures’, all hedge fund strategies experienced a substantial increase in their exposure to the manager-specific alpha after the rebound in the middle of 2020. Apparently, the managers could use the bull market to their advantage and make investment decisions that can only be attributed to their skill and not to the systematic risk factors used in the regression model. This proves the assumption that alphas rise during bull markets and decrease in bear markets. The other numbers are instead encouraging when looking at how close and consistent the forecasted values are compared to the hedge fund index returns. The RMSE, tracking error and Theil’s inequality coefficients are not bad, at 0.0085, 0.85% and 0.2754, respectively.

Second is the Eurekahedge Long/Short Equities strategy index (EHI LSEquity):

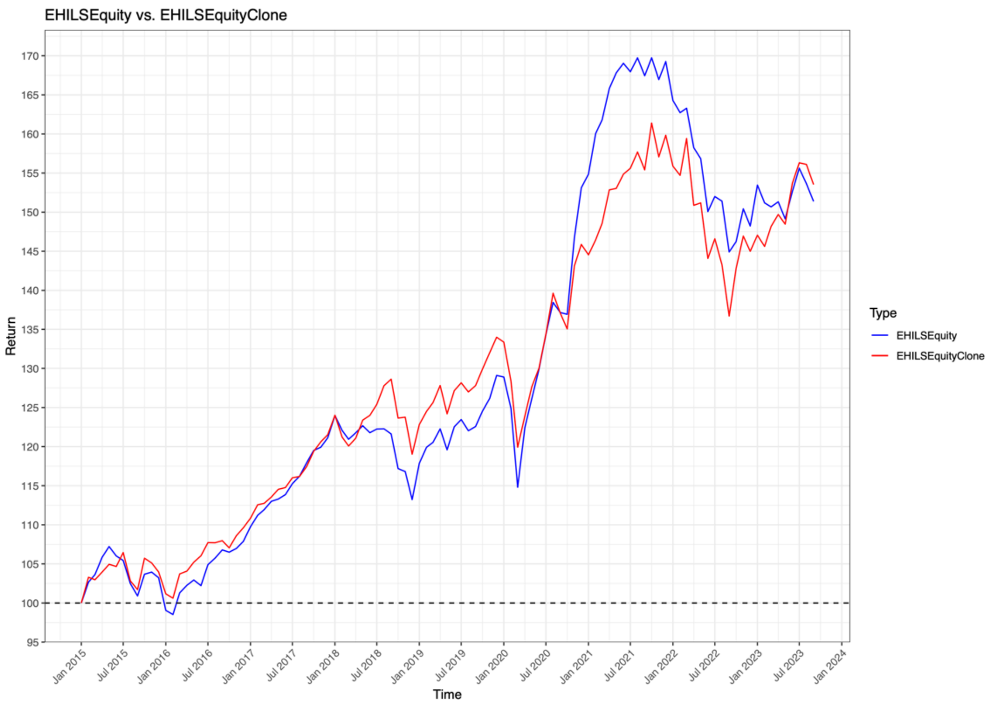


Figure 14: Performance of EHI LSEquity vs. EHI LSEquityClone

The clone for this strategy is the best-performing one among all strategies. While the mean return is just below the mean return of the hedge fund index, at the end of the period, the clone stands at 153.5 compared to 151.4. The correlation is, unsurprisingly, very high, with 0.87, as seen in the plot where the clone closely follows the index. Additionally, due to the almost identical returns and the same standard deviation, the Sharpe ratios are 0.19, which is the same. It must be noted that the clone is noticeably outperforming the hedge fund index until the beginning of 2020. Like the EHI Main index, the clone cannot fully grasp the bull market following the COVID-19 rebound. However, it is still much better at replicating the returns in this time frame than all other clones. This superiority above the other funds may be due to the strategy itself. As shown in figure 1, the S&P500 (SPX) significantly outperforms all hedge fund strategies. Also, in figure 6, the heatmap indicates a strong exposure of the strategy towards the SPX risk factor. This combination could be why the model successfully replicates the hedge fund index returns. Again, the numbers are acceptable when looking at how close and consistent the forecasted values are compared to the hedge fund index values. The RMSE, tracking error and Theil's inequality coefficients are all good, at 0.0110, 1.10% and 0.2459 respectively.

Last is the Eurekahedge Distressed Debt Strategy index (EHI DistressedDebt):

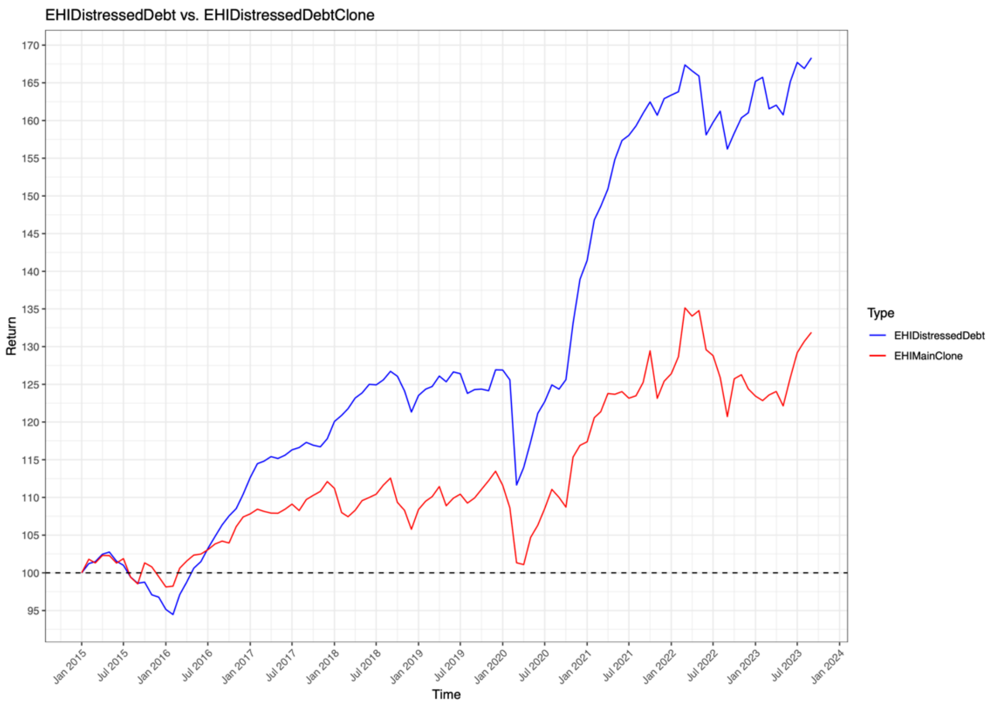


Figure 15: Performance of EHI DistressedDebt vs. EHI DistressedDebtClone

The clone for this strategy is the worst performing among all strategies. It is visible that the clone is underperforming the hedge fund index most of the time, especially during the rise after the COVID-19 crash. The clone's mean return is on a monthly average 0.24 percentage points lower than the hedge fund index returns, which leads to the clone ending at about 131.9 compared to the index at 168.3. The Sharpe ratio of the clone is also only about half of the indexes, with 0.14 compared to 0.27. This means that the hedge fund index can archive more excess returns but at the same time takes the same amount of risk as the clone. Again, part of the bad performance of the clone may come from the high exposure to alpha, as shown in figure 6 in section 5.3, ‘Alpha Exposures’. It also doubles from the middle of 2020, which could be the reason for the model's inability to capture the bull market in 2020/21. Since the clone is underperforming the index most of the time, not just after 2020, part of the bad performance could be that the regression model lacks the right risk factors. As with the two clones before, the measurements of how well the factor-based model handles the prediction of the clone returns regarding the performance and closeness of the clone compared to the actual hedge fund index returns, the numbers are still acceptable. The RMSE is slightly worse, with a value of 0.0143, showing that the predictive model is less accurate than in the other two strategies. Both the tracking error and Theil’s inequality coefficient are good, with values of 1.42% and 0.3738, respectively.

Overall, except for the Long/Short equities strategy, none of the generated clones has similar returns or can outperform its respective hedge fund index. The monthly returns of all clones are, on average, 0.12 percentage points lower than those of the hedge fund indices. This means a lower performance of about 1.5 percentage points annually, which is in line with the findings from Hartley, J. in 2019. He writes that “LAMFs underperform hedge funds on average by 1% to 2% per year on a net-of-fee basis when controlling for standard risk factors.”[69] As mentioned, the returns calculated in this study are not net-of-fees. However, it would probably still be somewhere in that range if they were. Regardless of the poor performance, the correlations are high, with seven out of ten strategies correlated at 0.70 and above. Four of them are even at 0.80 and above. This is, on average, just 0.06 less than the correlations of the out-of-sample replicated returns in section 6.1 ‘Out-Of-Sample Replications’. Plots for all investment strategies show a widening difference between the returns after the COVID-19 crash, the same way as all experience a sudden increase in exposure to alpha at this time. Despite the overall acceptable to good results on the RMSE, tracking error and Theil’s inequality coefficient tests, nine out of ten clones come up short compared to their counterparts at the end of the period.

[69] Cf. Hartley (2019), p. 2.

7 Conclusion

This thesis investigated the developments of linear factor-based hedge fund replication, which presents both challenges and opportunities, particularly highlighted during recent periods of significant market and economic change. Although the already extensive literature shows clear results and developments, an updated study incorporating the latest events was warranted.

In a first step, an up-to-date literature review was presented, including the theory behind hedge funds, the attribution of hedge fund returns and hedge fund replication. Following the theory, the methods behind the linear regression and the rolling window factor-based replication models were described, including several performance measurements. Next, hedge fund indices and common liquid risk factors for the implementation of the linear regression were identified, examined and, if necessary, adjusted. In the analysis of the linear regression, the first two research questions were answered. These include examining the structure of the hedge fund index factor exposures across different investment strategies and investigating the assumed decline of returns attributable to manager skill. Building upon this knowledge, out-of-sample replicated returns were calculated using the rolling window method. As the final part of this study, coming back to the title of the thesis itself, for all investment strategies, out-of-sample hedge fund index clones were constructed to evaluate whether the replication model can create clones with similar or superior returns compared to the actual hedge fund indices.

These investigations deliver several interesting and valuable results. One of the main findings is that hedge fund indices are highly exposed to systematic risk factors, as the linear regressions show R-squared values of up to 0.85 with an average of 0.68. These results are concerning for investors seeking to diversify their portfolios by investing in hedge funds. The high exposures to equities must be noted especially. The assumption of a decline of the manager-specific returns (alpha) until the beginning of 2020, shown in earlier studies, can be supported. However, investigations show that with the rebound from the COVID-19 crash and the subsequent strong bull market between June 2020 and June 2021, alphas for almost all investment strategies significantly rose. This leads to the assumption that hedge fund managers used their skills to deliver excess returns to their investors. As the alpha plays a more critical role in explaining hedge fund returns, constructing clones based solely on systematic risk factors is becoming more challenging. Although the replication model is, for most strategies, able to calculate returns that make the clones follow the index returns in a similar pattern, only the Long/Short Equities clone can show equal returns compared to the actual hedge fund index. All clones struggled to grasp the value increases in the bull market after the COVID-19 crash, which is very likely due to the sudden rise of alphas in the same period. The results of these investigations can be seen as significant. They reveal that while hedge funds are still highly exposed to systematic risk factors, the alphas show a substantial increase, leading to a worse performance of hedge fund clones.

However, a few changes can be made to improve the overall quality of the results and the performance of the hedge fund clones. First, other risk factors could be selected, and their number could change. This study's suit of risk factors tries to cover all asset classes and most parts of the markets. The selection for each investment strategy is rather ‘learning-by-doing’ as it is not possible to tell in advance which risk factors are significant. This work also struggled with the high exposures to the spread, which forced a significant leverage in the risk-free asset. A solution to this could be to constrain the betas, to sum up to one already during the calculations in the linear regression. It would mean there is no need to go short or long into the risk-free asset. However, using this method does not necessarily lead to an improved performance of the clones. Third, the linear regression method could be expanded by implementing more sophisticated tools like a Kalman filter or machine learning capabilities. They may be able to lower the difference between the expected and realized returns and improve the model's predictive capabilities, especially concerning the reaction time to significant market events.

Although results from this study show that a scenario of clones outperforming hedge funds while being cheaper, more transparent and more liquid is not within reach, the field of hedge fund replication should be further investigated. Further development of replication models and market changes may bring hedge fund replication closer to its intended goals.

References

Ackermann, Carl, Richard McAnally, and David Ravenscraft. "The performance of hedge funds: Risk, return, and incentives." The Journal of Finance. Vol. 54, No. 3, 1999: 833-874.

Asness, Clifford, Robert Krail and John Liew. "Do Hedge Funds Hedge?" The Journal of Portfolio Management, Vol. 28, No. 1, May 2001: 6–19.

Baker, H. Kent, Greg Filbeck. Hedge Funds: Structure, Strategies, and Performance. Oxford, England: Oxford University Press, 2017.

Bali, T. G., Stephen J. Brown and Mustafa Onur Caglayan "Do hedge funds' exposures to risk actors predict their future returns?" Journal of Financial Economics. Vol. 101, No. 1, 2011: 36–68.

BarkleyHedge. Backstop. July 2023. https://www.barclayhedge.com/solutions/assets-under-management/hedge-fund-assets-under-management/ (accessed January 07., 2024).

Blokhin Andriy. Investopedia. August 13., 2023. https://www.investopedia.com/ask/answers/060315/what-difference-between-linear-regression-and-multiple-regression.asp (accessed January 09., 2024).

Chen James. Investopedia. November 07., 2020. https://www.investopedia.com/terms/t/trackingerror.asp (accessed January 04., 2024).

Cherian, Joseph A., Christine Kon and Ziyun Li. "Replicas: Have Hedge Funds Re-Resurrected as Traditional Beta?" Social Science Research Network. August 2020.

Damodar, Gujarati, and Dawn Porter. Basic Econometrics. London, England: McGraw-Hill, 2009.

Eurekahedge Hedge Fund Index. Eurekahedge. December, 2023. https://www.eurekahedge.com/Indices/IndexView/Eurekahedge/473/Eurekahedge-Hedge-Fund-Index (accessed January 04., 2024).

Eurekahedge Equal Weighted Hedge Fund Index. Eurekahedge. https://www.eurekahedge.com/Indices/hedge-fund-index-methodology (accessed January 04., 2024).

Fama, Eugene and Kenneth French. "Common Risk Factors in the Returns of Stocks and Bonds." Journal of Financial Economics Vol. 33, 1993: 3-56.

Fung, W. and David Hsieh. "Hedge Fund Benchmarks: A Risk-Based Approach." Financial Analysts Journal, Vol. 60, No. 5, 2004: 65–80.

Getmansky, Mila, Peter Lee and Andrew Lo. "Hedge Funds: A Dynamic Industry in Transition." Annual Review of Financial Economics. No. 7, December 2015: 483-577.

Hasanhodzic, J. and Lo, A. "Can Hedge-Fund Returns be Replicated?: The Linear Case." Social Science Research Network. August 2006.

Hartley, Jonathan. “Liquid Alternative Mutual Funds versus Hedge Funds: Returns, Risk Factors and Diversification.” Journal of Alternative Investment Vol. 22, No. 1, 2019: 37–56.

Inglis Jack. The Hedge Fund Journal. “Managed Futures – Riding the wave.” Vol. 127, October 2017.

Kat, Harry, and Helder Palaro. “Who Needs Hedge Funds? A Copula-Based Approach to Hedge Fund Return Replication” Alternative Investment Research Centre. Working Paper No. 27. November 2005.

Liang, Bing. “On the Performance of Hedge Funds.” *Financial Analysts Journal* 55, No. 4, 1999: 72–85.

Metzger, Nicola, and Vijay Shenai. "Hedge Fund Performance during and after the Crisis: A Comparative Analysis of Strategies 2007–2017" *International Journal of Financial Studies.* Vol. 7, No. 1, 2019: 15.

SEC Office of Investor Education and Advocacy. “Investor Bulletin Hedge Funds.” SEC Pub. No. 139 (2/13) https://www.sec.gov/files/ib_hedgefunds.pdf (accessed January 01., 2024).

Sharpe, William. “The Sharpe Ratio.” The Journal of Portfolio Management. Vol. 21, Vol. 1, 1993: 49-58.

Soerensen, Michelle, and Hansen Emil. „Hedge Fund Replication Before, During and After the Financial Crisis.“ CBS Research Portal. January 2020.

Statistics How To. Independent Media Associates, Inc. https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/ (accessed January 04., 2024).

Ubide, Angel. “Demystifying Hedge Funds.” Finance and Development. Vol. 43, No. 2, June 2006.

Wadhwani Preeti. Global Markets Insights. December 2023. https://www.gminsights.com/industry-analysis/hedge-fund-market (accessed January 04., 2024).

Appendix

Hedge Fund Strategies	Statistic	(Intercept) in %	BAA10Y	SPX	IEF	DXY	VIXY	GSG	VALUE	MTUM	PHDG	Significance (R-Squared)
EHI Main Index	beta	0,26***	-1,91***	0,24***	0,06	0,00	0,00	0,03*	0,01	0,04	-0,04	0,82
	t-stat	5,07	-5,12	9,38	1,97	-0,12	0,16	2,59	0,34	1,88	-1,49	
EHI Arbitrage	beta	0,31***	-1,13***	0,11***	0,01	-0,02	0,00	0,00	-0,01	0,01	-0,02	0,67
	t-stat	8,24	-4,18	6,11	0,45	-0,89	-1,08	0,25	-0,60	0,74	-1,16	
EHI CTA/Managed Futures	beta	0,40***	0,53	0,08	0,07	0,09	0,01	0,04	0,01	0,05	-0,01	0,04
	t-stat	3,68	0.682	1,47	0,96	1,47	1,53	1,64	0,17	1,10	-0,20	
EHI Distressed Debt	beta	0,54***	-3,65***	0,19***	0,04	0,03	0,00	0,09***	0,08*	0,02	-0,05	0,68
	t-stat	6,24	-5,87	4,44	0,70	0,69	0,10	4,94	1,99	0,42	-1,16	
EHI Event Driven	beta	0,28***	-3,19***	0,30***	-0,05	0,02	0,00	0,05**	0,05	0,06	-0,09*	0,82
	t-stat	3,98	-6,30	8,55	-1,07	0,41	0,31	3,21	1,53	1,84	-2,41	
EHI Fixed Income	beta	0,30***	-3,30***	0,14***	0,16***	0,00	0,00	0,02*	0,04	0,01	-0,06*	0,76
	t-stat	6,50	-9,88	6,10	5,60	0,11	-0,92	2,53	1,71	0,37	-2,24	
EHI Long/Short Equities	beta	0,17*	-2,45***	0,35***	0,01	-0,07	-0,01	0,02	-0,01	0,04	-0,06	0,85
	t-stat	2,50	-4,92	10,19	0,21	-1,87	-1,23	1,61	-0,19	1,49	-1,49	
EHI Macro	beta	0,27***	-0,93*	0,13***	0,11**	0,02	0,00	0,02*	0,04	0,05*	-0,02	0,58
	t-stat	5,14	-2,45	5,03	3,31	0,73	0,79	2,18	1,60	2,18	-0,75	
EHI Multi Strategy	beta	0,26***	-1,85***	0,21***	0,06	0,00	0,00	0,03*	0,02	0,03	-0,04	0,80
	t-stat	5,27	-5,30	8,62	1,91	0,07	-0,24	2,56	0,82	1,47	-1,45	
EHI Relative Value	beta	0,27***	-1,18**	0,18***	0,01	-0,02	-0,01*	0,01	0,03	-0,01	0,00	0,74
	t-stat	5,54	-3,33	7,38	0,45	-0,67	-2,19	1,26	1,44	-0,60	0,05	

***, ** and * mean significance at the 0.1%, 1% and 5% levels respectively.

Table 3: Summary of the linear regressions

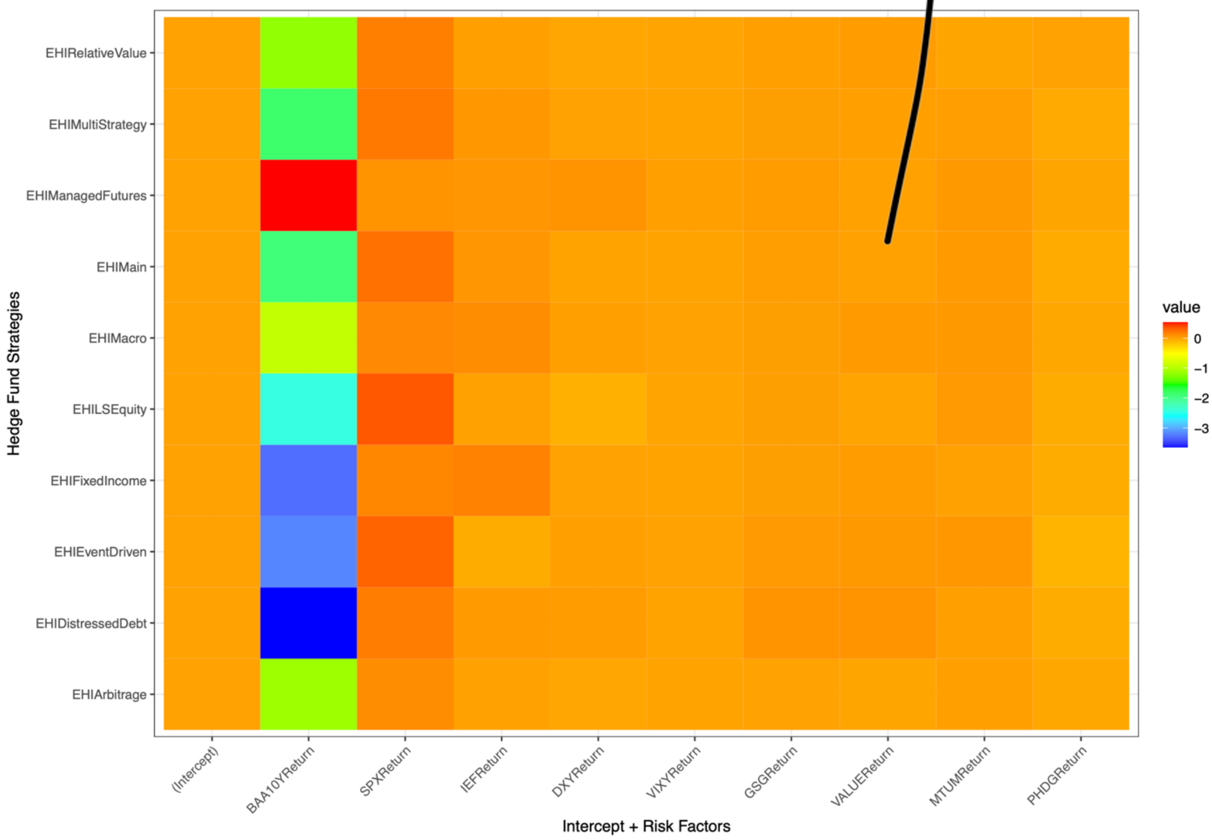
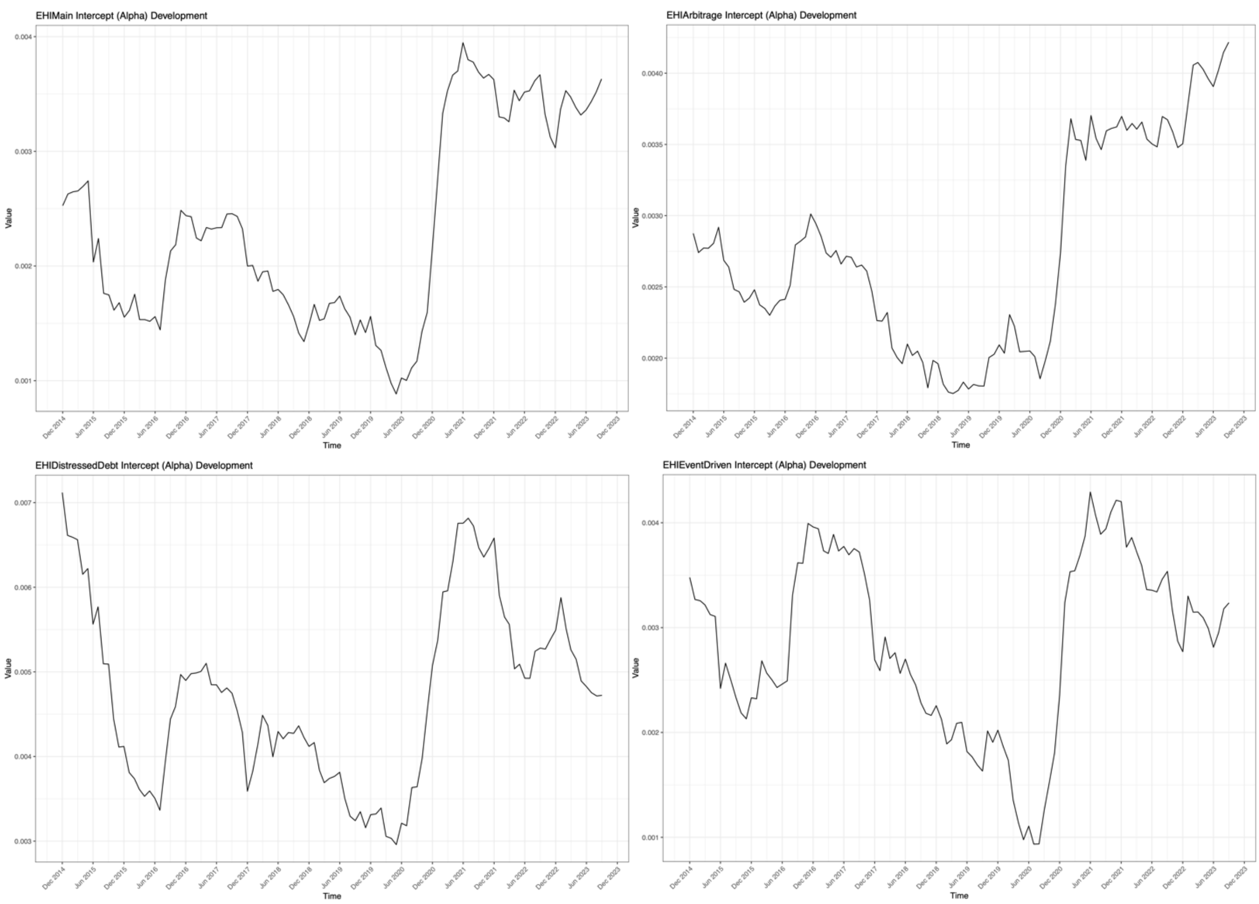


Figure 16: Heatmap of hedge fund index beta exposures with BAA10Y



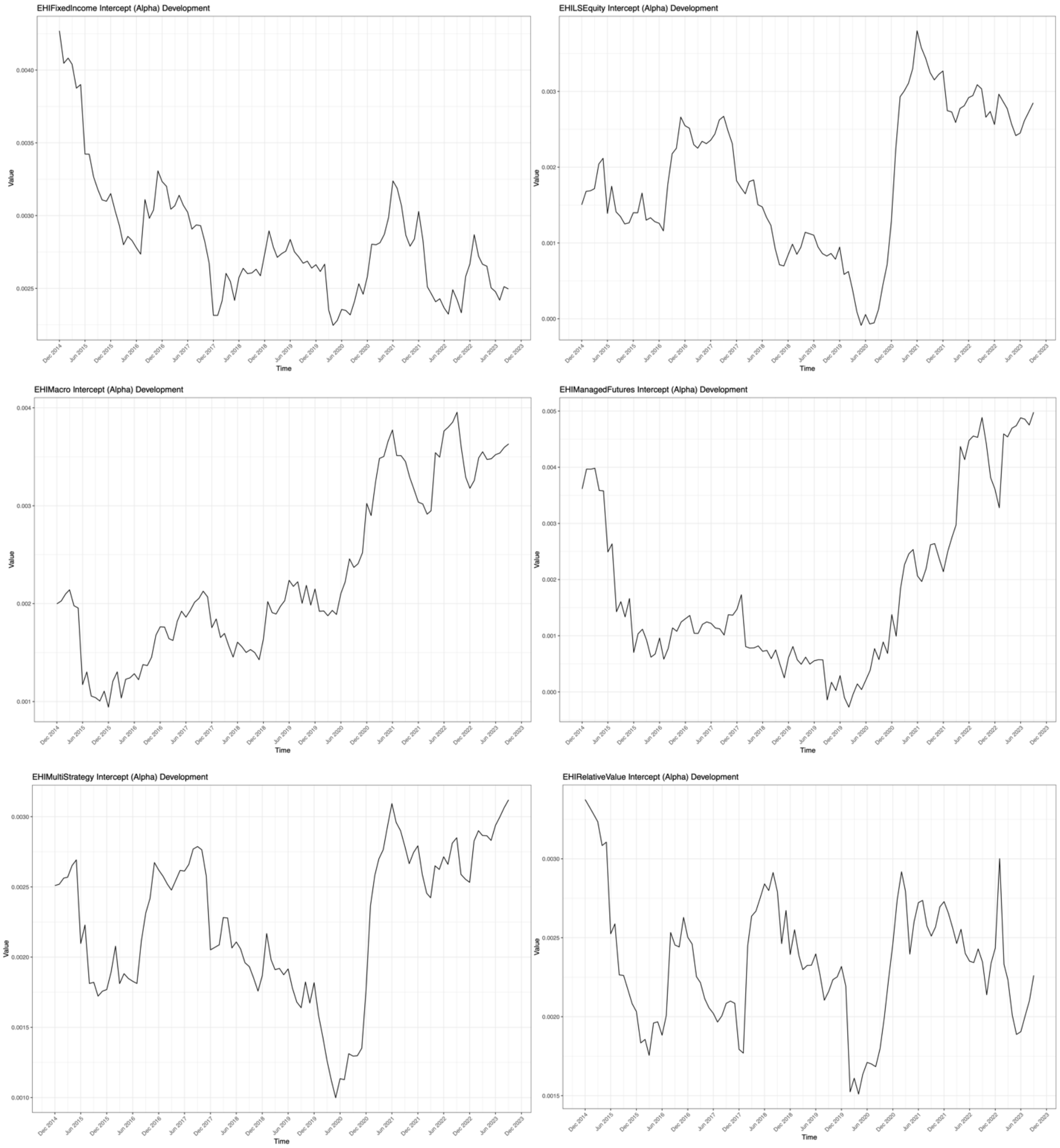
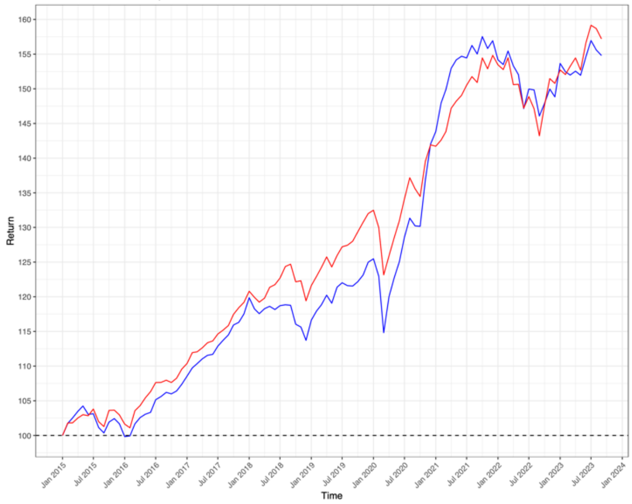


Figure 17: All alpha developments in separate plots

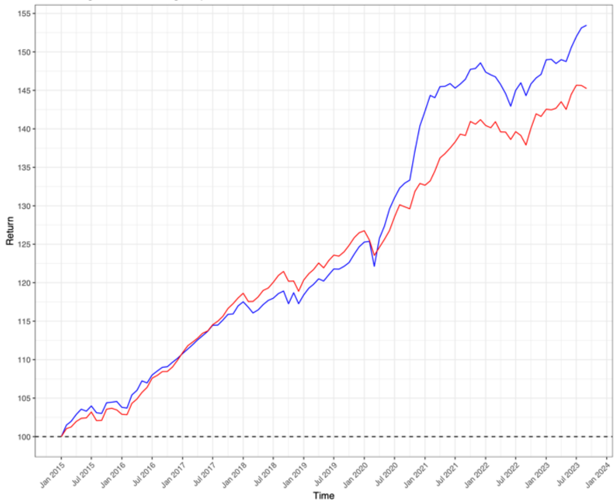
Hedge Fund Strategies	Realized Returns		Expected Returns				Correlation
	Average Monthly Realized Returns	Average Monthly SD	Average Monthly Expected Returns	Average Monthly SD	Average Difference with Realized Returns	T-Value of Paired T-Test	Correlation between RR and ER
EHI Main Index	0,44%	1,56%	0,44%	1,36%	0,00%	0,01	0,88
EHI Arbitrage	0,41%	0,81%	0,36%	0,66%	-0,06%	1,08	0,75
EHI CTA/Managed Futures	0,37%	1,30%	0,29%	0,89%	-0,08%	0,57	0,16
EHI Distressed Debt	0,56%	1,96%	0,59%	1,43%	0,03%	-0,22	0,81
EHI Event Driven	0,45%	2,07%	0,53%	1,71%	0,07%	-0,70	0,85
EHI Fixed Income	0,32%	1,25%	0,37%	0,92%	0,06%	-0,81	0,82
EHI Long/Short Equity	0,47%	2,25%	0,54%	2,03%	0,06%	-0,66	0,90
EHI Macro	0,36%	1,06%	0,32%	0,94%	-0,04%	0,52	0,73
EHI Multi Strategy	0,40%	1,37%	0,40%	1,19%	0,00%	-0,07	0,87
EHI Relative Value	0,39%	1,24%	0,38%	1,11%	0,00%	0,05	0,81

Table 4: Summary of out-of-sample expected returns vs. realized returns

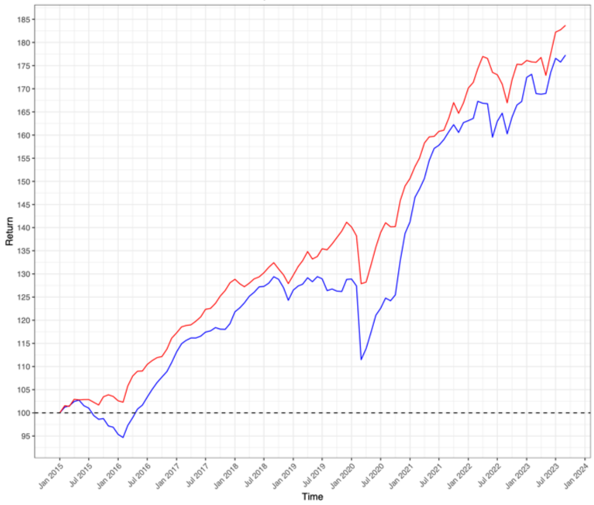
EHIMain vs. EHIMainReplication



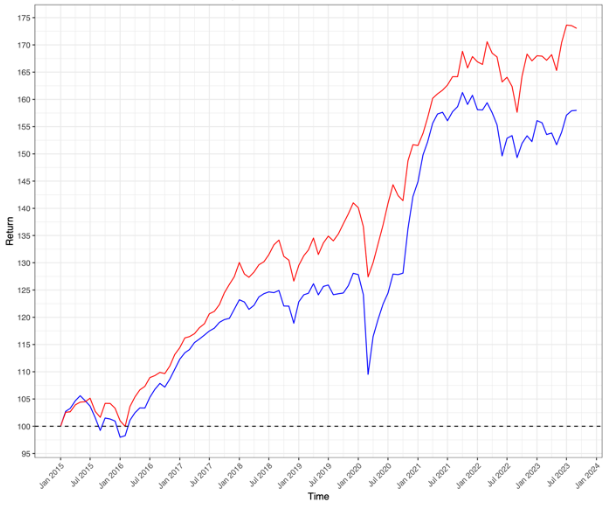
EHIArbitrage vs. EHIArbitrageReplication



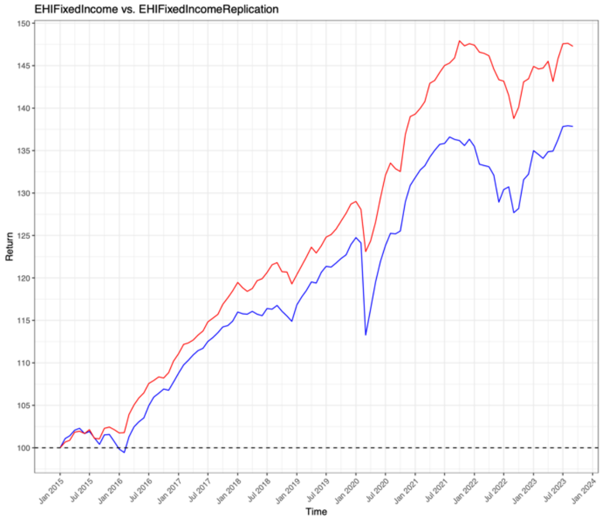
EHIDistressedDebt vs. EHIDistressedDebtReplication



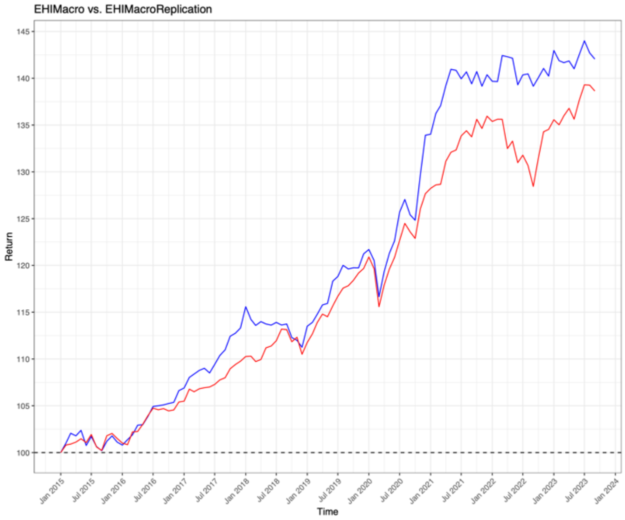
EHIEventDriven vs. EHIEventDrivenReplication

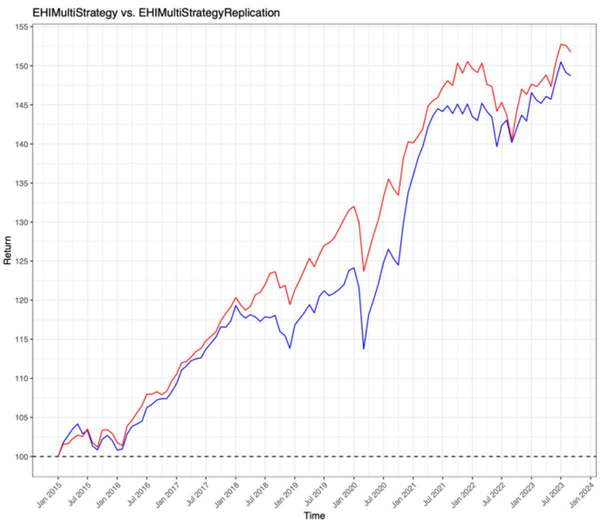


EHIFixedIncome vs. EHIFixedIncomeReplication



EHIMacro vs. EHIMacroReplication





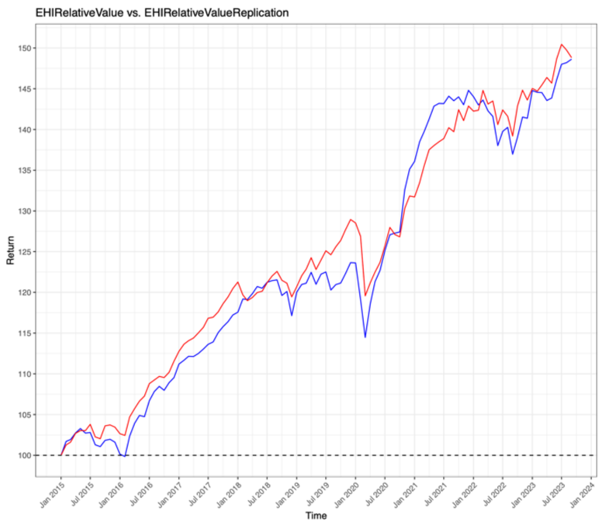


Figure 18: All expected returns vs. realized returns

HF Index Strategies	Realized Excess Returns			Clone Excess Returns					Performance		
	Average Monthly Realized Returns	Average Monthly SD	Sharpe Ratio	Average Monthly Clone Returns	Average Monthly SD	Sharpe Ratio	Average Difference with Realized Returns	Correlation between RR and CR	RMSE	Tracking Error	Theil's Inequality Coefficient
EHI Main Index	0,39%	1,52%	0,25	0,28%	1,52%	0,18	-0,11%	0,84	0,0085	0,85%	0,2754
EHI Arbitrage	0,36%	0,80%	0,45	0,15%	0,80%	0,18	-0,21%	0,68	0,0067	0,64%	0,3992
EHI CTA/Managed Futures	0,32%	1,36%	0,23	0,23%	1,36%	0,17	-0,09%	0,20	0,0171	1,72%	0,6209
EHI Distressed Debt	0,51%	1,88%	0,27	0,27%	1,88%	0,14	-0,24%	0,72	0,0143	1,42%	0,3738
EHI Event Driven	0,40%	2,01%	0,20	0,33%	2,01%	0,17	-0,07%	0,80	0,0125	1,26%	0,3084
EHI Fixed Income	0,27%	1,18%	0,23	0,22%	1,18%	0,18	-0,05%	0,60	0,0105	1,05%	0,4360
EHI Long/Short Equity	0,42%	2,20%	0,19	0,42%	2,20%	0,19	-0,01%	0,87	0,0110	1,10%	0,2459
EHI Macro	0,30%	1,05%	0,29	0,15%	1,05%	0,15	-0,15%	0,70	0,0083	0,82%	0,3849
EHI Multi Strategy	0,35%	1,32%	0,26	0,24%	1,32%	0,19	-0,10%	0,82	0,0080	0,79%	0,2955
EHI Relative Value	0,33%	1,21%	0,28	0,20%	1,21%	0,16	-0,14%	0,77	0,0082	0,81%	0,3328
Returns are calculated in excess of the risk-free rate.											

Table 5: Summary out-of-sample clone returns vs. realized returns

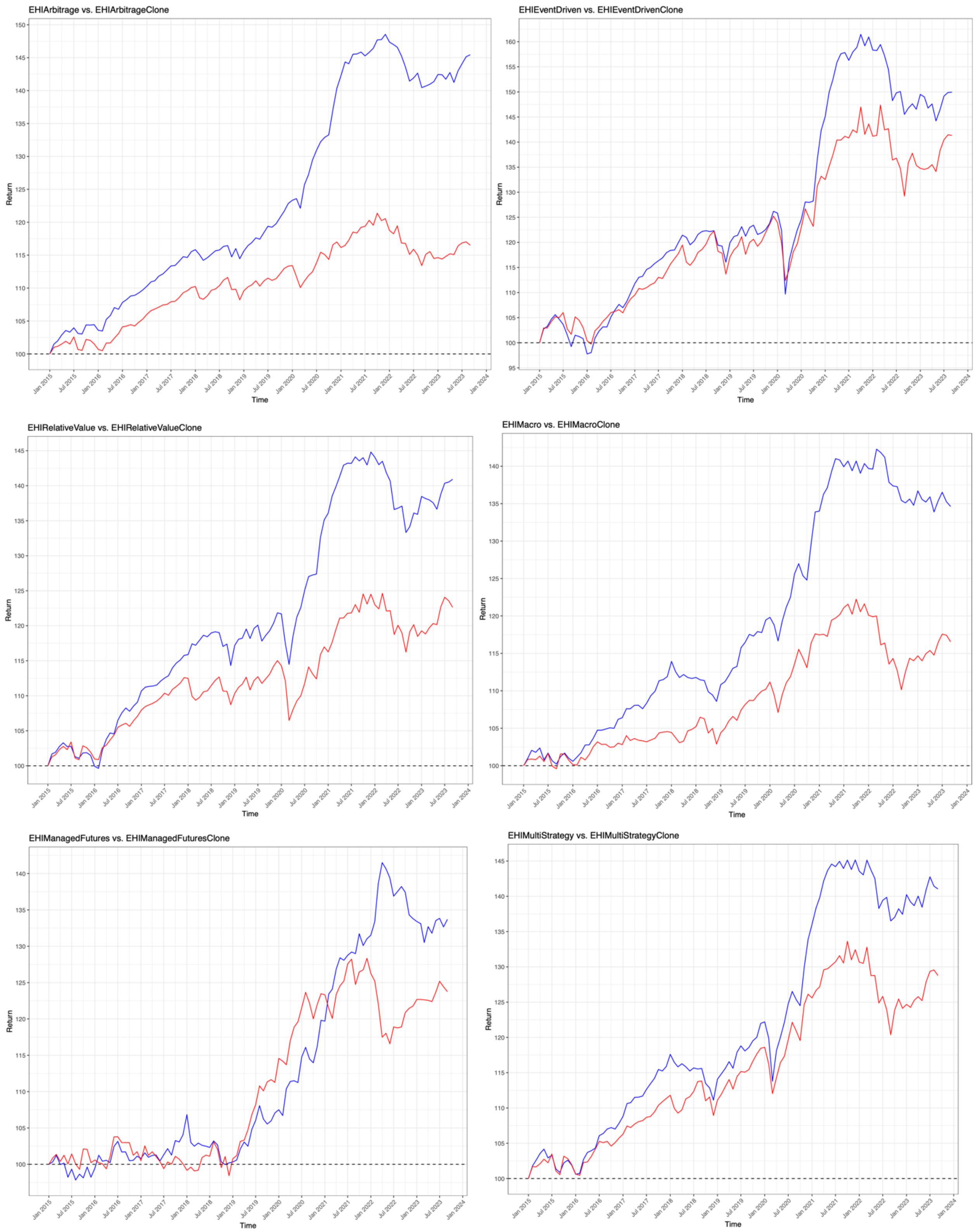


Figure 19: All clone returns vs. realized returns