8-2020

Capturing Hedge Fund Risk Factor Exposures: Hedge Fund Return Replication with ETFs

Jun Duanmu
Louisiana Tech University

Yongjia Li
Boise State University

Alexey Malakhov
University of Arkansas

Capturing Hedge Fund Risk Factor Exposures: Hedge Fund Return Replication with ETFs

Jun Duanmu
Louisiana Tech University

Yongjia Li
Boise State University

Alexey Malakhov*
University of Arkansas

Abstract

We develop a new factor selection methodology of spanning the space of hedge fund risk factors with all available exchange traded funds (ETFs). We demonstrate the efficacy of the methodology with out-of-sample individual hedge fund return replication by ETF clone portfolios. This is consistent with our interpretation of ETF returns as proxies to risk factors driving hedge fund returns. We further consider portfolios of "cloneable" and "noncloneable" hedge funds, defined as top and bottom in-sample R^2 matches, and demonstrate that our ETF clone portfolios slightly outperform cloneable hedge funds out of sample.

Keywords: hedge funds, risk factor exposures, factor selection, return replication

JEL Classification: G11, G23

1. Introduction

Hedge funds have experienced tremendous growth in recent years, with more than $3.2 trillion currently invested in hedge funds globally,[1] and are now considered an essential part of alternative investment strategies by institutional investors and financial institutions. Hedge funds have been able to produce returns with relatively low correlations with major asset classes, such as stocks and bonds, due to the multitude of investment opportunities available to fund managers. Unlike managers of more traditional mutual funds, hedge fund managers have the flexibility to invest in nontraditional asset classes (including derivative securities), employ leverage and engage in short sales. However, such strategies also expose investors to alternative risk factors that may not be easy to quantify, given the opacity of the hedge fund industry. It is then natural to question whether the returns earned by hedge fund managers are due to managerial skill or merely compensation for exposure to alternative risk factors.[2] If a significant portion of hedge fund returns comes from alternative risk factor exposures, then it is reasonable to presume that it is possible for investors to replicate that part of hedge fund returns at a lower cost by taking on these risk exposures themselves. However, such an exercise hinges on the investor's ability to identify and quantify these alternative risk factors via proxies of

* *Corresponding author*: Sam M. Walton College of Business, University of Arkansas, WCOB 343, Fayetteville, AR 72701; Phone: 479-575-6118; E-mail: amalakhov@walton.uark.edu.

We have benefited from comments by Richard Warr (editor), an anonymous referee, Carl Larsson, David Louton, Denitsa Stefanova, Marno Verbeek and seminar and conference participants at the University of Arkansas, the 2014 Financial Management Association Annual Meeting, the 2014 FMA Consortium on Research in Hedge Funds, Trading Strategies, and Related Topics, the 2015 Eastern Finance Association Annual Meeting, the 2015 Southwestern Finance Association Annual Meeting and the 2015 Financial Management Association European Conference.

[1] According to Hedge Fund Research, Inc., April 19, 2018, press release.

[2] For example, John H. Cochrane observes, "As I look across the hedge fund universe, 90% of what I see is not 'picking assets to exploit information not reflected in prices,' it is 'taking exposure to factors that managers understand and can trade better than clients'" (John H. Cochrane's "Hedge Funds" lecture notes at http://faculty.chicagobooth.edu/john.cochrane/teaching/35150_advanced_investments/hedge_notes_and_questions.pdf).

portfolios of tradable and liquid securities.[3] That is why the issue of choosing appropriate risk factors is central to any study of hedge fund performance, and currently there is no set of factors that is universally accepted across the literature.[4]

The main research objective of the paper is to focus on the factor-driven component of hedge fund returns[5] and capturing it with easily tradable investment instruments of exchange traded funds (ETFs). Our approach builds on the methodology of return attribution, and it relies on proper identification and selection of risk factors relevant for each individual hedge fund. We argue that passive ETFs can be interpreted as proxies for risk factors, as these return patterns are executed formulaically without human discretion. We argue that the full universe of ETFs currently provides comprehensive coverage of the space of risk factors that investors find attractive from the risk-and-return perspective. We span the space of potential risk factors with passively managed ETFs from 1997 to 2015. This time period saw an explosion in ETFs available, with the number of U.S. listed passively managed ETFs going from 19 in 1997 to 1,711 in 2015. Meanwhile, the ETF coverage of alternative risk factors went from almost nonexistent in 1997 to being comprehensive, with ETFs currently providing access to a great variety of alternative strategies that were previously available only to hedge funds or institutional investors.[6] This provides us with a unique opportunity to investigate how the expanding space of alternative risk factors affects the quality of hedge fund replication with ETFs available at the time.

While the large number of ETFs available in the later years of our study allows for more complete spanning of the space of risk factors, it also increases potential for spurious results due to excessive data mining. We develop a new methodology for linear hedge fund return replication that overcomes multicollinearity among ETFs and minimizes data mining bias while utilizing all ETFs available. Comparing the performance of hedge funds with their ETF clones in and out of sample, we find high accuracy of hedge fund replication with ETFs when there is sufficient number of ETFs available. We demonstrate that in the subperiod starting in 2005, the overall out-of-sample performance of the portfolio of all hedge funds is not statistically different from the portfolio of clones. We attribute this to the sufficiently large number of available ETFs in the later years, which allow us to successfully span the space of hedge fund risk factors.

In a departure from previous hedge fund replication studies, we go beyond considering replicating hedge fund indexes or average hedge fund performance. We consider portfolios of “cloneable” and “noncloneable” hedge funds, defined as top and bottom in-sample adjusted R^2 matches. Intuitively, we should not expect success in hedge fund return replication for a truly skilled hedge fund manager who pursues investment opportunities uncorrelated with risk factors, delivering true alpha to investors. On the other hand, we fully expect success in return replication for a manager who follows a rigid formulaic strategy, such as writing out of the money put options on the S&P 500 index, earning returns by exposing investors to an easily quantifiable alternative risk factor. An illustration of our success in out-of-sample return replication of a particular cloneable hedge fund is provided in Figure 1.[7]

Consistent with the above intuition, we demonstrate that the portfolio of clones created with our procedure provides better out-of-sample performance than the portfolio of cloneable hedge funds, which is likely due to the lower fee structure among the clones. Furthermore, the portfolio of cloneable hedge funds does not produce significantly positive risk-adjusted out-of-sample performance, measured by the Fung and Hsieh (2004) alpha. We conclude that there is no

[3] Notice that if there is no tradable option available to investors for a particular alternative risk factor, then it could be argued that hedge funds are valuable by merely providing access to that risk exposure. Such exposure through hedge funds comes at a high premium in the form of management and incentive fees.

[4] For example, return attribution studies (Fung and Hsieh, 2001, 2004; Agarwal and Naik, 2004) introduce new trend following and option-based risk factors in addition to Fama and French (1993) and Carhart (1997) factors. On the other hand, hedge fund replication studies (Hasanhodzic and Lo, 2007; Amenc, Martellini, Meyfredi and Ziemann, 2010; Giamouridis and Paterlini, 2010) employ liquid index portfolios available to investors.

[5] Such return patterns are commonly associated with either “passive” or “smart beta” investment styles (the difference between “passive” and “smart beta” investment styles is typically in the degree of sophistication in utilizing exotic risk factors).

[6] As an example of available ETF strategies, consider ALPS U.S. Equity High Volatility Put Write Index Fund (ticker HVPW) that tracks NYSE Arca U.S. Equity High Volatility Put Write Index with an annual expense ratio of 0.95%. The ETF benchmark tracks the performance of options sold on a basket of 20 stocks chosen from the largest capitalized equities that have the highest volatility, as determined by NYSE Arca Inc. Other examples include currency carry ETFs, volatility ETFs, value ETFs, momentum ETFs and so on.

[7] This particular (anonymous) hedge fund is in the “Multi Strategy” self-reported style, it has an inception year of 2007, and it was active at the end of our study period. Notice that the out-of-sample comparison begins in 2012, after dropping the first two years of observations to control for the backfill bias and after using another two years for the in-sample clone matching.

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Financial Review*, published by Wiley on behalf of The Eastern Finance Association. Copyright restrictions may apply. https://doi.org/10.1111/fire.12221.

statistical evidence of active managerial skill in the set of cloneable hedge funds and that their performance is primarily driven by exposure to ETF-quantifiable risk factors. Furthermore, we conclude that risk factor exposures among cloneable funds are relatively stable over time, hence these funds can be successfully replicated with ETF portfolios out of sample.

Finally, the out-of-sample portfolio of noncloneable hedge funds produces significantly positive mean excess returns along with a Fung and Hsieh (2004) alpha, outperforming the portfolio of clones out of sample. This can be interpreted as evidence of active managerial skill among the managers of noncloneable hedge funds.

We conclude that our methodology succeeds in comprehensively spanning the set of quantifiable risk factors that could be misinterpreted as “skill” by unsophisticated investors in hedge fund performance evaluation. This provides value in both identifying skilled managers of noncloneable hedge funds and also successfully replicating out-of-sample returns that are due to alternative risk exposures of cloneable hedge funds, thus providing a transparent and liquid alternative to investors who may find these return patterns attractive.

The rest of the paper proceeds as follows. Section 2 summarizes the related literature and discusses our contribution to the literature. Section 3 describes the hedge fund and ETF data. Section 4 explains the methodology on how we perform the in-sample matching analysis and how we construct the portfolios for the out-of-sample test. Section 5 discusses and analyzes the empirical results. Section 6 concludes.

2. Related Literature

Our methodology directly extends the factor-based hedge fund replication literature that goes back to Sharpe (1992)–style analysis approach. In its original form, it constructs a replicating portfolio by relying on constrained beta coefficients from a linear regression on a set of relevant factors. Hasanhodzic and Lo (2007) apply this methodology relying on six fixed factors to replicate hedge fund returns from TASS database, and demonstrate that replication works reasonably well for Dedicated Short Bias, Equity Market Neutral, Global Macro, Managed Futures, Fund of Funds, Convertible Arbitrage, Long/Short Equity Hedge and Multi-Strategy categories. However, their clones underperform in Event Driven and Emerging Market categories. Amenc, Martellini, Meyfredi and Ziemann (2010) extend Hasanhodzic and Lo (2007) by considering nonlinear and conditional hedge fund replication models. They do not find that going beyond linear models enhances the replication power. On the other hand, they find that selecting factors for each hedge fund category based on economic rationale yields a substantial improvement in out-of-sample replication quality. In addition, Bollen and Fisher (2013) utilize futures positions to clone ten Credit Suisse hedge fund indexes and find high correlation between clone returns and the hedge fund indexes. However, they find that the clones underperform the hedge fund indexes, and the high out-of-sample return correlation is largely driven by the matching quality for the equity index.

This is an intuitive result from the perspective of the literature on hedge fund risk and performance evaluation, as we do not have an equilibrium model of hedge fund performance evaluation and instead rely on risk-based factor models that approximate the true set of hedge fund risk factors. However, it is virtually impossible to observe the true set of hedge fund risk factors due to the myriad of available strategies to hedge fund managers and the opacity of the industry, and many hedge fund risk and performance evaluation studies[8] rely on statistical techniques, such as stepwise regression, to identify the dominant risk factors. More recently, Giamouridis and Paterlini (2010) and Weber and Peres (2013) employ statistical techniques in the factor-based hedge fund replication context, applying stepwise, as well as RIDGE, LASSO and LAR LASSO regressions[9] to sets of 16 and 30 risk-based factors.

Our contribution lies in expanding the universe of available risk factors by considering all available U.S. listed passively managed ETFs. We argue that a rigid formulaic strategy[10] can be viewed as a risk factor, and these ETFs represent reasonable proxies to a multitude of alternative risk factors that investors find attractive from the risk-and-

[8] See, for example, Fung and Hsieh (2001), Agarwal and Naik (2004), Vrontos, Vrontos and Giamouridis (2008) and Titman and Tiu (2011).

[9] See Hoerl and Kennard (1970), Tibshirani (1996) and Efron, Hastie, Johnstone and Tibshirani (2004) for descriptions of RIDGE, LASSO and LAR methodologies.

[10] For example, writing out of the money put options or covered call options on the S&P 500 index, earning returns by exposing investors to an easily quantifiable alternative risk factor.

return perspective.[11] We develop a methodology that successfully identifies a unique set of factors for each individual hedge fund based on cluster analysis and LASSO selection methodology that overcomes multicollinearity among ETFs, and also minimizes data mining bias, resulting in parsimonious factor selection. We test the performance of our hedge fund clones in and out of sample and find that the overall accuracy of hedge fund replication with ETFs increases with the number of ETFs available. Our out-of-sample portfolio approach allows minimizing the hedge fund attrition bias that Ben Dor, Jagannathan, Meier and Xu (2012) find to be a major driver of poor hedge fund index clone performance against hedge fund index benchmarks.

Another contribution is in separately considering risk-adjusted performance of cloneable and noncloneable hedge funds and their ETF clone portfolios out of sample, unlike previous studies that focus on replicating hedge fund indexes or average hedge fund performance. This also demonstrates potential effectiveness of our ETF-based methodology and contributes to the literature on hedge fund risk and performance evaluation.[12] Our result shows that ETF-based clones perform slightly better than their underlying hedge funds in the cloneable group, while for noncloneable funds, their ETF-based clones underperform their underlying hedge funds. Also, while comparing the out-of-sample performance of cloneable and noncloneable hedge funds, we demonstrate that cloneable funds fail to deliver significantly positive risk-adjusted out-of-sample performance, while we find superior out-of-sample risk-adjusted performance for noncloneable funds. The difference in the out-of-sample performance between cloneable and noncloneable hedge funds confirms the results in Titman and Tiu (2011) and Bollen (2013) with our ETF methodology. It also supports the previous literature on hedge fund performance evaluation and systematic risk factor exposures (e.g., Bali, Brown and Caglayan, 2011, 2012, 2014). While our approach in defining cloneability by adjusted R^2 is similar to that in Titman and Tiu (2011) and Bollen (2013), there is a major difference with Titman and Tiu (2011) and Bollen (2013) in covering the space of risk factors with ETFs in our study. While Titman and Tiu (2011), Bollen (2013) and Bali, Brown and Caglayan (2011, 2012, 2014) link hedge fund performance to exposure to systematic risk factors, our study, to our knowledge, is the first one to analyze the out-of-sample performance of hedge fund clones based on tradable proxies of systematic risk factors, sorted by adjusted R^2. Hence, our ETF-based methodology also provides value in hedge fund performance evaluation by identifying skilled managers who deliver superior out-of-sample risk-adjusted performance.

3. Description of Data

In this study, we utilize hedge fund data from Bloomberg for the period 1997–2015, which includes 18,135 unique hedge funds.[13] To minimize survivorship bias, we include both active and inactive funds. We partially offset the effects of backfill bias by eliminating the first 24 months of reported returns.[14] Since we require two years of data to create a hedge fund clone and at least a year to test the clone error, we only consider funds with inception dates prior to 2011, which leaves us with 6,822 unique funds with 2,393 active funds and 4,429 inactive funds (i.e., acquired, liquidated or chose to stop reporting).

Panel A of Table 1 reports summary statistics of fund returns, fees, investor liquidity measures and fund longevity. As medians are better measures of typical funds in our database, we find that the typical fund has a 1.5% management fee, a 20% incentive fee on all profits over an investor's high-water mark, a $250,000 minimum initial investment and a 30-day redemption period. Unsurprisingly, active funds display higher monthly returns and assets under management and greater longevity than inactive funds. Interestingly, however, inactive funds have longer redemption periods. Panels B and C of Table 1 report percentages of funds with certain characteristics and declared styles, respectively.

[11] On the other hand, actively managed ETFs involve human judgment on the part of their managers, hence they can hardly be interpreted as proxies for risk factors.

[12] See, for example, Jagannathan, Malakhov and Novikov (2010), Titman and Tiu (2011), Avramov, Kosowski, Naik and Teo (2011), Sun, Wang and Zheng (2012), Bali, Brown and Caglayan (2011, 2012, 2014), Avramov, Barras and Kosowski (2013), Jurek and Stafford (2015) and Duanmu, McCumber and Malakhov (2018).

[13] We do not include funds of hedge funds. The non-USD-dominated funds are converted into USD and are reported in Bloomberg.

[14] The 24 months backfill correction is in line with results in Jagannathan, Malakhov and Novikov (2010) and Titman and Tiu (2011) suggesting dropping the first 25 and 27 months of returns. As a robustness check, we drop only the first 12 monthly observations to address the backfill bias, which increases the data availability for the matching procedure. The results are quantitatively similar, and we decide not to report these results for brevity.

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Financial Review*, published by Wiley on behalf of The Eastern Finance Association. Copyright restrictions may apply. https://doi.org/10.1111/fire.12221.

Eighty-two percent of all funds have a high-water mark provision, although only 4% allow hurdle rates in addition to high-water marks. The most common declared style is long-short equity, at 28% of all funds, while activist is the least common style, accounting for 0.12% of hedge funds.[15]

We collect the ETF data from Morningstar for the period 1994–2015, which contains 1,904 unique U.S. listed ETF funds.[16] We manually check the description of each ETF and exclude all ETFs that are not passively managed index-tracking funds,[17] as well as ETFs that track hedge fund–style indexes; this leaves us with 1,799 unique ETFs. In our study, we require ETFs to have at least 24 monthly observations starting from January, and we drop those ETFs with missing management fee information. Finally, 1,711 unique ETFs for the period 1997–2015 are included in this study.[18]

4. Research Methodology

4.1. Style analysis with ETFs

In this study, we employ a factor selection model termed LASSO (least absolute shrinkage and selection operator) proposed in Tibshirani (1996). For a given parameter *t*, LASSO regression identifies an optimal set of factors with nonzero coefficients such that

$$\begin{aligned} &\hat{\beta}_{Lasso} = \arg\min_{\beta} \| \mathbf{r} - \mathbf{X}\beta \|^2, \\ &\text{such that } \sum_{j=1}^{m} |\beta_j| \le t. \end{aligned} \tag{1}$$

where $\mathbf{r}$ is the vector of hedge fund monthly returns in our research and $\mathbf{X}$ is the vector of ETF monthly returns.

Conceptually, provided a set of factors, LASSO regression determines the appropriate factors to be selected through an optimization approach. In the constrained form of ordinary least squares regressions, the sum of absolute values of the beta coefficients is estimated and constrained to be smaller than a specific parameter. For a given selection parameter *t*, some of the beta coefficients could be zero if the corresponding factors reveal little or no information about the dependent variable. As a result, LASSO regression "shrinks" the set of regression factors until the beta coefficients are the solution of the optimization problem. The degree of "shrinking" depends on the chosen value of the parameter *t*, with lower values of *t* resulting in fewer factors being selected for the model. We calculate LASSO regression solutions across a range of *t* values by employing a computationally efficient least angle regression (LAR) modification of the LASSO procedure introduced by Efron, Hastie, Johnstone and Tibshirani (2004). We then use Bayesian Information Criterion (BIC) as the model selection criterion, and we select the model with the lowest BIC value.

However, before adding all ETFs as explanatory variables into LASSO regression, we need to tackle the multicollinearity in the comprehensive set of ETFs. Although our ETFs database has factored in a broad set of trading strategies, it is not surprising that some ETFs are exposed to similar risk factors, therefore exhibiting similar or even the same return patterns. And even though LASSO regression could be a powerful selection method in dealing with collinearity, it is not feasible for LASSO regressions to handle collinearity for such a large number of closely correlated ETF factors in a meaningful way.

---

[15] Using inverse equity ETFs to replicate the performance of short-biased hedge funds could be a useful test to examine the efficacy of our methodology. Regrettably, due to data limitation, conducting such experiment is difficult in this study as short biased hedge funds only represent 0.21% of our sample and we require such funds to survive at least four years to be included in our analysis, which eventually results in an insufficient amount of funds for the test.

[16] Exchange Traded Notes (ETNs) are also included. Commodity, leveraged and volatility factors are often based on ETNs.

[17] Benchmark indexes that retained ETFs track may not be publicly available. Some funds track in-house indexes.

[18] The distribution of number of ETFs available and number of ETF selected across years is available in Online Appendix Figures A and B at the journal webpage (https://financialreview.poole.ncsu.edu).

To address this problem, we conduct cluster analysis among ETFs in order to reduce the number of ETF factors prior to running LASSO regressions. For every ETF in each cluster we calculate the distance away from the center of its cluster, as defined by the *SDI* measure from Sun, Wang and Zheng (2012). This distance measure for an ETF *i* is calculated as one minus the correlation of the ETF's return with the mean return of all ETFs from the same cluster *I*—that is,

$$SDI_i = 1 - corr(r_i, \mu_I), \\ \text{where } \mu_I = \frac{\sum_{i \in I} r_i}{count(i \in I)}. \qquad (2)$$

The lower the *SDI*, the closer the ETF is from the center of its cluster. We specify the ETF with the lowest *SDI* as a proxy for all the ETFs in the same cluster, and then we include this ETF as a replicating factor in LASSO regression. This approach allows efficient spanning of the space of potential risk factors, while mitigating multicollinearity by maximizing the distance between ETFs used.

Because the number of ETFs changes over time and we do not know the true number of clusters, we assume that the number of clusters ranges from 1 to 100. We set the maximum number to 100 since we believe it is an efficient and sufficiently large set of investment opportunities (because there are fewer than 100 ETFs for years before 2003, we set the maximum number of cluster as the number of ETFs during those years). We then iteratively run cluster analysis for 100 times and use the corresponding number of ETFs (each selected ETF is located at the center of its cluster) in LASSO regression. Consequently, after running cluster analysis and LASSO regressions, each fund would have 100 corresponding models. We then choose the model with the highest adjusted R^2 as our clone model. Such an approach minimizes data mining bias, resulting in parsimonious factor selection.

The basic model for LASSO regression is as follows:

$$r_{i,gross} - r_f = \beta_1(ETF_1 - r_f) + \beta_2(ETF_2 - r_f) + ... + \beta_{100}(ETF_{100} - r_f) + \varepsilon_i \qquad (3)$$

where $r_{i,gross}$ is the gross monthly return of fund *i*, and r_f is the risk-free rate proxied by the monthly return of the 30-day U.S. Treasury bill. We use gross hedge fund returns[19] on the left-hand side and gross ETF returns on the right-hand side, since we try replicating hedge fund return patterns that are driven by exposure to alternative risk factors.[20] Otherwise, the true factor risk-driven hedge fund returns would be altered if we consider them net of fees, and hence the matched ETF risk profile would not reflect the true factor risk exposures. We also suppress the intercept in regressions, following a common approach in the hedge fund cloning literature (see, e.g., Hasanhodzic and Lo 2007).[21] In a slight departure from Sharpe (1992)–style analysis methodology, we do not restrict beta coefficients to be positive or add up to one, as imposing such restrictions would likely result in model misspecification in the context of hedge funds that are free to take leverage and short positions.[22]

In order to quantify the dynamic nature of hedge funds' investment activities, we run the LAR LASSO methodology for Model 3 for every hedge fund in our data over a set of two-year windows, rolling them annually over the sample period. We consider adjusted R^2 and BIC values from these matching regressions as in-sample proxies of the "overall quality" of our matching procedure. We interpret higher adjusted R^2 and lower BIC values as indicators of our methodology's success in capturing hedge fund risk factors, and thus the potential for cloning hedge fund returns with ETFs.

[19] See Online Appendix A for details on the gross return calculations. The appendix is available at the journal webpage (https://financialreview.poole.ncsu.edu).

[20] Getmansky, Lo and Makarov (2004) show that hedge funds exhibit serial correlation in returns due to fund positions in illiquid assets and/or due to deliberate smoothing by managers. We therefore apply MA(2) smoothing correction to hedge fund returns in our baseline regressions (3).

[21] The economic rationale for omitting the constant coefficient in cloning regressions is that it is impossible to replicate a constant rate of return above the risk-free rate with commonly available financial instruments.

[22] Ter Horst, Nijman and de Roon (2004) demonstrate that imposing unwarranted style-based constraints can lead to biased risk exposure estimates.

However, the ultimate goal is to test the predictive power of the methodology so as to validate the in-sample explanatory power manifested by high R^2 and low BIC values. For each hedge fund, we consider the corresponding ETF matches selected through the previous two-year window LASSO regression and their coefficients, and then construct the hedge fund clone by loading selected ETFs with regression-determined weights. The hedge fund clone performance after the matching period is then given by

$$CloneRet_{i,t} = r_{f,t} + \sum_{j=1}^{n} \beta_{j,t-1}(ETF_{j,t} - r_{f,t}), \qquad (4)$$

where $\beta_{j,t-1}$ is the coefficient from the previous two-year window LASSO selected ETF j. We rely on net-of-fees returns for both hedge funds and their ETF matches in our out-of-sample analysis as we compare future returns from an investor perspective. Finally, we address the survivorship bias among hedge funds by constructing out-of-sample portfolios and rebalancing them when hedge funds drop out of the database.

4.2. Cloneable and Noncloneable Hedge Funds

In a departure from previous hedge fund replication studies, we go beyond exploring aggregate characteristics of clones versus hedge funds they replicate. Instead, we concentrate on comparing cloneable and noncloneable hedge funds, defined as top and bottom in-sample adjusted R^2 matches. We argue that the success in hedge fund replication depends on a hedge fund manager's style and that properly deconstructing that style is paramount for assessing the true value of a hedge fund for investors. For example, if a hedge fund manager has genuine ability and pursues a unique strategy uncorrelated with identifiable risk factors in a noncloneable fund, then we should not expect success in replicating such fund performance. On the other hand, if a manager pursues algorithmic strategies highly correlated with risk factors in a cloneable fund, then we expect success in out-of-sample replication, as our hedge fund clone would deliver a similar risk and return profile but at a lower cost compared to the cloneable fund.[23] Furthermore, it would be unlikely to find evidence of superior risk-adjusted managerial skill in cloneable funds in the context of a return attribution model, as their performance would be driven mostly by factor risk exposures.

5. Empirical Results

5.1. Matching Regressions

Our matching (or "cloning") procedure is based on in-sample LAR LASSO regressions for Model 3, with the best model chosen according to the BIC, as described in the previous section. Table 2 reports the results for annual rolling two-year matching regressions from 1997 to 2014. To highlight the effect of the broadened investment opportunity set for our matching procedure, we also consider subperiods of 1997–2003 and 2003–2014 separately.[24] The results confirm our expectation of better matching in later years, reflecting a greater degree of success in spanning the space of available risk factors as more ETFs become available. On average, in 1997–2003 there are only 45 ETFs available, and the average matching adjusted R^2 is 0.34, while in 2003–2014 there are on average 557 ETFs available for the matching regressions, and the average adjusted R^2 is 0.49. We also observe that the mean BIC has declined through time, from 60.07 in 1997–2003 to 43.38 in 2003–2014. This suggests that matching quality has improved along with the broadened investment opportunity set as more ETFs become available. Moreover, the average number of factors selected by the LAR LASSO procedure is 2.64 for the whole sample period (2.85 for the 2003–2014 period), which indicates that our methodology results in a parsimonious factor selection. Last, in order to track the effectiveness and the time pattern of the replication procedure, we calculate two ratios that relate the number of hedge funds and the number of ETFs. Coverage Ratio is the ratio of the number of ETFs available over the number of hedge funds, and Selection Ratio is the ratio of the number of ETFs selected over the number of hedge funds. As expected, the Coverage Ratio is increasing consistently through our sample period, which indicates that the ETF industry has witnessed a significant expansion and provided more coverage on the investment opportunity set through our sample. In addition, ETF industry is growing at a higher pace proportional to hedge fund industry as the Coverage Ratio increases from

[23] Unfortunately, this study does not account for ETF trading fees due to the fact that the data on ETF bid and ask prices is available in Bloomberg only since 2012. Incorporating trading costs via accounting for bid-ask spreads for ETFs over the limited time period of 2012–2015, when such data are available, could be the subject of a different study.

[24] We chose 2003 as the break year, since it is the first year when there are more than 100 ETFs available, which allows full utilization of our methodology based on a variable number of ETF clusters up to 100.

7% in our first two-year window to 36% in the last window period. Selection Ratio, on the other hand, confirms the effectiveness of ETF factor selection methodology. As the ETF industry expands to provide more coverage, our selection process successfully picks out the most relevant factors, which is proved by a set of stable Selection Ratios. Consistently across our sample, only around 5% of the most representative ETFs in each window are selected to explain and replicate hedge fund returns.

5.2. Out-of-Sample Matching Statistics

The ultimate test of our methodology lies in considering the out-of-sample performance of hedge fund clones versus the hedge funds they replicate. As described in the methodology section, we construct a hedge fund clone as a portfolio of model selected ETFs with the matching regression-determined weights. The out-of-sample performance of a hedge fund clone is given by Equation (4). It is important to reiterate that in the out-of-sample analysis, we use the net-of-fees returns for both hedge funds and their ETF clones, as we decide to compare the out-of-sample performance from an investor perspective.[25] Finally, we calculate tracking errors as the differences between the clone returns and the corresponding hedge fund returns—that is,

$$TrackingError_{i,t} = CloneRet_{i,t} - HedgeFundRet_{i,t}. \qquad (5)$$

Table 3 reports the results of the out-of-sample matching statistics for hedge funds and their respective clones in the year following the in-sample matching period.[26] Consistent with the in-sample results, reported in Table 2, the average out-of-sample accuracy has increased over the years with the average monthly tracking error increasing from –0.68% in 1999–2004 to –0.05% in 2005–2015, and the average monthly tracking error volatility decreases from 4.09% in 1999–2004 to 3.35% in 2005–2015.[27] This is consistent with our expectation of the improved matching quality when more ETFs become available to span the set of potential hedge fund risk factors.

5.3. Cloning Success and the Evolution of ETFs

We further investigate the ETFs that are selected based on the LASSO model to gain better insights into the efficacy of our methodology. Essentially, we want to confirm that the success of our methodology relies on the expanding nature of the ETF industry with more ETFs available and better coverage of different asset classes provided by such ETFs. We obtain the ETF reported styles from MorningStar and examine the distributions of the ETF styles for each window in our sample period. We have six broad ETF categories in our sample, which are Allocation, Alternative, Commodities, Equity, Fixed Income and Tax Preferred, and the results are reported in Table 4. In the first subsample period, only Equity ETFs are selected. While in the second period, we find that ETFs with different styles are matched based on our methodology. For example, in the window 2002–2003, 76 ETFs are selected and they are all equity ETFs. In the window 2011–2012, a total of 130 ETFs are picked among which only 41 are Equity ETFs, 66 are Alternative ETFs, 18 are Commodity ETFs, 2 are Allocation ETFs and 3 are Fixed Income ETFs. This is consistent with our argument that the better coverage of different asset classes provided by the expanding number of ETFs improves our matching and prediction quality.[28]

[25] Recall that the in-sample matching regressions rely on gross returns, because we want to get closest possible matches to "true" hedge fund strategies, as carried out by hedge fund managers.

[26] The out-of-sample results are subject to the survivorship bias since not all hedge funds survive throughout the one-year comparison period. Because we consider only hedge funds with at least four years of return history in our analysis, it is impossible to directly compare our attrition rates to previous literature on the matter, such as Liang (2000) and Liang and Park (2010), while the closest indirect comparison would be to "failure rates" in Liang and Park (2010). For the comparable time period from 1999 to 2004, the average attrition rate in our study is 4.24%, compared to the adjusted failure rate of 4.36% over the same time period from Liang and Park (2010).

[27] The choice of 2004 as the out-of-sample break year is consistent with 2003 being the in-sample break year, since it is the first year when out-of-sample predictions are based on more than 100 ETFs available. Given the different nature of the two time periods, 1999–2004 and 2005–2015, the exact statistical significance of the difference in average tracking error and tracking error volatility between the two time periods cannot be reasonably assessed and interpreted.

[28] To further study the cloning success and hedge fund styles, we follow Agarwal, Daniel and Naik (2009) and reclassify the hedge funds into four consolidated categories: Directional Traders, Relative Value, Security Selection and Multiprocess. Overall, it seems that our cloning methodology

In addition, we also investigate the persistence of those matched ETFs. We notice that from 2004 to 2014, the number of ETFs available increases from 119 to 1,196, while the number of ETFs selected only increases from 93 to 137. It is reasonable to question if a fixed set of ETFs are repeatedly selected in every window, which would weaken our argument of the importance of the ETF industry expansion. We then calculate the rate of persistence, which examines the overlap of the current selected ETF pool and that from the previous window. The results are reported in parentheses in Table 4. For example, in the first window, only 19 Equity ETFs are selected and the same 19 ETFs are used again in the second window. The rate of persistence therefore is 100%, which indicates a 100% overlap. In the window 2013–2014, 48 Equity ETFs are selected with a rate of persistence of 31.25%, which means only 31.25% of the ETFs are the same from the equity ETF pool in the window 2012–2013. We observe an overall decreasing trend of the rate of persistence from the first sample period to the second period and the average rate of persistence across all windows are reasonably low. This confirms our argument that the increasing number of ETFs is essential to the success of our methodology.[29]

5.4. Cloneable and Noncloneable Hedge Funds

Our prior results in Table 3 indicate that the performance of clones is comparable with the performance of hedge funds in aggregate. However, there is a wide discrepancy in replication success among individual funds. In this section, we consider two groups of hedge funds, formed based on the top and bottom quintile of the in-sample adjusted R^2. We define the funds with top quintile of adjusted R^2 as cloneable and the funds with the bottom quintile of the adjusted R^2 as noncloneable.[30]

As our methodology allows us to effectively span the space of potential risk factors, the adjusted R^2 could be viewed as a proxy for how quantifiable or "decipherable" the investment strategy of a hedge fund is. Arguably, there is a fundamental difference in risk profiles between the top and bottom adjusted R^2 groups of hedge funds. For example, it is plausible that a manager of a cloneable (i.e., high R^2) fund generates returns by simply loading up on risk factors, which is identifiable by our methodology, while a manager of a noncloneable (i.e., low R^2) fund likely has the genuine ability and pursues a truly unique strategy uncorrelated with identifiable risk factors. We identify cloneable and noncloneable hedge funds based on their in-sample LASSO adjusted R^2 rankings, on quintile basis. Table 5 reports in-sample matching results for cloneable and noncloneable funds, and Table 6 reports out-of-sample matching results for cloneable and noncloneable hedge funds and the clone portfolios.

Consistent with full sample results from Table 2, the overall quality of in-sample matches increases over time for both cloneable and noncloneable funds as more ETFs become available. On average, cloneable funds register larger magnitudes of increases in the matching R^2 and decreases in BIC compared to the results of noncloneable funds. Cloneable funds have a higher number of regressors of 3.93 compared with 1.48 for the noncloneable funds. The cloneable fund returns are more negatively skewed with a lower kurtosis. In addition, we demonstrate that cloneable funds exhibit higher extreme tail risk, which is captured by VaR.[31]

Table 6 reports the out-of-sample matching statistics for cloneable funds, noncloneable funds and their respective clones in the one-year period following the matching window. We find that cloneable funds yield higher quality out-of-sample matches with closer means and smaller volatilities of tracking errors compared to noncloneable funds. This difference is especially pronounced in the second subperiod, which is consistent with the previous results of increased effectiveness of our methodology when the number of available ETFs exceeds 100.

It is important to point out that we rely on gross returns for the in-sample matching with the objective to fully account for all the risk factors inherent in the strategies pursued by hedge fund managers. On the other hand, we use net-of-fees returns in our out-of-sample analysis as we compare returns from an investor perspective. Successful clone is

could be somewhat more successful for Relative Value and Multiprocess styles, compared with Directional Traders and Security Selection styles. The results are presented in Online Appendix C at the journal webpage (https://financialreview.poole.ncsu.edu).

[29] We perform additional robustness tests on the relation between the number of ETFs and cloning success. With an increased number of ETFs available, our methodology generates better in-sample matching quality and out-of-sample prediction accuracy. The results are presented in Online Appendix D at the journal webpage (https://financialreview.poole.ncsu.edu).

[30] In addition to the quintiles, we also use quartiles to identify cloneable and noncloneable funds. The results are quantitatively similar and are available upon request.

[31] This is consistent with our interpretation of cloneable funds as funds with overall risk exposure more easily attributed to systematic risk factors.

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Financial Review*, published by Wiley on behalf of The Eastern Finance Association. Copyright restrictions may apply. https://doi.org/10.1111/fire.12221.

expected to slightly outperform the hedge fund target since our ETF-based clone has a lower fee structure compared to the hedge fund being cloned. Consistently, we report positive average tracking errors for cloneable funds in the period 2005–2015, when there is a sufficient number of ETFs available for spanning the investment opportunity set.

With respect to the risk measures, cloneable funds and their clones are more negatively skewed, have lower kurtosis and higher VaR compared with noncloneable funds and their clones.[32] The negative skewness observed among the cloneable funds and their respective clones is consistent with the interpretation that cloneable hedge funds mostly load up on exotic risk factors with asymmetric payoffs.[33] Furthermore, the fact that the clones of cloneable hedge funds also demonstrate negative average out-of-sample skewness with comparable VaR and kurtosis values could be interpreted as our success in "deciphering" strategies of cloneable funds and producing clones with similar risk and return profiles. However, our methodology could not provide good in-sample matches for noncloneable funds, and the clones are not successful in delivering comparable out-of-sample performance.[34] This is consistent with the interpretation of the irreplicability of truly active hedge fund management for noncloneable funds whose alpha-driven return could be beneficial to potential investors.

As noted above, it is reasonable to attribute cloning success to individual hedge fund strategies, which may be partially reflected in individual fund characteristics. We then compare differences in fund characteristics between cloneable and noncloneable funds and report the results in Table 7. The results are consistent with our attribution of nonclonability to active hedge fund management, as we find that noncloneable fund managers charge higher management fees and performance fees compared with managers of cloneable funds. The difference is statistically significant at the 1% level. In addition, noncloneable funds exhibit higher minimum investment requirement and are more likely to impose high-water marks, which is again consistent with our interpretation of active fund management as a feature of noncloneable funds. We do not find any significant difference in the total assets under management and lockup requirement among these two groups.

Finally, we notice that in Table 6, the noncloneable hedge funds have higher average attrition rates than cloneable funds, and they are less negatively skewed with higher kurtosis and lower VaR. While it is not possible to unequivocally claim an underlying reason for this phenomenon, it could be indicative of other risks, not quantifiable with our methodology, among noncloneable hedge funds.[35] For example, noncloneable funds tend to have higher fees and more likely to have high-water mark provisions, while offering less restrictive lockup and redemption periods, according to Table 7. This may cause fund assets to be more sensitive to streaks of bad performance and thus result in funds dropping out of the database in order not to report instances of extremely bad performance.[36] However, Liang and Park (2010) argue that a decision to leave a database should not be equated with failure as a hedge fund, and it is possible that the higher attrition rate among noncloneable funds does not imply higher failure rate compared with cloneable funds.

5.5. Out-of-Sample Portfolio Analysis

We now concentrate on out-of-sample portfolio tests[37] for the following reasons. First, by considering all funds up until the moment of their disappearance from the database, we minimize the effects of the survivorship bias. Second, the portfolio approach allows for out-of-sample risk-adjusted performance evaluation of hedge funds and their clones over long periods of time.

[32] The pattern holds both in and out of sample, and it is especially pronounced during the time period when applying our ETF matching methodology yields the most meaningful results—that is, in 2005–2015.

[33] Payoffs from such strategies, like writing out of the money put options on the S&P 500 index, may look attractive from the point of not very sophisticated investors.

[34] Out-of-sample clones yielded negative average tracking errors and high tracking error volatility and could not match the skewness of noncloneable funds.

[35] This is consistent with Bollen's (2013) findings of higher probability of failure for zero-R^2 hedge funds.

[36] Hedge funds may stop reporting extremely bad performance to databases in an effort to prevent redemptions in case investors are not restricted by lockup and redemption provisions.

[37] We follow Duanmu, Malakhov and McCumber (2018) to conduct the out-of-sample portfolio tests.

We form portfolios on December 31, 1998. We invest the same dollar amount into each fund within a portfolio in the beginning and follow its net-of-fees performance until December 31, 2015, rebalancing it once a year based on the annually updated LASSO regression matches. When a portfolio fund disappears from the database, we redistribute the remaining capital in the fund equally among surviving portfolio funds.[38] This procedure produces a time series of 204 monthly returns for each portfolio, allowing us to evaluate long-term portfolio performance across various economic conditions, including the most recent financial crisis of 2008–2009. We then calculate end dollar values based on a $1 initial investment, mean excess monthly returns, Sharpe ratios, Sortino ratios, Fung and Hsieh (2004) alphas,[39] information ratios, skewness and attrition rates for each time series of monthly portfolio returns from January 1999 until December 2015. In addition, we also examine the out-of-sample performance in two different time spans so as to reflect the nature of the booming ETF industry. The first period is from 1999 to 2004, where we have fewer than 100 ETFs that could be used for the matching procedure, while the second period is from 2005 to 2015, where we have more than 100 ETFs, resulting in comprehensive coverage of the space of potential hedge fund risk factors. Hence, we expect to observe increased replicating quality in the second period.

Table 8 reports out-of-sample performance results for the portfolio of all hedge funds in our sample. For the whole sample period, our clones fail to compete with real hedge fund returns in every performance measure. However, when looking into the details, we observe that these unfavorable results are driven by the inferior clone performance in the first period, 1999–2004. This again confirms our discussion that the quality of replication is highly influenced by the number of available ETFs. Looking at the first-period performance alone, we demonstrate that hedge funds deliver significantly better returns than their respective clones, which is consistent with our previous observations of the matching quality in the first period being worse than in the second. In the second subperiod of 2005–2012, our result shows that the clones do reasonably well in terms of producing similar return patterns and skewness, almost the same monthly excess returns, as well as pretty close risk-adjusted measures—that is, Fung and Hsieh (2004) alphas, Sharpe ratios, Sortino ratios and information ratios. We therefore conclude that the clones constructed using our methodology deliver payoffs that are similar to payoffs of the hedge funds, given a broad selection of ETFs representing potential hedge fund risk factors.

5.6. Out-of-Sample Portfolio Analysis for Cloneable and Noncloneable Funds

We now apply the out-of-sample portfolio approach to analyzing portfolios of cloneable and noncloneable hedge funds, defined as top and bottom adjusted R^2 from in-sample LASSO regression matches. Tables 9 and 10 report top and bottom quintile portfolio comparisons for the whole sample period and the two subperiods, respectively. We find that the clone portfolios underperform both cloneable and noncloneable hedge fund portfolios over the whole sample period (Table 9) and the period of 1999–2004 (Table 10, Panel A), and this is mostly driven by the inferior matching quality resulted from the insufficient number of ETFs available.

The out-of-sample portfolio analysis for the second subperiod of 2005–2015 yields interesting results, which are presented in Panel B of Table 10.[40] Our results show that the portfolio of clones delivers slightly better out-of-sample performance, with a very similar risk and skewness profile, compared to the portfolio of cloneable hedge funds. However, both hedge funds and clones fail to deliver statistically significant Fung and Hsieh (2004) alphas. This implies that hedge fund managers of cloneable hedge funds mostly produce returns driven by risk factors and do not add value to their managed portfolios. Given the favorable fee structure of ETFs, compared with hedge fund fees, it is not surprising that our ETF clones can replicate, or even slightly outperform, the overall performance of cloneable hedge funds.

On the other hand, the portfolio of noncloneable hedge funds outperforms the portfolio of their ETF clones and produces a statistically significant Fung and Hsieh (2004) alpha, along with greater Sharpe, Sortino and information ratios. This is consistent with the interpretation that noncloneable hedge fund managers add value through actively managing their funds and deliver superior risk-adjusted performance. Furthermore, the managerial skills possessed by noncloneable fund managers seem to be truly unique and cannot be replicated with ETF clones that are driven by

[38] This is somewhat conservative as it is possible that a fund simply chooses to stop reporting to the database, which is likely for well-performing funds that are no longer accepting new investor flows. However, without returns data, we obviously cannot keep the fund in the portfolio.

[39] See Online Appendix B for details on Fung and Hsieh (2004) alpha calculation. The appendix is available at the journal webpage (https://financialreview.poole.ncsu.edu).

[40] Quartile portfolio results are qualitatively similar and are available upon request.

publicly tradable risk factors. However, as mentioned in Section 5.4, the noncloneable hedge funds have higher average attrition rate than cloneable funds, which could be indicative of other risks, not quantifiable with our methodology, associated with their active management style.

6. Conclusion

We develop a methodology of hedge fund return replication with ETFs based on cluster analysis and LAR LASSO factor selection that overcomes multicollinearity among ETFs and also minimizes data mining bias, resulting in parsimonious factor selection. We test the performance of our hedge fund clones in and out of sample and find that the overall out-of-sample accuracy of hedge fund replication with ETFs increases when there is sufficient number of ETFs available. This is consistent with our interpretation of ETF returns as proxies to a multitude of alternative risk factors that could be driving hedge fund returns. We further consider portfolios of cloneable and noncloneable hedge funds, defined as top and bottom in-sample adjusted R^2 matches. Our results show that the portfolio of clones created with our procedure provides better out-of-sample performance than the portfolio of cloneable hedge funds. We demonstrate superior risk-adjusted performance for noncloneable funds, while cloneable funds fail to deliver significantly positive risk-adjusted performance, which is consistent with our success in cloning them. This approach contributes to the literature on hedge fund risk and performance evaluation, enabling investors to identify skilled managers who deliver superior out-of-sample performance.

We conclude that our methodology provides value in both identifying skilled managers of noncloneable hedge funds and also successfully replicating out-of-sample returns that are due to alternative risk exposures of cloneable hedge funds, thus providing a transparent and liquid alternative to investors who may find these return patterns attractive.

References

Agarwal, V., N.D. Daniel and N.Y. Naik. 2009. Role of managerial incentives and discretion in hedge fund performance, *Journal of Finance* 64, 2221–2256.

Agarwal, V. and N.Y. Naik, 2004. Risk and portfolio decisions involving hedge funds, *Review of Financial Studies* 17, 63–98.

Amenc, N., L. Martellini, J.-C. Meyfredi and V. Ziemann, 2010. Passive hedge fund replication—beyond the linear case, *European Financial Management* 16, 191–210.

Avramov, D., L. Barras and R. Kosowski, 2013. Hedge fund return predictability under the magnifying glass, *Journal of Financial and Quantitative Analysis* 48, 1057–1083.

Avramov, D., R. Kosowski, N.Y. Naik and M. Teo, 2011. Hedge funds, managerial skill, and macroeconomic variables, *Journal of Financial Economics* 99, 672–692.

Bali, T.G., S.J. Brown and M.O. Caglayan, 2011. Do hedge funds' exposures to risk factors predict their future returns? *Journal of Financial Economics* 101, 36–68.

Bali, T.G., S.J. Brown and M.O. Caglayan, 2012. Systematic risk and the cross section of hedge fund returns, *Journal of Financial Economics* 106, 114–131.

Bali, T.G., S.J. Brown and M.O. Caglayan, 2014. Macroeconomic risk and hedge fund returns, *Journal of Financial Economics* 114, 1–19.

Ben Dor, A., R. Jagannathan, I. Meier and Z. Xu, 2012. What drives the tracking error of hedge fund clones? *Journal of Alternative Investments* 15, 54–74.

Bollen, N.P.B., 2013. Zero-R^2 hedge funds and market neutrality, *Journal of Financial and Quantitative Analysis* 48, 519–547.

Bollen, N.P.B. and G.S. Fisher, 2013. Send in the clones? Hedge fund replication using futures contracts, *Journal of Alternative Investments* 16, 80–95.

Brown, S.J., 2016. Why hedge funds? *Financial Analyst Journal* 72, 5–7.

Carhart, M., 1997. On persistence of mutual fund performance, *Journal of Finance* 52, 57–82.

Duanmu, J., A. Malakhov and W.R. McCumber, 2018. Beta active hedge fund management, *Journal of Financial and Quantitative Analysis* 53, 2525–2558.

Efron, B., T. Hastie, I. Johnstone and R. Tibshirani, 2004. Least angle regression, *Annals of Statistics* 32, 407–499.

Fama, E.F. and K.R. French, 1993. Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

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

Fung, W. and D.A. Hsieh, 2004. Hedge fund benchmarks: A risk-based approach, *Financial Analysts Journal* 60, 65–80.

Getmansky, M., A.W. Lo and I. Makarov, 2004. An econometric model of serial correlation and illiquidity in hedge fund returns, *Journal of Financial Economics* 74, 529–609.

Giamouridis, D. and S. Paterlini, 2010. Regular(ized) hedge fund clones, *Journal of Financial Research* 3, 223–247.

Hasanhodzic, J. and A.W. Lo, 2007. Can hedge-fund returns be replicated? The linear case, *Journal of Investment Management* 5, 5–45.

Hoerl, A.E. and R. Kennard, 1970. Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics* 8, 27–51.

Jagannathan, R., A. Malakhov and D. Novikov, 2010. Do hot hands exist among hedge fund managers? An empirical evaluation, *Journal of Finance* 65: 217–255.

Jurek, J. and E. Stafford, 2015. The cost of capital for alternative investments, *Journal of Finance* 70, 2185–2226.

Liang, B., 2000. Hedge funds: The living and the dead, *Journal of Financial and Quantitative Analysis* 35, 309–326.

Liang, B. and H. Park, 2010. Predicting hedge fund failure: A comparison of risk measures, *Journal of Financial and Quantitative Analysis* 45, 199–222.

Marois, M.B., 2014. CalPERS to exit hedge funds, divest $4 billion stake, *Bloomberg: https://www.bloomberg.com/news/articles/2014-09-15/calpers-to-exit-hedge-funds-citing-expenses-complexity*.

Sharpe, W.F., 1992. Asset allocation: Management style and performance management, *Journal of Portfolio Management* 18, 7–19.

Sun, Z., A. Wang and L. Zheng, 2012. The road less traveled: Strategy distinctiveness and hedge fund performance, *Review of Financial Studies* 25, 96–143.

ter Horst, J.R., T.E. Nijman and F.A. de Roon, 2004. Evaluating style analysis, *Journal of Empirical Finance* 11, 29–53.

Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso, *Journal of the Royal Statistical Society B* 58, 267–288.

Titman, S. and C. Tiu, 2011. Do the best hedge funds hedge? *Review of Financial Studies* 24, 123–168.

Vrontos, S.D., I.D. Vrontos and D. Giamouridis, 2008. Hedge fund pricing and model uncertainty, *Journal of Banking and Finance* 32, 741–753.

Weber, V. and F. Peres, 2013. Hedge fund replication: Putting the pieces together, *Journal of Investment Strategies* 3, 61–119.

Figure 1

An Example of Hedge Fund and Clone Out-of-Sample Returns

The figure presents the out-of-sample comparison of an anonymous hedge fund and its clone, constructed according to our in-sample matching methodology. This hedge fund is in the "Multi-Strategy" self-reported style, it has an inception year of 2007, and it was active at the end of our study period. The out-of-sample comparison begins in 2012, after dropping the first two years of observations to control for the backfill bias and after using another two years for the in-sample clone matching.

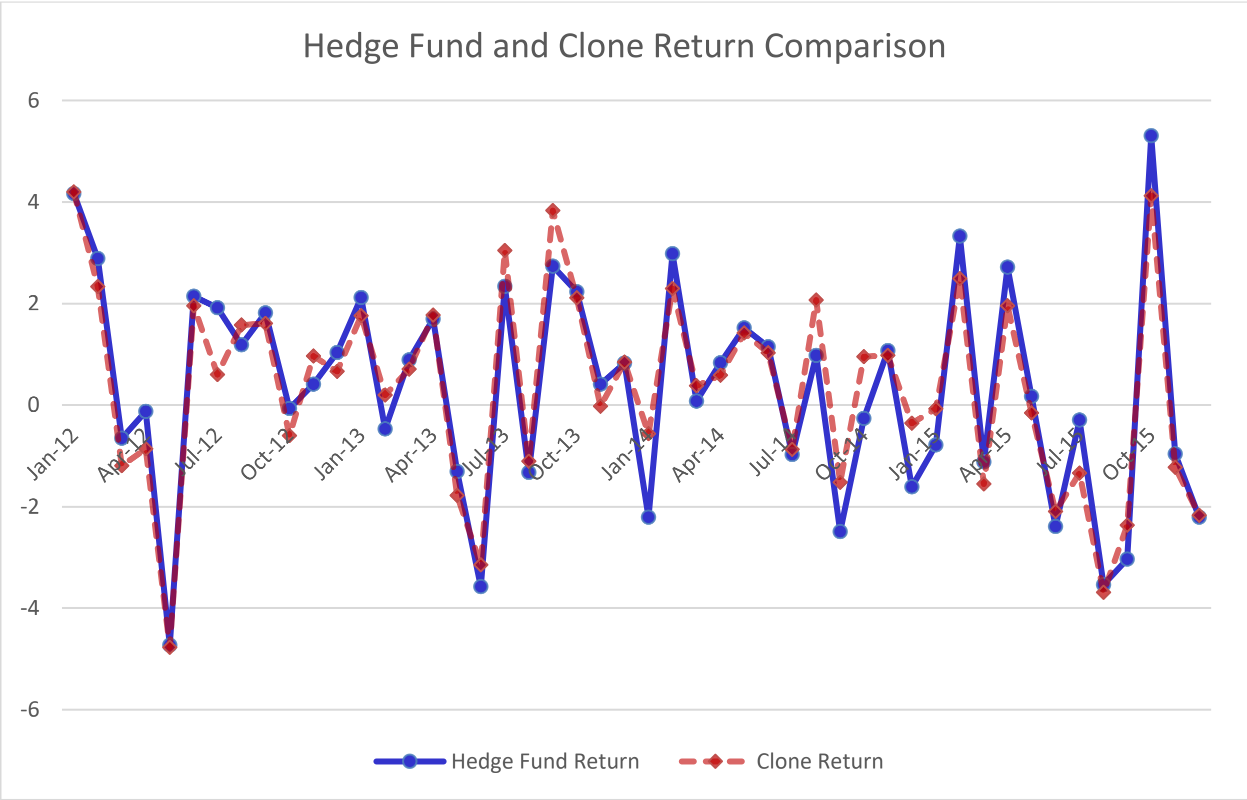

Table 1

Summary Statistics of Hedge Funds

Summary statistics of all hedge funds 1997–2015, reporting as of May 2016. Panel A reports returns, fees, investor liquidity measures and fund longevity. Panel B reports means of indicator variables for fund characteristics while Panel C reports self-declared fund styles. Significance at the 10%, 5% and 1% levels are designated by *, ** and ***, respectively.

Panel A

	Full Sample (6,822 unique funds)				
	Mean	Median	10th pct	90th pct	Std dev
Monthly Return	0.37	0.41	-0.44	1.14	1.32
Assets ($M)	264.61	62.47	6.59	489.36	1594.55
Min Invest ($M)	1.49	0.25	0.03	1	15.13
Mgmt Fee (%)	1.51	1.50	0.75	2	0.66
Perf Fee (%)	17.25	20	0	20	7.02
Hurdle Rate (%)	0.31	0	0	0	1.57
Lockup Period (days)	80.25	0	0	360	278.81
Redemption Notice (days)	14.05	0	0	45	27.76
Redemption Period (days)	53.09	30	30	90	53.80
Total Redemption (days)	67.87	44	30	135	64.69
Longevity (months)	81.50	70	37	144	44.33
	Active Funds (2,393 unique funds)				
	Mean	Median	10th pct	90th pct	Std dev
Monthly Return	0.46	0.49	-0.20	1.16	0.87
Assets ($M)	313.03	99.04	12.00	708.07	799.81
Min Invest ($M)	1.42	0.25	0.03	1	12.17
Mgmt Fee (%)	1.51	1.50	0.77	2	0.62
Perf Fee (%)	17.14	20	0	20	6.94
Hurdle Rate (%)	0.35	0	0	0	1.65
Lockup Period (days)	86.55	0	0	360	371.13
Redemption Notice (days)	21.61	2	0	65	32.49
Redemption Period (days)	49.39	30	7	90	51.50
Total Redemption (days)	71.51	55	16	150	66.85
Longevity (months)	97.66	84	45	176	49.55
	Inactive Funds (4,429 unique funds)				
	Mean	Median	10th pct	90th pct	Std dev
Monthly Return	0.33	0.36	-0.56	1.13	1.51
Assets ($M)	238.02	47.67	5.25	367.54	1,893.75
Min Invest ($M)	1.54	0.25	0.03	1	16.58
Mgmt Fee (%)	1.51	1.50	0.75	2	0.68
Perf Fee (%)	17.31	20	0	20	7.06
Hurdle Rate (%)	0.28	0	0	0	1.52
Lockup Period (days)	76.54	0	0	360	205.94
Redemption Notice (days)	9.79	0	0	30	23.65
Redemption Period (days)	55.27	30	30	90	55.00
Total Redemption (days)	65.73	35	30	120	63.30
Longevity (months)	72.77	62	35	129	38.51

Table 1 (*continued*)

Summary Statistics of Hedge Funds

Panel B - Indicator

	% of Funds			
	Full Sample	Active Funds	Inactive Funds	Active - Inactive
High Water Mark	81.84%	87.00%	79.05%	7.96%***
Hurdle Rate	4.27%	5.35%	3.68%	1.67%***
Offshore (non-US)	36.84%	37.48%	36.49%	1.00%
Closed to New Inv	5.32%	5.31%	5.33%	-0.02%
Liquidated	29.99%	0.00%	46.20%	-46.20%***
Acquired	2.59%	0.00%	4.00%	-4.00%***

Panel C - Fund Styles

	% of Funds			
	Full Sample	Active Funds	Inactive Funds	Active - Inactive
Long-Short	27.69%	29.04%	26.96%	2.08%*
Multi Strategy	10.73%	10.91%	10.63%	0.27%
Undisclosed	6.46%	0.04%	9.93%	-9.89%***
Market Neutral	6.44%	5.73%	6.82%	-1.09%*
Long Biased	5.69%	7.44%	4.74%	2.70%***
Systematic	4.94%	7.90%	3.34%	4.56%***
Discretionary	4.49%	3.59%	4.97%	-1.37%***
Fixed Income Diversified	4.22%	5.06%	3.77%	1.29%**
Discretionary Thematic	4.18%	4.72%	3.88%	0.84%
Emerging Market	3.27%	3.43%	3.18%	0.24%
Fixed Income Arbitrage	2.23%	2.01%	2.35%	-0.34%
Macro Diversified	2.08%	1.63%	2.33%	-0.70%**
Distressed Securities	2.01%	1.50%	2.28%	-0.78%**
Convertible Arbitrage	1.93%	1.63%	2.10%	-0.47%
Event Driven Diversified	1.73%	2.01%	1.58%	0.43%
Statistical Arbitrage	1.72%	0.63%	2.30%	-1.68%***
Systematic Diversified	1.63%	2.80%	0.99%	1.81%***
Cap Structure/Credit Arbitrage	1.44%	1.46%	1.42%	0.04%
Emerging Market Debt	1.25%	0.96%	1.40%	-0.44%*
Merger Arbitrage	1.23%	1.13%	1.29%	-0.16%
Equity Hedge Diversified	1.08%	1.34%	0.95%	0.39%
Asset-Backed Securities	0.94%	1.71%	0.52%	1.19%***
Currency	0.89%	1.00%	0.84%	0.17%
Mortgage-Backed	0.76%	1.17%	0.54%	0.63%**
Special Situation	0.66%	0.96%	0.50%	0.46%**
Short Biased	0.21%	0.04%	0.29%	-0.25%***
Activist	0.12%	0.17%	0.09%	0.08%

Table 2

LASSO Matching Regression Results

LASSO matching regression results are reported. Regressions are run over a 24-month window. ETFs available represent all ETFs available for LASSO regressions, while ETFs selected represent ETFs that are selected by LASSO as regressors for individual hedge funds. LASSO adjusted R^2, BIC and number of matched LASSO regressors are reported for each matching window. Coverage Ratio is Number of ETFs Available over Number of Hedge Funds; Selection Ratio is Number of ETFs Selected over Number of Hedge Funds.

Year	Number of Hedge Funds	Number of ETFs Available	Number of ETFs Selected	Coverage Ratio	Selection Ratio	Adj. R^2	BIC	Number of Regressors	Adj. R^2	BIC	Number of Regressors
1997-1998	287	19	19	0.07	0.07	0.34	63.42	2.10			
1998-1999	381	19	19	0.05	0.05	0.33	68.02	2.22			
1999-2000	499	29	29	0.06	0.06	0.34	65.48	2.39			
2000-2001	653	30	30	0.05	0.05	0.30	62.65	2.07			
2001-2002	830	75	66	0.09	0.08	0.36	51.08	2.35			
2002-2003	1099	97	76	0.09	0.07	0.35	49.75	2.32			
2003-2004	1331	107	93	0.08	0.07				0.46	42.21	2.77
2004-2005	1629	119	93	0.07	0.06				0.45	34.29	2.59
2005-2006	1993	153	106	0.08	0.05				0.47	32.91	2.72
2006-2007	2310	201	113	0.09	0.05				0.45	38.39	2.86
2007-2008	2381	332	117	0.14	0.05				0.47	54.72	2.84
2008-2009	2598	539	129	0.21	0.05				0.49	60.83	2.94
2009-2010	2913	680	129	0.23	0.04				0.54	51.78	3.16
2010-2011	3116	786	131	0.25	0.04				0.56	43.93	3.08
2011-2012	3031	937	130	0.31	0.04				0.50	43.27	2.73
2012-2013	3032	1078	141	0.36	0.05				0.51	38.82	2.88
2013-2014	2755	1196	137	0.43	0.05				0.47	36.06	2.77
Average				0.16	0.05	0.34	60.07	2.24	0.49	43.38	2.85

Table 3

Out-of-Sample Individual Matches

Summary statistics of out-of-sample individual matching of hedge funds and clones are reported. Attrition rate, mean tracking error (monthly, in %) and tracking error volatility (monthly, in %) are reported for each one-year predicting window.

Year	Number of ETFs Available	Number of Hedge Funds		Attrition Rate	Tracking Error		Tracking Error	
		Start	End		Mean	Volatility	Mean	Volatility
1999	19	287	278	3.14%	-1.06	4.76		
2000	19	381	372	2.36%	-0.57	4.95		
2001	29	499	491	1.60%	-0.74	4.60		
2002	30	653	618	5.36%	-0.26	3.90		
2003	75	830	782	5.78%	-1.37	3.38		
2004	97	1099	1020	7.19%	-0.06	2.97		
2005	107	1331	1237	7.06%			0.03	2.68
2006	119	1629	1497	8.10%			-0.03	2.46
2007	153	1993	1786	10.39%			-0.16	2.89
2008	201	2310	1851	19.87%			0.07	5.31
2009	332	2381	2029	14.78%			-0.95	4.29
2010	539	2598	2270	12.63%			-0.23	3.49
2011	680	2913	2465	15.38%			0.35	3.49
2012	786	3116	2618	15.98%			0.13	2.87
2013	937	3031	2502	17.45%			-0.27	2.98
2014	1078	3032	2513	17.12%			0.29	2.99
2015	1196	2755	2206	19.93%			0.22	3.40
Average				10.83%	-0.68	4.09	-0.05	3.35

Table 4

Matched ETF-Style Distribution and the Rate of Persistence

Matched ETF-style distribution and the rate of persistence are reported. Based on the style category from MorningStar, ETFs are identified into the following styles: Allocation, Alternative, Commodities, Equity, Fixed Income and Tax Preferred. The number of LASSO matched ETFs in each style for every window is reported. In each window, the rate of persistence is the percentage of overlapped ETFs from the previous window. The rate of persistence is reported in parentheses.

Year	Number of ETFs Available	Number of ETFs Selected	Allocation	Alternative	Commodities	Equity	Fixed Income	Tax Preferred
1997-1998	19	19				19 (0.00%)		
1998-1999	19	19				19 (100%)		
1999-2000	29	29				29 (65.52%)		
2000-2001	30	30				30 (96.67%)		
2001-2002	75	66				66 (45.45%)		
2002-2003	97	76				76 (72.37%)		
2003-2004	107	93				89 (80.90%)	4 (0.00%)	
2004-2005	119	93				89 (84.27%)	4 (50.00%)	
2005-2006	153	106			1 (0.00%)	100 (68.00%)	5 (80.00%)	
2006-2007	201	113		1 (0.00%)	1 (0.00%)	108 (63.89%)	3 (100.00%)	
2007-2008	332	117		10 (10.00%)	5 (20.00%)	99 (46.46%)	3 (66.67%)	
2008-2009	539	129	2 (0.00%)	31 (16.13%)	11 (9.09%)	77 (49.35%)	7 (28.57%)	1 (0.00%)
2009-2010	680	129	3 (0.00%)	51 (29.41%)	20 (30.00%)	48 (41.67%)	5 (20.00%)	2 (0.00%)
2010-2011	786	131	2 (50.00%)	55 (49.09%)	26 (57.69%)	43 (23.26%)	5 (20.00%)	
2011-2012	937	130	2 (50.00%)	66 (56.06%)	18 (66.67%)	41 (29.27%)	3 (33.33%)	
2012-2013	1078	141		70 (52.86%)	16 (31.25%)	47 (29.79%)	8 (0.00%)	
2013-2014	1196	137		66 (66.67%)	16 (68.75%)	48 (31.25%)	7 (42.86%)	
Average Number of ETFs			2.25	43.75	12.67	60.47	4.91	1.50
Average Rate of Persistence			25.00%	35.03%	31.49%	54.59%	40.13%	0.00%

Table 5

Cloneable and Noncloneable Funds: Matching Regression Results

Summary statistics of in-sample matching regressions are reported. LASSO Adj. R^2, BIC, number of matched LASSO regressors, skewness, kurtosis and VaR are reported for each matching window. Panel A reports the matches with LASSO Adj. R^2 on the top quintile. Panel B reports the matches with LASSO Adj. R^2 on the bottom quintile.

Panel A: In-Sample Matches, Cloneable Funds (Top R^2 Quintile)

Year	Number of ETFs Available	Number of Hedge Funds	Adj. R^2	BIC	Number of Regressors	Skewness	Kurtosis	VaR	Adj. R^2	BIC	Number of Regressors	Skewness	Kurtosis	VaR
1997-1998	19	58	0.73	53.95	3.45	-0.37	2.08	8.06%						
1998-1999	19	77	0.69	65.28	3.90	-0.12	2.01	8.13%						
1999-2000	29	100	0.67	66.11	3.85	0.42	0.76	8.16%						
2000-2001	30	131	0.59	64.86	3.06	0.35	1.72	7.83%						
2001-2002	75	166	0.66	52.90	3.58	-0.17	0.75	6.51%						
2002-2003	97	220	0.71	42.89	3.61	-0.03	0.51	5.19%						
2003-2004	107	267							0.76	25.95	4.08	0.19	0.63	3.58%
2004-2005	119	326							0.76	21.05	3.86	-0.06	0.14	3.94%
2005-2006	153	399							0.79	24.90	4.19	-0.05	0.13	3.96%
2006-2007	201	462							0.73	28.28	4.33	-0.01	0.47	3.61%
2007-2008	332	477							0.77	45.64	4.13	-0.63	1.59	8.99%
2008-2009	539	520							0.78	55.17	4.25	-0.36	0.92	10.38%
2009-2010	680	583							0.86	40.18	4.43	0.06	0.57	6.82%
2010-2011	786	624							0.88	28.98	4.32	-0.19	0.33	7.79%
2011-2012	937	607							0.80	31.99	3.78	-0.55	0.89	6.82%
2012-2013	1078	607							0.80	25.13	4.16	-0.41	1.23	4.55%
2013-2014	1196	551							0.76	20.64	3.90	-0.13	0.27	4.78%
Average			0.68	57.67	3.57	0.01	1.30	7.31%	0.79	31.63	4.13	-0.19	0.65	5.93%

Panel B: In-Sample Matches, Noncloneable Funds (Bottom R^2 Quintile)

Year	Number of ETFs Available	Number of Hedge Funds	Adj. R^2	BIC	Number of Regressors	Skewness	Kurtosis	VaR	Adj. R^2	BIC	Number of Regressors	Skewness	Kurtosis	VaR
1997-1998	19	58	0.03	73.28	1.10	-0.19	1.66	6.39%						
1998-1999	19	77	0.04	69.42	1.12	-0.05	2.10	6.94%						
1999-2000	29	100	0.05	52.03	1.16	0.26	1.23	6.70%						
2000-2001	30	131	0.06	55.91	1.16	0.24	1.77	6.30%						
2001-2002	75	166	0.10	53.87	1.38	0.07	1.10	5.40%						
2002-2003	97	220	0.06	59.32	1.20	0.03	0.86	4.66%						
2003-2004	107	267							0.16	51.83	1.57	0.21	0.97	4.04%
2004-2005	119	326							0.13	42.12	1.43	0.09	0.82	4.01%
2005-2006	153	399							0.14	33.66	1.51	0.01	0.65	3.77%
2006-2007	201	462							0.16	48.23	1.63	-0.07	1.02	3.89%
2007-2008	332	477							0.14	53.89	1.58	-0.38	1.67	6.70%
2008-2009	539	520							0.15	58.87	1.69	-0.18	1.58	7.48%
2009-2010	680	583							0.18	58.24	1.76	0.13	1.21	5.47%
2010-2011	786	624							0.18	51.47	1.75	-0.12	1.08	5.83%
2011-2012	937	607							0.16	49.81	1.67	-0.15	1.36	5.62%
2012-2013	1078	607							0.18	52.86	1.68	-0.02	1.23	4.66%
2013-2014	1196	551							0.18	52.21	1.72	-0.02	1.04	4.90%
Average			0.06	60.64	1.19	0.06	1.45	6.07%	0.16	50.29	1.64	-0.05	1.15	5.12%

Table 6

Cloneable and Noncloneable Funds: Out-of-Sample Individual Matches

Summary statistics of out-of-sample individual matching of hedge funds and clones formed on the basis of LASSO Adj. R^2 are reported. Attrition rate, mean tracking error (monthly, in %), tracking error volatility (monthly, in %), skewness, kurtosis and VaR are reported for each one-year predicting window. Panel A reports the matches with LASSO Adj. R^2 on the top quintile. Panel B reports the matches with LASSO Adj. R^2 on the bottom quintile.

Panel A: Out-of-Sample Matches, Cloneable Funds (Top R^2 Quintile)

Year	Number of Hedge Funds		Attrition Rate	Tracking Error		Skewness		Kurtosis		VaR		Tracking Error		Skewness		Kurtosis		VaR	
	Start	End		Mean	Volatility	Hedge Funds	Clones	Hedge Funds	Clones	Hedge Funds	Clones	Mean	Volatility	Hedge Funds	Clones	Hedge Funds	Clones	Hedge Funds	Clones
1999	58	57	1.72%	-0.86	4.39	0.38	0.23	-0.08	-0.65	6.38%	4.64%								
2000	77	76	1.30%	-0.86	4.98	0.39	0.49	0.34	0.36	9.02%	8.00%								
2001	100	99	1.00%	-0.60	4.90	-0.08	-0.41	0.17	-0.32	7.93%	10.64%								
2002	131	125	4.58%	-0.09	4.07	-0.02	0.10	-0.05	-0.40	8.27%	7.05%								
2003	166	157	5.42%	-1.03	3.47	0.25	0.06	-0.01	-0.35	4.10%	3.52%								
2004	220	205	6.82%	0.26	2.54	-0.02	-0.52	0.48	0.93	4.20%	3.32%								
2005	267	253	5.24%									0.09	2.00	-0.21	-0.24	-0.37	-0.95	3.92%	3.70%
2006	326	305	6.44%									0.19	1.93	-0.10	-0.48	0.29	0.99	3.77%	2.91%
2007	399	362	9.27%									-0.07	2.71	-0.20	-0.15	0.08	-0.12	4.80%	4.03%
2008	462	392	15.15%									0.34	5.18	-0.54	-0.47	0.77	0.18	14.47%	13.01%
2009	477	409	14.26%									-0.94	4.15	0.02	-0.06	0.70	0.29	7.42%	6.78%
2010	520	468	10.00%									0.30	3.26	-0.11	-0.21	-0.04	-0.54	8.24%	5.49%
2011	583	494	15.27%									0.41	3.20	-0.05	-0.17	0.50	1.06	9.86%	10.42%
2012	624	535	14.26%									0.16	2.27	-1.10	-1.33	2.49	3.14	5.88%	5.51%
2013	607	492	18.95%									-0.06	2.39	-0.11	-0.19	0.01	-0.12	4.01%	3.26%
2014	607	548	9.72%									0.53	2.47	0.00	-0.26	0.16	-0.12	5.56%	4.34%
2015	551	462	16.15%									0.30	2.76	0.19	0.30	0.23	0.16	7.52%	6.12%
Average			9.15%	-0.53	4.06	0.15	-0.01	0.14	-0.07	6.65%	6.19%	0.11	2.94	-0.20	-0.30	0.44	0.36	6.86%	5.96%

Panel B: Out-of-Sample Matches, Noncloneable Funds (Bottom R^2 Quintile)

Year	Number of Hedge Funds		Attrition Rate	Tracking Error		Skewness		Kurtosis		VaR		Tracking Error		Skewness		Kurtosis		VaR	
	Start	End		Mean	Volatility	Hedge Funds	Clones	Hedge Funds	Clones	Hedge Funds	Clones	Mean	Volatility	Hedge Funds	Clones	Hedge Funds	Clones	Hedge Funds	Clones
1999	58	52	10.34%	0.15	5.37	-0.27	0.31	0.79	1.53	8.62%	0.65%								
2000	77	75	2.60%	-0.16	5.05	0.38	-0.07	0.74	0.50	7.87%	0.98%								
2001	100	100	0.00%	-0.53	3.49	0.03	0.09	0.84	-0.03	4.93%	1.15%								
2002	131	121	7.63%	-0.61	3.42	-0.21	0.03	0.92	-0.38	4.87%	1.22%								
2003	166	156	6.02%	-1.43	3.05	0.28	0.01	0.70	-0.39	3.35%	1.09%								
2004	220	205	6.82%	-0.34	3.46	0.07	-0.10	0.27	0.73	4.97%	0.89%								
2005	267	241	9.74%									-0.15	3.03	-0.04	-0.19	0.10	-0.44	4.43%	1.41%
2006	326	298	8.59%									-0.21	2.96	0.12	-0.29	0.67	0.53	4.08%	0.69%
2007	399	338	15.29%									-0.36	2.77	-0.08	-0.08	0.57	-0.13	3.67%	0.80%
2008	462	339	26.62%									-0.57	5.35	-0.21	-0.44	0.96	0.44	8.27%	3.89%
2009	477	395	17.19%									-0.68	3.63	0.18	0.05	0.82	0.53	4.93%	2.25%
2010	520	436	16.15%									-0.64	3.49	-0.03	-0.13	0.61	-0.44	4.94%	1.46%
2011	583	475	18.52%									0.33	3.80	-0.06	-0.09	0.53	0.07	6.62%	2.84%
2012	624	496	20.51%									0.16	3.17	0.04	-0.19	0.92	0.70	5.06%	1.44%
2013	607	499	17.79%									-0.31	3.12	-0.16	-0.16	0.63	0.24	4.61%	1.53%
2014	607	463	23.72%									-0.14	3.96	0.09	-0.14	0.82	0.15	6.10%	1.65%
2015	551	409	25.77%									0.25	3.92	0.07	0.04	0.55	0.56	6.45%	2.38%
Average			13.73%	-0.49	3.97	0.05	0.05	0.71	0.33	5.77%	1.00%	-0.21	3.56	-0.01	-0.15	0.65	0.20	5.38%	1.85%

Table 7

Fund Characteristics, Cloneable versus Noncloneable Funds

Summary statistics of cloneable and noncloneable hedge funds formed on the basis of LASSO Adj. R^2 are reported. The table reports the summary statistics of hedge funds in the top and bottom quintile of LASSO Adj. R^2. Significance at the 10%, 5% and 1% levels are designated by *, ** and ***, respectively.

	Cloneable Funds	Noncloneable Funds	Diff t-stat
Assets ($M)	213.83	343.20	-1.58
Min Invest ($M)	0.85	1.55	-2.33**
Mgmt Fee (%)	1.36	1.63	-11.25***
Perf Fee (%)	16.07	18.59	-10.71***
Hurdle Rate (%)	0.38	0.18	5.17***
Lockup Period (days)	102.62	85.09	1.40
Redemption Notice (days)	10.48	9.79	0.49
Redemption Period (days)	69.14	57.84	2.13**
Total Redemption (days)	80.39	68.20	2.53**
High Water Mark	0.76	0.82	-1.86*
Hurdle Rate	0.05	0.03	4.91***
Offshore (non-US)	0.32	0.17	2.87***
Directional Traders	0.18	0.34	-7.58***
Relative Value	0.10	0.21	-8.01***
Security Selection	0.49	0.20	12.09***
Multiprocess	0.15	0.16	-0.68

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Financial Review*, published by Wiley on behalf of The Eastern Finance Association. Copyright restrictions may apply. https://doi.org/10.1111/fire.12221.

Table 8

Comparisons of Hedge Fund Portfolios and Clone Portfolios

Comparisons of hedge funds portfolios and clone portfolios 1999–2015 are reported. Portfolios are formulated as of December 31, 1998, and rebalanced annually. Annual returns and cumulative risk-adjusted performances are reported. End value is as of December 31, 2015. Skewness reports the mean skewness of out-of-sample portfolio net returns for one-year predicting window. Significance at the 10%, 5% and 1% levels are designated by *, ** and ***, respectively.

Year	Number of ETFs Available	Adj. R^2	Annual Return		Annual Return		Annual Return	
			Hedge Funds	Clones	Hedge Funds	Clones	Hedge Funds	Clones
1999	19	0.344	26.62	9.38	26.62	9.38		
2000	19	0.329	8.45	0.80	8.45	0.80		
2001	29	0.342	7.73	-2.79	7.73	-2.79		
2002	30	0.296	3.65	0.66	3.65	0.66		
2003	75	0.362	24.74	6.79	24.74	6.79		
2004	97	0.354	10.57	10.40	10.57	10.40		
2005	107	0.465	8.46	8.48			8.46	8.48
2006	119	0.450	14.25	14.25			14.25	14.25
2007	153	0.467	12.26	9.96			12.26	9.96
2008	201	0.454	-16.36	-16.34			-16.36	-16.34
2009	332	0.475	24.16	10.48			24.16	10.48
2010	539	0.487	10.55	8.46			10.55	8.46
2011	680	0.543	-6.64	-2.54			-6.64	-2.54
2012	786	0.555	5.25	7.43			5.25	7.43
2013	937	0.498	9.12	6.18			9.12	6.18
2014	1078	0.510	-1.44	2.45			-1.44	2.45
2015	1196	0.473	-4.67	-1.62			-4.67	-1.62
End Value			3.40	1.95	2.11	1.27	1.61	1.53
Monthly Return			0.47***	0.19	0.82***	0.10	0.28	0.24
(t-stat)			(3.13)	(1.51)	(3.59)	(0.64)	(1.44)	(1.36)
alpha			0.22***	-0.02	0.50***	-0.04	0.10	0.04
(t-stat)			(2.88)	(-0.26)	(5.42)	(-0.55)	(1.15)	(0.53)
Sharpe Ratio			0.22	0.11	0.42	0.08	0.12	0.12
Sortino Ratio			0.32	0.13	1.05	0.10	0.18	0.15
Info Ratio			0.21	-0.02	0.55	-0.07	0.10	0.05
Skewness			-0.20	-1.00	0.80	-0.70	-0.46	-0.99
Attrition Rate			10.83%		4.24%		14.43%	
Mean Adj. R^2			0.435		0.338		0.489	

Table 9

Cloneable and Noncloneable Funds: Portfolio Comparisons, 1999–2015

Annual returns and cumulative risk-adjusted performances of portfolios 1999–2015 formed on the basis of LASSO Adj. R^2. Portfolios of hedge funds and clones are formed as December 31, 1998, and rebalanced annually for funds in the top and bottom quintile of LASSO Adj. R^2. End value is as of December 31, 2015. Skewness reports the mean skewness of out-of-sample portfolio net returns for one-year predicting window. Significance at the 10%, 5% and 1% levels are designated by *, ** and ***, respectively.

	Cloneable Funds, Top R^2 Quintile			Noncloneable Funds, Btm R^2 Quintile		
Year	Adj. R^2	Annual Return		Adj. R^2	Annual Return	
		Hedge Funds	Clones		Hedge Funds	Clones
1999	0.733	31.56	18.74	0.026	4.59	4.24
2000	0.694	4.13	-5.96	0.042	9.47	5.29
2001	0.673	0.70	-8.05	0.054	9.77	2.91
2002	0.594	-3.69	-5.10	0.057	9.19	2.27
2003	0.657	35.61	21.52	0.096	19.71	1.33
2004	0.712	12.73	16.35	0.059	6.79	2.68
2005	0.765	10.45	11.46	0.157	8.07	5.79
2006	0.757	15.31	18.40	0.132	10.95	8.55
2007	0.795	17.08	15.45	0.138	11.35	6.84
2008	0.731	-27.89	-25.45	0.157	1.03	-4.37
2009	0.767	36.57	21.37	0.142	8.68	0.84
2010	0.783	10.96	15.99	0.149	8.64	1.24
2011	0.863	-10.75	-6.69	0.180	-4.49	0.30
2012	0.882	9.58	11.57	0.176	-0.07	2.69
2013	0.803	13.46	13.28	0.160	5.98	0.89
2014	0.803	-4.61	1.75	0.179	3.11	1.84
2015	0.757	-8.65	-5.17	0.176	-3.25	0.50
End Value		3.20	2.51		2.82	1.53
Monthly Return		0.47**	0.35		0.36***	0.06
(t-stat)		(2.02)	(1.50)		(3.58)	(1.50)
α		0.08	-0.03		0.22***	0.01
(t-stat)		(0.81)	(-0.37)		(2.76)	(0.38)
Sharpe Ratio		0.14	0.11		0.25	0.10
Sortino Ratio		0.19	0.14		0.35	0.10
Info Ratio		0.06	-0.03		0.21	0.03
Skewness		-0.42	-0.74		0.12	-1.09
Attrition Rate		9.15%			13.73%	
Mean Adj. R^2		0.751			0.122	

Table 10

Cloneable and Noncloneable Funds: Portfolio Comparisons, 1999–2004 and 2005–2015

Annual returns and cumulative risk-adjusted performances of portfolios 1999–2015 formed on the basis of LASSO Adj. R^2. Portfolios of hedge funds and clones are formed as December 31, 1998, and rebalanced annually for funds in the top and bottom quintile of LASSO Adj. R^2. End value is as of December 31, 2015. Skewness reports the mean skewness of out-of-sample portfolio net returns for one-year predicting window. Panel A reports the comparisons of performances 1999–2004. Panel B reports the comparisons of performances 2005–2015. Significance at the 10%, 5% and 1% levels are designated by *, ** and ***, respectively.

Panel A: Year 1999 to 2004

	Cloneable Funds, Top R^2 Quintile			Noncloneable Funds, Btm R^2 Quintile		
Year	Adj. R^2	Annual Return		Adj. R^2	Annual Return	
		Hedge Funds	Clones		Hedge Funds	Clones
1999	0.733	31.56	18.74	0.026	4.59	4.24
2000	0.694	4.13	-5.96	0.042	9.47	5.29
2001	0.673	0.70	-8.05	0.054	9.77	2.91
2002	0.594	-3.69	-5.10	0.057	9.19	2.27
2003	0.657	35.61	21.52	0.096	19.71	1.33
2004	0.712	12.73	16.35	0.059	6.79	2.68
End Value		2.03	1.38		1.75	1.20
Monthly Return		0.79**	0.25		0.56***	0.01
(t-stat)		(2.24)	(0.67)		(2.58)	(0.55)
α		0.43***	-0.06		0.24	-0.01
(t-stat)		(3.55)	(-0.51)		(1.61)	(-0.32)
Sharpe Ratio		0.27	0.08		0.31	0.06
Sortino Ratio		0.50	0.12		0.43	0.38
Info Ratio		0.42	-0.06		0.19	-0.04
Skewness		0.24	-0.51		-0.16	0.57
Attrition Rate	3.47%			5.57%		
Mean Adj. R^2	0.677			0.056		

Panel B: Year 2005 to 2015

	Cloneable Funds, Top R^2 Quintile			Noncloneable Funds, Btm R^2 Quintile		
Year	Adj. R^2	Annual Return		Adj. R^2	Annual Return	
		Hedge Funds	Clones		Hedge Funds	Clones
2005	0.765	10.45	11.46	0.157	8.07	5.79
2006	0.757	15.31	18.40	0.132	10.95	8.55
2007	0.795	17.08	15.45	0.138	11.35	6.84
2008	0.731	-27.89	-25.45	0.157	1.03	-4.37
2009	0.767	36.57	21.37	0.142	8.68	0.84
2010	0.783	10.96	15.99	0.149	8.64	1.24
2011	0.863	-10.75	-6.69	0.180	-4.49	0.30
2012	0.882	9.58	11.57	0.176	-0.07	2.69
2013	0.803	13.46	13.28	0.160	5.98	0.89
2014	0.803	-4.61	1.75	0.179	3.11	1.84
2015	0.757	-8.65	-5.17	0.176	-3.25	0.50
End Value		1.58	1.82		1.61	1.27
Monthly Return		0.30	0.41		0.26**	0.08
(t-stat)		(0.98)	(1.35)		(2.49)	(1.42)
α		-0.02	0.05		0.21***	0.04
(t-stat)		(-0.18)	(0.49)		(2.64)	(0.95)
Sharpe Ratio		0.09	0.12		0.21	0.12
Sortino Ratio		0.12	0.15		0.44	0.14
Info Ratio		-0.02	0.04		0.25	0.09
Skewness		-0.57	-0.83		0.15	-0.85
Attrition Rate	12.25%			18.17%		
Mean Adj. R^2	0.791			0.159		