Machine Learning in Hedge Fund Classification: Systematic vs. Discretionary Strategies and Their Performance Implications

Hui-Ching Chuang, Chung-Ming Kuan

Department of Statistics, National Taipei University

Department of Finance, National Taiwan University

Abstract

This paper applies machine learning to classify hedge funds into systematic and discretionary categories. Leveraging textual analysis and advanced methods, our approach eliminates subjective judgment in analyzing investment strategies. We find that systematically classified funds, on average, yield higher excess returns than discretionary ones. Additionally, after applying the false discovery rate test for linear asset pricing models, a higher portion of positive alpha is observed in the systematic category. The alpha average for outperforming systematic funds surpasses that of discretionary funds across various risk factor models.

Keywords: Machine learning; Textual analysis; False discovery rate; Fund performance
JEL: C63; G11; G14; G23

*We would like to express our appreciation for the constructive comments and valuable insights provided by anonymous referees, Po-Hsuan Hsu, O-Chia Chuang, Rui-Chung Yang, and the participants of the Quantitative Finance Workshop 3: Asset Pricing and Risk Management, IMS, Singapore; and the 26th Conference on the Theories and Practices of Securities and Financial Markets.
**Corresponding author: Hui-Ching Chuang. Address: Department of Statistics, National Taipei University. No. 151, University Rd., Sanxia Dist., New Taipei City, Taiwan 237. Phone: 886-9727-35021. We gratefully acknowledge the financial support from the National Science and Technology Council, Taiwan.

Email addresses: hcchuang@gm.ntpu.edu.tw (Hui-Ching Chuang), ckuan@ntu.edu.tw (Chung-Ming Kuan)

1. Introduction

Investment strategies are complex decision processes involving quantitative and qualitative market information assessments. Such strategies play a crucial role in hedge fund performance. A fund's investment strategy is usually disclosed in its private placement memorandum or fund prospectus. Fund managers tend to avoid specific descriptions of their strategies to permit investment flexibility. As such, it is not straightforward to categorize hedge funds based on their disclosed statements. On the other hand, the Hedge Fund Research (HFR) database offers its fund classification systems; for example, HFR strategy classification in 2017 includes five major categories: Equity Hedge, Event-Driven, Macro, Relative Value and Fund of Funds, each with several sub-strategy groups. While the HFR classification provides valuable information about the fund characteristics, much effort is still needed if one would like to categorize hedge funds under different criteria.

With the advancement of analytical tools and computational technology, more fund managers now rely on models, algorithms, and various learning methods to make investment decisions. Thus, it would be interesting to classify hedge funds into "systematic" and "discretionary" funds and study how these two groups of funds perform in practice. By systematic funds, we refer to the funds with strategies depending mainly on quantitative models without human intervention by discretionary funds we refer to those require primarily managers' professional skills and experience. Such classification is also in line with the HFR categorization for the sub-categories of Macro funds: Systematic Diversified funds and Discretionary Thematic funds.[1] Similarly, hedge/mutual

[1]HFR defines Systematic Diversified funds as funds with "investment processes that typically are functions of mathematical, algorithmic and technical models, with little or no influence from individuals over the portfolio positioning," and Discretionary Thematic funds are those "primarily reliant on the evaluation of market data, relationships, and influences, as interpreted by an individual or group of individuals who make decisions on portfolio positions."

funds are classified as "man" and "machine" in Harvey et al. (2017) and "quantitative" and "discretionary" ("non-quantitative") in Abis (2022) (Beggs et al., 2021), or "quantitative" and "fundamental" in Evans, Rohleder, Tentesch and Wilkens (2023).

In this paper, we first introduce an approach to building classifiers that bifurcate hedge funds into systematic and discretionary funds. This approach makes use of textual analysis and statistical learning methods. Specifically, we convert strategy descriptions into numeric data and extract features from these data. Taking the features of Systematic Diversified and Discretionary Thematic funds as the training sample, classifiers are trained using various learning methods, such as linear discriminant analysis, k-nearest neighbor, support vector machine, classification tree, random forest, and gradient boosting; see Hastie, Tibshirani and Friedman (2009) and James, Witten, Hastie and Tibshirani (2013) for details. With the resulting classifiers, funds in other categories can be classified in conformity with the strategy features of Systematic Diversified/Discretionary Thematic funds. This approach is similar to that applied by Abis (2022), Beggs et al. (2021), and Harvey et al. (2017), differing mainly in that we use the guidance from the HFR's internal expert classification as a training label, and do not require selecting keywords to classify funds.

We then compare whether positive alpha funds (we use outperforming funds for positive alpha funds interchangeably) belong to the systematic funds (or discretionary funds) and the magnitude of their performance difference. Examining the positiveness of thousands of individual funds' alpha is a multiple-testing question. Multiple testing is easy to suffer from the data-snooping bias, i.e., we are likely to identify the outperforming funds purely due to chance. The false discovery fallacy is critical when searching for positive alpha funds. Benjamini and Hochberg (1995) (BH) are pioneers proposing the test to examine the multiple hypotheses while controlling the false discovery rate (FDR), which is defined as the expected value of the false rejected number to the rejected number of the hypotheses. FDR and related multiple testing approaches have attracted more and more finance researchers recently, see, e.g., Harvey et al. (2020), Chordia et al. (2020), Hsu et al. (2023) and others.[2]

We evaluate fund performance using the FDR-based test proposed by Giglio et al. (2021) to test multiple alphas in the linear asset pricing models. Under the conventional Fama-MacBeth two-pass regression framework, Giglio et al. (2021) propose a rigorous multiple-test framework that accommodates missing data and omitted risk factors. Hedge fund data is known for its short life span (unbalanced panel return structure, missing values), herding trade (cross-sectional dependence), and highly nonlinear payoff structures (possibly the existence of latent risk factors). These characteristics and the generated variable bias from the two-pass procedure threaten the underlying independence assumptions of Benjamini and Hochberg (1995) test. Giglio et al. (2021) propose the adjustment to the conventional two-pass methods and FDR test, which mitigate those issues' threat to the validity of BH test and further improve the power of the test while maintaining the FDR control.

Our classification task includes the two sub-strategies of the Macro fund (training sample) and four sub-strategies of the Equity Hedge funds (testing sample) in the HFR database. We construct 3,494 common textual features in both training and testing samples. We apply various machine/statistical learning approaches to 85% of the training sample and use the ten-fold nested cross-validation framework to select the algorithm's hyper-parameters and find the best classifier. Results show that the best classifier, the random forest, yields as high as 90% , 88%, and 93% in

[2]Barras et al. (2010), Cuthbertson et al. (2012), Bajgrowicz and Scaillet (2012), Bajgrowicz et al. (2016) use the Bayesian FDR control test. Another strand of literature focuses on controlling the family-wise error rate (FWER), which is the probability of committing more than one false discovery, see White (2000), Hansen (2005), Romano and Wolf (2005), and Hsu and Kuan (2005). The applications of FWER control include the profitability of trading strategies: Kuang et al. (2014), Goyal and Wahal (2015). Harvey et al. (2016) and Chordia et al. (2020) also promote the FDP, false discovery proportion introduced in Romano and Wolf (2007), and Romano et al. (2008).

terms of accuracy, area under the ROC curve, and F1 scores. The random forest also performs consistent results on the hold-out 15% sample. It reaches 89%, 87%, and 92% in those classification scores. The random forest provides information on the feature's importance, ranking the feature's ability to predict the style of funds. The bi-gram (two consecutive words): *emerge market, fix income, investment process, invest opportunity* are the top essential features. This finding shows that the textual information in the trading strategies provides valuable and distinct key terms than the researchers commonly used words, such as machine, man, and quantitative, to classify Systematic and Discretionary funds.

We study the performance of those classified funds surviving at least 36 months from 1994 to 2015. We find that systematic funds have higher raw and factor-adjusted returns than those classified as discretionary. These results hold for all four sub-strategies of the Equity Hedge funds and are robust to one, three, five, and seven risk factor models. The FDR-based multiple alphas test shows consistent results. Overall, an average (across different factor models) of 4.27% more skilled funds as systematic funds than discretionary funds. The proportion of authentic positive alpha funds as systematic funds is 3.50% (0.77%) larger than that of authentic positive alpha funds as discretionary funds in Equity Hedge (Macro). Regarding the alpha value difference of the positive alpha funds, overall, the average alpha of the outperforming funds for systematic funds is higher than that of the discretionary ones. Take the Fung and Hsieh (1997) 7-factor model as an example; the average skilled funds' alpha of the discretionary funds is 85.4 bp per month, while the average outperforming funds' alpha of the systematic funds is 88.4 bp. The difference is 3.03 bp per month. The macro category generates a higher alpha difference of 14.70 bp, and the Equity Hedge is 1.51 bp. Our results show that systematic outperforming funds generate higher alphas than the outperforming discretionary funds, regardless of which factor model or categories we use in this research.

Our research makes two contributions to the literature. First, we propose a novel approach to classify the style of hedge fund investing strategies. The quantitative (non-quantitative) investment style of funds draws the researcher's attention to its impact on the market liquidity, tail risk, the economy of scale, and others. (Abis, 2022; Beggs et al., 2021; Evans et al., 2023) Our machine learning approach helps extract the textual information from the funds with well-defined classification styles and predict the less clearly defined styles of funds. It reduces researcher-dependant judgment effort while keeping the classification consistent with well-defined styles. Our tree-based random forest approach offers collateral benefits by providing information on the key terms essential to distinguish different fund styles that are overlooked by previous studies. Second, we identify the proportion of the authentic outperforming funds in systematic and discretionary funds using the test without data-snooping bias. Our results add to the research of Giglio et al. (2021) on the performance of Hedge fund's style investment and also give rigorous statistical evidence on the profitability of the systematic and discretionary funds in addition to Chincarini (2014), and Harvey et al. (2017).

This paper proceeds as follows. Section 2 discusses the methods for extracting features from fund strategy descriptions and for building fund classifiers from these features. In Section 3, we evaluate the performance of the classified systematic and discretionary funds and compare their overall performance via FDR tests. The last section concludes the paper.

2. Classification of Hedge Funds

In this section, we discuss our approach to training classifiers for systematic and discretionary funds, based on the documents of fund investment strategies. Following Harvey et al. (2017), we consider the two largest groups in the HFR classification system: Macro funds and Equity Hedge funds, where the former includes two sub-strategy groups (Systematic Diversified funds and Discretionary Thematic funds), and the latter contains four sub-strategy groups (Equity Market Neutral funds, Fundamental Growth funds, Fundamental Value funds, and Quantitative Directional funds).[3] Given that Macro funds have already been classified into systematic and discretionary funds, it is quite natural to use the information of Macro funds to train classifiers. All strategy descriptions are taken from the HFR database; we include the graveyard database to mitigate the survivorship bias.

We first follow standard practice in textual analysis to process the text of strategy descriptions. We exclude digital numbers, punctuation, symbols, and the stop-words (e.g., is, at, and, the) in all documents that are of little value for classification. The remaining words are then lemmatized, i.e., different forms of a word is converted to one single word, from which documents are tokenized based on "bigrams" (two consecutive words).[4] To ensure a bigram in Macro funds (the training sample) is also relevant in Equity Hedge funds; we set the ratio of the percentage of Equity Hedge funds with a particular bigram to the percentage of Macro funds with the same bigram to be greater than or equal to 0.2; see Harvey et al. (2017). This results in a total of 3,494 tokens.

We then construct the feature matrix of a given fund category as follows. For the token j in the document i, its "term frequency" (tf) is:

$$\mathrm{tf}_{ij} = \frac{\text{Number of times that token } j \text{ appear in the document } i}{\text{Total number of all tokens in the document } i}, \quad i = 1, \ldots, N, j = 1, \ldots, M,$$

[3]As Harvey et al. (2017), we ignore sector-specific funds and those with "multistrategy".

[4]We use the *R* package *textstem* for lemmatization and the *R* package *tidytext* for tokenization.

and every tf is weighted by the inverse-document frequency (idf):

$$\text{idf}_j = \log \frac{\text{Total number of documents}}{\text{Number of the documents that contain token } j}, \quad j = 1, \ldots, M,$$

where N is the number of funds in a category (Macro funds or Equity Hedge funds), and $M = 3,494$ is the number of tokens. Note that the larger the idf, the less frequently the token j is observed in these documents; such token is considered more informative for classification and hence receives more weight; see, e.g., Manning and Schütze (1999). The feature matrix is an $N \times M$ matrix with the (i, j)-th element:

$$f_{ij} = \text{tf}_{ij} \cdot \text{idf}_j, \quad i = 1, \ldots, N, \; j = 1, \ldots, M.$$

In our study, the feature matrix of Macro funds is $2,222 \times 3,494$, and that of Equity Hedge funds is $7,158 \times 3,494$.

We define the binary target variable as taking the value 1 if it is a Systematic Diversified fund and 0 if it is a Discretionary Thematic fund and the training sample are the feature matrix of Macro funds. The following statistical learning methods are employed: Linear regression, logistic regression, linear discriminant analysis (LDA), k-nearest neighbor (KNN), support vector machine (SVN) with the Gaussian kernel, classification tree, bagging, gradient boosting, as well as random forests.[5] For detailed discussions of these learning methods, we refer to Hastie et al. (2009) and James et al. (2013). Recall that, while Chincarini (2014) subjectively selects certain keywords for fund classification, Harvey et al. (2017) rely on the "systematic words" determined by the frequency count of single words. By contrast, our training approach utilizes text mining and

[5]These learning methods are implemented using various *R* packages: *stats* for linear and logistic regressions, *lda* for LDA, *caret* for KNN, *e1071* for SVM, *tree* for classification tree, *gbm* for gradient boosting, and *ranger* for bagging and random forest.

statistical learning and hence avoids subjectivity to a large extent. Moreover, the use of bigrams for tokenization also alleviates the problem of misinterpreting single words. Note that Abis (2022) and Beggs et al. (2021) adopt a similar approach to bifurcate mutual funds, yet the HFR's existing categories determine our training label.

We select the trained classifier with the best classification performance. To this end, we consider four performance measures: Accuracy, area under the receiver operating characteristic curve (AUC), precision, and F1 score; for details of these measures we refer to James et al. (2013) and Chinchor (1992). We evaluate the performance of classifiers using the nested 10-fold cross-validation. This cross-validation involves two layers: the inner 10-fold cross-validation determines the best hyper-parameters for each learning method, and the outer 10-fold cross-validation evaluates the classification ability of different classifiers with the best hyper-parameters. In our study, we split Macro funds into two sub-samples, one with 85% (1,888) funds and the other with 15% (334) funds, by the stratified sampling on the strata of the Systematic Diversified dummy. The nested 10-fold cross-validation is then applied to the sub-sample of 85% Macro funds to search for the best classifier; the remaining 15% of Macro funds is reserved for out-of-sample evaluation of the best classifier. An illustration of the nested cross-validation approach is given in the Appendix A with illustration.

Table 1 contains four panels, where each panel summarizes the performance results of all learning methods under a particular measure. We report the summary statistics (median, mean, maximum, minimum, and standard deviation) based on the outer 10-fold samples. It can be seen that random forest dominates other classifiers for all measures in terms of these statistics, with gradient boosting as the second-best classifier. On the other hand, linear regression, logistic regression, and LDA perform pretty poorly. For example, the mean of accuracy is 0.86 for random forest and 0.84 for gradient boosting. On the other hand, linear regression, logistic regression, and LDA have respective means of 0.70, 0.70, and 0.68. Applying the selected random forest with the best hyper-parameters to the 15% validation sample, the resulting out-of-sample accuracy, AUC, precision, and F1 score are, respectively, 0.89, 0.87, 0.90, and 0.92, which are all greater than the corresponding medians and means in Table 1.

Figure 1 illustrates the "importance" measure of Breiman (2001) for the top 30 features in the selected random forest classifier. It can be seen that the top five features (bigrams) are: *emerge market, fix income, investment process, invest opportunity*, and *macro economic.* These features are very different from those of Chincarini (2014) and Harvey et al. (2017). For example, the keywords used by Harvey et al. (2017) for classifying machine-based funds are: *algorithm, approx, computer, model, statistical*, and *system*, which do not appear in our list of leading features. Although the top features here may not be as intuitive as one would like, they are helpful for classification. This shows that our approach can capture some characteristics in fund strategies that conventional approaches overlook.

3. Comparison of Fund Performance

We assess the relative performance of these funds by comparing their excess returns and alphas derived from various factor models and across the different categorized groups. In line with prior studies on hedge funds, our analysis is limited to funds that report monthly returns and adopt the "Net of All Fees" reporting style, as seen in Cao et al. (2013). Furthermore, we narrowed our dataset to include only those funds with at least 36 consecutive monthly returns to ensure robust regression outcomes. The final dataset for our performance evaluation consists of 4,051 Equity Hedge funds and 1,194 Macro funds from January 1994 to November 2015.

3.1. Performance Based on Excess Returns and Alphas

We consider the following factor model:

$$E(r_i) = \alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\lambda}, \quad i = 1, \cdots, N, \tag{1}$$

where r_i is the excess return of fund i (excess of the risk-free rate), α_i is the pricing error (alpha) of fund i, $\boldsymbol{\beta}_i$ is the $S \times 1$ vector of risk exposures to S risk factors, and $\boldsymbol{\lambda}$ is the $S \times 1$ vector factor risk premia (reward for risk exposure), and N is the number of funds; see, e.g., Cochrane (2009). A fund is considered superior if its alpha is greater than zero, suggesting positive abnormal return.

Assuming that all common risk factors are observable, we apply the conventional two-pass regression to estimate alpha. First, we conduct time-series regression of each fund's excess return on risk factors:

$$r_{i,t} = a_i + \boldsymbol{\beta}_i'\mathbf{f}_t + \epsilon_{i,t}, \quad t = 1, \cdots, T,$$

where $\mathbf{f}_t$ is the $S \times 1$ vector of risk factors at time t. This yields the risk exposure estimates for N funds: $\{\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2', \cdots, \hat{\boldsymbol{\beta}}_N'\}$. In the second stage, we regress the time-series average of excess returns ($\bar{r}_i = \sum_t r_{i,t}/T$) on the estimated risk exposure $\hat{\boldsymbol{\beta}}_i$ to obtain the estimates of risk premia:

$$\bar{r}_i = \hat{\boldsymbol{\beta}}_i'\lambda + \alpha_i, \quad i = 1, \cdots, N.$$

The residuals of this cross-section regression are the estimates of alphas:

$$\hat{\alpha}_i = \bar{r}_i - \hat{\boldsymbol{\beta}}_i'\hat{\boldsymbol{\lambda}}, \quad i = 1, \ldots, N,$$

where $\hat{\boldsymbol{\lambda}} = (\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\bar{\boldsymbol{r}}$, and $\boldsymbol{B} := \{\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \cdots, \hat{\boldsymbol{\beta}}_N\}'$, and $\bar{\boldsymbol{r}} = \{\bar{r}_1, \cdots, \bar{r}_N\}'$.

In this study, we opted for models with 1, 3, 5, 7, and 11 factors, denoted as S=1,3,5, 7, and 11, respectively. For the 1-factor model, the only risk factor considered is the market factor (MKT), calculated as the value-weighted return of all CRSP firms in excess of the risk-free rate. For the 3-factor model, the risk factors include MKT, SMB (small minus big), and HML (high minus low). SMB and HML represent size and book-to-market equity mimicking portfolios in stock returns, as defined by Fama and French. For the 5-factor model, according to Fung and Hsieh (2001, 2004)Fung and Hsieh (2001, 2004), the risk factors are PTFSBD, PTFSFX, PTFSCOM, PTFSIR, and PTFSSTK. These factors represent the returns from the long position of the lookback straddle of bonds, currencies, commodities, short-term interest rates, and stocks. Finally, for the 7-factor model, we consider MKT, SMB, CS (credit spread), Δ10Y, PTFSBD, PTFSFX, and PTFSCOM as risk factors. CS is the monthly change in the difference between a BAA bond yield and a 10-year constant maturity Treasury yield (GS10). Δ10Y represents the long-term interest rate, specifically the monthly change of GS10.[6] The 11-factor model adds additional four factors to the 7-factor model: HML, MOM, PTFSIR, and PTFSSTK, where MOM is the momentum factor of Carhart (1997).[7]

Table 2 summarizes the risk factors' statistics. In Panel A, we present the means, medians, standard deviations, minimums, and maximums of the monthly percentage returns of the risk factors over the sample period. On the other hand, Panel B displays these risk factors' correlation matrix. Our findings are closely aligned with the statistics presented in Table II of the online appendix of Bali et al. (2014), especially when we adjust our sample period to match theirs.

[6]Risk-free rate and factors MKT, SMB, and HML are sourced from Kenneth R. French's website. The five hedge fund factors can be found on Professor David A. Hsieh's website: `http://faculty.fuqua.duke.edu/~dah7`. Data for BAA and GS10 are available through the Federal Reserve Economic Data of the Federal Reserve Bank of St. Louis.

[7]We retrieve the risk-free rate and the factors MKT, SMB, HML, and MOM from Professor Kenneth R. French's website.

Table 3 summarizes the performance outcomes for various fund categories. In Panel A, we have divided the funds into three primary categories: all funds, Equity Hedge funds, and Macro funds. Moving on to Panel B, we have further classified the Equity Hedge funds into four sub-strategy groups. We have segregated the funds into systematic and discretionary types for each of these categories and sub-strategies. Equity Hedge funds are classified using the optimal random forest classifier, as explained in the preceding section. In contrast, the classification of Macro funds into Systematic Diversified funds and Discretionary Thematic funds follows the HFR system criteria. For systematic and discretionary funds across each category and sub-strategy, we have meticulously reported their average excess returns and the differential of these averages. Additionally, we have provided the average alpha estimates derived from the 1-, 3-, 5-, and 7-factor models, along with the differences between these averages.

Referring to Table 3, we first note that the average excess returns and average alphas are positive across the board. Systematic funds consistently outperform their discretionary counterparts regarding average excess returns across all fund categories. Regarding the difference in returns, Macro funds exhibit a more considerable disparity between systematic and discretionary funds (0.114) than Equity Hedge funds. Within the four sub-strategy groups of Equity Hedge funds, Quantitative Directional funds boast the highest return difference (0.203), while Fundamental Value funds have the lowest (0.064). Regarding alpha values derived from factor models, systematic funds outperform discretionary funds in almost all scenarios. The most significant alpha differences are observed in the 5-factor model for Equity Hedge funds (0.072) and the 1-factor model for Macro funds (0.125). The sole exception to this trend is found in Macro funds when applying the 7-factor model, which results in a negative alpha difference (-0.018).

In addition to analyzing average performance, we compare the proportions of systematic and discretionary funds with positive excess returns and positive alphas within each category, as detailed in Table 4. The data reveals that, across all fund categories, systematic funds more consistently achieve positive excess returns than their discretionary counterparts. Furthermore, the percentage of funds with positive alphas ranges from 70% to 86%. In most cases, systematic funds outnumber discretionary funds in this regard. Specifically, the difference in these proportions ranges from 2.4% to 3.9% for Equity Hedge funds. The difference spans from −1.2% to 7.1% for Macro funds. These results reinforce the conclusion that systematic funds generally outperform discretionary funds, aligning with the findings presented by Chincarini (2014).

3.2. Significant Performance under False Discovery Rate Control

In addition to comparing the average performance of systematic and discretionary funds, we also delve into the significance of each fund's performance. To achieve this, we formulate the following multiple hypotheses:

$$H_{0,i} : \alpha_i \leq 0, \quad i = 1, \cdots, N. \tag{2}$$

Refuting the null hypothesis $H_{0,i}$ implies that the superior performance (positive alpha) of fund i is statistically significant and cannot merely be attributed to chance. Instead, it may indicate the fund manager's genuine investment acumen.

We used the test from Giglio et al. (2021) to identify funds with positive alpha in each category. The test by Giglio et al. (2021) includes several steps. First, they use observable risk factors to calculate risk exposures and residuals for each fund through time-series regression. Second, they employ matrix completion on the unbalanced residual matrix, Hastie et al. (2015), and use PCA to identify latent risk factors and exposures. Then, they perform a cross-sectional regression of the mean excess return on the concatenated observed and unobserved exposures to estimate risk premiums and fund alphas. To account for potential estimation errors in alpha, the alpha estimates are debiased before applying the alpha-screening B-H test, a power-enhanced version of the original B-H test that accounts for inequality in hypotheses. Detailed information on the alpha estimation algorithm and alpha-screening B-H test can be found in Appendices B and C, respectively.[8]

Table 5 displays the ratio of rejected hypotheses (i.e., funds with positive alpha) to the total number of hedge funds, categorized accordingly. The FDR is maintained below the 5% level. Columns (1) to (4) represent results from observable 3-, 5-, 7-, and 11-factor models, referred to as Fs models where s equals 3, 5, and 7. Columns (5), (6), and (7) combine a 7-factor model with 2-, 3-, and 4-unobservable factors, respectively, labeled as F7+Uk where k equals 2, 3, or 4. The final three columns consider pure unobservable factor models with 3, 5, and 7 factors, denoted as Uk. Panel A outlines the proportion of positive alpha funds within discretionary or systematic funds across various main categories, while Panel B focuses on the sub-strategies of Equity Hedge funds. The table simplifies the comparison by displaying the proportion of positive alpha funds relative to the total number of funds. For instance, considering the F7+U3 model in column (6), out of 5,243 funds, 1,572 (29.97%) are identified as positive alpha funds. Among these, 12.28% are discretionary, and 17.69% are systematic. Furthermore, within the positive alpha discretionary funds, 13.84% (calculated as 1.7% divided by 12.2%) belong to Macro funds, while 86.15% are Equity Hedge funds.

Results in Table 5 deliver the critical implication of performance difference between systematic and discretionary funds. We find that more positive-alpha funds belonging to the systematic funds than the discretionary ones, no matter in which factor model or in which Main (Sub) categories.

[8]The alpha-screening FDR test was conducted using a program developed by Giglio et al. (2021), available at `https://dachxiu.chicagobooth.edu/`.

Overall, there are 3.51% to 5.41% more positive-alpha funds as systematic funds than discretionary funds when considering different factor specifications. (3.51% is the difference in the F11 model, and 5.41% is for the F7+U3 model.) For the Macro category, the proportion of positive-alpha funds between systematic and discretionary funds is smaller, as the average difference is 0.77% across different models. The proportion difference is more considerable in the Equity Hedge category. (average is 3.50%) Besides, Table 5 also shows that more positive-alpha funds come from the Equity Hedge category than the Macro category. Fundamental Value generated the most skilled funds when comparing four Sub-category of the Equity Hedge funds.

We report the average alpha value of the positive-alpha funds in Table 6. This table also shows similar pattern of performance difference between systematic and discretionary funds as shown in Table 5. Overall, we find that the average alpha of the positive-alpha funds for systematic funds is higher than that of the discretionary funds. Take F7+U3 model as example, the average positive-alpha funds' alpha of the discretionary funds is 76.5 bp per month, while the average positive-alpha funds' alpha of the systematic funds is 82 bp. The difference is 6.2 bp per month. As the Macro category is concerned, even though Table 5 shows the the proportion difference of the positive-alpha funds between systematic and discretionary is subtle, the average alpha difference is large in Table 6. The differences of the average alpha ranges from 22.64 bp to 1.26 bp (F3 and F7+U2 respectively) with mean of 10.24 bp. We also observed similar results for Equity Hedge funds in Panel A and its 4 Sub-Categories in Panel B. Systematic positive-alpha funds generate, on average, higher alphas than the positive-alpha discretionary funds, regardless of which factor model or which categories we use in this research.

4. Conclusions

This paper studies hedge fund classification and performance and contributes to the hedge fund literature. First, we introduce a machine learning approach to classifying hedge funds into systematic and discretionary funds that differ from existing methods. Second, we use the false discovery control test to examine whether factor-adjusted returns (alphas) of the classified systematic and discretionary funds' performance. Our empirical results show that systematic funds are preferred to discretionary funds across all categories of funds we considered. We identify more of the authentic positive alpha funds in the systematic-type funds than discretionary ones while controlling the multiple test bias. Also, the average alpha estimates of the outperforming funds are higher for the systematic-type funds than for the discretionary ones. All these results support that systematic funds perform better than discretionary funds.

Appendix A: Nested 10-fold cross-validation

The nested cross-validation adds a layer of cross-validation to the original cross-validation approach to choose the hyper-parameters, see e.g., Kuhn and Johnson (2013) for more detailed information. To simplify the illustration, we take the k-NN approach as an example and consider only two hyper-parameter candidates as the number of nearest neighbors, $k = k_1, k_2$. The performance measure is chosen as the accuracy ratio. The $n_1 \times n_2$ nested cross-validation splits the whole data into n_1 outer folds and for each training set of the outer fold, nested cross-validation further splits it into n_2 folds to select the best hyper-parameters used in the corresponding outer fold.

Figure A.1 illustrates the 3×2 nested cross-validation procedure. Assume, in the first place, we have three randomly (or stratified sampled) equal-sized data sets, denoted as $S1, S2$, and $S3$. The outer fold 1 contains the training sets of $S1$ and $S2$, and the validation set $S3$ and the outer fold 2 contains the training sets of $S1$ and $S3$, and the validation set $S2$, and so on. For the training set of the outer fold, we perform the 2-fold cross-validation. Specifically, the inner fold 1 (of the outer fold 1) has the training set of $S1$, and a validation set of $S2$ and the inner fold 2 (of the outer fold 1) has the training set of $S2$ and validation set of $S1$. Given each hyper-parameter k, we use the training set to estimate the model and evaluate model performance on the validation set. For example, given parameter $k1(k2)$, inner fold 1(2) of the outer fold 1 has the accuracy ratio as 0.82 (0.86) and we have an average accuracy ratio of 0.84 of the inner two folds. Then, we compare the average accuracy ratios among different hyper-parameters and select the highest parameter as the parameter used in the corresponding outer fold. For outer fold 1, 2, and 3, the best-performed parameters in the inner cross-validation are k_2, k_2, and k_1 respectively. Finally, for each outer fold and given the hyper-parameter selected in the corresponding inner folds, we estimate the model using the training set of the outer fold and evaluate its performance by the validation set of the outer fold, then obtain the average the of accuracy among all outer folds. discretionary funds

Appendix B: Estimate alpha under the unbalanced panel and observable and unobservable mixture factor models

This appendix follows the Algorithm 6 and 7 in Giglio et al. (2021). Assume the general factor model with S observable factors and K unobservable factors:

$$E(r_i) = \alpha_i + \boldsymbol{\beta}'_{i,o}\boldsymbol{\lambda}_o + \boldsymbol{\beta}'_{i,u}\boldsymbol{\lambda}_u, \quad i = 1, \cdots, N,$$

where $\boldsymbol{\beta}_{i,o}$ and $\boldsymbol{\beta}_{i,u}$ are $S \times 1$ and $K \times 1$ risk exposure to the observable and unobservable risk factors. $\boldsymbol{\lambda}_o$ and $\boldsymbol{\lambda}_u$ are the risk premium of the asset for bearing observable and unobservable risks respectively. Assume that the excess return of fund i at time t is $r_{i,t}$, $i = 1, \cdots, N$; and $t \in \mathcal{T}_i$, which is the time indices set which of fund i has excess return. $\mathcal{N}_t$ is the fund's indices set which includes the existing funds at time t.

Step 1. Time series regression. For each fund, estimate the time-series regression of excess return on the observable risk factors with the same range to obtain the observable risk exposure $\hat{\boldsymbol{\beta}}_{i,o}$ and residual $e_{i,t}$ for $t \in \mathcal{T}_i$. Let $E_{N\times T}$ be the residual matrix (with missing values).

Step 2. Matrix completion of the residual matrix. Suppose $E = M + U$, where M is a $N \times T$ low rank matrix, and U is the noise. Let Ω indicate the existing status of the matrix E, i.e., $\omega_{i,t} = 1$ if $e_{i,t}$ is observed, and 0 if missing. The projection matrix, $P_\Omega(E)$ imputes zeros on the missing entries of matrix E as

$$\left[P_\Omega(E)\right]_{i,t} = \begin{cases} e_{i,t}, & \text{if } \omega_{i,t} = 1; \\ \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

We want to find a low-rank matrix, M, such that minimizes the

$$\min_{M \in R^{N \times T}} \left\|(E - M) \circ \Omega\right\|_F^2 + c\|M\|_*,$$

where $\circ$ is the element-wise product of matrices; c is the tuning parameter; $\|M\|_F$ is the Frobenius norm. $\|M\|_F^2 := \sqrt{\sum_i \sum_t |m_{i,t}|^2}$; and $\|M\|_*$ is the nuclear norm. $\|M\|_* := \sum_{j=1}^{\min\{N,T\}} \sigma_j(M)$, where $\sigma_1(M) \geq \sigma_2(M) \geq \cdots$ are the ordered singular values of M. The iterative approach to obtain estimates of M, $\hat{M}$, see Hastie et al. (2015) and Giglio et al. (2021).

Step 3. Unobservable factors and exposure estimate. Apply singular value decomposition on the matrix $\hat{M}$, and define the unobservable $K \times 1$ factors and their exposures as:

$$\hat{f}_{u,t} = \left(\sum_{i \in \mathcal{N}_t} u_i u_i'\right)^{-1} \sum_{i \in \mathcal{N}_t} u_i e_{i,t},\ t = 1, \cdots, T,$$

$$\hat{\boldsymbol{\beta}}_{u,i} = \left(\sum_{t \in \mathcal{T}_i} \hat{f}_{u,t}\hat{f}_{u,t}'\right)^{-1} \sum_{t \in \mathcal{T}_i} \hat{f}_{u,t} e_{i,t},\ i = 1, \cdots, N,$$

where $u_1, \cdots, u_K$ is the top K left singular-vector of $\hat{M}$. Define all risk exposures as $\hat{\boldsymbol{\beta}} := (\hat{\boldsymbol{\beta}}_o, \hat{\boldsymbol{\beta}}_u)$ and all observable and unobservable risk factor as $\hat{f}_t := \left(f_{o,t} - \bar{f}_o, \hat{f}_{u,t}'\right)'$, where $f_{o,t}$ is the observable $S \times 1$ risk factors for $t = 1, \cdots, T$, and $\bar{f}_o = \frac{1}{T}\sum_{t=1}^T f_{o,t}$.

Step 4. Estimate risk premium. Run a cross-section regression of $\bar{r}_i$ on $\hat{\boldsymbol{\beta}}$ to obtain the slope $\hat{\boldsymbol{\lambda}}$ as the risk premium.

Step 5. De-biased alpha estimates.

$$\hat{\alpha}_i = \bar{r}_i - \hat{\boldsymbol{\beta}}_i'\hat{\boldsymbol{\lambda}} + \hat{A}_i.$$

where $\hat{A}_i$ is the (de-)biased term for the unbalanced data, see Giglio et al. (2021).

Step 6. Construct the t-statistics and its p-values. The t-statistics is the standard asymptotic normal for one-side test.

$$t_i = \frac{\hat{\alpha}_i}{se(\hat{\alpha}_i)}, \quad p_i = 1 - \Phi(t_i), \quad i = 1, \cdots, N,$$

$\Phi(\cdot)$ is the standard normal CDF, and $se(\hat{\alpha}_i) = \frac{1}{|\mathcal{T}_i|}\sqrt{\sum_{t\in\mathcal{T}_i} \hat{\epsilon}_{i,t}^2 \left(1 - \hat{f}_t'\hat{\Sigma}_f^{-1}\hat{\boldsymbol{\lambda}}\right)^2}$, where $\hat{\epsilon}_{i,t} = r_{i,t} - \bar{r}_i - \hat{\boldsymbol{\beta}}_i'\hat{f}_t$, $\hat{\Sigma}_f = \frac{1}{T}\sum_{t=1}^{T} \hat{f}_t\hat{f}_t'$.

Appendix C: Alpha-screening Benjamini and Hochberg (1995) FDR control

It is well known that simultaneously testing multiple hypotheses is easy to suffer from the false discovery problem. Suppose t_i is a test statistics to examine $H_{0,i}$ in equation (2). A null hypothesis is rejected when $t_i > c_i$ for a threshold c_i. Let $\mathcal{H}_0$ is the set of indices of the true null hypotheses, $\mathcal{R}$ is the set of indices of the rejected hypotheses, and $\mathcal{F}$ is the indices of false rejected hypotheses, i.e.,

$$\mathcal{R} = \bigcup_{1\le i\le N} \{i : t_i > c_i\};$$

$$\mathcal{F} = \bigcup_{1\le i\le N} \{i : t_i > c_i \,\text{and}\, \alpha_i \le 0\}.$$

The false discovery proportion is defined as the number of falsely rejected hypotheses to the total number of rejections. As the number of false rejection is unobservable, the false discovery rate (FDR) is then defined as the expectation of false discovery proportion, i.e.,

$$FDR := E\left(\frac{|\mathcal{F}|}{|\mathcal{R}|}\right).$$

where $|A|$ denotes the number of elements in the set A. If the number of rejections is zero, then FDR is defined as zero. Benjamini and Hochberg (1995) proposed the following procedures to control the FDR under q level. Let

$$p_{(1)} \leq \cdots \leq p_{(N)}$$

be the ordered p-values corresponding to the null hypotheses $H_{0,(1)}, \cdots, H_{0,(N)}$. Rejects the hypotheses $H_{0,(1)}, \cdots, H_{0,(j^*)}$, where j^* is the number such that

$$j^* := \max_{1 \leq j \leq N} \left\{ j : p_{(j)} \leq \gamma_j \right\}.$$

where $\gamma_j := \frac{j}{N} q$ be the rejection criteria. Giglio et al. (2021) suggest modify the method of Benjamini and Hochberg (1995) by precluding the extremely negative alpha funds in advance (in fact fund's t statistics). Define the reduced set of funds indices as

$$\tilde{N} := \bigcup_{1 \leq i \leq N} \left\{ i : t_i > -\log(\log(T))\sqrt{\log N} \right\}$$

and the rejection criteria γ_j is therefore change to $\frac{j}{|\tilde{N}|} q$. They show by theoretic inference and Monte Carlo simulation that this alpha screen procedure improves test power while remaining controlling for the FDR under q level. In our sample, aka, drop the funds whose t statistics is lower than -5.03 before applying the BH procedure. ($T = 263, N = 5,245$)

References

Abis, S., 2022. Man vs. machine: Quantitative and discretionary equity management. Working Paper .

Bajgrowicz, P., Scaillet, O., 2012. Technical trading revisited: False discoveries, persistence tests, and transaction costs. Journal of Financial Economics 106, 473–491.

Bajgrowicz, P., Scaillet, O., Treccani, A., 2016. Jumps in high-frequency data: Spurious detections, dynamics, and news. Management Science 62, 2198–2217.

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

Barras, L., Scaillet, O., Wermers, R., 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. The Journal of Finance 65, 179–216.

Beggs, W., Brogaard, J., Hill-Kleespie, A., 2021. Quantitative investing and market instability. Available at SSRN 3281447 .

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57, 289–300.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Cao, C., Chen, Y., Liang, B., Lo, A.W., 2013. Can hedge funds time market liquidity? Journal of Financial Economics 109, 493–516.

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

Chincarini, L., 2014. The impact of quantitative methods on hedge fund performance. European Financial Management 20, 857–890.

Chinchor, N., 1992. Muc-4 evaluation metrics, in: Proceedings of the 4th Conference on Message Understanding, Association for Computational Linguistics. p. 22–29.

Chordia, T., Goyal, A., Saretto, A., 2020. Anomalies and false rejections. The Review of Financial Studies 33, 2134–2179.

Cochrane, J.H., 2009. Asset Pricing: Revised edition. Princeton University Press.

Cuthbertson, K., Nitzsche, D., O'Sullivan, N., 2012. False discoveries in UK mutual fund performance. European Financial Management 18, 444–463.

Evans, R.B., Rohleder, M., Tentesch, H., Wilkens, M., 2023. Diseconomies of scale in quantitative and fundamental investment styles. Journal of Financial and Quantitative Analysis 58, 2417–2445.

Fung, W., Hsieh, D.A., 1997. Empirical characteristics of dynamic trading strategies: The case of hedge funds. The Review of Financial Studies 10, 275–302.

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

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

Giglio, S., Liao, Y., Xiu, D., 2021. Thousands of alpha tests. The Review of Financial Studies 34, 3456–3496.

Goyal, A., Wahal, S., 2015. Is momentum an echo? Journal of Financial and Quantitative Analysis 50, 1237–1267.

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business & Economic Statistics 23, 365–380.

Harvey, C.R., Liu, Y., Saretto, A., 2020. An evaluation of alternative multiple testing methods for finance applications. The Review of Asset Pricing Studies 10, 199–248.

Harvey, C.R., Liu, Y., Zhu, H., 2016. ... and the cross-section of expected returns. The Review of Financial Studies 29, 5–68.

Harvey, C.R., Rattray, S., Sinclair, A., Van Hemert, O., 2017. Man vs. machine: Comparing discretionary and systematic hedge fund performance. The Journal of Portfolio Management 43, 55–69.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Hastie, T., Tibshirani, R., Wainwright, M., 2015. Statistical Learning with Sparsity: the Lasso and Generalizations. CRC press.

Hsu, P.H., Kuan, C.M., 2005. Reexamining the profitability of technical analysis with data snooping checks. Journal of Financial Econometrics 3, 606–628.

Hsu, P.H., Kyriakou, I., Ma, T., Sermpinis, G., 2023. Mutual funds' conditional performance free of data snooping bias. Journal of Financial and Quantitative Analysis , forthcoming.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer.

Kuang, P., Schröder, M., Wang, Q., 2014. Illusory profitability of technical analysis in emerging foreign exchange markets. International Journal of Forecasting 30, 192–205.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer.

Manning, C.D., Schütze, H., 1999. Foundations of Statistical Natural Language Processing. MIT Press.

Romano, J.P., Shaikh, A.M., Wolf, M., 2008. Formalized data snooping based on generalized error rates. Econometric Theory 24, 404–447.

Romano, J.P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. Econometrica 73, 1237–1282.

Romano, J.P., Wolf, M., 2007. Control of generalized error rates in multiple testing. The Annals of Statistics 35, 1378–1408.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.

Figure 1: Importance of Features. This figure displays the first 30 bi-grams with the highest variable importance of the random forest.

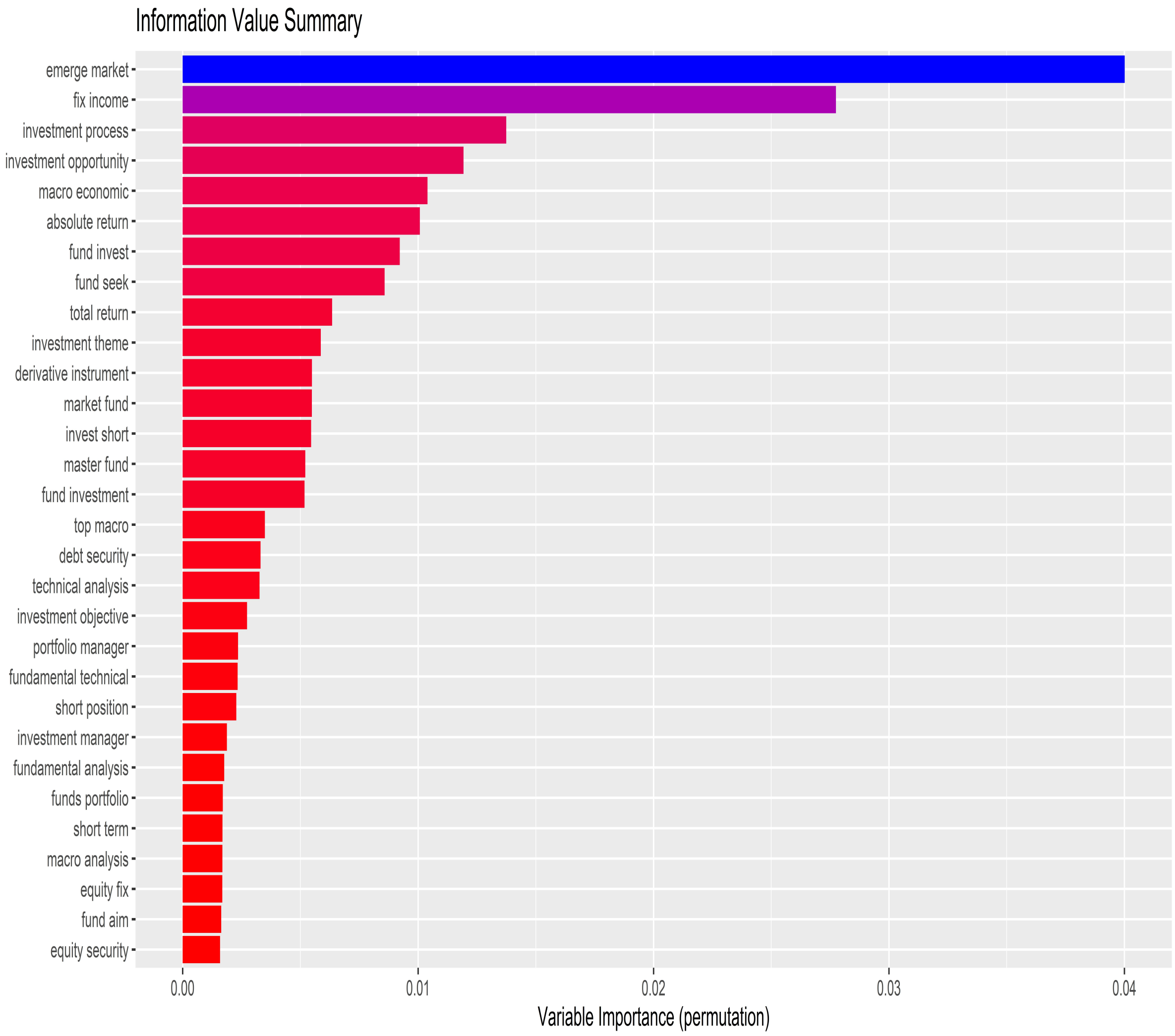

Figure A.1 Illustration of the Nested Cross-Validation

Data S1, S2, S3

3-fold cross-validation (outer)

Outer folds	Training set	Validation set	Hyperparameter	Accuracy ratio	Avg. of accurcay ratios
Fold 1	S1, S2	S3	k2	0.81	
Fold 2	S1, S3	S2	k2	0.82	0.82
Fold 3	S2, S3	S1	k1	0.83	

2-fold cross-validation (inner)

	Inner folds	Training set	Validation set	Hyperparameter	Accuracy ratio	Avg. of accurcay ratios
Fold 1	Fold 1	S1	S2	k1	0.82	0.84
	Fold 2	S2	S1	k1	0.86	
	Fold 1	S1	S2	k2	0.87	0.86
	Fold 2	S2	S1	k2	0.85	
Fold 2	Fold 1	S1	S3	k1	0.82	0.84
	Fold 2	S3	S1	k1	0.86	
	Fold 1	S1	S3	k2	0.87	0.86
	Fold 2	S3	S1	k2	0.85	
Fold 3	Fold 1	S2	S3	k1	0.87	0.86
	Fold 2	S3	S2	k1	0.85	
	Fold 1	S2	S3	k2	0.82	0.84
	Fold 2	S3	S2	k2	0.86	

Table 1: Classification Performances Measures of Nested 10-fold Cross Validation

	Linear	Logit	LDA	KNN	SVM	RF	Tree	GB
Panel A: Accuracy								
Median	0.70	0.70	0.68	0.80	0.84	0.86	0.76	0.84
Mean	0.70	0.70	0.68	0.81	0.84	0.86	0.77	0.84
Max	0.77	0.75	0.76	0.87	0.86	0.90	0.80	0.88
Min	0.65	0.65	0.65	0.79	0.80	0.82	0.73	0.81
SD	0.04	0.03	0.03	0.02	0.02	0.02	0.03	0.02
Panel B: AUC								
Median	0.70	0.70	0.66	0.73	0.80	0.82	0.73	0.81
Mean	0.70	0.70	0.66	0.74	0.80	0.82	0.74	0.80
Max	0.78	0.75	0.74	0.81	0.84	0.88	0.80	0.84
Min	0.65	0.65	0.61	0.72	0.75	0.79	0.70	0.74
SD	0.04	0.03	0.04	0.03	0.03	0.02	0.03	0.03
Panel C: Precision								
Median	0.82	0.82	0.77	0.79	0.85	0.86	0.81	0.84
Mean	0.82	0.82	0.77	0.79	0.85	0.86	0.82	0.85
Max	0.89	0.86	0.84	0.84	0.88	0.91	0.86	0.89
Min	0.78	0.76	0.73	0.77	0.79	0.82	0.80	0.82
SD	0.03	0.03	0.03	0.02	0.03	0.02	0.02	0.03
Panel D: F1 Score								
Median	0.75	0.75	0.75	0.87	0.88	0.90	0.82	0.88
Mean	0.75	0.75	0.75	0.87	0.88	0.90	0.83	0.88
Max	0.82	0.80	0.81	0.91	0.90	0.93	0.85	0.91
Min	0.70	0.71	0.71	0.86	0.86	0.87	0.79	0.86
SD	0.04	0.03	0.03	0.01	0.01	0.02	0.02	0.01

This table reports the classification performance measures: Median, Mean, Max, Min, and SD, which are the median, average, maximum, minimum, and standard deviation of the outer 10-folds measures. Statistical learning methods include the following. Linear: linear regression; Logit: logistic regression; LDA: linear discrimination analysis; KNN: k-nearest neighbor approach; SVM: support vector machine with the Gaussian kernel; RF: random forest; Tree: classification tree; GB: gradient boosting.

Table 2: Summary Statistics of Risk Factors: 1994-2015

Panel A: Risk Factors

	N	Mean	STD	Min	Median	Max
MKTRF	263	0.619	4.421	-17.230	1.320	11.350
SMB	263	0.146	3.393	-17.170	0.000	22.080
HML	263	0.192	3.094	-11.250	-0.020	12.910
MOM	263	0.489	5.161	-34.580	0.580	18.380
CS	263	-0.053	0.208	-0.804	-0.060	1.532
Δ10Y	263	-0.013	0.227	-1.110	-0.030	0.650
PTFSBD	263	-1.492	15.251	-26.630	-3.860	68.860
PTFSFX	263	-0.700	19.492	-30.130	-5.180	90.270
PTFSCOM	263	-0.366	14.255	-24.650	-3.010	64.750
PTFSIR	263	-0.891	25.708	-35.130	-6.040	221.920
PTFSSTK	263	-4.876	14.056	-30.190	-6.990	66.620

Panel B: Correlation Matrix

	MKTRF	SMB	HML	MOM	CS	Δ10Y	TFSBD	PTFSFX	PTFSCOM	PTFSIR	PTFSSTK
MKTRF	1.000										
SMB	0.208	1.000									
HML	-0.157	-0.323	1.000								
MOM	-0.271	0.104	-0.199	1.000							
CS	-0.140	-0.066	-0.056	0.043	1.000						
Δ10Y	0.101	0.101	-0.039	-0.072	0.629	1.000					
PTFSBD	-0.255	-0.057	-0.092	0.017	-0.026	-0.180	1.000				
PTFSFX	-0.206	-0.001	-0.005	0.120	0.047	-0.177	0.298	1.000			
PTFSCOM	-0.179	-0.054	-0.057	0.192	0.040	-0.115	0.194	0.341	1.000		
PTFSIR	-0.269	-0.098	0.006	-0.004	0.141	-0.170	0.204	0.249	0.228	1.000	
PTFSSTK	-0.238	-0.089	0.099	0.000	0.056	-0.198	0.223	0.254	0.174	0.332	1.000

This table reports the summary statistics for the risk factors for the sample period from January 1994 to November 2015. SMB, HML, and MOM are the mimicking portfolios for size, book-to-market, and momentum in stock returns, respectively. PTFSBD, PTFSFX, PTFSCOM, PTFSIR, and PTFSSTK are the returns for the long position of the look back straddles of bonds, currencies, commodities, short-term interest rates, and stocks, respectively. CS is the credit risk factor that is defined as the monthly change in the difference between the BAA bond yield and the 10-year constant maturity Treasury yield (GS10). Δ10Y is the difference between the current GS10 and the lag one GS10.

Table 3: Summary Statistics of Fund Performance: 1994–2015

Panel A: Main Strategy

	All Strategy			Macro			Equity Hedge		
	Dis.	Sys.	Diff.	Dis.	Sys.	Diff.	Dis.	Sys.	Diff.
N	2161	3084	923	390	804	414	1771	2280	509
Ret	0.398	0.494	0.096	0.349	0.463	0.114	0.409	0.505	0.096
F1	0.353	0.430	0.077	0.313	0.438	0.125	0.362	0.428	0.066
F3	0.317	0.385	0.068	0.289	0.381	0.092	0.323	0.386	0.063
F5	0.420	0.489	0.069	0.349	0.435	0.086	0.435	0.508	0.072
F7	0.334	0.354	0.020	0.268	0.250	-0.018	0.348	0.391	0.042
F11	0.317	0.341	0.024	0.244	0.228	-0.016	0.334	0.381	0.048

Panel B: Sub Strategy of Equity Hedge

	Market Neutral			Fundamental Growth			Fundamental Value			Quantitative Directional		
	Dis.	Sys.	Diff.	Dis.	Sys.	Diff.	Dis.	Sys.	Diff.	Dis.	Sys.	Diff.
N	198	398	200	655	689	34	835	972	137	83	221	138
Ret	0.203	0.324	0.121	0.403	0.547	0.1405	0.465	0.529	0.064	0.395	0.598	0.203
F1	0.173	0.309	0.136	0.288	0.430	0.143	0.461	0.469	0.009	0.398	0.449	0.050
F3	0.165	0.281	0.116	0.264	0.364	0.100	0.403	0.439	0.036	0.373	0.415	0.041
F5	0.208	0.327	0.119	0.453	0.558	0.105	0.481	0.516	0.035	0.387	0.640	0.253
F7	0.176	0.294	0.118	0.314	0.393	0.080	0.418	0.427	0.009	0.328	0.396	0.067
F11	0.174	0.275	0.102	0.297	0.374	0.077	0.399	0.425	0.026	0.348	0.405	0.057

Disc. and Sys. stand for discretionary and systematic funds, respectively; Diff. is the difference between the results of Disc. and Sys. N is the number of funds in each category; Ret is the average excess return of each category; F1, F3, F5 and F7 denote the 1-, 3-, 5-, and 7-factor models, respectively, with the corresponding entries being alphas.

Table 4: Proportion of Positive-alpha Fund Performance: 1994–2015

Panel A: Main Strategy

	All Strategy			Macro			Equity Hedge		
	Dis.	Sys.	Diff.	Dis.	Sys.	Diff.	Dis.	Sys.	Diff.
Ret	0.801	0.851	0.050	0.805	0.862	0.057	0.800	0.846	0.046
F1	0.782	0.828	0.047	0.772	0.843	0.071	0.784	0.823	0.039
F3	0.765	0.800	0.035	0.751	0.790	0.039	0.768	0.803	0.035
F5	0.820	0.848	0.028	0.800	0.840	0.040	0.824	0.850	0.026
F7	0.774	0.784	0.010	0.726	0.714	-0.012	0.784	0.808	0.024
F11	0.766	0.778	0.011	0.751	0.690	-0.061	0.770	0.808	0.039

Panel B: Sub Strategy of Equity Hedge

	Market Neutral			Fundamental Growth			Fundamental Value			Quantitative Directional		
	Dis.	Sys.	Diff.	Dis.	Sys.	Diff.	Dis.	Sys.	Diff.	Dis.	Sys.	Diff.
Ret	0.707	0.812	0.104	0.774	0.846	0.072	0.840	0.862	0.023	0.831	0.842	0.010
F1	0.692	0.804	0.112	0.728	0.803	0.074	0.850	0.844	-0.007	0.771	0.828	0.057
F3	0.662	0.781	0.120	0.722	0.772	0.050	0.829	0.844	0.015	0.771	0.760	-0.011
F5	0.697	0.819	0.122	0.824	0.859	0.035	0.856	0.858	0.002	0.807	0.846	0.039
F7	0.697	0.774	0.077	0.736	0.801	0.065	0.838	0.836	-0.002	0.831	0.769	-0.062
F11	0.692	0.791	0.100	0.740	0.795	0.055	0.813	0.829	0.016	0.747	0.787	0.040

Disc. and Sys. stand for discretionary and systematic funds, respectively; Diff. is the difference between the results of Disc. and Sys. Ret is the proportion of the positive excess return of each category; F1, F3, F5 and F7 denote the 1-, 3-, 5-, and 7-factor models, respectively, with the corresponding entries being proportion of positive alphas.

Table 5: Proportion of Positive-alpha Funds

		(1) F3	(2) F5	(3) F7	(4) F11	(5) F7+U2	(6) F7+U3	(7) F7+U4	(8) U3	(9) U5	(10) U7
Panel A: Main strategy											
All											
	Dis.	10.64%	9.34%	11.19%	12.41%	11.78%	12.28%	12.30%	11.13%	11.59%	10.79%
	Sys.	15.25%	12.87%	14.59%	15.92%	16.26%	17.69%	17.18%	15.54%	15.86%	15.04%
Macro											
	Dis.	1.60%	1.41%	1.49%	1.81%	1.56%	1.70%	1.79%	1.54%	1.54%	1.54%
	Sys.	2.06%	2.15%	1.68%	2.12%	2.76%	3.18%	2.99%	2.35%	2.15%	2.27%
Equity Hedge											
	Dis.	9.04%	7.93%	9.70%	10.60%	10.22%	10.58%	10.51%	9.59%	10.05%	9.25%
	Sys.	13.19%	10.72%	12.91%	13.80%	13.50%	14.51%	14.18%	13.19%	13.71%	12.77%
Panel B: Sub-Strategy of Equity Hedge											
Equity Market Neutral											
	Dis.	1.01%	1.03%	1.07%	1.14%	1.09%	1.09%	1.09%	1.05%	1.11%	1.12%
	Sys.	2.36%	2.29%	2.50%	2.57%	2.38%	2.54%	2.48%	2.40%	2.50%	2.52%
Fundamental Growth											
	Dis.	2.54%	2.35%	2.84%	2.96%	3.24%	3.28%	2.78%	2.82%	3.07%	2.99%
	Sys.	3.24%	2.59%	3.30%	3.53%	3.85%	4.04%	3.64%	3.36%	3.83%	3.91%
Fundamental Value											
	Dis.	5.19%	4.31%	5.41%	6.10%	5.01%	5.40%	5.15%	5.21%	5.68%	6.01%
	Sys.	6.44%	5.01%	6.08%	6.73%	6.02%	6.25%	5.76%	5.83%	6.14%	6.56%
Quantitative Directional											
	Dis.	0.31%	0.25%	0.38%	0.40%	0.25%	0.29%	0.23%	0.34%	0.36%	0.38%
	Sys.	1.14%	0.82%	1.03%	0.97%	0.93%	0.88%	0.90%	1.01%	1.03%	1.20%

This table reports the proportion of the positive-alpha funds to total fund number. We use the false discovery rate (FDR) test of Giglio et al. (2021) to identify the superior funds and control the FDR under 5% level. Fs is short for the observable s-factor model, where $s = 3, 5, 7$, and 11. Uk is short for the unobserved k-factor model, where $k = 3, 5$, and 7. We also consider the mixture model of observable seven-factor models to two, three, and four unobserved k-factor model, which is denoted as $Fs + Uk$ model.

Table 6: Alpha of Positive-alpha Funds

		(1) F3	(2) F5	(3) F7	(4) F11	(5) F7+U2	(6) F7+U3	(7) F7+U4	(8) U3	(9) U5	(10) U7
Panel A: Main strategy											
All											
	Dis.	0.853	0.942	0.854	0.832	0.783	0.765	0.740	0.803	0.789	0.776
	Sys.	0.888	1.058	0.884	0.900	0.839	0.827	0.807	0.854	0.844	0.832
Macro											
	Dis.	0.838	0.917	0.853	0.807	0.814	0.816	0.769	0.869	0.824	0.832
	Sys.	1.064	1.046	1.000	0.993	0.826	0.849	0.867	0.972	0.890	0.854
Equity Hedge											
	Dis.	0.855	0.947	0.854	0.836	0.778	0.757	0.735	0.793	0.783	0.767
	Sys.	0.861	1.060	0.869	0.886	0.841	0.824	0.797	0.839	0.835	0.827
Panel B: Sub-Strategy of Equity Hedge											
Equity Market Neutral											
	Dis.	0.584	0.632	0.598	0.569	0.585	0.579	0.568	0.605	0.601	0.588
	Sys.	0.658	0.716	0.666	0.656	0.677	0.662	0.660	0.667	0.663	0.661
Fundamental Growth											
	Dis.	1.034	1.158	1.000	1.027	0.884	0.855	0.806	0.907	0.890	0.876
	Sys.	0.946	1.233	0.961	0.968	0.854	0.849	0.811	0.887	0.863	0.860
Fundamental Value											
	Dis.	0.820	0.905	0.831	0.788	0.752	0.738	0.734	0.777	0.768	0.752
	Sys.	0.881	1.050	0.875	0.901	0.868	0.850	0.821	0.857	0.863	0.851
Quantitative Directional											
	Dis.	0.859	0.981	0.801	0.918	0.747	0.679	0.706	0.657	0.678	0.666
	Sys.	0.925	1.537	1.033	1.087	1.031	0.988	0.964	0.980	0.987	0.939

This table reports the average alphas of the positive-alpha funds. We use the false discovery rate (FDR) test of Giglio et al. (2021) to identify the superior funds and control the FDR under 5% level. Fs is short for the observable s-factor model, where $s = 3, 5, 7$, and 11. Uk is short for the unobserved k-factor model, where $k = 3, 5$, and 7. We also consider the mixture model of observable seven-factor models to two, three, and four unobserved k-factor model, which is denoted as $Fs + Uk$ model.