Systematic 13F Hedge Fund Alpha*

Luca Angelini[§] Mobeen Iqbal[†] Farouk Jivraj[‡]

First version: October, 2019

*The authors wish to thank Randy Cohen, the participants of the London Quant Group (LQG) Seminar (November 2019) and the Barclays Quantitative Investment Strategies (QIS) group, specifically Benedict Redmond and David Haefliger for their helpful comments and suggestions. A version of this paper has been published by Barclays under the title "Systematic 13F Hedge Fund Alpha" on the 9th of October 2019.

[§]Luca Angelini is from Novus Partners Inc. (langelini@novus.com).

[†]Mobeen Iqbal (corresponding author) is from Quantitative Investment Strategies (QIS), Barclays Investment Bank (mobeen.iqbal@barclays.com).

[‡]Farouk Jivraj is Head of Quantitative Investment Strategies (QIS) Research, Barclays Investment Bank, (farouk.jivraj@barclays.com) and a Visiting Researcher at Imperial College Business School, London.

Systematic 13F Hedge Fund Alpha

Abstract

Institutions holding greater than $100 million in securities are required to disclose their holdings in US listed stocks to the Securities and Exchange Commission (SEC) no later than 45 days after the quarter-end, in a form known as 13F. We show that the "best ideas" of hedge funds produce economically meaningful and statistically significant risk-adjusted returns that outperform the S&P500, following tests identified in Cohen, Polk, and Silli (2010). We construct alternative measures that are more suitable for hedge funds, rather than mutual funds: conviction and consensus. We find that to systematically identify hedge fund alpha in the 13F filings, one must select the *right group* of managers that have longer-term views on stock picks. Based on our findings, we find that a strategy that combines conviction and consensus of such managers that outperforms the S&P500 by 3.80% on average and delivers a Sharpe ratio of 0.75 over the period May 2004 to June 2019.

JEL Classification: G11, G23

Keywords: 13F trading strategies, hedge fund alpha, copycat investing

I. Introduction

A 1978 amendment to the Securities and Exchange Act of 1934 required all institutions, including hedge fund managers, that hold greater than $100 million in securities under discretionary management to file their holdings in US-listed stocks with the Securities and Exchange Commission (SEC) no later than 45 days after the end of each calendar quarter under a form known as '13F'. These filings have been a popular source of hedge fund intelligence. Professional and amateur investors are known to attempt to use the 13F filings to pursue 'copycat investing' strategies (i.e. cloning the stock picks of hedge funds, for instance, in order to piggyback on their 'alpha' ability). There also exists a vast academic literature studying the nature, behaviour, and performance of investment managers - particularly hedge funds, whose holdings and stock picking skills are purposefully secretive. Several studies have used holdings data to uncover features of fund performance and construct trading strategies; however, hedge fund holdings data is notoriously difficult to use without thorough cleaning. In particular, most sources of SEC holdings data fail to correctly classify fund types and therefore are unable to specifically focus on hedge fund holdings.[1] Additionally, a common argument against using 13F filings as a source of data for such copycat strategies is that there is a 45-day lag in between the reporting date (quarter-end) and the filing date, on which the holdings become publically available, and so the value of the stock picks may have disappeared in that 45-day period.

This paper aims to answer several questions. First, does holdings data allow one to construct strategies that harness the stock-picking skills of hedge funds? Second, what is the best methodology to identify the best ideas of hedge funds? Third, once a methodology has been selected, can a copycat strategy be improved by tracking a subset of hedge funds that focus on stock picking alone? Finally, does the fact that the holdings data are released with a lag erode the value in the stock picks? In this paper, we use proprietary data provided by Novus to answer these

[1] Many types of filers appear in the SEC 13F filings. These include banks, insurance companies, pension funds, mutual funds, investment advisors, and hedge funds. As the SEC does not require a filer to disclose the type of filer that they are, it is notoriously difficult to subset out hedge funds. Further, even if one was able to extract hedge funds based on a list of names of known hedge funds, it is likely that one would not select all hedge funds and, additionally, without a proprietary algorithm, one would find it difficult to classify fund styles to select those funds that behave like stock pickers rather than funds that use equity to hedge exposures in other asset classes or in equity derivatives, for example.

questions.[2] The data allows us to specifically address the third question: the 'who' matters in portfolio construction. This is because Novus systematically identifies funds within the 'Hedge Fund Universe' (HFU) that have a longer-term view on equity. We classify these funds as 'Fundamental Equity Hedge Funds' (FEHF) and describe them in more detail below. Broadly speaking, however, these funds are equity long/short, equity market neutral, equity long-only, and equity event driven funds. That means we are able to differentiate between the holdings of funds that specifically focus on stock picking rather than, for instance, credit funds hedging out exposures in their corporate bond holdings or their derivative positions. Doing so allows one to harness off the ability of funds that have identified stocks and expressed a direct directional view in the stock price.

We start our analysis by following tests identified in Cohen, Polk, and Silli (2010)'s seminal paper 'Best Ideas' but with modifications to fit the universe of managers that we track. The authors outline four measures of 'best ideas' and construct trading strategies based on the holdings of mutual funds. We show that at least two of the 'best ideas' strategies, rebalanced with a 47-day lag after the quarter-end on the HFU deliver statistically significant six-factor alphas and can outperform the S&P 500 Total Return Index, with Sharpe ratios of 0.63 and 0.69. The best performing test is one that identifies positions that a manager is overweight relative to a value-weighted portfolio of stocks that they hold at the quarter-end. We describe the tests in more detail at a later stage in the paper but the tests in Cohen, Polk, and Silli (2010) make two strong assumptions: that managers are maximising an information ratio and that the model accurately captures the factor structure of returns. We relax these assumptions. In particular, for our group of managers - that is, hedge funds - the assumption that the manager maximises an information ratio is not entirely valid. Such an assumption is perfectly plausible for mutual funds, who are typically benchmarked and so face a tracking error constraint that they will bear in mind when selecting portfolio weights of their stock selections. Hedge funds, on the other hand, are unlikely to be benchmarked; hence, we turn to alternative measures of 'best ideas'.

Our argument relies on the fact that hedge funds tend to express directional views on stock prices on an absolute weight basis rather than in a benchmarked manner, particularly for stocks that they

[2]Novus provides data analytical services to the investment management community, specifically hedge funds, being founded in 2007.

have high conviction in. On the back of this, we construct two alternative measures to run tests on: (i) conviction, which ranks stocks in each manager's portfolio into quintiles by their position sizes and then takes long positions in all stocks appearing in the top quintile; and (ii) consensus, which ranks stocks by how many unique managers are holding them at each quarter-end and goes long the top quintile of stocks by this measure. We find that both measures, when tested on HFU, perform similarly, delivering Sharpe ratios of 0.65. As the 'how' does not stand out from running the conviction and consensus tests on the HFU, we repeat the tests on FEHF and HFU excluding FEHF (henceforth known as HFU ex FEHF). The FEHF outperforms both the HFU and HFU ex FEHF strategies for both measures. Hence, we argue that the 'who' is just as important as the 'how'. Returning to the 'how', we construct a single strategy that combines conviction and consensus. This is because we note that conviction and consensus are weakly positively correlated in the cross-section when looking at FEHF managers' positions. We note then that overlaying consensus with a measure of conviction results in returns that are monotonically increasing with the position size that managers have in various stocks.

We find that a strategy that overlays consensus with a conviction threshold performs well. However, the strategy so far is not entirely practical as it is based on quintiles. To test a practical tradable strategy, we test a simple strategy that invests in the top 50 high conviction positions in the aggregate FEHF portfolio at each quarter-end. The strategy tracks FEHF managers' stock picks in which at least 7.5% of their portfolio at each quarter-end is invested. Once we remove positions with holdings of less than 7.5% for each manager, we rank stocks based on how many FEHF managers are holding them and take long positions in the top 50 stocks by this ranking. If stocks are tied in this ranking, the strategy selects the stock with the greater aggregate market value in the FEHF manager portfolio. The strategy delivers a six-factor alpha of 0.32% per month and an annual Sharpe ratio of 0.75. The alpha drops to -0.003% and is not statistically significant when tracking HFU ex FEHF managers, again reiterating that the 'who' is just as important as the 'how' particularly when constructing a practical trading strategy.

Overall, in this paper, we make five main contributions. First, in a similar spirit to Cohen, Polk, and Silli (2010), we show that the "best ideas" of hedge funds indeed deliver economically meaningful and statistically significant risk-adjusted returns that outperform the S&P 500. Second, we construct alternative measures of "best ideas" that are more suitable to studying hedge fund holdings, rather than mutual funds, that incorporate conviction and consensus. Third, and most importantly, we demonstrate the power of the 'who' in constructing 13F trading strategies. To the best of our knowledge, we are the first in the academic literature to use qualitative fund classifications to separate out those funds that tend to have longer-term views. We show that constructing a strategy that tracks these funds' positions outperforms others and delivers systematic 13F hedge fund alpha. This is not to say that these managers outperform other managers on an absolute basis - we are silent on fund performance as a whole as our dataset does not allow us to compare fund performance. This is because our data is quarterly and only shows the long positions of funds. Many funds can and will take short positions as well as changing their holdings intra-quarter. Thus, arguing that one set of managers outperforms another on an absolute basis using quarterly holdings data is unfeasible. Our results speak to the construction of trading strategies based on publically available, but hard to *correctly* identify, 13F hedge fund holdings. Fourthly, based on our first three main contributions, we find that a strategy that combines conviction and consensus of such managers delivers returns that outperform the S&P500 by 3.80% on average over the period Q1 2004 to Q2 2019. Finally, we show that our findings are also robust to the lag in between the filing date (the quarter-end) and the date on which the holdings become publically available (45 days later).

II. Relevant literature

There exists a very large literature studying the performance and holdings of professional money managers. A lot of this literature attempts to understand whether the performance of such funds justifies the fees, whether the performance can be attributed to luck or skill, or, amongst a vast range of other questions, whether the performance of such funds can be explained away by standard asset pricing factors such as the Fama and French (2015) five-factor model (i.e. excess market return, size (SMB), high-minus-low (HML), robust-minus-weak (RMW), and conservative-minus-aggressive (CMA)). On the mutual fund aspect of this literature, a very closely related paper to ours, which we use as a benchmark for the tests we run, is Cohen, Polk, and Silli (2010). In their study, the authors advocate for the identification of the "best ideas" in mutual fund managers' portfolios, which they find not only generates economically and statistically significant risk-adjusted returns but also outperforms the rest of the positions in their portfolios. We describe the methodology in their paper in more detail when we discuss the benchmark tests but, broadly, the authors attempt to identify ex-ante the positions that active managers display the most conviction towards measuring the difference between stock weights in the portfolio against measures of a benchmark portfolio. The most overweight stock for each manager is classed as their 'best idea' and the strategy takes long positions in these stocks. In their setting, if a stock is the 'best idea' of more than one manager, it is overweight in the portfolio. We modify this point in our paper so that we can compare like-for-like strategies.

Similarly, Verbeek and Wang (2013) find that, by constructing hypothetical copycat fund portfolios, certain managers are indeed able to significantly outperform their peers (net of trading costs and expenses) and as such copycat investing can be an attractive strategy. It is important to note however, that the hedge fund literature has not converged to a universal view on whether hedge fund managers have superior skill. This, in part, may be due to the data constraints in studying hedge funds that are notoriously secret in the positions they have taken in various securities. The majority of studies in this literature find that those hedge funds that deliver significant abnormal returns seem to display the behaviour of skill in stock-picking (e.g. see Ackermann, McEnally, and Ravenscraft (1999), Brown, Goetzmann, and Ibbotson (1999) and Kosowski, Naik, and Teo (2007)). However, several studies also find the opposite, such as Asness, Krail, and Liew (2001), Amin and Kat (2003), Malkiel and Saha (2005), and Kat and Palaro (2006). The divergence in views here typically stems from the fact that evaluating hedge funds using returns is difficult; most databases are constructed by self-reporting from the hedge funds thereby giving rise to sample selection, survival, and back-fill biases. Therefore, holdings data can be seen as a better source of evaluating hedge fund skill.

Brunnermeier and Nagel (2004) is an early example of using holdings data from the SEC's 13F filings during the 1998-2000 technology bubble to uncover whether hedge funds had an informational advantage. Griffin and Xu (2009) use holdings data to study long-only equity hedge funds and find weak evidence of the perceived superior skill of hedge fund managers relative to mutual fund managers. However, their data collection procedure means that their sample period stops at

2004 and they identify only 307 hedge fund companies. Bae, Baik, and Kim (2011) also use holdings data over a more recent sample period relative to Griffin and Xu (2009) and identify a large number of hedge fund companies. They find that hedge funds do have superior forecasting ability (almost four times as great) as compared to other institutional investors and uncover a positive link between the level of hedge fund ownership in a stock and the stock's subsequent performance. Titman and Tiu (2011) also use a range of holdings data sources to argue that better-informed hedge funds have less exposure to systematic factor risk and have higher Sharpe ratios. On the other hand, Brown and Schwarz (2013) use 13F filings to investigate whether market participants utilize the holdings data. Overall, they find significant evidence that market participants attempt to take advantage of the information in the 13F filings as there is an increase in the volume traded in securities that appear in the filings. Additionally, they find that these securities display positive excess returns up to two days following such disclosures and that these results are unique to hedge funds disclosures as opposed to those of mutual funds. However, the authors argue that there is no evidence to suggest that copycat trading strategies based on 13F disclosed positions provide any benefit to long-term investors. As evident, the literature has not converged to a universal view of whether one can successfully extract alpha from hedge fund holdings.

This divergence stems from the fact that some studies suffer from not being able to identify enough or all hedge fund companies; others suffer from sample biases. Additionally, some are unable to achieve an appropriate level of granularity to be able to classify hedge funds by style and disentangle positions that are primarily for hedging purposes, for instance, in the case of funds that have derivatives positions and are delta-hedging their exposures. Therefore, a key contribution of this paper is that we use a novel dataset of scraped and thoroughly cleaned fund-level holdings from the SEC 13F forms supplied by Novus. This dataset allows us to leverage the findings in the literature discussed that demonstrates the superior stock picking ability of certain funds. Specifically, because we are able to classify funds as FEHF, we track funds that tend to have a long-term view in picking stocks. Therefore, a key contribution of this chapter is that a qualitative classification, such as the one that Novus provides us, combined with a robust measure of conviction from quarterly positions allows one to extract hedge fund alpha from information on holdings. Thus far, in the literature, as far as we are aware, this sort of qualitative classification and portfolio construction methodology has not been tested.

The remainder of this paper is structured as follows. We first describe the data we use in detail and the classifications that Novus has provided. We then run our benchmark empirical tests on our data following the tests specified in Cohen, Polk, and Silli (2010). Following this, we show measures of conviction and consensus that are suitable for our manager universe. Next, we find that a strategy that combines conviction and consensus, taking the top 50 stock picks of FEHF managers as measured by these measures, is able to extract value from 13F filings and is simple to execute point-in-time. We demonstrate the power of fund classifications by showing the improvement in alpha by tracking the right managers.

III. Data

Our stock return data comes from the Center for Research in Security Prices (CRSP) and covers all listed US common stocks inclusive of Real Estate Investment Trusts (REITs) but excluding American Depositary Receipts (ADRs).[3] The holdings data is from the SEC 13F filings and is obtained on a quarterly basis over the period March 2004 to June 2019. The SEC requires *all* institutions with greater than \$100 million in SEC 13F eligible securities and positions greater than \$200,000 or 10,000 shares in such securities to disclose their holdings through the 13F form. However, given this form does not require the nature of the institution to be disclosed (e.g. whether it is a bank, insurance company, hedge fund, mutual fund, pension fund or corporate), any importance in the information contained in the holdings of the type of filer, cannot be naturally investigated using the holdings data directly from the SEC.

As such, we use a 13F database compiled and maintained by Novus, who classify each of the 13F filings by institutional type and style (e.g. equity hedge fund versus credit hedge fund), thus allowing for an informed study on the holdings of investor types - we outline details on the classification process below. Additionally, Novus also take several validation steps to ensure accuracy of the 13F data collected from the SEC - this is particularly important given that the SEC does not take ownership for ensuring data accuracy.[4] For example, Novus systematically detect

[3] An ADR is a negotiable security that represents securities of a non-US company that trades in US financial markets. For more information, see https://www.sec.gov/investor/alerts/adr-bulletin.pdf.

[4] More can be found on this here: https://www.sec.gov/fast-answers/answers-form13fhtm.html

outliers by looking for anomalous quantities of market values and identifying incorrect position mappings (e.g. by crosschecking the reference CUSIP versus the text stated security), which occurs more frequently than one would expect.[5] Additionally, and very crucially, the database that Novus provides us holds historical holdings data of funds that are now closed. Therefore, our dataset is free from survivorship bias, which tends to be a major problem in research on hedge funds.

A. Fund classification

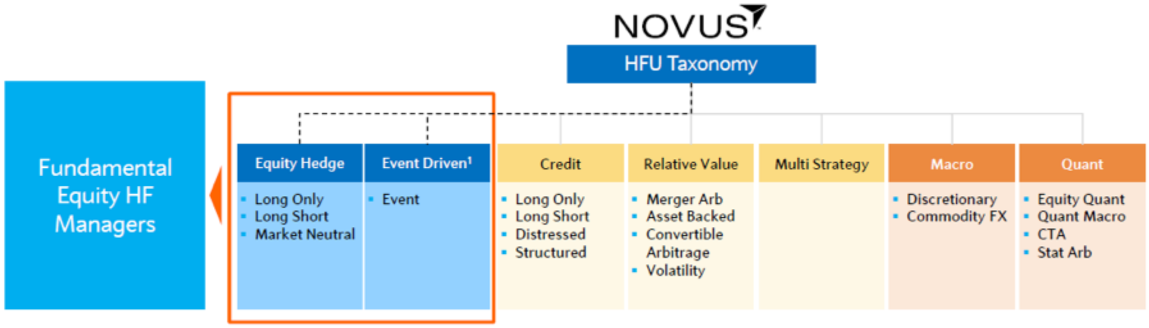
Our study relies on the qualitative classification of filer types provided by Novus to separate out the HFU from the wider set of 13F filers from the SEC. As a second step, Novus also classifies the style of the hedge fund in accordance with the taxonomy of Figure 1, whereby a subset of HFU managers whose mandates require them to be more long-term focused for their equity positions can be formed. Such a subset is named the 'Fundamental Equity Hedge Fund' (FEHF) managers and is made up of the equity hedge and event driven categories. The only other source of holdings data which includes a classification of manager type that we are aware of is the CDA/Spectrum data maintained by Thomson Reuters - however it is well documented that the manager classification in this dataset is particularly unreliable.[6] As such, Novus' curated and maintained dataset, to the best of our knowledge, is the ideal dataset to investigate the informational benefit of holdings data by fund type.

Examples of FEHF managers are Point72 Asset Management, Bayesian Capital Management, Avesta Capital Advisors, Empire Capital Management, and Sigma Capital Management. The

[5] Novus began parsing the SEC's public regulatory filings on a quarterly basis since 2007, with Perl scripts that parsed SEC Edgar's then txt file formats (EDGAR changed to Online XML in March 2014). The system starts by looking for unique CIKs (the identifier the SEC uses to classify an institution) and the specific filings associated with 13F. Those files are loaded into Novus' database, from which positions are connected via their identifier (CUSIP) to a pricing server. It is at this stage that various outlier detection methods are employed to check and clean the data. Additionally, the SEC website is scraped on a continuous basis to capture any late or amended filings (13-FA) each quarter.

[6] There exists a well-documented error in the manager type code in the Thomson Reuters database from 1998 onwards due to an error when integrating data from the former Technimetrics. This error meant that many institutions were incorrectly classified as 'Other' and resulted in the number of investment advisors in 1998 drop from 1,156 to 244 whilst institutions identifying as mutual fund management companies disappeared entirely. These errors have not been fixed and more details can be found at https://wrds-www.wharton.upenn.edu/pages/support/applications/institutional-ownership-research/introduction-thomson-reuters-13f-tr-13f-database-and-its-classification-institutional-investors/.

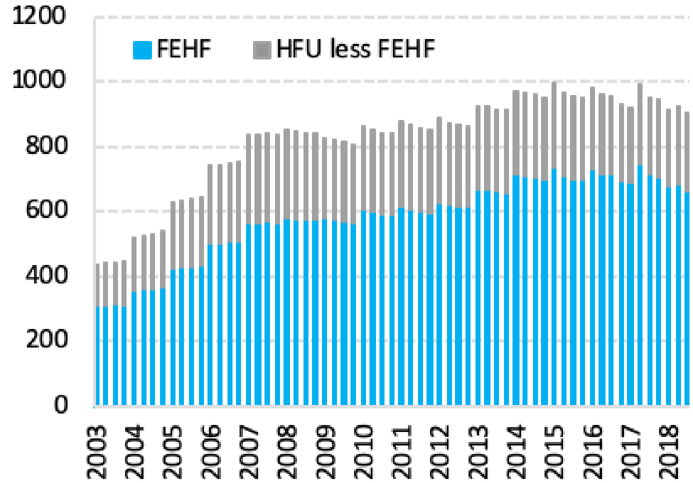
Figure 1. Novus 13F manager classification taxonomy



HFU ex FEHF category includes a broader set of hedge funds including AQR Capital Management, Balyasny Asset Management, and Millennium Capital Management. One may argue now that these hedge fund managers may manage funds that also have long-term views on stocks and thus should be classed as FEHF managers. However, the nature of the SEC filings is such that positions are aggregated on a *manager* level rather than *fund* level. Therefore, if these managers run quantitative funds - as many of them do - and one was to copycat those positions, the effect of the lag between reporting date and filing date is likely to mean that the strategy would suffer significant alpha decay due to the likelihood that these quantitative funds will trade in and out of stocks a lot more frequently than, for instance, equity long only funds. As such, to construct a 13F trading strategy that piggybacks off stock picks with a lag between the reporting date (the quarter-end) and the data release date (45 days after the quarter-end), we believe that identifying FEHF managers is crucial.

B. Summary statistics

To give a flavour of what the holdings data looks like, Tables A1 and A2 (in the Appendix) provide summary statistics for the FEHF (Panel A) and HFU ex FEHF managers (Panel B) over time. It is evident from that the number of fund managers identified in both the FEHF group and the HFU has grown over time (Figure 2, top panel) and so has their total market value (Figure 2, bottom panel). Hedge fund assets under management (AuM), measured by the market value of the long positions appearing in the filings, grew from \$525.77 billion in Q4 2003 to over \$2.3 trillion in Q2 2019. Whist this may in large part be due to the growth in the equity market itself, it highlights



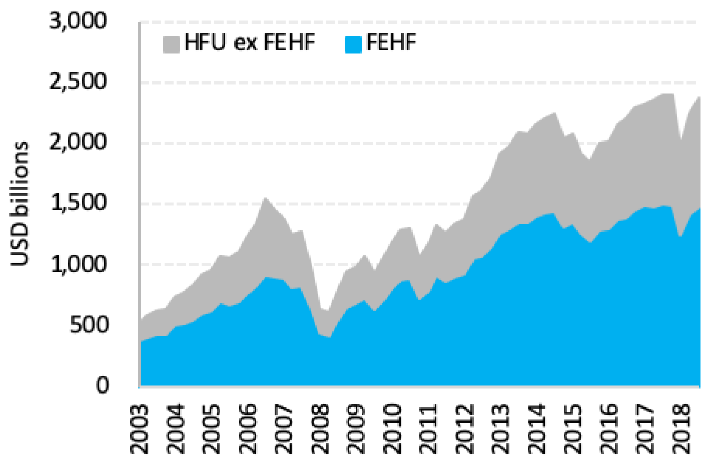


Figure 2. Number of managers (top panel) and total assets under management of managers (bottom panel)

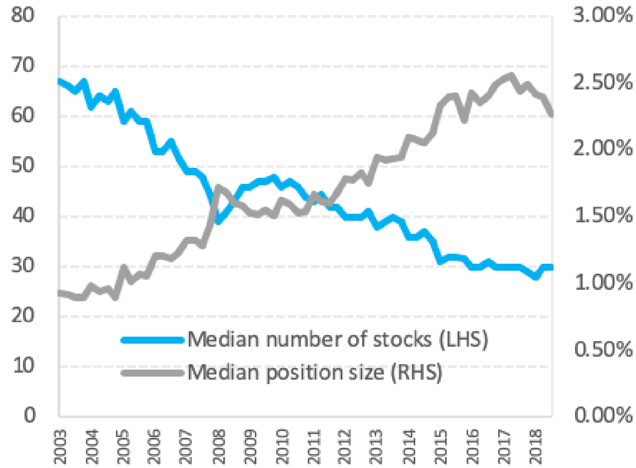
that hedge funds are an influential group of investors.

These summary statistics are also helpful in pointing out some differences between the FEHF and HFU ex FEHF manager groups. Figure 3 plots the median number of securities held by managers against the median position size across managers for each of the manager groups.[7] The overall trend is similar between both manager groups in that portfolio concentration is, in general, rising. That is because we can see that larger position sizes are being taken with a small number of securities. However, the difference between the two manager groups is that the trend is a lot more

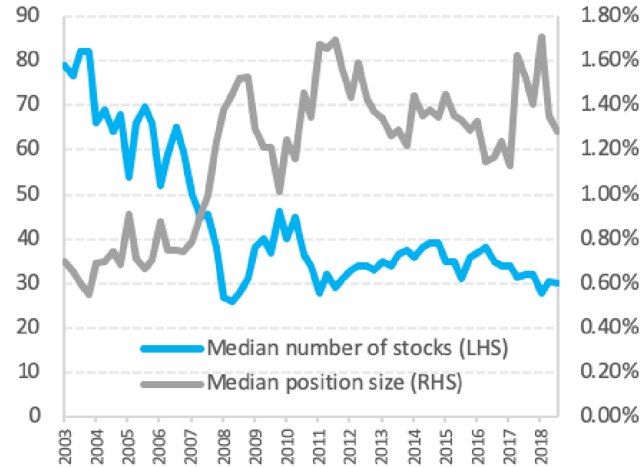
[7]We calculate median position size by first calculating the median position size for each manager at each quarter end and then taking the median across managers

stable for FEHF managers, whereas this fluctuates substantially for HFU ex FEHF managers. In fact, the HFU ex FEHF managers behave somewhat cyclically: during the 2008/09 financial crisis, for example, median position size spiked and the median number of securities dipped drastically to a low. It is unclear why this happened. We do not address this issue in detail here as it is beyond the scope of this chapter. We note however that Khandani and Lo (2007) discuss the behaviour of quantitative funds during the crisis, starting in August 2007, where they find that during the week of August 6, 2007 - which aligns with the dip in median number of securities for HFU ex FEHF managers in Figure 3 (right panel) - a number of quantitative long/short equity funds experienced drastic losses caused by forced liquidation, possibly due to margin calls or risk reduction. These sudden unwinds are likely what explains the dip that we observe in Figure 3 (right panel).

Two other interesting differences between the two groups of managers which are revealed by Tables A1 and A2 are: One, we note that there was an increase in the unique number of securities in the HFU ex FEHF group versus a decrease in the FEHF manager group. This shows that portfolios indeed became a lot more concentrated into fewer names for FEHF manages in aggregate. Two, the total number of securities appearing in the HFU ex FEHF filings almost doubled, whereas the increase was not as drastic for the FEHF group. Additionally, and very importantly, portfolio concentration, although rising for both sets of managers, is a lot more stable through time for FEHF managers. Thus, if one is to build strategies based on conviction, as we do, it is important to base such a strategy on a stable trend such as the one shown in Figure 3, which clearly shows that portfolio concentration is a lot more prevalent through time for FEHF managers. Additionally, despite the median number of stocks being similar in magnitude for both manager groups, the median position sizes are quite different. For instance, the median position size climbs to just under 2.5% for FEHF managers at the end of the sample period whereas it is approximately 1.3% for HFU ex FEHF managers. One reason why this may happen is because FEHF managers, as we suspect, tend to display independent conviction towards their stock picks and will do this by taking larger position sizes. Another reason is that even though the median number of stocks are similar in magnitude across manager groups, the total number of stocks is very different as evident in Tables A1 and A2. This means that HFU ex FEHF managers (which include quantitative funds) allocate smaller portfolio weights across a larger number of stocks compared to FEHF managers.



(a) FEHF



(b) HFU ex FEHF

Figure 3. Portfolio concentration of FEHF managers (left panel) versus HFU ex FEHF managers (right panel)

Figure 4 plots the average holding period on a position level across FEHF and HFU ex FEHF managers. We calculate holding periods by tracking stocks that appear in manager's portfolios through consecutive quarters. If a stock appears in the manager's filings in Q1, disappears in Q2, but reappears in Q3 we assign that stock as two separate positions. We then calculate how many quarters each position is held for and calculate the average holding period per manager per position. We find that FEHF managers tend to have longer holding periods as compared to HFU ex FEHF managers. The median holding period on a position level for FEHF managers is 12 months compared to 9 months for HFU ex FEHF managers. This is line with our intuition as to why tracking FEHF managers is important: these are the managers that tend to have longer-term views on stock picks and so for any quarterly holdings data that allow for an information advantage in stock-picking taking into account the lag between the reporting date and filing date, one must track those positions that are unlikely to be sold within shorter time periods. As such, we are comforted by the FEHF manager group who hold stocks on average for very long periods, some even holding stocks for longer than 54 months. This is not the case for the HFU ex FEHF manager group.

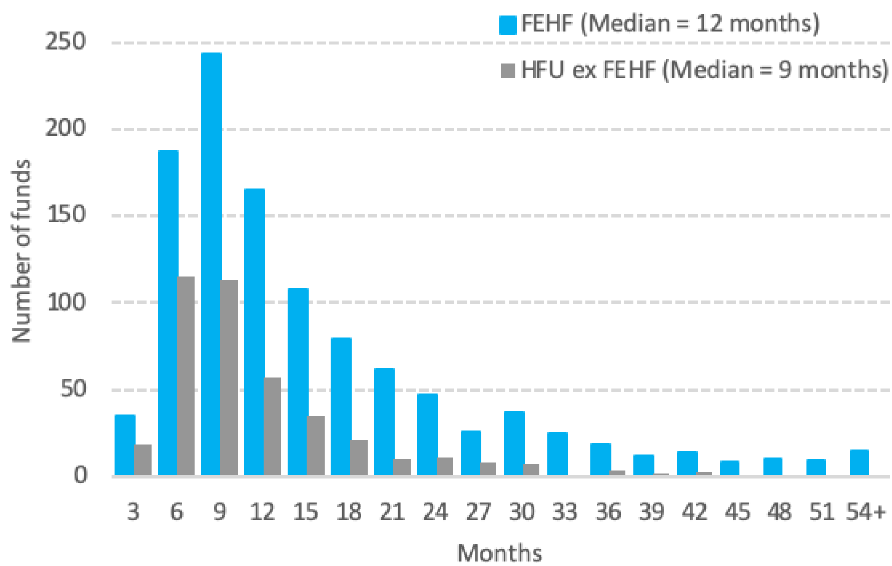


Figure 4. Average position level holding period for FEHF versus HFU ex FEHF

C. Benchmark methodology: Best Ideas

Our benchmark tests closely follow the methodology introduced in Cohen, Polk, and Silli (2010). The authors have four ways to measure the best idea of managers and subsequently construct a 13F strategy. We slightly modify the measures and explain the way in which we test them below.

Market tilt. The first test compares the weights of stocks in each manager's portfolio to the weights of the same stocks in the S & P 500 index. Denote the weight of stock i as λ_{ijt} in manager j's portfolio at time t and as λ_{iMt} in the S & P 500 index at time t. We then compute the difference between the weights for each stock i at time t

$$MarketTilt_{ijt} = \lambda_{ijt} - \lambda_{iMt}, \tag{1}$$

for each manager j and rank each stock in the manager's portfolio. The stock with the maximum tilt is the manager's 'best idea'. Then, for each stock, we count how many times it appeared as the 'best idea' across managers at time t and rank the stocks again. We take long positions in these stocks and rebalance quarterly, 47 days after the quarter end. This is one difference in our tests compared to the original tests in Cohen, Polk, and Silli (2010) as the authors run their tests with a zero lag (i.e. executing on the quarter-end day). This is not possible in practice as the filings are only made public 45 days after the quarter-end as outlined above.

CAPM tilt. The second test is similar to the first test in that it compares weights of stocks in managers' portfolios to the weights of the same stocks in the S & P 500 index. However, as the authors argue, the first test, market tilt, is not motivated by finance theory and is fairly simple in nature. The second test therefore is centred on the Capital Asset Pricing Model (CAPM) and captures the return generating process of equity returns. In this test, the idiosyncratic risk component of each stock is estimated by computing the mean square error (MSE) from a regression of a daily time series of stock i's excess returns over the risk-free rate on market excess returns over a period of 60 days. There are two implicit assumptions here that the authors point out: (i) the model captures the factor structure of returns; and (ii) the goal of each manager is to select portfolios that maximise the information ratio - that is, she maximises expected excess returns relative to volatility. This is a plausible assumption. Money managers, whether those in the hedge fund world or the mutual fund world, tend to focus a lot on Sharpe ratios as performance metrics and therefore this assumption need not be very restrictive. Therefore, in this test, the manager's weighting in each stock relative to the benchmark market portfolio - the S & P 500 index - is given by the expected risk-adjusted return. Following the same procedure as the market tilt test, we rank stocks in each manager's portfolio by computing

$$CAPMTilt_{ijt} = \sigma_{it}^2(\lambda_{ijt} - \lambda_{iMt}) \tag{2}$$

and take the stock with maximum value of $CAPMTilt_{ijt}$ for each manager j and proceed in a similar way to the first test.

Portfolio tilt. The third test is similar to the first test in that it compares weights of stocks in managers' portfolios to the weights of stocks in a benchmark portfolio. However, a key problem with the first test is that managers may not be picking high conviction stocks that also appear in the S & P 500 index. This is particularly true for our case as we study hedge funds rather than mutual funds; though the same problem may arise when studying mutual funds as managers may not be benchmarked at all or may not be benchmarked against the S & P 500 and thus choosing the relevant index to compare weights to becomes difficult. Therefore, to counter this issue, we construct a value weighted portfolio for each manager's positions and compare the weights appearing in their 13F filings to that portfolio. Again denoting λ_{ijt} as the weight of stock i in manager j's portfolio at time t and, similarly, denoting λ_{ijVt} as the weight of stock i in the value-weighted portfolio made up of the stocks in manager j's portfolio at time t, we compute

$$PortfolioTilt_{ijt} = \lambda_{ijt} - \lambda_{ijVt} \tag{3}$$

for each manager j and take the stock with the maximum tilt by this measure. We proceed with the same steps on portfolio construction as the above tests.

CAPM portfolio tilt. The final test is the CAPM equivalent of the portfolio tilt test. Therefore, we repeat the same steps as the portfolio tilt test but adapt it so that it is based on the CAPM. This means we compute

$$CAPMPortfolioTilt_{ijt} = \sigma_{it}^2(\lambda_{ijt} - \lambda_{ijVt}) \tag{4}$$

and proceed similarly to the earlier tests.

D. *The distribution of best ideas*

Cohen, Polk, and Silli (2010) argue that there is very little overlap in the best ideas of managers. They suggest that in theory the number of best ideas that exist in the industry at any point in time may well be as large as the number of the managers (no overlap) or as few as one (100% overlap). They report results for the market tilt measure and find that 62% of all best ideas over the sample period were held by only one manager at the same time. Similarly, we look at the distribution of best ideas of the Figure 5. We find that 60% of managers' best ideas, when measured by their maximum portfolio tilt, do not overlap in the entire sample period (Q1 2004 - Q2 2019). When measured by market tilt, on the other hand, this number drops to just over 35%.

Figure 6 looks at the distribution of best ideas for FEHF managers (left panel) and HFU ex FEHF managers (right panel). We find that for both measures - portfolio and market tilt - there is less

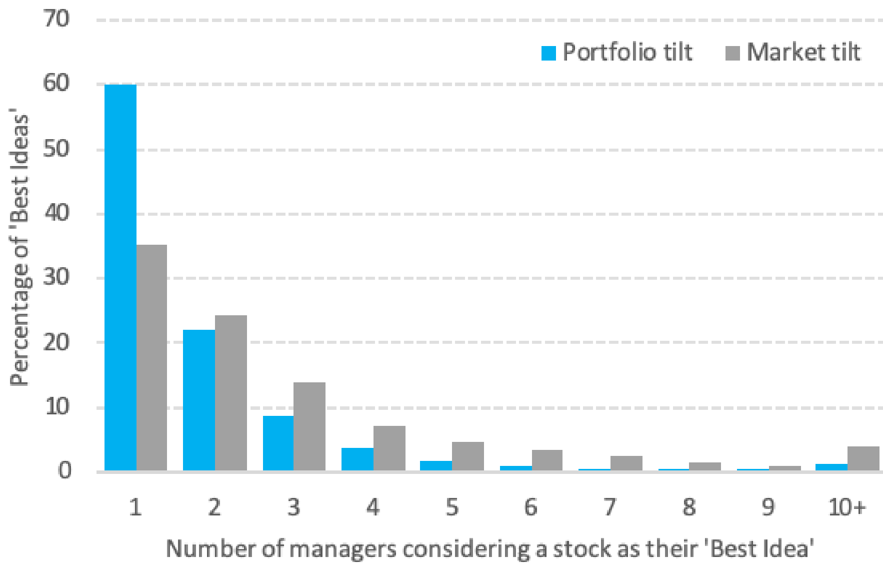
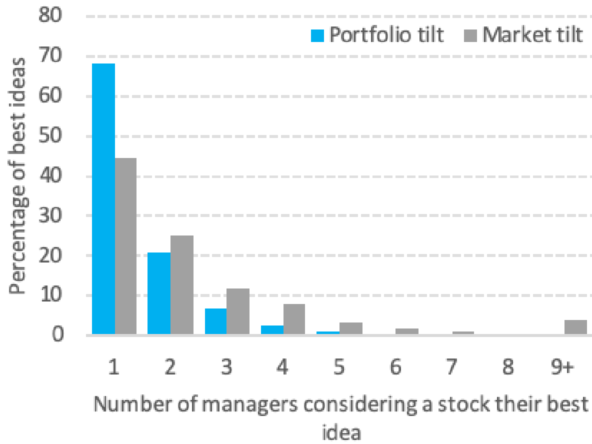
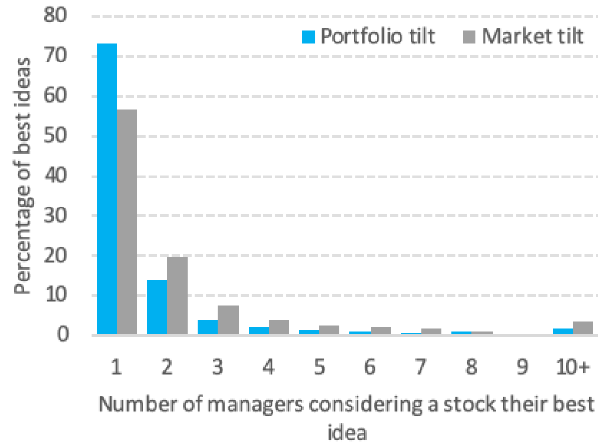


Figure 5. The distribution of Best Ideas across HFU managers

overlap compared to on the entire HFU. Hence, similar to Cohen, Polk, and Silli (2010), we find that the best idea portfolios identified using these two tests do not consist of a handful of popular names. Rather, these names are based on the picks of hundreds of hedge fund managers that have somewhat independently displayed high conviction in at least one stock. The fact that we find even less overlap compared to mutual fund managers in Cohen, Polk, and Silli (2010) means that the hedge fund managers in our sample tend to display even more independent conviction compared to their mutual fund counterparts. Finally, the fact that measuring overlap by portfolio tilts, rather that by market tilts, shows less overlapping best ideas indicates that it the value-weighted benchmark may be a better measure to identify conviction in our sample. This may be due to the fact that hedge fund managers tend not to be benchmarked against a market portfolio. We discuss this later in more detail.



(a) FEHF



(b) HFU ex FEHF

Figure 6. The distribution of Best Ideas across FEHF managers (left panel) versus HFU ex FEHF managers (right panel)

IV. Empirical results

A. *Benchmark tests on HFU*

Figure 7 and Table I show the results of the benchmark methodology on HFU. The tests have all been run with a 47-day lag after the quarter-end. The outperformance of each strategy relative to the S&P 500 total return index shows that, despite the lag, there is still value to be extracted from the holdings data. Out of the benchmark tests executed on the HFU manager group, the portfolio tilt measure delivers the best risk-adjusted performance with a Sharpe ratio of 0.69 and annual outperformance of 3.80%. Our tests are all run with a 47-day lag after the quarter-end and we equally weight unique names by each measure. This is slightly different to the methodology in Cohen, Polk, and Silli (2010) who execute their tests at the quarter-end date (when the filings are not public) and they overweight names that is selected as a manager's best idea. Our tests are more strict in this regard. First, we accept that there is bound to be a decay in alpha between the quarter-end and the day our strategy rebalances (47 days later). Second, we do not overweight the most popular names in these tests so as to not piggyback on a select few stocks' performance. Despite this, the strategy delivers strong positive outperformance.

Table II shows the loadings on the five Fama and French (2015) and momentum factors using

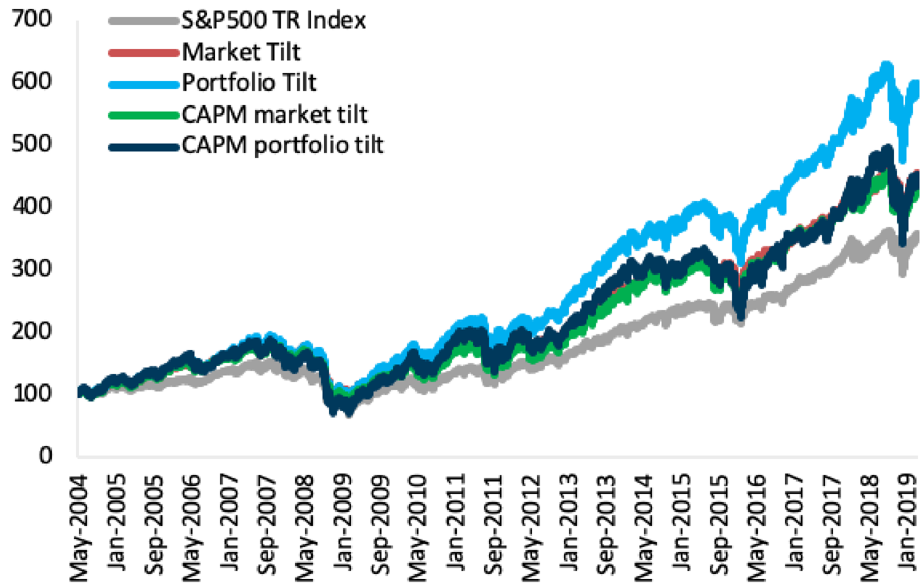


Figure 7. Best Ideas tests on HFU (cumulative returns), May 2004 - June 2019

Table I. Performance statistics of Best Ideas on HFU

	S&P 500	Market tilt	Portfolio tilt	CAPM market tilt	CAPM portfolio tilt
Ann. avg rets	10.34%	12.19%	14.14%	12.31%	12.74%
Ann. vol	18.31%	19.24%	20.39%	21.96%	23.85%
Sharpe ratio	0.56	0.63	0.69	0.56	0.53
Max drawdown	55.25%	54.07%	57.24%	58.15%	63.02%
Relative out-performance					
Ann. outperformance		1.85%	3.80%	1.97%	2.40%
Tracking error		2.87%	6.29%	5.91%	9.98%
Information ratio		64.47%	60.32%	33.31%	24.02%
Return correlation		0.99	0.95	0.97	0.92

monthly returns. Despite the strategies delivering annual outperformance of at least 1.85% relative to the S&P 500, only two strategies deliver statistically significant and positive alpha - the market tilt portfolio tilt strategies. All strategies load positively and significantly on the market risk premium (SP500 ER) and the size factor (SMB). All strategies, except for CAPM market tilt, load negatively on the investment factor (CMA). Cohen, Polk, and Silli (2010) find that the best ideas of mutual fund managers covary with small, high-beta, growth stocks that have recently performed well (i.e. positive momentum). Our findings on the HFU manager group agree on the result of size and high-beta, but we do not find any covariance with momentum, in contrast with the same tests run on mutual fund managers. In their paper, the authors argue that a loading on momentum reflects a failure to rebalance on the part of managers because stocks that have a substantial tilt (in terms of the Best Idea measures) tend to be those that have performed well over the past year. Therefore, the high conviction score for those stocks tends to be driven by the past growth in stock price indicating a failure to rebalance after a stock has performed well. The fact that we do not find this in hedge funds is testament to our intuition that the 'who' is an incredibly important factor forming copycat strategies based on 13F filings. In fact, as we show later, this effect disappears when running the tests on the HFU ex FEHF sample alone thereby further providing support for our intuition. All strategies, except for CAPM market tilt, have a negative, statistically significant, tilt to CMA. This means that the best ideas of the HFU managers covary with stocks that invest aggressively and thus are likely to be high beta stocks, which is in line with our findings.

B. Identifying the Best Ideas of hedge fund managers

The benchmark tests showed that alpha can indeed be extracted from 13F hedge fund holdings. The tests indicated that the portfolio tilt measure performs best. A key assumption in Cohen, Polk, and Silli (2010) to motivate the tests they use is that managers are maximising their information ratios. This is typically the case when a manager is benchmarked against an index with which they attempt to minimise tracking error. Hedge funds, unlike mutual funds that are considered in Cohen, Polk, and Silli (2010), are highly unlikely to be benchmarked or consider tracking error as a constraint in their portfolio selection. Therefore, one may question whether there is a better way to identify the ex-ante best ideas of hedge fund managers. The portfolio tilt measure is a good starting point: it finds positions in which the manager is overweight relative to a benchmark portfolio made up of all the names in the manager's portfolio with with weightings proportional to their market capitalizations. As we would like to drop the notion of benchmarking entirely, the most natural test is to identify those positions with the highest absolute weights per manager. Another natural test for the aggregate hedge fund universe is to simply count how many hedge funds hold a particular stock. Qian and Zhong (2018)find that hedge funds have special

Table II. Best Ideas on HFU Fama-French 5 factor and Momentum analysis

Notes: The regression estimated here is $r_{i,t} - r_f = \alpha_i + \beta_{Mkt}(SP500TR - R_f)_t + \beta_{SMB}(SMB)_t + \beta_{HML}(HML)_t + \beta_{RMW}(RMW)_t + \beta_{Mom}(Mom)_t$ and is based on monthly returns data. The factors have been sourced from Kenneth French's website. SP500 ER in the table below denotes the excess return of the S&P 500 Total Return index in excess of the risk-free rate provided by Kenneth French. Stars indicate statistical significance: * $p<0.1$, ** $p < 0.05$, *** $p < 0.001$. Returns have been computed for the period May 2004 to June 2019. .

	Market tilt	Portfolio tilt	CAPM market tilt	CAPM portfolio tilt
Ann. avg rets	12.19%	14.14%	12.31%	12.74%
Alpha	0.12%***	0.28%***	0.10%	0.12%
	[2.357]	[4.185]	[1.066]	[1.195]
SP500 ER	1.001***	0.999***	1.103***	1.093***
	[69.800]	[52.661]	[41.613]	[38.316]
SMB	0.247***	0.702***	0.438***	1.043***
	[9.793]	[21.099]	[9.428]	[20.841]
HML	0.0241	-0.0351	0.0805	-0.0546
	[0.811]	[-0.893]	[1.465]	[-0.922]
RMW	0.050	-0.0451	-0.1210*	-0.197**
	[1.346]	[-0.918]	[-1.764]	[-2.666]
CMA	-0.0958**	-0.228***	0.999*	-0.338***
	[-2.324]	[-4.192]	[-1.836]	[-4.120]
Mom	0.011	0.0235	-0.0215	-0.0183
	[0.889]	[1.383]	[0.366]	[0.476]
Adj.-R^2	0.98	0.97	0.95	0.95

connections across the industry that gives them, as a group, an information advantage over the rest of the market. Therefore, the more hedge funds that hold a name in their portfolio may be an indication that the aggregate hedge fund group has information on a stock that they believe will rise in value.

To test conviction, we first sort stocks in each manager's portfolio at each quarter-end into quintiles based on their position sizes (i.e. the market value of the stock in the portfolio at each quarter-end divided by the total portfolio of the manager at each quarter-end). We then select stocks that appear in the top quintile across managers and construct an equally weighted portfolio of unique names. The portfolio is rebalanced with a 47-day lag after the quarter-end. To test consensus, we simply count the number of managers holding each stock and then sort stocks into quintiles by their count. We construct an equally weighted portfolio of the unique stocks appearing in the top

Table III. Conviction and consensus on HFU, May 2004 - June 2019

	Conviction	Consensus
Ann. avg rets	13.20%	13.05%
Ann. vol	20.28%	19.99%
Sharpe ratio	0.65	0.65
Max drawdown	58.21%	55.02%
Relative outperformance		
Ann. outperformance	2.86%	2.71%
Tracking error	6.63%	5.46%
Information ratio	43.15%	49.57%
Return correlation	0.95	0.96

quintile by this measure and rebalance 47 days after the quarter-end. We report results for our conviction and consensus tests on HFU in Table III. We find that both conviction and consensus deliver Sharpe ratios of 0.65, with conviction outperforming on a returns-basis only marginally. Conviction has a slightly higher volatility and higher maximum drawdown, but the differences are minor.

Table IV reports the results from a regression of each strategy's returns on the Fama and French (2015) 5-factors and a momentum factor based on monthly returns. The results are again very similar for both conviction and consensus, which deliver statistically significant six-factor alpha of 0.22% and 0.18% respectively. Consensus has a slightly higher loading on the market risk premium, which is not surprising given that it purely tracks what the majority of hedge funds in the HFU are holding - hence, consensus picks up the market risk premium for the most part but has an additional loading on stocks that hedge funds, in aggregate, believe some additional alpha is to be found. Both conviction and consensus have a significant tilt to SMB, which is similar to the result found in our benchmark methodology as well as Cohen, Polk, and Silli (2010).

Neither conviction nor consensus was a clear winning strategy when running the tests on the HFU manager group. Although we have seen that changing the 'how' matters, we now see that the 'who' matters as well. We run the same tests on the FEHF and HFU ex FEHF manager groups and see that the results are indeed different. Table V reports the performance statistics and shows that both conviction (panel A) and consensus (panel B) outperform for FEHF managers as compared

Table IV. Fama-French 5 factor and Momentum analysis for conviction versus consensus on HFU

Notes: The regression estimated here is $r_{i,t} - r_f = \alpha_i + \beta_{Mkt}(SP500TR - R_f)_t + \beta_{SMB}(SMB)_t + \beta_{HML}(HML)_t + \beta_{RMW}(RMW)_t + \beta_{Mom}(Mom)_t$ and is based on monthly returns data. The factors have been sourced from Kenneth French's website. SP500 ER in the table below denotes the excess return of the S&P 500 Total Return index in excess of the risk-free rate provided by Kenneth French. Stars indicate statistical significance: * p<0.1, ** $p < 0.05$, *** $p < 0.001$. Returns have been computed for the period May 2004 to June 2019. .

	Conviction	Consensus
Ann. avg rets	13.20%	13.05%
Alpha	0.22%***	0.18%***
	[3.995]	[2.878]
SP500 ER	0.982***	1.01***
	[63.401]	[57.038]
SMB	0.782***	0.632***
	[28.740]	[20.035]
HML	-0.017	-0.040
	[-0.513]	[-1.099]
RMW	-0.039	0.014
	[0.329]	[0.312]
CMA	-0.118***	-0.103**
	[-2.650]	[-2.033]
Mom	0.0115	-0.016
	[0.827]	[0.310]
Adj.-R^2	0.98	0.97

to HFU ex FEHF managers.

We next study the loadings on the Fama and French (2015) and momentum factors for conviction and consensus on both FEHF and HFU ex FEHF. The results are presented in Table VI. We see that although alpha is statistically significant for both FEHF and HFU ex FEHF, it is larger in magnitude for FEHF. The results show that the loadings on the Fama and French (2015) and momentum factors are similar to when the strategies are run on HFU. The exception is when running consensus on FEHF, the strategy has a statistically significant negative tilt to value and quality. This is interesting as it shows that the most popular stocks held by FEHF managers are neither value nor growth stocks. If the strategy was to covary positively with HML and RMW, one could argue that managers in our sample are simply picking up factor premia. We find that this is not the case for FEHF managers when looking at consensus.

Table V. Performance statistics for Conviction (Panel A) and Consensus (Panel B) on FEHF versus HFU ex FEHF, May 2004 - June 2019

Panel A: Conviction	FEHF	HFU ex FEHF	*Panel B: Consensus*	FEHF	HFU ex FEHF
Ann. avg rets	13.53%	12.65%	Ann. avg rets	13.30%	12.65%
Ann. vol	20.44%	20.01%	Ann. vol	20.26%	19.73%
Sharpe ratio	0.66	0.63	Sharpe ratio	0.66	0.64
Max drawdown	57.56%	57.05%	Max drawdown	55.38%	55.39%
Relative outperformance			**Relative outperformance**		
Ann. outperformance	3.19%	2.31%	Ann. outperformance	2.97%	2.31%
Tracking error	6.71%	5.72%	Tracking error	5.73%	5.16%
Information ratio	47.61%	40.35%	Information ratio	51.77%	44.76%
Return correlation	0.95	0.96	Return correlation	0.96	0.97

C. Combining conviction and consensus

The results shown above indicate that both conviction and consensus are promising measures to extract the best ideas of hedge fund managers without relying on any assumptions relating to benchmarking (i.e. that the manager maximises an information ratio). Conviction and consensus both yield similar results regardless of which manager group we track. To determine a more economically robust measure, we note that, as shown in Figure 8, that conviction and consensus are positively correlated in the cross-section of stocks but that this correlation is higher for HFU ex FEHF managers as compared to FEHF managers. This follows the analysis we presented in Figure 6, where we studied the overlap of best ideas measured by market and portfolio tilt. The reasoning there was we would ideally not like to pick up the most common stocks held by managers. Rather, we would like to select names that managers' have independently displayed a strong positive long-term view. We do similar analysis here. We calculate the correlation by assigning a percentile score based on our measure of conviction, as detailed above, and based on consensus for each stock held by the respective manager group. We then compute the cross-sectional correlation coefficient for all stocks. There are two main points to notice from the correlation result. First, that, as

Table VI. Fama-French 5 factor and Momentum analysis for Conviction and Consensus on FEHF versus HFU ex FEHF

Notes: The regression estimated here is $r_{i,t} - r_f = \alpha_i + \beta_{Mkt}(SP500TR - R_f)_t + \beta_{SMB}(SMB)_t + \beta_{HML}(HML)_t + \beta_{RMW}(RMW)_t + \beta_{Mom}(Mom)_t$ and is based on monthly returns data. The factors have been sourced from Kenneth French's website. SP500 ER in the table below denotes the excess return of the S&P 500 Total Return index in excess of the risk-free rate provided by Kenneth French. Stars indicate statistical significance: * p<0.1, ** p < 0.05, *** p < 0.001. Returns have been computed for the period May 2004 to June 2019. .

Panel A: Conviction	FEHF	HFU ex FEHF	*Panel B: Consensus*	FEHF	HFU ex FEHF
Ann. avg rets	13.53%	12.65%	Ann. avg rets	13.30%	12.65%
Alpha	0.26%***	0.17%***	Alpha	0.20%***	0.16%***
	[4.534]	[2.912]		[2.994]	[2.435]
SP500 ER	0.982***	0.987***	SP500 ER	1.016***	1.008***
	[61.256]	[60.382]		[56.447]	[56.447]
SMB	0.798***	0.634***	SMB	0.643***	0.589***
	[28.357]	[22.101]		[19.833]	[18.803]
HML	0.003	-0.008	HML	-0.090**	-0.002
	[0.097]	[-0.231]		[-2.340]	[-0.044]
RMW	-0.815**	-0.025	RMW	-0.090**	0.034
	[1.964]	[-0.602]		[-2.340]	[0.731]
CMA	-0.170***	-0.122**	CMA	-0.149***	-0.037
	[-3.697]	[-2.594]		[-2.808]	[-0.714]
Mom	0.0284**	0.0178	Mom	-0.005	-0.017
	[1.974]	[0.227]		[-0.302]	[-1.076]
Adj.-R^2	0.98	0.98	Adj.-R^2	0.97	0.97

mentioned, the correlation for FEHF managers is lower than for the HFU ex FEHF managers. This shows that there is less overlap in high conviction stocks and high consensus stocks for FEHF managers and thus the distribution of names is larger in the FEHF group. Second, the correlation is a lot more stable for FEHF managers as compared to HFU ex managers through time. For the latter group, the correlation starts at roughly 34% at the start of our sample period (Q1 2004) and ends at just over 50% in Q2 2019. On the other hand, for FEHF managers, the correlation stays stable – albeit with a brief spike during the global financial crisis - around the 20% mark.

Based on our correlation analysis, we explore the impact of combining conviction and consensus in a single strategy. There are understandably many ways that this can be done. We overlay our consensus tests with a measure of conviction by removing stocks that do not satisfy a conviction threshold for each manager. That is, we compute the position size of each stock in each manager's

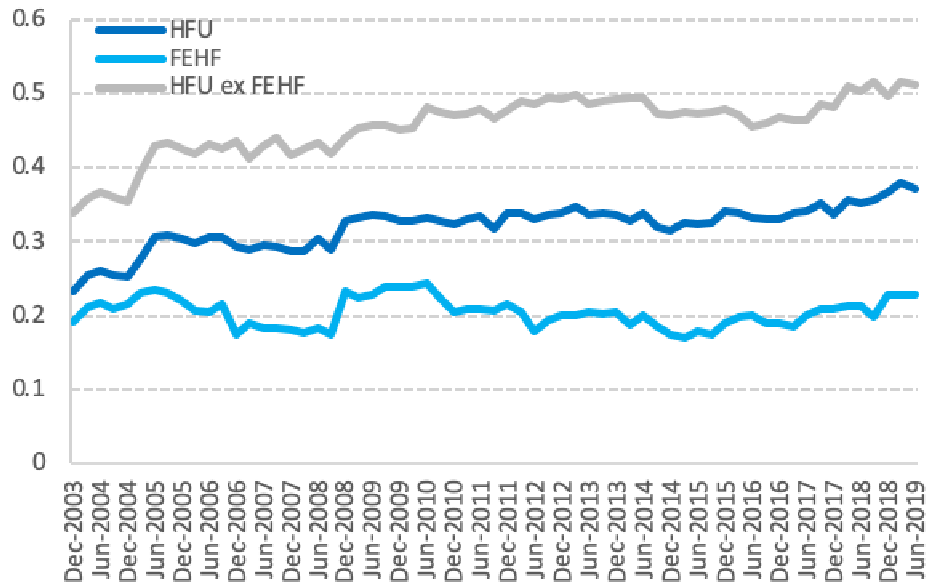


Figure 8. Cross-sectional correlation of conviction against consensus.

portfolio at the quarter-end and remove those that are under a threshold (ranging from 2.5% to 10%). Figure 9 plots the outperformance of each strategy (relative to the S&P 500 total return index) and the associated Sharpe ratio. It is clear that the returns are monotonically increasing in the conviction threshold. Cohen, Polk, and Silli (2010) conduct similar analysis by considering the returns to their tests when picking only the top percentiles of active tilts. They similarly find that increasing this threshold improves performance.

D. Occam's razor: simple execution, improved alpha

Strategies combining conviction and consensus delivered promising outperformance, particularly by tracking FEHF managers. However, these strategies were based on selecting stocks by quintiles of their consensus scores meaning that we may not have a stable number of stocks in our portfolio through time. We now test a simple strategy that is able to improve the alpha available in 13F filings and is realistically tradable with respect to the number of positions in the portfolio.

We start by removing FEHF managers that have fewer than five stocks in their portfolios for two reasons to ensure that we are not picking up managers who behave like activist investors, for

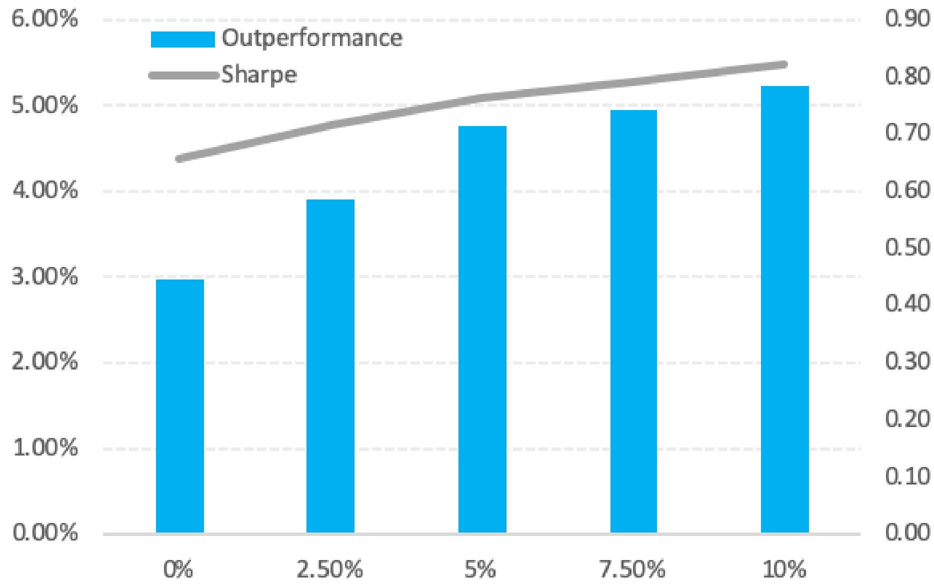


Figure 9. Returns to strategies combining conviction and consensus on FEHF.

instance, or those managers that may have been classed as FEHF but are not pure stock pickers. We similarly filter out the positions that we select in the 13F filings. First, we select stocks whose portfolio weight within the eligible FEHF funds is greater than 7.5%. We choose 7.5% as we see that, based on Figure 9, the outperformance starts to plateau beyond the 5% conviction threshold. Although a 10% conviction threshold is the best performer, we believe 7.5% is sensible so as to somewhat improve portfolio diversification and to not rely on one or two managers that display conviction more aggressively than others. Next, we rank these high conviction stocks by the number of unique managers that hold them. The number of unique managers holding these stocks is the conviction score. We then construct a portfolio by taking the top 50 (by conviction score) of the remaining stocks after all filters are applied. The portfolio is rebalanced quarterly 47 days after the quarter-end. The portfolio's cumulative returns are shown in Figure 10 and the performance statistics in Table VII. As shown, the FEHF strategy outperforms the HFU ex FEHF strategy. The strategy that tracks FEHF managers has an annualized average return of 14.14% with a Sharpe ratio of 0.75 whereas the strategy that tracks HFU less FEHF managers has an annualized return of 10% (thereby underperforming the S&P 500 by 0.34%) and a Sharpe ratio of 0.52.

In Table VIII, we report results of a six-factor Fama and French (2015) and momentum regression.

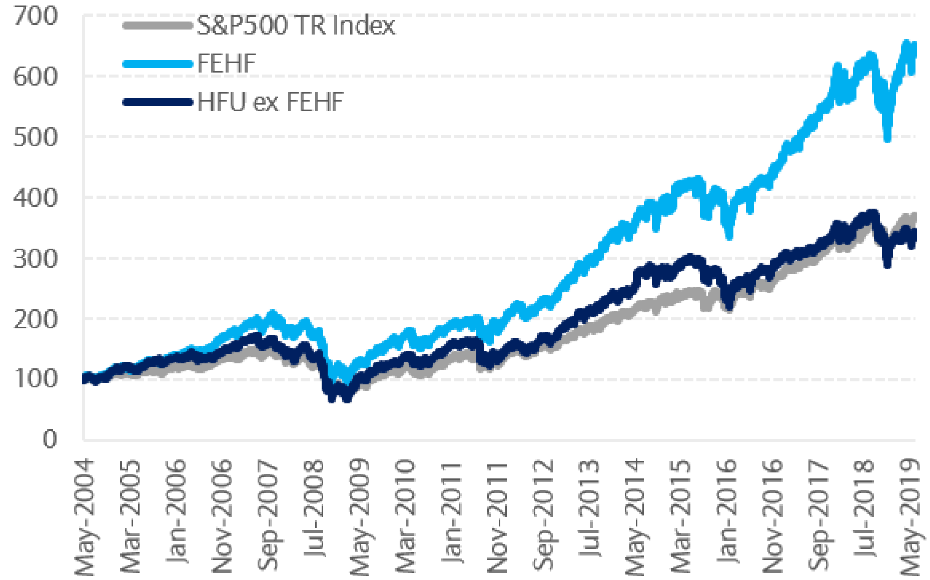


Figure 10. Cumulative returns of a strategy that combines conviction and consensus, May 2004 - June 2019

Table VII. Performance statistics for a strategy that combines conviction and consensus, May 2004 - June 2019

	FEHF	HFU ex FEHF
Ann. avg rets	14.14%	10.00%
Ann. vol	18.87%	19.12%
Sharpe ratio	0.75	0.52
Max drawdown	54.40%	61.43%
Relative outperformance		
Ann. outperformance	3.80%	-0.34%
Tracking error	5.15%	6.70%
Information ratio	73.80%	-5.04%
Return correlation	0.96	0.94

The FEHF strategy delivers a statistically significant six-factor alpha of 0.32%. However, this is not the case for the HFU ex FEHF strategy, delivering a negative and statistically insignificant six-factor alpha. The FEHF strategy loads positively on the market risk premium, with a beta close to 1, positively and significantly on size, negatively and significantly on investment and negatively, but not significantly, on value and quality. The strategies do not have tradability filters when it comes to stock selection and so the loadings on SMB are plausible. Implementing filters on market capitalization and liquidity in the stock selection process does not materially affect our results (see

Table VIII. Fama-French 5 factor and Momentum analysis for our trading strategy

Notes: The regression estimated here is $r_{i,t} - r_f = \alpha_i + \beta_{Mkt}(SP500TR - R_f)_t + \beta_{SMB}(SMB)_t + \beta_{HML}(HML)_t + \beta_{RMW}(RMW)_t + \beta_{Mom}(Mom)_t$ and is based on monthly returns data. The factors have been sourced from Kenneth French's website. SP500 ER in the table below denotes the excess return of the S&P 500 Total Return index in excess of the risk-free rate provided by Kenneth French. Stars indicate statistical significance: * p<0.1, ** p < 0.05, *** p < 0.001. Returns have been computed for the period May 2004 to June 2019. .

	FEHF	HFU ex FEHF
Ann. avg rets	14.14%	10.00%
Alpha	0.32%***	-0.003%
	[3.444]	[0.979]
SP500 ER	0.984***	0.958***
	[37.724]	[26.255]
SMB	0.142***	0.413***
	[3.094]	[6.456]
HML	-0.003	0.094
	[-0.059]	[1.244]
RMW	-0.023	0.084
	[0.735]	[0.891]
CMA	-0.412***	-0.475***
	[-5.499]	[-4.532]
Mom	0.009	-0.108***
	[0.370]	[-3.309]
Adj.-R^2	0.93	0.88

Appendix). The HFU ex FEHF strategy's loading on SMB is much larger in magnitude indicating that it is much more likely to pick up very small cap stocks. An interesting point to note is that the HFU ex FEHF strategy has a statistically significant negative tilt away from momentum whereas the FEHF strategy does not have a statistically significant coefficient. This confirms our intuition that the HFU rebalances out of stocks that have positive momentum, which is contrary to what Cohen, Polk, and Silli (2010) find for mutual funds.

Our results therefore clearly show that a simple strategy constructed using a combination of conviction and consensus delivers outperformance and statistically significant six-factor alpha when tracking FEHF managers. In summary, the 'who' and the 'how' must be considered in equal weight when attempting to systematically extract alpha from 13F filings.

V. Conclusion

In this paper, we study the value in 13F filings through the lens of investment strategies that piggyback off hedge funds' highest conviction stock picks. We first show that hedge fund alpha can be extracted systematically from the filings, despite the 45-day lag in the holdings being made publicly available after the quarter end. We do this by running benchmark tests identified in the academic literature on a universe of hedge fund managers identified by a proprietary data provider, Novus, which identifies hedge fund managers in the 13F filings and detects outliers in the filings. Further, Novus identifies a subset of managers that are classed as FEHF. These managers tend to have longer-term views on equity. We show that tracking these funds' best ideas improves alpha. These tests, however, are based on selecting stock picks that managers are overweight relative to either a market or value-weighted benchmark portfolio, under the assumption that the manager is maximising an information ratio in their portfolio selection. For hedge fund managers, this is not necessarily the best way to identify conviction owing to the fact that hedge funds tend not to be benchmarked at all. We thus study two alternative measures: (i) conviction, which takes the positions in the top quintile of each manager's portfolio (by position size); and (ii) consensus, which ranks stocks based on the number of managers holding them and takes the top quintile. We find that both these measures improve returns. However, we also find that returns to consensus are monotonically increasing in conviction. That is, if we set a conviction threshold (based on position size) and then rank stocks by consensus, there is an improvement in Sharpe ratios that is strictly increasing in conviction threshold. Based on this, we test a strategy that combines conviction and consensus. The strategy is rebalanced with a 47-day lag after the quarter-end and takes the top 50 positions after ranking for conviction and consensus. We show that this strategy delivers a statistically significant six-factor alpha when tracking FEHF manager's positions. When tracking HFU ex FEHF managers' positions, there is no statistical significance and, in fact, no outperformance relative to the S&P 500. Thus, we show that in constructing a 13F strategy, one must think about the 'who' just as much as the 'how' to be able to systematically extract statistically significant hedge fund alpha in a point-in-time way. This, however, relies strongly on being able to identify the 'who' accurately, which is difficult to do without a reliable data provider.

References

Ackermann, C., R. McEnally, and D. Ravenscraft (1999): "The Performance of Hedge Funds: Risk, Return, and Incentives," *Journal of Finance*, 54(3), 833–874.

Amin, G. S., and H. Kat (2003): "Hedge Fund Performance 1990-2000: Do the "Money Machines" Really Add Value?," *Journal of Financial and Quantitative Analysis*, 38(2), 251–274.

Asness, C., R. Krail, and J. M. Liew (2001): "Do Hedge Funds Hedge?," *Journal of Portfolio Management*, 28(1), 6–19.

Bae, K.-H., B. Baik, and J.-M. Kim (2011): "Do Hedge Funds Have Information Advantages?," Working paper.

Brown, S., W. Goetzmann, and R. G. Ibbotson (1999): "Offshore Hedge Funds: Survival and Performance, 1989-95," *The Journal of Business*, 72(1), 91–117.

Brown, S., and C. Schwarz (2013): "Do market participants care about portfolio disclosure?," Working paper.

Brunnermeier, M. K., and S. Nagel (2004): "Hedge Funds and the Technology Bubble," *Journal of Finance*, 59(5), 2013–2040.

Cohen, R., C. Polk, and B. Silli (2010): "Best Ideas," Working paper.

Fama, E. F., and K. R. French (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116, 1–22.

Griffin, J. M., and J. Xu (2009): "How Smart Are the Smart Guys? A Unique View from Hedge Fund Stock Holdings," *The Review of Financial Studies*, 22(7), 2531–2570.

Kat, H., and H. Palaro (2006): "Replication and Evaluation of Fund of Hedge Funds Returns," Working paper.

Khandani, A., and A. Lo (2007): "What happened to the quants in August 2007?," Working paper.

Kosowski, R., N. Y. Naik, and M. Teo (2007): "Do hedge funds deliver alpha? A Bayesian and bootstrap analysis," *Journal of Financial Economics*, 84(1), 229–264.

Malkiel, B. G., and A. Saha (2005): "Hedge Funds: Risk and Return," *Financial Analysts Journal*, 61(6), 80–88.

Qian, H., and Z. K. Zhong (2018): "Do Hedge Funds Possess Private Information about IPO Stocks?," *The Review of Asset Pricing Studies*, 8(1), 117–152.

Titman, S., and C. Tiu (2011): "Do the Best Hedge Funds Hedge?," *The Review of Financial Studies*, 24(1), 123–168.

Verbeek, M., and Y. Wang (2013): "Better than the original? The relative success of copycat funds," *Journal of Banking and Finance*, 37(9), 3454–3471.

VI. Appendix

A. Summary statistics

Table A1. Summary statistics for FEHF managers

Date	Unique managers	Unique securities	Total securities	AuM ($bn)	Median securities	Median position size
Q4 2003	310	4,833	31,223	375.76	67	0.93%
Q4 2004	356	5,096	36,164	501.49	62	0.98%
Q4 2005	425	5,296	40,414	626.86	58	1.12%
Q4 2006	500	5,409	45,592	763.36	53	1.21%
Q4 2007	564	5,367	47,005	887.16	49	1.33%
Q4 2008	580	4,886	40,243	446.46	39	1.72%
Q4 2009	579	4,842	47,708	680.18	46	1.53%
Q4 2010	608	4,931	49,383	801.25	46	1.62%
Q4 2011	617	4,794	48,677	794.98	43	1.67%
Q4 2012	629	4,689	48,065	929.36	40	1.79%
Q4 2013	669	4,783	51,971	1,266.94	38	1.94%
Q4 2014	714	4,945	51,259	1,406.48	36	2.10%
Q4 2015	735	4,877	50,034	1,351.21	31	2.34%
Q4 2016	731	4,681	48,317	1,295.15	30	2.42%
Q4 2017	691	4,644	49,033	1,487.51	30	2.54%
Q4 2018	680	4,526	44,034	1,252.77	28	2.42%
Q2 2019	666	4,557	47,917	1,484.37	30	2.27%

Notes: The sample period is 31/12/2003 - 30/06/2019. AuM is the total market value of the holdings appearing in the 13F filings and not necessarily representative of the actual AuM of the funds.

Table A2. Summary statistics for HFU ex FEHF managers

Date	Unique managers	Unique securities	Total securities	AuM ($bn)	Median securities	Median position size
Q4 2003	126	4,132	25,825	150.04	69	0.70%
Q4 2004	162	4,532	33,058	237.62	63	0.69%
Q4 2005	201	4,982	42,559	330.53	57	0.91%
Q4 2006	241	5,264	51,008	463.58	53	0.88%
Q4 2007	272	5,331	50,508	491.50	49	0.79%
Q4 2008	272	4,845	38,870	193.23	36	1.38%
Q4 2009	247	4,609	43,971	304.91	45	1.30%
Q4 2010	253	4,661	47,605	375.27	45	1.24%
Q4 2011	262	4,626	46,388	373.27	40	1.68%
Q4 2012	258	4,648	48,072	450.04	38	1.43%
Q4 2013	257	4,702	50,719	645.49	37	1.34%
Q4 2014	257	4,933	54,777	755.06	36	1.44%
Q4 2015	260	4,877	56,404	727.65	32	1.45%
Q4 2016	250	4,778	58,018	724.63	31	1.33%
Q4 2017	227	4,726	51,984	831.56	30	1.13%
Q4 2018	232	4,804	48,170	705.21	28	1.71%
Q2 2019	235	4,782	52,195	827.34	30	1.35%

Notes: The sample period is 31/12/2003 - 30/06/2019. AuM is the total market value of the holdings appearing in the 13F filings and not necessarily representative of the actual AuM of the funds. The tables show a snapshot of the data at the quarter-ends and the statistics reported have not been aggregated over the year.

B. Impact of tradability filters

Our tests thus far have not been subject to any filters on tradability. Such filters are important for practical reasons; for instance, to ensure that the stocks being selected by the strategy are liquid enough to be traded. We show in Figure B1 and Table B1 that imposing a restriction that stocks must have a market capitalization of greater than $2 billion and an average daily volume traded (over 60 days) of at least $20 million does not materially affect our results. The majority of the difference in the strategy comes from the post-2012 period.

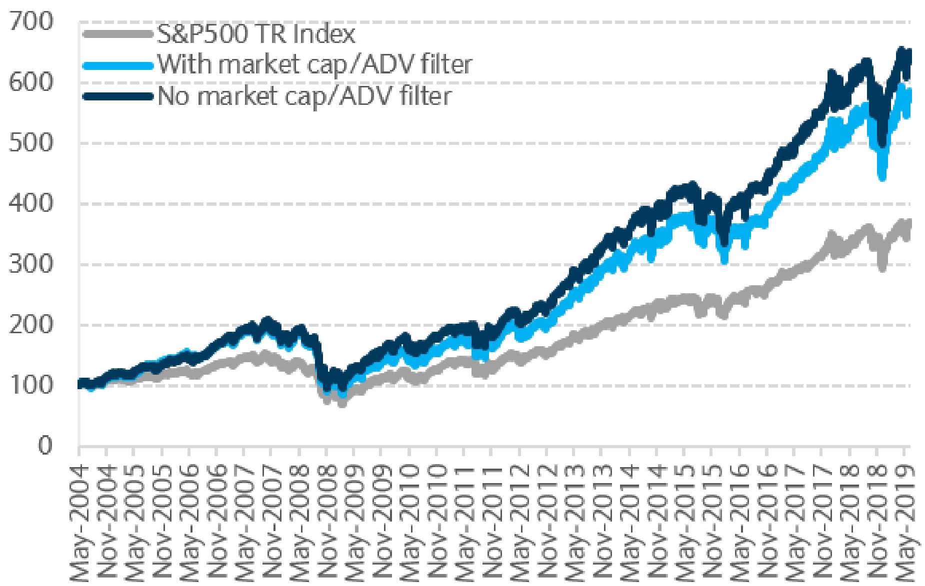


Figure B1. Impact of tradability filters, May 2004 - June 2019

Table B1. Performance statistics for market capitalization and ADV filter analysis on a strategy that combines conviction and consensus tracking the positions of FEHF managers, May 2004 - June 2019

	S&P 500 TR Index	Market cap/ADV filter	No filter
Ann. avg rets	10.34%	13.64%	14.14%
Ann. vol	18.31%	19.77%	18.87%
Sharpe ratio	0.56	0.69	0.75
Max drawdown	55.25%	57.61%	54.40%
Relative outperformance			
Ann. outperformance		3.30%	3.80%
Tracking error		5.03%	5.15%
Information ratio		65.58%	73.80%
Return correlation		0.97	0.96

C. Lag analysis

A key argument against using 13F filings to construct trading strategies is that the filings are only made public 45 days after the quarter-end. We have shown that there is still hedge fund alpha even when executing a strategy 47 days after the quarter-end, as in all the tests so far, which is restrictive. Figure C1 shows that the impact is marginal, with only a slight deterioration in performance from the zero lag to the 47-day lag. Hence, our results underestimate the true value of the informational content of 13F filings; however, we test all strategies at this lag to ensure a like-for-like comparison and to test a strategy that can be implemented in practice.

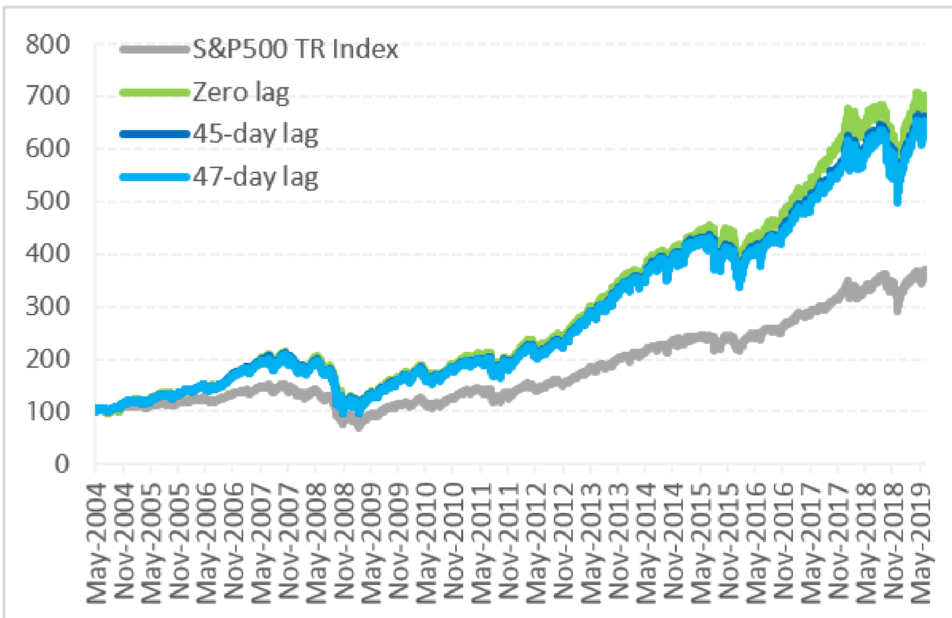


Figure C1. Impact of executing with different lags, May 2004 - June 2019

Table C1. Performance statistics for lag analysis on a strategy that combines conviction and consensus tracking the positions of FEHF managers, May 2004-June 2019

	S&P 500	Zero-day lag	45-day lag	47-day lag
Ann. avg rets	10.34%	14.55%	14.24%	14.14%
Ann. vol	18.31%	18.46%	18.83%	18.87%
Sharpe ratio	0.56	0.79	0.76	0.75
Max drawdown	55.25%	53.23%	54.32%	54.40%
Relative outperformance				
Ann. outperformance		4.22%	3.90%	3.80%
Tracking error		5.21%	5.16%	5.15%
Information ratio		80.98%	75.58%	73.80%
Return correlation		0.96	0.96	0.96