Mikko Kauppila

WHITHER ALPHA? HEDGE FUND PERFORMANCE IN VOLUNTARY VERSUS REGULATORY DATA SETS
MIKKO KAUPPILA

WHITHER ALPHA? HEDGE FUND PERFORMANCE IN VOLUNTARY VERSUS REGULATORY DATA SETS

Academic dissertation to be presented with the assent of The Doctoral Programme Committee of Human Sciences, University of Oulu for public defence in the Arina auditorium (TA105), Linnanmaa, on 25 November 2022, at 12 noon.

UNIVERSITY OF OULU, OULU 2022
Supervised by
Professor Petri Sahlström
Assistant Professor Juha Joenvääät

Reviewed by
Professor Florian Weigert
Professor Olga Kolokolova

Opponent
Professor Timo Rothovius


ISSN 1455-2647 (Printed)
ISSN 1796-2269 (Online)

Cover Design
Raimo Ahonen

PUNAMUSTA
TAMPERE 2022
Kauppila, Mikko, Whither alpha? Hedge fund performance in voluntary versus regulatory data sets.
University of Oulu Graduate School; University of Oulu, Oulu Business School
*Acta Univ. Oul. G 130, 2022*
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

**Abstract**

Recent years have witnessed growing pessimism from investment practitioners regarding hedge funds’ ability to produce abnormal returns, or “alpha”. The academic research has confirmed this view: since the global financial crisis, hedge fund managers have been unable to produce abnormal returns in aggregate. However, such research mostly relies on commercially available data sets based on voluntary reporting by fund managers, and thus the conclusions depend crucially on the correct understanding of the biases in these data.

My dissertation increases and deepens our understanding on how conclusions around hedge fund performance depend on the choice of data, and especially between voluntary and regulatory data sets. Essay I explores the benefit of aggregating voluntary data across several databases and shows that performance estimates derived from too few databases suffer from upward bias. After correcting for such “single-vendor bias”, we show that the average fund produces no abnormal return after fees. Essay II strengthens this finding by showing that starting from 2008, even previously proposed predictors for picking top hedge funds cannot add value for investors. We find evidence for decreasing returns to scale, central bank operations, and increased regulatory discretion all playing a role in this decline.

Essay III studies the self-selection bias arising from the omission of the funds that choose to never report voluntarily. We combine our voluntary data set with the first-ever systematic regulatory data set on hedge funds based on confidential Form PF reports. We find that funds not voluntarily reporting their returns have significantly better performance compared to reporters. Despite their lower performance, the reporting funds do gather more investor assets, highlighting the signaling role of voluntary reporting as a marketing vehicle. Essay IV shows via regulatory 13F holdings reports that hedge funds exhibit option picking skills and purchase options with characteristics ideal for informed trading. The overarching conclusion of my dissertation is thus that the hedge fund industry is still thriving, but the best funds do not report voluntarily to commercial data vendors.

*Keywords:* hedge funds, managerial skill, option picking skills, sample selection bias
Tiivistelmä


Asiassanat: hedge-rahastot, optioopiminnan taito, salkunhoitajan taito, valikoitumisharha
Acknowledgements

I would like to thank my supervisors Professor Petri Sahlström and Assistant Professor Juha Joenvääärä for their support and guidance. I also thank my follow up group, Professor Juha-Pekka Kallunki and Professor Mikko Puhakka, as well as my pre-examiners, Professor Florian Weigert and Professor Olga Kolokolova. I thank Dr. Mikko Leppämäki of the Graduate School of Finance for coordinating the excellent study program which formed the backbone of my doctoral studies. I would like to express my gratitude to Professor Timo Rothovius for accepting to serve as my opponent.

I thank my co-authors Assistant Professor Juha Joenvääärä, Professor Robert Kosowski, Dr. Pekka Tolonen, Professor Nicolas Bollen, Dr. Daniel Barth, and Professor Russ Wermers; as well as my co-workers at the Department of Economics, Accounting and Finance, Dr. Hamed Salehi and Jari Karppinen; for their excellent collegiality. I thank Suomen Arvopaperimarkkinoiden Edistämissäätiö, who provided financial assistance towards finishing my dissertation. Finally, I thank my friends and family for offering me a life outside of work and studies.

October 2022

Mikko Kauppila
List of original essays

This thesis is based on the introductory chapter and the following essays, which are referred throughout the text by their Roman numerals:


III  Barth, D., Joenväärä, J., Kauppila, M., & Wermers, R. The hedge fund industry is bigger (and has performed better) than you think. Manuscript.

Contents

Abstract
Tiivistelmä
Acknowledgements 7
List of original essays 9
Contents 11
1 Introduction 13
  1.1 Background ................................................................. 13
  1.2 The research aims and hypotheses ................................... 14
2 Theoretical motivation and methodology 17
  2.1 Data sources and their biases ....................................... 17
    2.1.1 Survivorship bias ................................................. 17
    2.1.2 Backfill and delisting bias .................................... 18
    2.1.3 Self-selection bias .............................................. 20
  2.2 Measuring hedge fund performance ............................. 21
    2.2.1 Performance measures ........................................ 21
    2.2.2 Empirical evidence ............................................ 21
  2.3 Predicting hedge fund performance .............................. 22
    2.3.1 Performance persistence ...................................... 23
    2.3.2 Trading activity ............................................... 23
3 Thesis contribution 25
  3.1 Essay I: Hedge fund performance: Are stylized facts sensitive to
    which database one uses? ............................................ 25
  3.2 Essay II: Hedge fund performance: End of an era? .......... 27
  3.3 Essay III: The hedge fund industry is bigger (and has performed
    better) than you think ............................................... 29
  3.4 Essay IV: Information content of hedge fund equity option
    holdings ................................................................. 31
4 Implications 35
List of references 37
Original essays 41
1 Introduction

1.1 Background

Hedge funds are investment vehicles that seek to achieve absolute return, or positive returns uncorrelated with standard asset classes. Such uncorrelated investments provide a useful diversification tool to institutional investors. Unlike the more traditional delegated portfolio managers in the mutual fund industry, hedge funds are not regulated by the Investment Company Act of 1940, which allows them more freedom in designing their trading strategies, as well as in their compensation contracts and share restriction clauses. For example, whereas mutual funds generally charge a fixed annual management fee as a percentage of assets, hedge funds usually charge an additional performance-based incentive fee, which in theory should help align the managerial incentives with those of the investors.

In a similar vein, mutual funds must provide prompt investment redemptions, but hedge fund investors can contractually agree to longer redemption periods and lock-up periods: in theory, this should give the manager more predictability and freedom to trade illiquid investments, especially during bear markets, where excess redemptions could trigger fire sales.

Because hedge funds do not generally publicize their returns (and in some jurisdictions have even been barred from doing so), our scientific understanding of the hedge fund industry relies largely on commercial databases that have gathered voluntarily reported hedge fund returns over time. Such voluntary data are subject to biases, which academics have spent great energy unveiling. For example, if a database does not store information on defunct funds, performance estimates derived from the database will be biased upwards, a phenomenon known as survivorship bias. Similar upward bias will result if the database allows a newly reporting fund to backfill its early performance history, as this selects for funds with higher incubation-period returns.

State-of-the-art bias correction method of Jorion and Schwarz (2019) suggests that such backfill bias could increase performance estimates across all funds by as much as 4% per year. As recent years have witnessed a growing pessimism from investment practitioners regarding hedge funds’ ability to produce abnormal returns—with some major pension funds slashing their hedge fund investments—such large biases call into question the ability of the hedge fund industry to produce positive risk-adjusted returns. This is important to both investors (who desire such
“alpha”) and academics (who seek to understand market efficiency). The academic research (e.g., Dichev and Yu 2011; Bali, Brown, and Demirtas 2013; Sullivan 2021) has largely confirmed this view: since the global financial crisis, hedge fund managers have been unable to produce abnormal returns in aggregate.

However, commercial databases are also subject to self-selection bias: who are the funds that choose to never report to the databases? Whether the pessimism towards hedge funds is justified depends crucially on the answer to this question. The issue is currently poorly understood, because addressing it would require access to previously unknown data on the non-reporting funds. The theory of whether a fund chooses to report or not is based mainly on signaling: reporting allows a fund to advertise its operations to potential new investors, who then generate future fees for the fund managers. Intuitively, a poorly performing fund may choose to not advertise its returns to avoid bad reputation. Less obviously, a fund with outstanding performance may also choose to not report. This is because such a fund may already have enough investors, or be closed to new investors, and nearing its capacity constraints, thus it has no need to attract new investors via advertising through voluntary reporting—in fact, reporting returns could put the fund at risk of copycats trying to reverse engineer its returns.

The question of selection bias is empirically particularly interesting because the theory offers competing answers. The existing evidence (Agarwal, Fos, and Jiang 2013; Aiken, Clifford, and Ellis 2013; Edelman, Fung, and Hsieh 2013) based on special data sets on non-reporting managers suggests that the selection bias is either zero or positive. That is, voluntarily reporting managers are at least as good as non-reporting managers. For the hedge fund investor, this paints a gloomy image: not only are the reporting funds poorly performing (when disregarding the non-investable backfilled returns), but access to non-reporting funds would not improve investment performance either. Has market efficiency reached a level where even the most sophisticated traders are unable to produce abnormal returns; is the era of hedge fund performance over?

1.2 The research aims and hypotheses

My dissertation seeks to increase and deepen our understanding on how hedge fund performance, and the “stylized facts” around it, depend on the choice and treatment of data. While many hedge fund studies employ just a single commercial database for simplicity, others aggregate multiple databases to increase their coverage of funds. However, while such aggregation is intuitively appealing, its benefits have
not been thoroughly documented. Essay I rectifies this situation by combining the seven most prominent commercial databases—Lipper TASS, HFR, BarclayHedge, Eurekahedge, Morningstar Direct, Prequin, and eVestment—to produce the widest cross-section of funds to date from 1994 through 2016, while employing the state-of-the-art backfill bias correction by Jorion and Schwarz (2019). We show that database aggregation is important because poorly performing tend to report to fewer databases, consistent with signaling theory: using too few databases risks missing out these poor performers, which biases performance estimates upwards. After correcting such “single-vendor bias”, we show that the average fund produces no abnormal return after fees. Starting from 2006, even pre-fee performance begins to falter.

After these rather negative results on hedge fund industry’s performance, essay II seeks to further understand the performance decline from 2008 through 2016. Instead of looking at only equal-weighted or value-weighted hedge fund indices, we construct realistically sized hedge fund portfolios based on prediction models proposed in previous literature. (Generally speaking, these prediction models rule out the worst funds and should therefore give the remaining funds a fairer chance to show their potential.) We then study the diversification benefit of these hedge fund portfolios by adding them to a multi-asset portfolio.

Our results show that hedge funds do add diversification benefit in the form of higher Sharpe ratio in the early period 1997–2007, but not from 2008 through 2016. When exploring potential explanations for this performance erosion, we find some evidence to decreasing returns to scale at the industry level; that is, the industry has grown beyond the capacities of the common strategies. Our switch point analysis reveals a marked shift in performance during the depths of the 2008–09 financial crisis and the passage of the Dodd–Frank reform. The consequent central bank interventions and heightened regulatory scrutiny may therefore also partly explain the decline in hedge fund performance.

Essay III turns to the issue of selection bias. This research was conducted in collaboration with the U.S. Office of Financial Research (OFR), who had access to the confidential Form PF filings, available from 2013 onwards. These filings are mandatory for hedge funds above a size threshold with at least one U.S. investor, and include detailed fund characteristics, assets under management, as well as monthly returns (both gross and net of fees). Unlike the selective samples of non-reporting funds used by Aiken et al. (2013) and Edelman et al. (2013), these Form PF filings offer the first comprehensive data set of non-reporting funds.
After merging these Form PF data with our previously used data set of reporting funds, we find that the non-reporting funds perform significantly better than the reporting funds, in striking contrast to Aiken et al. (2013) and Edelman et al. (2013). The performance difference is due to abnormal return, not systematic risk exposures. Indeed, fund leverage, risk exposures, and share restrictions are similar for reporters vs non-reporters. This is important from a regulatory, financial stability standpoint, because it shows that the non-reporters pose no heightened risk of systemic contagion. Despite their lower performance, reporting funds gather more investor inflows, underlining the role of commercial databases as marketing vehicles.

In sum, essays I–III show that whereas voluntarily reporting funds are struggling to produce abnormal returns in the modern market, the hedge fund industry as a whole is still thriving, because the non-reporting funds are performing so much better. These essays were based on either voluntarily reported commercial data, or mandatory but confidential regulatory data. Essay IV looks at the intersection of the two: mandatory yet publicly available regulatory data. More specifically, we look at quarterly 13F holdings reports of hedge fund management firms. We exploit a relatively common reporting error, wherein the firms accidentally reveal more information about their equity options than necessary. This allows us for the first time to study the characteristics and the performance of the equity options purchased by hedge funds.

We find that hedge funds (compared to the average option market participant) prefer options with characteristics that previous theoretical research has found to be especially suited for informed trading: higher embedded leverage, more liquidity (which lessens the price impact), and lower lottery-like skewness (which carries a price premium). Furthermore, even though the average option on the market unexpectedly has a negative return due to its insurance-like premium, option portfolios that replicate hedge funds’ option purchases produce positive returns. In sum, the findings of essay IV propose one possible mechanism for the continuing outperformance of hedge funds found in essay III.
2 Theoretical motivation and methodology

2.1 Data sources and their biases

As covered in the introduction, our scientific knowledge of the hedge fund industry relies mostly on commercially available databases such as Lipper TASS, HFR, and BarclayHedge, which gather voluntarily reported information on hedge funds’ monthly performance, assets under management, and fund characteristics (e.g., compensation structure and share restrictions). The choice of a database can affect conclusions about the industry performance. This is known in the mutual fund literature, where Elton, Gruber, and Blake (2001) demonstrate systematic return differences between the popular Morningstar and CRSP mutual fund databases; and in the private equity sphere, where Harris, Jenkinson, and Kaplan (2014) show that venture capital funds in the Burgiss database outperform those in the Thomson Venture Economics database. No such comparative study exists for hedge fund databases with regards to fund performance, although Liang (2000) shows that Lipper TASS and HFR differ in their characteristics coverage. Still, researchers are aware of the potential downsides of using just a single database, as evidenced by several hedge fund studies that consolidate multiple databases.

2.1.1 Survivorship bias

A database generally provides its data either as a single data set, or as separate live and graveyard data sets, the latter of which contains information on defunct funds (or funds that otherwise stopped reporting). However, few vendors kept records of defunct funds prior to 1994 (e.g., Brown, Goetzmann, and Ibbotson 1999). Assuming that funds are liquidated after bad performance, exclusion of such funds results in upward survivorship bias in performance estimates. As a result, the sample period in most empirical hedge fund studies is from January 1994 forwards. Survivorship bias can also arise spuriously from database mergers (Aggarwal and Jorion 2010a), or very subtly from varying collection patterns: for example, BarclayHedge started collecting share restriction characteristics only in 2000, so all analyses relying on these characteristics will be survivorship-biased prior to 2000.

Bhardwaj, Gorton, and Rouwenhorst (2014) demonstrate that some databases suffer from graveyard bias, wherein funds are retroactively removed from the
graveyard data set, possibly due to pressure from the hedge fund managers. This is related to an emerging literature on database revisions, wherein fund returns or characteristics are revised across database vintages. Patton, Ramadorai, and Streatfield (2015) report that historical fund returns are routinely revised; furthermore, these revisions do not reflect mere error corrections, but are predictable by fund characteristics, and predictive of future underperformance. While Agarwal, Daniel, and Naik (2009) argue that fund contract terms are difficult to change after inception, Agarwal and Ray (2011) find that as much as 7.8% of hedge funds change their compensation characteristics across vintages. Getmansky, Liang, Schwarz, and Wermers (2015) show that between quarterly vintages, 1%–9% of funds change their share restrictions. Correcting the effect of such revisions requires access to multiple vintages.

2.1.2 Backfill and delisting bias

Figure 1 demonstrates the lifespan of a hedge fund and its relation to voluntary data reporting. Inception date and liquidation date denote when the fund begins and ceases its operations. Listing date and delisting date denote when the fund begins and ceases reporting to database. Upon listing to a database, the fund generally backfills its incubation period returns to the database, and thereafter begins reporting monthly returns in real-time.¹ Funds with good incubation period performance are more likely to list, and therefore the incubation period returns are generally biased upwards, resulting in backfill bias.

¹ Listed funds may also delay reporting of individual monthly returns, possibly strategically (Aragon and Nanda, 2009).
If the listing dates are available, backfill bias can be eliminated by simply removing the pre-listing returns (e.g., Jagannathan, Malakhov, and Novikov, 2010; Bhardwaj, Gorton, and Rouwenhorst 2014). Unfortunately, not all databases provide these listing dates. Studies employing such databases have resorted to ad hoc cut-off rules, such as removing the first 12 or 24 months or each fund’s returns (e.g., Kosowski, Naik, and Teo 2007; Joenväärä, Kosowski, and Tolonen 2019). Recently, Jorion and Schwarz (2019) proposed an algorithm for inferring the approximate listing date from numeric fund identifiers. Using the proper listing-date based correction, they estimate the backfill bias across all funds to be around 2.91%–4.34% per year depending on database, far larger than estimates based on simple ad-hoc cutoff rules.

After delisting from a database the fund returns until liquidation are no longer observable, which results in delisting bias. However, the direction of delisting bias is not obvious, because funds may stop reporting due to either bad performance (which they may not want to advertise) or good performance (having been closed to new investors). Aiken, Clifford, and Ellis (2013) utilize observed holdings of a subset of funds of hedge funds to deduce quarterly hedge fund returns from the holdings’ valuations. They find that the delisting bias is on average positive: compared to funds that continue reporting, delisted funds have 1.84% lower returns per quarter. Hodder, Jackwerth, and Kolokolova (2014) instead deduce the holdings of funds of hedge funds using principal component analysis on their database-reported returns, which yields a more comprehensive sample of monthly delisting returns on the constituent funds. They find no statistically significant delisting bias.
for the average fund, but they show that funds with poor pre-delisting returns and no clearly stated delisting reasons have inferior post-delisting performance.

### 2.1.3 Self-selection bias

The exclusion of funds that never choose to report to databases is termed self-selection bias. Signaling theory would suggest that hedge funds employ commercial databases as a marketing tool to attract investment capital. Jorion and Schwarz (2014) find that managers strategically list their small, best-performing funds in multiple databases immediately, while preserving the option to list their other funds in additional databases later. If the performance of a fund is positively related to the number of databases it reports to, we might naïvely expect funds reporting to zero databases to fare the worst. However, exceptionally good funds may be closed to new investors to avoid diseconomies of scale, and any extra advertisement of their returns could only attract copycats.

Uncovering the direction and magnitude of the self-selection bias is difficult because it requires access to non-voluntarily reported fund data, which typically must be specially constructed by the researchers. Aiken, Clifford, and Ellis (2013) gather quarterly holdings reports of funds of hedge funds, which allows them to calculate the quarterly returns on their underlying funds. They find that reporting funds outperform non-reporting funds: in other words, that the selection bias on the performance estimates of publicly reporting funds is positive. Agarwal, Fos, and Jiang (2013) find a similar result using fund returns derived from 13F holdings. Edelman, Fung, and Hsieh (2013) gather a data set of non-reporting funds manually from industry surveys, regulatory filings, fund managers and fund investors. They concentrate their effort on the largest “mega” hedge funds that collectively manage over 50% of (known) industry assets, which gives their results economic significance (in dollar-weighted terms). Ultimately, they find that the selection bias is statistically insignificant: the reporting and non-reporting funds perform roughly the same.

The major downside in these three studies is the representativeness of their data sets: 13F reports are restricted to long equity holdings, funds of hedge funds are a specialized investor subset, and the comprehensiveness of manually collected data is unclear (especially for the most secretive funds). Proper understanding of self-selection bias would require a data set that is not itself selection biased.
2.2 Measuring hedge fund performance

2.2.1 Performance measures

Investors are not willing to pay hefty hedge fund fees for returns stemming from exposure to standard asset classes such as stocks or bonds, or well-known risk factors such as size and value factors. Therefore, the performance of hedge funds should be measured in risk-adjusted terms. The seven-factor benchmark by Fung and Hsieh (2004) has emerged as the standard workhorse in the literature. It employs two equity factors, two bond factors, and three option-based trend-following factors. Fung and Hsieh (2004) show that the benchmark captures well the common return variation across the heterogeneous hedge fund styles.

Like all benchmark models, the Fung and Hsieh (2004) benchmark suffers from misspecification risk, that is, the possibility of omitted risk factors, which could result in upward-biased alphas. A common solution is to introduce additional factors such as an emerging market factor (Edelman, Fung, Hsieh, and Naik 2012) or call and put option factors (Agarwal and Naik 2004). However, Bollen (2013) shows that even with seven additional factors, common risk still exists. Specifically, portfolios of funds with zero \( R^2 \) against the seven-factor or the 14-factor benchmark have substantial volatility. Aside from the incompleteness of the Fung and Hsieh (2004) benchmark, Bhardwaj et al. (2014) show that the benchmark produces upward-biases estimates of alpha, because the trend-following factors are inefficient replications of actual trend-following styles.

Alternatively, researchers can use benchmark-free measures of risk-adjusted performance, the simplest of which is the Sharpe ratio (Sharpe 1966). Goetzmann, Ingersoll, Spiegel, and Welch (2007) propose a manipulation-proof performance (MPPM), which they argue cannot be “gamed” via dynamic trading, and which according to Bali, Brown, and Demirtas (2013) addresses the non-normality seen in hedge fund returns (e.g., Fung and Hsieh 2001; Mitchell and Pulvino 2001; Agarwal and Naik 2004). Another alternative is to use utility-based certainty equivalents as in Fleming, Kirby, and Ostdiek (2001).

2.2.2 Empirical evidence

How well are hedge funds succeeding in their mandate of producing positive risk-adjusted returns after fees? Across all investors, active management must be a losing game after fees (Sharpe 1991), but for a subset of skilled managers such as
hedge funds. This may not be the case. Early research on the issue was generally positive. Fung and Hsieh (2004) find that over the period 1994–2002, the four hedge fund indices they studied produced statistically significant annualized alpha of 7.92%–11.34% over their seven-factor benchmark. These estimates reflect portfolios of hedge funds, whereas Kosowski et al. (2007) study the performance of hedge funds in the cross-section across the same 1994–2002 period. They find that the average fund produced an annualized alpha of 5.04%, but with an insignificant average $t$-statistic of 1.42. However, by applying a bootstrap approach, they show that the top managers were quite skillful: for example, the manager at the top 90th percentile produced an annualized alpha of 14.88% with a $t$-statistic of 3.78.

More recent studies offer a far more pessimistic view. Dichev and Yy (2011) show that in dollar-weighted terms, hedge funds underperform the S&P 500 index. Using utility-based performance metrics, Bali, Brown, and Demirtas (2013) find that only 2 of 11 hedge fund indices outperform the S&P 500 index. Most recently, Sullivan (2021) reported that the alpha of a broad hedge fund index flipped from positive to negative in the decade following the 2008 financial crisis.

There are various theories to explain the decrease in hedge fund returns. First, as the industry grows, both diseconomies of scale (Fung, Hsieh, Naik, and Ramadorai 2008) and increased competition (Cao and Velthuis 2017) can erode performance. Second, post-crisis central bank interventions can create distortions in the market, such as increased correlation between asset classes (Cotter, Gabriel, and Roll 2018), that hamper trading strategies. Third, the increased regulatory oversight from 2010 Dodd–Frank Act has imposed new compliance costs and potentially chilled some hedge fund trading reporting activity (Cumming, Dai, and Johan 2020; Dimmock and Gerken 2016; Honigsberg 2019). Finally, as in McLean and Pontiff (2016), publicized research about return predictability may erode performance.

### 2.3 Predicting hedge fund performance

Understanding the determinants of hedge fund performance is important to both their potential investors as well as academics seeking to understand market efficiency and investor behaviour. Also, the soundness of hedge fund regulation depends on whether the contract terms are related to fund performance in their purported manner. Aragon (2007) and Barth and Monin (2019) report that tighter share restrictions are associated with higher returns, consistent with their proposed,
discretion-enhancing nature. Agarwal, Daniel, and Naik (2009) find that managerial incentives—the level of incentive fee, the existence of a high-water mark provision, and the “delta” of manager’s implied call option—are positively associated with future performance, suggesting that these contracts successfully reduce the agency problems inherent in delegated portfolio management.

Aggarwal and Jorion (2010b) and Teo (2009, 2011) show that smaller funds outperform larger ones, consistent with funds’ trading strategies having capacity constraints. Finally, both hedge fund domicile (Aragon, Liang, and Park 2013) and investment style (Brown and Goetzmann 2003) are known to drive differences in returns.

2.3.1 Performance persistence

A major question in active management is whether individual managers’ performance persists over time. The theoretical model of Berk and Green (2004) suggests that skillful managers attract investor flows until the manager’s alpha erodes under capacity constraints. This results in non-persistent equilibrium returns as seen in mutual funds (Carhart 1997). In other words, managers extract all the rents via fees. Like mutual fund investors, hedge fund investors too “chase” previously winning funds by offering them more capital (Fung, Hsieh, Naik, and Ramadorai 2008). In addition, this flow-performance sensitivity is stronger for offshore funds, who have fewer restrictions on advertisement (Argon, Liang, and Park 2014).

However, hedge fund managers have a theoretical incentive to reject excess investor flows, as proposed by Glode and Green (2011). Specifically, because the secret trading strategies are shared with existing investors, managers cannot afford to extract all the rents, lest the investors exit with the secret knowledge. Under such conditions, hedge fund performance can be persistent, which is supported by widely cited empirical studies (Brown, Goetzmann, and Ibbotson 1999; Kosowski, Naik, and Teo 2007; Jagannathan, Malakhov, and Naik 2010). Joenväärä, Kosowski, and Tolonen (2019) show that this persistence is driven by smaller funds, as it only appears in equal-weighted portfolios, and disappears in value-weighted portfolios.

2.3.2 Trading activity

Investment advisors with assets under management of at least $100 million in qualified securities—in practice, public U.S. companies—must report their quarter-
end long positions in these companies, including long interest via convertible bonds and plain vanilla options, to the U.S. Securities and Exchange Commission on a form 13F holdings report. Although these holdings suffer from several shortcomings, such as the omission of short sales and written options, hedge fund researchers have still applied them successfully to study the information content of hedge funds’ holdings, that is, how their trading activity reflects in the performance of the funds or the underlying stocks.

Brunnermeier and Nagel (2004) employ 13F holdings to show that hedge funds successfully “rode” the dot-com bubble, investing heavily in technology stocks during the upturn, and reducing their positions in stocks that were about to decline. An early study by Griffin and Xu (2009) shows that hedge fund managers possess no stock picking ability above mutual fund managers, despite placing more active share bets and having higher portfolio turnover. Cao, Chen, and Liang (2018) show that hedge funds tend to hold undervalued stocks whose mispricing subsequently dissipates, resulting in abnormal stock returns. Cao, Goldie, Liang, and Petrasek (2016) show that hedge fund managers’ risk arbitrage trades outperform a naïve risk arbitrage benchmark, attributable to their ability to manage downside risk.

Regarding the information content of option holdings, Aragon and Martin (2012) show that hedge funds’ directional option trades predict future stock performance, such that stocks with high call (resp. put) option volume tend to rise (resp. fall). Similarly, they show that options held in straddle (i.e., as a combination of call and put options on the same stock) predict an increase in the stock volatility. Both findings are consistent with informed use of options by hedge funds. Cao, Gempesaw, and Simin (2018) find that aggregate informed options trading is useful for predicting market returns, but also that the amount of informed trading is on the decline.


3 Thesis contribution

3.1 Essay I: Hedge fund performance: Are stylized facts sensitive to which database one uses?

My first essay seeks to understand the effect of database selection on hedge fund research: how does the choice of database, or the aggregation of multiple databases, affect basic research conclusions on hedge fund performance—the aggregate performance level, persistence, and determinants of performance? To make our results extensive, we employ the five most commonly used databases (BarclayHedge, Eurekahedge, HFR, Lipper TASS, and Morningstar Direct) as well as two previously unused ones (eVestment and Preqin) over the period 1994–2016. We also use the state-of-the art backfill bias correction method of Jorion and Schwarz (2019) throughout to ensure that our results do not suffer from standard biases unrelated to database choice.

A single hedge fund database generally contains various share classes (e.g., currency classes, onshore vs offshore classes) for the same fund. In studies that employ a single database, this issue is typically ignored, and the share classes are treated as separate funds. (Alternatively, the researcher can limit the data to USD-denominated and US-domiciled share classes.) When multiple databases are aggregated, the standard approach in hedge fund literature is to select a single “representative” share class across the databases to represent the fund; for example, the share class with longest return history. We instead follow the standard in mutual fund literature of properly aggregating the share classes across and within databases to the fund level, mostly by using the medians of observations. We show that such aggregation increases data coverage across variables, as different databases may vary in their coverage of specific fund characteristics as well as periods of reported time series.

We start by analyzing how databases differ in their coverage of funds, and how much coverage can be increased by aggregating specific subsets of databases. Individual databases contain between 5,241 (Preqin) to 11,826 (BarclayHedge) funds, whereas the full seven-database aggregate contains 26,432 funds. As aggregating seven databases is a major undertaking for most researchers, we also for search for reasonably simple aggregates involving fewer, but high-quality databases (i.e., ones with high coverage of fund characteristics and bias-free time series). We find that the combination of BarclayHedge, HFR, and Lipper TASS
produces a reasonably good aggregate in terms of number of funds and coverage of industry assets, and in most later results this three-database aggregate behaves roughly similar to the seven-database aggregate.

Next, we seek to understand the effect of database choice and aggregation on the level of fund performance. We find that compared to the seven-database aggregate, the average fund performance across individual databases is 0.58%–1.25% per year higher. We then form seven portfolios, each portfolio \( k = 1, \ldots, 7 \) consisting of all fund-month observations that were reported to exactly \( k \) databases and find that the portfolio performance increases monotonically in \( k \). This suggests that the better funds report to multiple databases, and that databases differ in their coverage of the worse funds. Performance estimates derived from single databases, or too few databases, therefore suffer from an upward bias, which is mitigated by database aggregation.

To study risk-adjusted fund performance, we employ two factor benchmarks. The first is the standard seven-factor Fung and Hsieh (2004) benchmark. The second is the global Carhart (1997) model augmented with a time-series momentum factor (Moskowitz, Ooi, and Pedersen 2012) as well as the Pastor–Stambaugh (2003) liquidity risk factor and a betting-against-beta factor (Frazzini and Pedersen 2014). The second benchmark addresses the issues covered in Section 2.2.1, and in general produces higher coefficients of determination, but also serves as a robustness check. We find that equal-weight and value-weight portfolios of hedge funds consistently produce statistically significant alpha before fees (ranging from 3.11% to 4.50% per year), but not after fees. In other words, fund managers extract all the economic rents.

Lastly, we revisit the effect of database aggregation on performance persistence and performance determinants. In short, our results are consistent with the previous findings covered in Section 2.3, but database aggregation generally strengthens the results, as the aggregated data cover a wider cross-section of funds.

To summarize, my first essay contributes to our understanding of hedge fund database selection and aggregation, and their effects on the stylized facts of hedge fund performance. We propose a novel aggregation procedure and demonstrate its benefits in increasing not just the number of funds, but also the number of characteristics and time series observations within funds. We contribute to the literature on data biases by reporting that performance estimates derived from too few databases suffer from upward bias. As a guideline for fellow researchers, we find that aggregating just three databases (BarclayHedge, HFR, and Lipper TASS) can mitigate most of the upward bias while providing good data coverage.
Finally, we contribute to our understanding of the level of alpha in the hedge fund industry, and how it’s distributed between managers (in the form of fees, which are partly performance-based) and investors (in the form of post-fee alphas). Across benchmarks, periods, and weighting schemes, pre-fee alphas are statistically significant and positive, whereas post-fee alphas are zero, suggesting that managers extract all the economic rents. However, the level of pre-fee alpha begins to falter later in the sample: whereas in 1995–2005, the pre-fee alpha was 2.92%–6.61% per year \( (t = 1.96–6.15) \), in the 2006–2016 it was only 2.00%–2.41% per year \( (t = 1.68–2.56) \), with an average reduction of 2.49% per year.

### 3.2 Essay II: Hedge fund performance: End of an era?

As covered in Section 2.1.2, and as corroborated by my first essay, it’s unclear whether hedge funds add value to their investors anymore. Given the importance of the subject to investors, regulators, as well as academics, my second essay seeks to further understand this decline in hedge fund performance, including its causes. We employ an aggregate hedge fund database as in essay I over the period 1997–2016, split into two subperiods 1997–2007 and 2008–2016. The 2007–2008 demarcation point coincides with Sullivan’s (2021) breakpoint, but more importantly both subperiods contain a full economic cycle.

Our main benchmark is a 50/50 stock/bond portfolio consisting of equal parts of S&P 500 and the VBTIX ETF. This portfolio produced a cumulative return of 125% in 1997–2007, and 70% in 2008–2016. The respective cumulative returns for our equal-weighted portfolio of hedge funds saw a massive drop from 225% to 25%. The percentage of funds with positive and significant Fung and Hsieh (2004) alpha dropped from about 20% to 10%, whereas the percentage with significantly negative alpha increased from 5% to about 20%. These initial findings confirm the suspicion of a considerably weakening in hedge fund performance at the aggregate level.

However, investors do not, nor cannot invest in the equal-weighted hedge fund index. The literature has proposed plenty of predictors of hedge fund performance, which can be used to construct potentially superior portfolios. For example, the Bayesian alpha of Kosowski et al. (2007) and the style-relative alpha of Jagannathan et al. (2010) were used by the respective authors to demonstrate positive performance persistence, and therefore constitute reasonable predictors of future performance. We replicate seven such predictors. To simulate a realistic investor experience, at each year-end we select 15 random funds from the top
predicted quintile of funds. We then repeat this simulation 1,000 times to produce a cross-section of realized hedge fund portfolios.

As institutional investors generally use hedge funds as a diversification tool rather than a standalone investment, we measure the benefit of an allocation to hedge funds by comparing the performance of the 50/50 stock/bond portfolio against a multi-asset-class, 30/50/20 stock/bond/hedge fund portfolio. Having 1,000 multi-asset-class portfolios allows calculating an empirical p-value against the null hypothesis of the stock/bond and multi-asset-class portfolios being equal (with respect to a selected performance measure, such as Sharpe ratio).

Over the full 1997–2016 sample, two of the predictors increased the multi-asset-class portfolio’s Sharpe ratio relative to that of the stock/bond portfolio: the t-statistic of Fung and Hsieh (2004) seven-factor alpha and the macroeconomic timing skill measure of Bali, Brown, and Caglayan (2014). In both cases, the improvement in Sharpe ratio was achieved by a substantial reduction in volatility. However, when looking at subperiods, Sharpe ratio is only improved in 1997–2007. In 2008–2016, average returns decrease together with volatility, leaving the Sharpe ratio unchanged. These findings were unchanged in various robustness checks. We conclude that deterioration in hedge fund performance cannot be overcome by using predictors, at least for investors with reasonable levels of risk-aversion.

Finally, we turn to economic explanations of the performance decline, which were covered in Section 2.2.2. First, to gauge the effects of decreasing returns to scale both at fund-level and industry-level, we regress monthly fund alphas on lagged fund size (logarithmic) and lagged hedge fund industry size (fraction of global equity market capitalization) using the recursive demeaning procedure of Pástor, Stambaugh, and Taylor (2015). We find that for every 1% increase in relative industry size, estimated fund alpha drops by 28 basis points, whereas we find no evidence for fund-level decreasing returns to scale, consistent with Cao and Velthuis (2017).

Rest of the economic explanations yield testable hypotheses related to the dating of the performance decline. First, if central bank interventions reduced arbitrage opportunities, the decline should begin soon after the first round of quantitative easing was announced by U.S. Federal Reserve on 25 November 2008. Second, if regulatory reform increased compliance costs and created a chilling effect on misbehaviour such as return-smoothing, the decline should begin after the enactment of the Dodd–Frank reforms on 21 July 2010. Third, if academic research on hedge fund performance predictors drives excess capital flows to previously
superior strategies, the decline in the performance of funds with top predicted performance should begin soon after the relevant research was first published.

To test these hypotheses, we collect the realized 1997–2016 returns on the top-quintile funds as ranked on each of the seven predictors. For each of the seven portfolios, we perform switch point regressions as in Bollen and Whaley (2009), which find the optimal date on which the portfolio’s alpha (i.e., regression intercept) switched between two levels—for our purposes, this switch point defines the timing of the start of the performance decline. We find that the switch points are clustered around either early 2008 (five predictors) or May 2011 (two predictors). These dates are consistent with the beginning of central bank interventions and the enactment of the regulatory reforms, respectively. This points to economic policies having a hand in the post-crisis decline of hedge fund performance.

To summarize, essay II contributes to our understanding of the declining hedge fund performance. First, it strengthens the case that, post-crisis, institutional investors would not have seen a diversification benefit from realistically sized hedge fund portfolios, even when utilizing published research to pick funds. Second, it reports evidence for three mechanisms contributing to the performance decline: the growth of the hedge fund industry, central bank interventions, and increased regulatory scrutiny, the first two relating to diminishing arbitrage opportunities.

### 3.3 Essay III: The hedge fund industry is bigger (and has performed better) than you think

Essays I and II present strong evidence for performance decline of hedge funds voluntarily reporting to commercial data vendors, when properly controlling for all but one bias: the self-selection bias arising from the omission of the non-vendor listed funds. As described in Section 2.1.3, existing research points to this selection bias being either zero or positive, which suggests that access to non-listed funds would not improve investor returns.

However, the data sets for the non-reporting funds utilized in the existing research are either hand-collected or based on selective data sources such as 13F holdings reports, which calls their representativeness into question. In essay III, we combine our consolidated hedge fund database with the first-ever systematic regulatory collection of large hedge funds over the period 2013–2016. The regulatory data come from Form PF reports, which are mandatory for all hedge fund advisors with at least $150 million in private fund assets, and which contain
information similar to commercial databases (fund characteristics and time series) for funds with at least one U.S. investor. These data only omit the smallest hedge funds and funds without U.S. investors and should thus be highly representative of the hedge fund industry, especially in value-weighted terms.

This combined data set allows us to address several fundamental questions relating to the hedge fund industry. First, we find that the total net assets under management at the end of 2016 is at least $5.0 trillion, which is around 37% than any previous estimate. Around half of these assets are reported to commercial databases, whereas the other half comes from non-vendor listed managers. The average leverage factor is around 1.6 for both vendor-listed and non-listed funds, which puts the total economic footprint of the hedge fund industry at above $8.3 trillion. The hedge fund industry grew by around 34.0% between 2013–2016, with vendor-listed assets growing more than non-listed assets (51.3% vs 20.5%).

Next, we assess the industry performance. We find that the total value-weighted return after fees on non-listed funds over the sample period is 29%, dramatically higher than the 10% return on vendor-listed funds. This pattern holds within almost all strategy types. At fund-level, we find that the systematic risk exposures (against the global seven-factor benchmark introduced in essay 1) are similar between vendor-listed and non-listed funds; thus, the performance difference is driven by alpha. The equal-weighted average alpha before fees is 5.64% per year for non-listed funds, but −1.46% for vendor-listed funds—a 7.39% difference. When applying the bootstrap methodology of Fama and French (2010) to analyse the cross-section of alphas, we find that only the top 20% of vendor-listed managers produce statistically significant positive alpha, compared to top 70% of non-listed managers.

Although confidentiality requirements prevented us from sharing raw data with our co-authors, we could nevertheless perform Fama–MacBeth (1973) regressions, as these involve sharing only the covariance structure. When regressing month-ahead alpha on observable fund characteristics, the annualized alpha was still 3.36%–5.40% higher in non-listed vs vendor-listed funds. To understand whether the listing decision affects performance persistence, we regressed future returns on past returns across different horizons, controlling for fund characteristics as well as return-smoothing (Getmansky, Lo, and Makarov 2004) and censoring (Heckman 1973). In these persistence regressions, we adjust each return against the fund’s strategy index, which has the benefit of controlling for omitted risk factors, to the extent that these factors are shared across the funds in the strategy. We find that the
persistence coefficients are consistently higher in the non-listed vs vendor-listed funds, suggesting that these funds possess (more) true skill.

Our finding of inferior returns yet superior asset growth of vendor-listed funds is inconsistent with previous research flow-performance relationship (Section 2.3.1). When conducting a regression of quarterly flow on previous quarter’s performance, we find that the flows of vendor-listed funds are around five times more sensitive to past performance, compared to non-listed funds. This suggests that the previous estimates of flow-performance relationship are significantly biased upwards, as they rely only on vendor-listed data.

To summarize, essay III radically revises our understanding of the hedge fund industry and elucidates the differences between vendor-listed and non-listed funds. In contrast to previous studies relying on potentially unrepresentative data, we find that the selection bias of hedge fund performance in commercial databases is strongly negative: that is, non-listed funds outperform vendor-listed funds. This result is strengthened by us not correcting the vendor-listed data for backfill bias, as this means that the negative selection bias overrides the positive backfill bias. The outperformance of non-listed funds stems from abnormal returns, not systematic risk exposures or leverage, which is important from a financial stability point of view. Lastly, our results suggest that previous estimates of flow-performance sensitivity are biased upwards.

3.4 Essay IV: Information content of hedge fund equity option holdings

Essay III established that hedge fund performance is superior in regulatory Form PF data compared to voluntarily reported data in commercial databases. However, these regulatory data are confidential, and thereby inaccessible to investors. Essay IV looks at a different form of regulatory data that are publicly available: quarterly 13F holdings reports of hedge funds. Existing research has documented that selected stock and option holdings in these data are informative of future stock returns (Section 2.3.2), and therefore suggestive of (some) mechanisms through which hedge funds generate their abnormal returns.

Our study contributes to this line of research, but instead of looking at future stock returns, we study future option returns. This has previously been impossible because, per 13F rules, each option position must be reported in terms of four variables: 1) the option class (call or put), 2) an identifier for the underlying stock, 3) the underlying number of shares, and 4) the notional value of the position, i.e.,
number of shares times the share price. Consequently, an option position does not reveal anything about the strike price or maturity of the option itself.

In a 13F report filed according to the above rule, the unit price (valuation divided by number of shares) for a given underlying stock should be constant across all instrument types (shares, calls, puts). However, we find that this is often not the case. Instead, over our sample period 2005 Q1–2013 Q2, a large minority of hedge fund advisors disregard the above 13F rule and report not the notional value but the market value of their option positions. From the market value we obtain the unit option price, which we match to OptionMetrics, and thereby get access to the individual option-level data. We demonstrate the representability of our sample by showing that such erroneous reporters are indistinguishable from correctly-reporting firms along several dimensions. This suggests that erroneous reporting is merely an accident, and not, for example, a signal of incompetence.

We begin by studying the characteristics of the options purchased by hedge funds, as the theoretical literature suggests several interesting characteristics related to option trading behavior. Overall, we find that hedge funds’ option holdings exhibit characteristics ideal for informed trading (e.g., Easley, O’Hara, and Srinivas 1998). Compared to the average option market participant, hedge funds prefer options with greater embedded leverage, which could offer more “action” for informed trades (Black 1975). Hedge funds also prefer liquid options of liquid stocks, which are more easily traded without large market impact or increased adverse selection risk (Glosten and Milgrom 1985), which are particularly harmful to asymmetrically informed traders.

Next, motivated by these attractive-looking option characteristics, we verify whether hedge funds are skilled at selecting equity options. We form option portfolios based on hedge funds’ aggregate holdings on each quarter-end 2005 Q1–2013 Q2 and track their daily returns through the subsequent quarter. At one-month horizon, the all-option portfolio generates a positive return of 0.25% per day ($t = 2.19$), which is impressive given that market-aggregate option returns are negative (e.g., Frazzini and Pedersen 2012). A portfolio of directional options (which are least likely to be used for hedging purposes) produces returns that persist over one quarter (0.17% per day, $t = 2.68$).

These results are robust to adjusting for common risk factors as well as option characteristics, and part of the returns are obtainable by an investor subject to the 45-day 13F reporting deadline, despite the early outperformance not being available to these investors. Finally, we verify that the option positions present an
economically significant portion of the total 13F portfolios as well as contribute positively to their risk-adjusted performance.

To summarize, essay IV contributes to our understanding of the information content of hedge funds’ holdings. Our novel contribution is to focus on option returns and characteristics, which have previously been inaccessible. We show that hedge funds prefer to hold equity options with characteristics consistent with rational models of informed trading (e.g., Easley, O’Hara, and Srinivas 1998), and that their option portfolios do generate positive returns, more so when concentrating on directional trades. The findings of essay IV propose one possible mechanism for the continuing outperformance of hedge funds found in essay III, and echo the idea that alpha is best found in regulatory data free from self-selection bias.
4 Implications

Overall, my thesis shows that the hedge fund industry is still thriving, but that the best funds do not report voluntarily to commercial data vendors. Based on our flow-performance analyses in essay III, it seems that some investors are still making their investment decisions based on the voluntary data, which according to our results is ill-informed. If the data vendors want to retain their investor clientele, they need to step up their game and gain access to the currently non-listed, superior funds. The hedge fund consultant PivotalPath has gathered extensive information on the non-listed funds, which they originally used to provide insights to their allocator clients, but which is now finding its way to academic researchers through PivotalPath’s partnership with the Institute for Private Capital at UNC Kenan–Flagler Business School.

For the financial regulator interested in systemic risk, the results of essay III should prove soothing. We find no differences in systematic risk exposures or levels of leverage between voluntary and regulatory data. However, vendor-listed funds have more lax share restrictions, which could pose a higher contagion risk in the form of fire sales. Essay II suggests that central bank interventions may have decreased arbitrage opportunities, and that increased regulatory scrutiny may have successfully decreased reporting misbehaviour.

For us academics, essay I shows that consolidating databases is important to ensure good coverage of the worse funds. We should note that half of the assets are currently non-listed, and this includes the better funds. Hedge fund research based on voluntary data may not be representative of the entire industry. Luckily, in essay III we find that many characteristics such as risk exposures and leverage are similar between voluntary and regulatory data. However, performance-related results will undoubtedly miss the higher quantiles. Also, studies on flow-performance relations are likely biased.

Although essay III set its scope mostly around the issue of self-selection bias, its data set provides a fertile setting for future studies on the strategic listing behaviour of hedge funds. Understanding the drivers of the superior performance of non-listed funds also warrants further research.
List of references


Original essays


III Barth, D., Joenväärä, J., Kauppila, M., & Wermers, R. The hedge fund industry is bigger (and has performed better) than you think. Manuscript.


Reprinted with permission from NOW Publishers (essay I), Taylor & Francis (essay II), and Portfolio Management Research (essay IV).

Original essays are not included in the electronic version of the dissertation.
113. Salehi, Hamed (2020) The use of ETFs and protective option strategies by delegated asset managers
114. Hatami, Akram (2020) Keep Others in mind : a way to proceed with ethical decisions under uncertainty
115. Pohjosenperä, Timo (2020) Value co-creation in health care logistics services
116. Gomes, Julius Francis (2020) Exploring connected health business ecosystems through business models
118. Heikka, Eija-Liisa (2021) Communicating value in knowledge-intensive business relationships
120. Teirilä, Juha (2021) Studies on integrating solar and wind power into electricity markets
121. Beyer, Bianca (2021) Executive compensation and corporate insiders’ personal traits
122. Sarome-Nykänen, Sari (2021) Corporation transformation through network changing agents in the context of outsourcing and automation
123. Likon, Branka (2022) Teacher teaming as a process of a changing network: cross-case analysis of Finnish and Slovenian schools
124. Yang, Jialei (2022) Dynamics in appropriating value from innovation and inter-organizational collaborations
125. Ullah, Zeeshan (2022) Determinants and performance-outcome of sustainable product attributes
126. Huhtala, Tero (2022) Data-based value creation in healthcare service delivery networks
127. Vaisänen, Marjo (2022) Enabling control in post-acquisition integration
128. Tran, Thuy-Van (2022) The relationships between performance evaluation practices and fairness perceptions
129. Okkonen, Hanna (2022) From planning to unexpected outcomes: a process-relational practice approach to understanding the emergence of strategy

Book orders:
Virtual book store
http://verkkokauppa.juvenesprint.fi
Mikko Kauppila

WHITHER ALPHA? HEDGE FUND PERFORMANCE IN VOLUNTARY VERSUS REGULATORY DATA SETS