A Bias-Free Assessment of the Hedge Fund Industry: A New Evaluation of Total Assets, Alphas, and the Flow-Performance Relation *

Daniel Barth† Juha Joenväärä‡ Mikko Kauppila§ Russ Wermers¶

October 8, 2023

Abstract

We combine data from six leading commercial hedge fund vendor databases with confidential regulatory filings to provide a comprehensive evaluation of hedge fund industry size, performance, and investor flows. We estimate that, as of 2019, the industry managed $6.0 trillion in worldwide net assets, 67% larger than the largest vendor estimate. Funds that report only via regulatory filings exhibit better risk-adjusted performance ("alpha"), stronger performance persistence, and a lower sensitivity of investor flows to past returns, relative to those reporting to vendor databases. Our results suggest that "non-vendor-listed" hedge funds have less fragile capital and higher alphas than publicly-marketed funds.

Key words: Hedge funds, net assets, gross assets, strategy, domicile, returns, flows
JEL: G23, G28

*We thank Vikas Agarwal, George Aragon, Nick Bollen, Greg Brown, Randy Cohen, Phillip Howard, Philippe Jorion, Robert Kosowski, Pete Kyle, Christian Lundblad, Phillip Monin, Veronika Pool, Christian Tiu, and seminar participants at the 2021 Western Finance Association Annual Meeting, the Q Group Fall 2021 Seminar, the 4F Symposium at Aalto University, UNC Chapel Hill, University of Maryland, UC Irvine, the Junior European Finance Seminar, the Active Management Research Alliance, Varma Pension Insurance Company, Keva, Veritas, the Federal Reserve Board of Governors, and the Office of Financial Research for helpful comments. Views and opinions expressed are those of the authors and do not necessarily represent official positions or policy of the Federal Reserve Board of Governors of the Office of Financial Research. All remaining errors are the those of the authors and the authors alone. A version of this paper using data through 2016 was circulated under the title "The Hedge Fund Industry is Bigger (and Has Performed Better) Than You Think" (SSRN link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3544181).

†Federal Reserve Board of Governors and the Office of Financial Research, U.S. Department of Treasury. Email: Daniel.J.Barth@frb.gov

‡Aalto University School of Business. Email: juha.joenvaara@aalto.fi

§University of Oulu. Email: mikko.kauppila@oulu.fi

¶Smith School of Business, University of Maryland at College Park. Email: wermers@umd.edu

1 Introduction

Due to limited regulatory constraints and the ability to restrict investor capital, hedge funds serve as a primary arbitrageur in financial markets.[1] Hedge funds can take large short positions, make heavy use of leverage and derivatives, and lock up investor capital in order to invest in illiquid assets and pursue convergence trades with uncertain payoff horizons. These activities promote price efficiency and provide liquidity to less liquid markets. Hedge funds are, in turn, important for understanding modern financial markets; as evidence, between 2010 and 2020, 99 articles in the top three finance journals alone have included the term "hedge fund" in either the title or the abstract.[2] Further, Stulz (2007) predicts that "some hedge funds will choose their investors and how they organize themselves so that they will be less affected by the increasing institutionalization and regulation of the industry", consistent with a potentially increasing role of hedge funds in providing price discovery to markets outside of the view of the general investing public.

At least since the forced wind-down of Long-Term Capital Management in September 1998, hedge funds have been recognized as posing a potential threat to financial stability—largely because of their investment flexibility and the opacity of their trading activities.[3] Since then, the role of hedge funds in financial markets has grown considerably, as the scale of the industry has expanded and its interconnectedness with other major financial market participants has increased. These interconnections plausibly lead to contagion in hedge fund returns that may amplify stress during periods of instability (Boyson, Stahel, and Stulz (2010)). More recently, during the severe March 2020 market turbulence, hedge funds have been tied to volatility in Treasury markets (see Schrimpf, Shin, and Sushko (2020), Barth and Kahn (2021), and Kruttli, Monin, Petrasek, and Watugala (2022)). However, data limitations make a thorough and complete assessment of hedge fund activities difficult. As hypothesized by Stulz (2007), and which we confirm in this paper, much of the hedge fund industry may be outside the view of investors and market observers, except for those few who have privileged access.

In this paper, we present new evidence on hedge funds' aggregate size, performance, and investor flows that includes, for the first time, all hedge funds of a significant size that are offered to at least one U.S. investor. Because most prominent hedge funds around the world have U.S. investors, this represents a dramatic expansion of hedge

[1]Studies that consider hedge funds as a primary arbitrageur include Liu and Mello (2011), Mitchell and Pulvino (2012), Brunnermeier and Pedersen (2008), Hombert and Thesmar (2014), Gabaix, Krishnamurthy, and Vigneron (2007), and Brunnermeier and Nagel (2004).

[2]Here, we follow the generally accepted classification of the Journal of Finance, Journal of Financial Economics, and Review of Financial Studies as the "top three" finance journals.

[3]For discussions of the potential systemic importance of hedge funds, see Dixon, Clancy, and Kumar (2012), Chan, Sherman, Hass, and Lo (2006), and Aragon and Strahan (2012)

fund data.[4] Specifically, we combine information from six major commercial hedge fund data vendors with confidential SEC Form PF regulatory filings for the seven year period of 2013–2019. We use these combined data to provide the first bias-free assessment of hedge fund net (of borrowing and short positions) and gross asset values, risk factor exposures, risk-adjusted performance (alpha), performance persistence, and the investor flow-performance relationship. Our main message is that our results overturn many findings common to the existing hedge fund literature. Compared to prior studies, we find that worldwide hedge fund assets under management (AUM)—both net and gross of leverage—are significantly larger, performance is substantially better on both an unadjusted and risk-adjusted basis, and flows are much less sensitive to performance.

Our combined sample provides numerous advantages compared to the data employed in prior studies. For instance, Edelman, Fung, and Hsieh (2013) (EFH) augment vendor data with information on very large non-vendor-listed funds collected from alternative industry sources. In a replication exercise, we find that, at the end of 2016, our data include $1.3 *trillion* more in net assets under management (AUM) than is captured by the EFH methodology.[5] Agarwal, Fos, and Jiang (2013) use 13F regulatory filings to examine performance differences between vendor-listed and non-listed funds. However, 13F filings include only long equity-holdings and certain derivatives, and are aggregated and reported at the level of the adviser rather than the fund. In contrast, our data include *fund-level* performance associated with the complete portfolio of fund investments, and are not limited to returns to a specific set of securities. Alternatively, Form ADV, a public regulatory filing, is reported at the fund-level but contains no information on net assets, investment strategy, performance, or investor flows. Our unique merger of publicly-available vendor data and confidential regulatory data provides the most informative and comprehensive view of the hedge fund industry to date.

Hedge fund data vendors collect information on a purely voluntary basis, and hedge funds' endogenous decisions to list with a data vendor have significant consequences for the hedge fund literature, as the vast majority of previous studies rely solely on commercial vendor databases (henceforth, we alternatively refer to funds that list with one or more data vendors as *vendor-listed* or *listed* funds). This is particularly problematic because many of the very largest hedge funds choose not to list in any vendor database (we refer to funds that don't list with any

[4]For instance, only one fund in the Institutional Investor Billion Dollar club does not register with the SEC, and that fund is captured in the vendor data.

[5]EFH collect historical printed volumes, over a 10-year period, from Institutional Investor's annual "Hedge Fund 100," which is a list of the 100 largest hedge fund firms, and Absolute Return+Alpha magazine's semiannual "Billion Dollar Club," which is a list of all firms managing $1 billion or more in assets. As we will show, our regulatory data provides a much more inclusive and granular view of the entire hedge fund industry's AUM, performance, and flows than is covered in EFH because there are many large funds that are not captured by either the vendor data or alternative industry collections, but are captured by regulatory data.

vendor as *non-listed* funds).[6] The incomplete nature of hedge fund vendor data is immediately evident from the variation in estimates of total industry size from prominent vendors: Eurekahedge ($2.29 trillion); BarclayHedge ($3.14 trillion); eVestment ($3.26 trillion); Hedge Fund Research (HFR) ($3.32 trillion); and Preqin ($3.61 trillion).[7] The largest publicly available estimate as of the end of 2019 is $4.14 trillion, which comes from the SEC's Form PF — the same regulatory data used here — but does not include any hedge funds not registered with the SEC.[8]

In sharp contrast, our merged dataset covers $6.0 trillion in hedge fund net AUM as of the end of 2019, indicating that the industry is at least 45% larger than the SEC's estimate of industry size, and 67% larger than the largest vendor estimate.[9] At the end of 2019, funds constituting 52.9% of the industry's total net AUM did not report to any of the six vendor databases, which helps to explain the significant underestimates of the size of the industry by commercial database vendors. Further, we calculate that aggregate gross assets under management, roughly equal to equity capital plus borrowing, exceed $11.3 trillion. Moreover, aggregate hedge fund net assets grew by 52% between the first quarter of 2013 and the end of 2019. Notably, consistent with the conjecture of Stulz (2007), growth was faster in non-listed funds; vendor-listed net assets grew by only 30%, whereas non-listed net assets grew by 79%. There is also considerable heterogeneity in the representativeness of the vendor data across various investment strategies and fund domiciles. For Multi-strategy funds and those outside standard classifications ("Other" strategies), the combined vendor data constitutes only 33% and 12% of net AUM, respectively; for Macro and Credit strategies, the vendor data account for 78% and 66% of net AUM, respectively. Overall, the combined vendor data contain only 41% of net assets of hedge funds domiciled in the United States, and less than 37% of net assets of hedge funds domiciled in the Cayman Islands.

The voluntary nature of reporting to a hedge fund data vendor induces a potential self-selection bias that mani-

[6]See EFH and the analysis herein for evidence on the size of non-listed funds.

[7]Hedge fund industry size estimates sourced from: Eurekahedge, as of August 2019 (https://www.eurekahedge.com/Research/News/1930/Global-Hedge-Funds-Infographic-August-2019); BarclayHedge, as of the fourth quarter of 2019 (https://www.barclayhedge.com/solutions/assets-under-management/hedge-fund-assets-under-management/hedge-fund-industry); HFR, as of the end of 2019 (https://www.pionline.com/hedge-funds/hedge-fund-assets-reach-record-high-fourth-quarter-hfr); eVestment, as of October 2019 (https://hedgenordic.com/wp-content/uploads/2019/11/eVestment-Hedge-Fund-Asset-Flows-ReportOctober-2019.pdf); and Preqin, as of November 2019 (https://docs.preqin.com/samples/2020-Preqin-Global-Hedge-Fund-Report-Sample-Pages.pdf). Additionally, vendor estimates of industry size tend to be much larger than the total AUM of funds listed in their respective datasets. Generally, these data vendors use a combination of statistical techniques and industry outreach to infer the size of the industry based on their universe of funds and additional industry information.

[8]See SEC's Private Fund Statistics: https://www.sec.gov/divisions/investment/private-funds-statistics/private-funds-statistics-2021-q3.pdf.

[9]Strictly speaking, the identities and data in vendor datasets are often available only to accredited investors; however, to our knowledge, there is no mechanism to ensure that such data is not dispersed to other investors. Many academic and regulatory institutions also have regular access to these vendor databases. In addition, our paper seeks to understand the investment barriers and rewards available via hedge funds, which are, by their very nature, only available to such accredited investors. Therefore, we often characterize the vendor data as being "publicly available" because of its widespread use by prospective investors, academic researchers, regulators, and market observers, and to distinguish it from regulatory data that is confidential and available only to a select few within the official sector.

fests in multiple ways: attrition bias that results from poorly performing funds exiting the vendor data, backfill bias that arises from funds reporting backdated (generally good) performance upon first listing, and the related incubation bias that results from investment advisers listing only the few funds (of potentially many) that have an encouraging track record.[10] While the literature has sought to address these biases through corrections to the vendor data, adjusting for the bias due to non-listing has proven much more elusive, as many funds never report to a vendor database.[11] If vendor-listed and non-listed funds differ in important ways, due either to observable or unobservable traits, then measures of industry performance, risk exposures, and flows based on vendor data alone may be significantly affected by this bias. And, if these different characteristics are correlated with the decision to list, then ancillary biases (e.g., the backfill bias) can be especially troublesome to researchers seeking to provide unbiased economic results.

Our unique merged data allow us to assess these potential biases directly by comparing funds that list in vendor databases to those that report only in confidential regulatory filings. We find significant differences across a number of key characteristics. First, we document that the aggregate total returns earned by non-listed funds, both on an equally-weighted and value-weighted basis, are notably *larger* than for vendor-listed funds. This performance difference is evident both in aggregate and within nearly every fund strategy category. The total return to the AUM-weighted portfolio of non-listed funds is 63% over the period 2013–2019; comparatively, the total return to the value-weighted portfolio of vendor-listed funds is only 20% over the same period. Similar differences are found within almost all strategy types.

The better performance of non-listed funds may be attributed either to greater risk-adjusted returns (alphas), or greater exposure to systematic risks, such as value, liquidity, or momentum. Here, we find that exposures to a range of systematic risk factors are largely similar between listed and non-listed funds. In fact, non-listed funds tend to have smaller and less-varied factor exposures. However, alphas differ considerably. We estimate alphas through three separate approaches. First, we compare the intercepts from a time-series factor regression using the "Global 7" factor model of Joenväärä, Kauppila, Kosowski, and Tolonen (2021). We find that monthly gross-of-fee alphas for vendor-listed funds have a mean of -0.063% and a median of -0.032%, compared with a mean and median of 0.410% and 0.201%, respectively, for non-listed funds. This translates to an implied annual (compounded) difference

[10]Bias in hedge fund vendor data associated with voluntary reporting was identified as early as 20 years ago (Ackermann, McEnally, and Ravenscraft (1999); Brown, Goetzmann, and Ibbotson (1999); and Fung and Hsieh (2000). For attrition bias, see Liang (2000), Amin and Kat (2003), and Malkiel and Saha (2005); for backfill bias, see Fung and Hsieh (2002), Posthuma and van der Sluis (2003), and Aggarwal and Jorion (2010).

[11]For instance, in an attempt to address survivor bias, many data vendors now include both a "live" data set of active funds, and a "graveyard" data set of historical information retained on non-active funds. To address backfill bias, scholars have advocated excluding the first 12-24 months of performance data reported for a given fund, or all returns before the listing date i.e., the date the fund was added to the database. See, for example, Bhardwaj, Gorton, and Rouwenhorst (2014) and Jorion and Schwarz (2019)

in mean alpha of nearly 5.8 percentage points. We find similar differences when we examine net-of-fee returns, and also find that non-listed funds deliver positive net-of-fee alpha on average.

The significantly larger alphas of non-listed funds suggest that the amount of economic value added through their investment decisions may also be larger. We follow Berk and van Binsbergen (2015) and construct a measure of value-added for each fund, equal to gross-of-fee risk-adjusted return multiplied by prior-month net assets under management. We find that non-listed funds generated an aggregate value-added of $600 billion between 2013 and 2019; in comparison, vendor-listed funds have generated a negative value-added of nearly $200 billion over this period.

The second approach to estimating differences in alphas employs the bootstrap methodology of Fama and French (2010), which allows us to conduct nonparametric hypothesis tests. We create a synthetic sample of hedge fund returns by subtracting each fund's estimated alpha from its actual observed returns. This set of returns maintains all of the empirical characteristics of actual realized returns but has zero alpha by construction. We then compare the t-statistics of the actual, realized alphas to those sampled from the zero-alpha distribution. We find that, for vendor-listed funds, we are unable to reject the null hypothesis that the gross-of-fee alpha is zero at the 5% statistical significance-level for percentiles below the 80^{th} percentile of the alpha t-statistic distribution. Alternatively, for non-listed funds, we are able to reject the null of zero alpha for percentiles beginning at the 30^{th} percentile of the alpha t-statistic distribution. That is, we find little statistical evidence for alpha in the vendor data, except for a minority of right-tail funds, but find substantial evidence of alpha across the empirical alpha distribution of non-listed funds. These findings are qualitatively similar for net-of-fee alphas.

A natural question is whether differences in observable fund characteristics can explain the difference in risk-adjusted performance between listed and non-listed funds. To assess the importance of fund observables, we estimate cross-sectional Fama-MacBeth regressions of pricing errors (alphas plus residuals) estimated from a first-stage factor model on various controls. Controls include fund size, strategy, investor redemption restrictions, fees, indicators for leverage and offshore domicile, and lagged performance. These controls span virtually all fund-level information available from the vendor data. Indeed, many of these characteristics are associated with performance. However, they are unable to explain the *difference* in risk-adjusted returns between listed and non-listed funds. Coefficient estimates indicate that, conditional on characteristics, residual net-of-fee alphas for non-listed funds are 0.13% – 0.38% higher *per month* than for vendor-listed funds, depending on the model, exceed 25 basis points per month in the majority of specifications, and are statistically significant in all cases.

These findings suggest that, contrary to the standard view, self-selection into listing with a data vendor biases estimates of aggregate hedge fund performance *downward*. The conventional wisdom is that funds with strong historical performance, or funds that expect strong performance in the future, have an incentive to report their returns to data vendors to garner interest from prospective investors (indeed, EFH find that hedge funds that newly list in vendor datasets have positive lagged alphas). In support of this prediction, Aiken, Clifford, and Ellis (2013) and Agarwal, Fos, and Jiang (2013) find evidence that selection into vendor-listing biases performance upward. Instead, our results, which are derived directly from data on non-listed funds, show the opposite effect; funds with strong performance tend to avoid disclosing their returns to vendors, implying that estimates of industry performance based on vendor data alone are substantially attenuated.[12] Thus, our evidence suggests that a wholesale reevaluation of the biases inherent in commercial vendor databases is necessary. Our paper provides evidence to help start this reevaluation.

Our final analysis of hedge fund performance examines persistence. Previous studies have found mixed evidence of persistence in hedge fund returns, although most find some support for it (Jagannathan, Malakhov, and Novikov (2010), Kosowski, Naik, and Teo (2007), Boyson (2008), and Fung, Hsieh, Naik, and Ramadorai (2008)). We find strong empirical support for risk-adjusted performance persistence — but only for non-listed funds. Funds that list with a data vendor show limited persistence over short horizons, and no persistence over medium or long horizons. Non-listed funds display strong persistence, with economically large and highly significant coefficients, throughout almost all estimation and prediction horizons. These results remain when we account for return smoothing that may artificially inflate return persistence. Our findings are at odds with the model of Berk and Green (2004), which suggests that competition for manager skill should eliminate performance persistence, but are consistent with Glode and Green (2011), who show that persistence in hedge fund returns may be endogenously related to secrecy surrounding a novel but replicable profit-generating strategy.

Lastly, we examine investor capital flows. There is an extensive literature that examines the flow-performance relation among hedge funds, and in asset management more generally.[13] Our data allow us to examine both aggregate industry flows and whether vendor-listed funds differ in the sensitivity of investor flows to performance. Despite worse performance, we find that investor flows into vendor-listed funds are substantially *higher*. We find that the flow-performance relationship for listed funds is roughly 87% stronger than for non-listed funds. For vendor-listed

[12] Agarwal, Fos, and Jiang (2013) use quarterly 13F holdings data with known limitations, and Aiken, Clifford, and Ellis (2013) use quarterly fund-of-funds holdings with relatively few non-listed funds to deduce quarterly returns of their underlying funds. We have replicated their methodologies, and found that their result of an upward bias in returns remains consistent in our newer 2013-2019 period. Thus, our opposite finding of a downward bias is unlikely to stem from a change in the direction of the bias in recent years, but rather from our use of a more comprehensive dataset.

[13] See Fung, Hsieh, Naik, and Ramadorai (2008), Agarwal, Green, and Ren (2018), Getmansky (2012), and Liang, Schwarz, Sherman, and Wermers (2019) as some recent examples of the existing work on the flow-performance relation among hedge funds.

funds, a 10-point improvement in the percentile rank of performance in the previous quarter is associated with a 0.91 percentage point greater expected flow during the following quarter; for non-listed funds, we find only a 0.12 percentage point increase in flows for a similar improvement in performance. When we estimate the flow-performance relationship separately for positive and negative returns, we find a stronger association for vendor-listed funds for both positive and negative performance, although the association is larger in magnitude and more statistically significant in the positive-return region. One explanation for the better performance but lower flows of non-listed funds is that these funds maintain a gross-of-fee efficient size by returning capital to investors (or refusing further capital injections). Indeed, aggregate flows for non-listed funds are actually *negative* in most quarters. Even so, we show that non-listed funds with positive net flows continue to substantially outperform listed funds, suggesting that non-listed funds near their (gross-of-fee) efficient scale are not solely responsible for the observed differences in performance.

Our results are consistent with hedge fund manager skill associating with the decision to list with commercial hedge fund data services. Our evidence suggests that hedge fund managers with uncertain or difficult-to-forecast skills list in vendor databases to garner attention from investors and to raise capital. Because skill is scarce, managers in the vendor data produce zero alpha on average, and demonstrate little performance persistence. Further, investor flows to these funds are highly sensitive to performance, as investors appear to strongly rely on past returns to update their beliefs about managers' diffuse abilities. Conversely, our results are consistent with managers with more certain and established skills having a greater ability to raise capital without reporting to data vendors, as there is less uncertainty around their skills, and the relationship between performance and flows is therefore smaller as a result. Thus, while we show that the hedge fund industry is substantially larger than previously estimated, the "unreported" capital is more permanent and earns significantly higher alphas. Previous research has highlighted the fragility associated with more volatile and performance-sensitive flows (Chen, Goldstein, and Ji (2010), Goldstein, Jiang, and Ng (2017), Aragon, Nanda, and Zhao (2020)). Our results imply aggregate hedge fund capital may be less fragile than implied by publicly-available commercial vendor data.

Our paper proceeds as follows. In Section 2, we describe the databases used in our study, as well as the matching procedures we use to link them together. Section 3 provides empirical results on the size of the worldwide hedge fund industry, including the distribution of assets under management by strategy and fund domicile. Section 4 examines fund performance. Section 5 studies investor flows. Section 6 concludes.

2 Data

Our data come from a combination of publicly available information from hedge fund data vendors and confidential regulatory filings from SEC Form PF. Below, we describe how we construct both data sets separately and then our methodology for combining them. We then provide summary statistics for relevant variables for vendor-listed and non-listed funds.

2.1 Hedge Fund Vendor Data

To assemble the hedge fund vendor data (what we call the *vendor data* or *public data*, and which include *vendor-listed* or simply *listed* funds), we consolidate six major commercial hedge fund databases: BarclayHedge, Eurekahedge, Hedge Fund Research (HFR), Lipper TASS, Morningstar, and Hedge Fund Management (HFM) Global.[14] Each database contains fund characteristics (e.g., investment style, compensation structure, and redemption restrictions), which we harmonize across databases.

We assign to each database-level fund a harmonized common fund and investment-adviser identifier across and within each of the six vendor datasets. Harmonization across databases is important because each database has a fund identifier that is unique to that database. In order to avoid multiple counting of funds that report to more than one database, we use fund name, returns, AUM, and other information to establish a unique identifier across the union of the vendor databases. Harmonization within databases is important because, except for Preqin, individual databases assign distinct fund identifiers to share classes of the same fund (e.g., onshore and offshore classes). For additional details about the construction of the merged vendor data, we refer the reader to Joenväärä, Kauppila, Kosowski, and Tolonen (2021).

Vendor data is generally reported monthly. To avoid double-counting of any of the listed-fund statistics we present, we take great care to ensure we include only one observation per fund in each period, and exclude all funds-of-funds. In cases where a fund reports different values of returns or other characteristics to different vendor databases, we use the median. For different values of AUM, we use the maximum. This is because sometimes a single share class constitutes a master share class whose AUM represents the total AUM of the fund across all share classes; in this case, using the median would underestimate the AUM of the fund.[15]

[14] In unreported results, we have also included information from two additional vendors for the sub-period 2013-2016: eVestment and Preqin. The results are unchanged by including these databases, which have little independent information given the six databases included in our main sample.

[15] In unreported results, we use a more involved AUM aggregation procedure that sums across distinct share-class level AUM values unless a subset of these values adds up to the maximum (i.e., the potential master share class AUM), with manual verification for the largest funds. This produces near-identical aggregate AUMs for the 2013-2016 sub-period. Joenväärä, Kauppila, Kosowski, and Tolonen (2021) document that most fund-level AUM observations are missing from the eVestment database, and regarding fund coverage, Preqin adds less than 4% to

For share classes that are not denominated in U.S. dollars, we convert their returns and AUM values into U.S. dollars using end-of-month spot rates that are downloaded from Bloomberg. To calculate fund-level flows, we use the value-weighted average of their share-class level flows.

2.2 Non-Listed Data: Form PF

For non-listed hedge fund data, we use fund-level regulatory data from SEC *Form PF* filings ("Private Fund").[16] Form PF was established in a joint rule-making by the SEC and the Commodity Futures Trading Commission in 2011 to fulfill a mandate in the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 to collect data that supports monitoring of large private funds. Except for a short period in 2006, when hedge funds were required to register with the SEC (Brown, Goetzmann, Liang, and Schwarz (2008)), Form PF constitutes the first detailed regulatory data collection of hedge fund information for funds of a minimum size and with at least one U.S. investor.[17] Because many prominent hedge funds globally have U.S. investors, this constitutes a substantial expansion of analytical hedge fund data.

Form PF has different reporting requirements for hedge fund advisers of different sizes, and no reporting requirement for advisers who manage less than $150 million in aggregate private fund assets. All investment advisers registered with the SEC who manage at least $150 million in private fund assets (which include hedge funds, private equity funds, and liquidity funds), must file Form PF at least annually, and report information on total gross and net assets, gross and net returns, total borrowing, strategy classification, investor composition, and the identity of their largest derivative and credit counterparties.[18] *Large Hedge Fund Advisers* — those with at least $1.5 billion in gross assets managed in hedge funds — are required to report this information quarterly for each of the hedge funds they advise. Large Hedge Fund Advisers must report additional information for each of their *Qualifying Hedge Funds* — funds with total net assets of at least $500 million — such as measures of portfolio, investor, and financing liquidity, asset class exposures, collateral posted, risk metrics, and more.[19]

the aggregate database.

[16]Form PF data are confidential. The Office of Financial Research has access to the data through an agreement with the SEC. The form itself is publicly available and can be downloaded here: https://www.sec.gov/rules/final/2011/ia-3308-formpf.pdf.

[17]Form ADV, also filed with the SEC, is a public source of some limited hedge fund data. See https://www.sec.gov/fast-answers/answersformadvhtm.html. Form PF, however, requires much more extensive information with much more granularity in reporting than that required by Form ADV.

[18]Non-U.S.-domiciled advisers are not required to report private fund assets that are not organized in the U.S. and are not offered to U.S. investors. Detailed investment adviser registration requirements can be found at: https://www.sec.gov/about/offices/oia/oia_investman/rplaze-042012.pdf.

[19]The thresholds for filing Form PF and for the "Large Hedge Fund Adviser" classification are on a gross basis, but the threshold for "Qualifying Hedge Fund" status is on a net basis, and is as of the last day in any month in the fiscal quarter immediately preceding the adviser's most recently completed fiscal quarter. Moreover, when determining whether a reporting threshold is met, advisers must aggregate the asset values of the funds themselves, associated parallel funds, dependent parallel managed accounts, and master-feeder funds. Advisers must also include these items for related persons that are not separately operated. Finally, while reporting thresholds are determined on an

Because many large funds do not report to any of the hedge fund data vendors, as documented in EFH and also shown here, and such data services predominantly comprise smaller funds, Form PF offers an unprecedented view of the activities of the largest hedge funds offered to U.S. investors. The lack of vendor listing by many large funds is the primary reason estimates of the size of the hedge fund industry based on vendor databases are too small.

We impose various data restrictions on the Form PF data. Our goal is to preserve as much data as possible while limiting the potential for data errors. We exclude funds that report negative values for either net or gross assets under management. Further, funds are instructed that they may exclude from their reported net assets any equity investments in other private funds; however, they are not *mandated* to exclude these assets. To report the most conservative estimates possible, for any fund with a reported AUM that exceeds their reported equity investment in other private funds, we use AUM minus equity investments in other private funds. This is surely too restrictive, as many funds will report AUM that already exclude investments in other private funds. Unfortunately, there is no way to determine for certain which funds' net assets under management already exclude other private fund investments.

Additionally, we drop funds that report a primary strategy of "Invests in Other Funds" or that have missing strategy data.[20] We also exclude observations with monthly returns greater that 100% or less than -50%. Quarterly flows are winsorized at the 1st and 99th percentiles for both listed and non-listed funds. Form PF does not explicitly collect information on management and incentive fees separately. Instead, for funds that report only to Form PF, we estimate fees following the method described in Barth and Monin (2020), and the necessary restrictions and other details can be found in that paper.

The reporting frequency of Form PF data poses certain challenges. Only funds advised by Large Hedge Fund Advisers file quarterly, while the remaining Form PF filers file annually. Some data, such as returns, are reported at a monthly frequency regardless of whether the fund files quarterly or annually. For variables such as net and gross asset values or strategy allocation information, however, information is reported only at the time of filing — quarterly for quarterly filers and annually for annual filers. To preserve as much information as possible, we fill forward net and gross asset values, reported domicile (U.S., Cayman Islands, etc.), and reported strategy allocations for up to 11 months. Similarly, we fill forward quarterly variables for up to two months.[21]

aggregate basis, advisers are permitted to report fund-level data on either an aggregated or disaggregated basis. Thus, some qualifying hedge funds in our sample have a AUM less than the associated threshold of $500 million. See Flood, Monin, and Bandyopadhyay (2015) and Flood and Monin (2016) for more information on the structure and history of Form PF. The additional information filed by Qualified Hedge Funds is reported in Section 2b of Form PF.

[20] On Form PF, funds report gross asset value allocated to 22 investment strategies that are contained within 8 broad strategy categories: Credit, Equity, Event Driven, Macro, Relative Value, Managed Futures, Invests in Other Funds, and Other. We classify a fund as pursuing a given broad strategy if 75% or more of its assets are allocated to that strategy. If there is no broad strategy category that contains 75% or more of the fund's assets, we classify the fund as Multi-strategy.

[21] Fund domicile is reported on Form ADV, not Form PF. We merge Form PF and Form ADV data by SEC identifier and date to collect

2.3 Matching Procedure

To assemble our comprehensive merged data set, we combine the union of funds that report to one or more of the vendor databases and funds that report to Form PF, without double-counting funds that report to both. Because the Form PF data collection began in 2012, and due to the constraints on the merged vendor data, the period of overlap between the vendor and regulatory data is 2013–2019.[22] This gives us 84 months of data for more than 20,000 unique funds across the fully merged data.

To combine the vendor and regulatory data, we first match the vendor data to regulatory data contained in SEC's Form ADV. Form ADV is a publicly available filing that requires investment advisers to report, on an aggregated basis, information on the private funds that they advise. Form ADV also requires advisers to list the SEC identifier (an "805-" number) of each private fund advised by the adviser. To match SEC identifiers to funds in the vendor data, we first collect the names and SEC identifiers of all private funds identified by their adviser as "hedge funds" on Form ADV. Next, we automatically match each ADV-reporting hedge fund name with its closest vendor-listed fund name (from any of the six vendor databases). We then manually verify and correct these automatic matches, using the vendor adviser name and Form ADV adviser name as context clues. If an ADV fund cannot be matched to a reasonably similar vendor fund cluster, we mark it as a non-matched fund during the manual verification. For these non-matched funds, our only fund-level data come from Form PF.

The value of matching the funds in the vendor data to funds in ADV is that the SEC identifiers are common between Form ADV and Form PF. This allows us to determine whether a given fund reporting to Form PF has also reported to one of the vendor databases. We construct our "non-listed" data set as the set of funds that report on Form PF but were not matched to any fund in any of the vendor databases.

Many funds that report returns and other information to vendor databases often do not provide information on assets under management.[23] If a fund reports to a vendor but does not provide AUM information, we consider it a non-listed fund for the purposes of calculating aggregate industry assets under management. Thus, the AUM we calculate from Form PF are associated with funds that either (i) do not report to any vendor database, or (ii) report to a vendor database but do not provide data on their AUM. We impose a similar methodology for returns. Importantly, due to data restrictions associated with confidential Form PF data, for funds that appear in both Form PF and at least one of the vendor databases, we use the information reported to the vendor database rather than to Form PF. While

fund domiciles.

[22]While the Form PF data began in 2012, we begin our sample in 2013 to ensure data quality and consistency of reporting.

[23]The 2013-2019 vendor data cover 16,427 funds, of which 13,653 (83.11%) report both returns and AUM, 2,751 (16.75%) report only returns, and 23 (0.14%) report only AUM.

the regulatory data are likely more accurate in cases where information is reported to both Form PF and a vendor database, confidentiality requirements make using the regulatory data infeasible in such cases.

Finally, we note that there is no formal definition of a "hedge fund." In order to avoid including private funds that are not properly considered hedge funds, we only include funds that list with hedge fund vendor databases, or funds that categorize themselves as a hedge fund on Form PF.[24]

2.4 Summary Statistics

In our combined sample, funds can fall into one of three categories: (1) those that report only to one or more vendor databases, (2) those that report only to the Form PF regulatory data, or (3) those that report to both a vendor database and to Form PF. This distinction is complicated by the fact that some funds report returns to vendor databases but not AUM, and in a few instances funds in Form PF report gross-of-fee or net-of-fee returns, but not both.[25] For the purposes of summary statistics, we therefore define category (2) as only funds that report neither returns nor AUM to any vendor database, and report either returns or AUM on Form PF. For category (3), we include funds that report either returns or AUM to a vendor and report either returns or AUM to Form PF.

Figure 1 shows the number of funds that fall into each of these three classifications at the end of each quarter and year in our sample.[26] The number of funds reporting to vendor databases (including those that report on Form PF) decreases each year, from a high of 10,348 in March of 2013 to 8,103 in December 2019. Conversely, the number of funds reporting to only Form PF increases each year from 2013 to 2019. In March of 2013, 4,479 funds in our sample reported to Form PF and not to any vendor database; by December 2019, that number grew to 6,229. The number of funds reporting information to both a vendor database and to Form PF drops from 1,622 in March 2013 to 1,247 by December 2019, consistent with the decrease in vendor-listed funds in general. The shrinking number of funds in the vendor databases highlights the importance of regulatory data; funds reporting *only* via Form PF constituted 27% of the total number of funds in our combined sample in early 2013, but constituted 40% by the end of 2019.

Table 1 provides summary statistics, separately for listed and non-listed funds, for variables used in analyses throughout the paper. For vendor-listed funds, all variables are reported at a monthly frequency. For non-listed

[24] There are a tiny fraction of funds in Form PF that describe themselves as "Other private funds" but that file Section 2b of Form PF, which is only to be filed by large hedge fund advisers for each of their qualifying hedge funds. We include "Other" funds that report on Form PF Section 2b in our sample as well.

[25] 16.75 % of vendor-listed funds report returns but not AUM for the period 2013–2019.

[26] There are some hedge funds that report to a vendor database and report to the SEC's Form ADV, but not to Form PF. This could happen for fund advisers that are not large enough to meet the Form PF reporting threshold, for instance. While those funds report to a vendor database and to Form ADV, our convention is to not include them in the calculation of the intersection, and instead include them in the set of funds that report only to a vendor database.

funds, monthly returns as well as (estimated) fees are reported monthly, but the remainder are reported at a quarterly frequency. Because each variable in Table 1 is used in various sets of analyses, we provide statistics for the largest set of observations used in our study. For instance, AUM is the primary variable of interest in section 3.1, and we place only limited restrictions on the sample in that section. But AUM is also used as a control in sections 4.5.4 and 5.2, where the sample is much smaller for non-listed funds due to inclusion of additional variables that are only reported by quarterly filers. In this case, we report summary statistics for AUM based on the observations used in section 3.1. Funds that report both to a vendor database and to Form PF are included with vendor-listed funds.

The first row of Table 1 reports net assets under management. Funds that report to at least one vendor database are considerably smaller than funds that report only on Form PF. The average size of funds that list with a data vendor is $370 million, with a median size of $50 million. For funds that report only on Form PF, the average size is $610 million, with a median of $130 million. This difference in size is partly mechanical; funds that are required to report to Form PF must be advised by registered investment advisers with sufficiently large private fund assets under management (see section 2.2). However, because every fund in the non-listed sample does not report to any vendor database *by construction*, Table 1 also highlights that many large funds do not report to any vendor database (this is consistent with Edelman, Fung, and Hsieh (2013)).

Monthly net-of-fee returns for vendor-listed funds are reported without any data restrictions, such as corrections for backfill bias or the exclusion of very small funds. For non-listed Form PF funds, we require that monthly returns (either gross or net of fee) are between -50% and 100% to eliminate possible data errors.[27] To be conservative, we exclude funds that may be reporting internal rates of return rather than monthly returns. We do so by only including returns for funds with return standard deviations that are greater than 0.25% per month. Vendor-listed funds have a mean net-of-fee monthly return of 0.35%, a median of 0.36%, and a standard deviation of 6.08%. Net-of-fee performance for non-listed funds has an average of 0.49% per month, a median of 0.40% per month, and a standard deviation of 4.96%. While the average net-of-fee performance is higher for non-listed funds, as we show in section 4.1 the differences in performance become much more pronounced when conservative restrictions are imposed, such as value-weighting and minimum observation thresholds. Non-listed funds also demonstrate better gross-of-fee performance by a similar margin. We conduct a thorough examination of returns and risk-adjusted returns (alphas) in section 4.1.

The median management and performance fees in vendor-listed funds is 1.50% and 20.00%, respectively, and

[27] Form PF annual filers report a full 12 months of monthly returns on their filings. So while return data is only reported once per year for annual filers, return data is provided at monthly increments.

for non-listed funds is 0.96% and 19.04%. This suggests that funds that report only to Form PF may charge slightly lower fees on average, although the upper tail of performance fees is substantially larger for non-listed funds.

Due to differences in the way investor share restrictions, such as lockups and redemption notice periods, are reported in Form PF and in the vendor databases, we measure the illiquidity of investors' shares by constructing indicator variables for funds with highly liquid and highly illiquid shares.[28] We define highly liquid funds as those for which all investors are contractually permitted to redeem their capital within at least seven days, including lockups, notice periods, and redemption frequencies. Highly illiquid funds are defined as those for which the most restricted investors are unable to redeem capital within one year or longer. Over 28% of vendor-listed funds are highly liquid, while 16% are highly illiquid. For non-listed funds, only around 15% of funds have highly liquid shares, whereas over 40% of funds are classified as having highly illiquid shares.

We also provide summary statistics for indicator variables for whether the fund employs leverage in their strategy, and whether the fund is domiciled in the Caribbean ("offshore"). Roughly half of vendor listed funds use leverage in their investment portfolio, whereas nearly 80% of non-listed funds use leverage.[29] Just over one-quarter of vendor listed funds are domiciled in the Caribbean compared to half of non-listed funds.

For non-listed funds, flows can only be calculated quarterly because net assets under management is reported quarterly on Form PF. For consistency, we calculate flows at a quarterly frequency for vendor-listed funds as well. As a further preview of our findings, vendor-listed funds have significantly larger investor flows than non-listed funds. The mean quarterly flow for vendor-listed funds is 3.78%, but is -0.10% for non-listed funds. At the 75th percentile, the flows for vendor-listed funds are more than double that of non-listed funds. This finding is unlikely to be driven entirely by closely-held funds available only to insiders; among the funds in the Form PF sample, the median manager ownership is 3% and the 75$^{\text{th}}$ percentile is 16%.[30] We study flows and the flow-performance relationship in more detail in section 5.

Table 1 highlights the differences in characteristics between vendor-listed and non-listed funds. In the following sections, we conduct a thorough analysis of performance and investor flows for the industry in aggregate, as well as separately for listed and non-listed funds. Motivated by the differences in attributes, during much of the analysis

[28]Vendor data sets generally provide information on the separate components of share restrictions, which include the lengths of initial lockups, redemption notice periods, and redemption frequencies (usually in days). Form PF asks funds to report the fraction of investor capital that can be redeemed within seven pre-specified horizons, where the redemption length reflects the total days from all components combined. The Form PF redemption horizons are: 0-1 days, 2-7 days, 8-30 days, 31-90 days, 91-180 days, 181-365 days, and more than 365 days. Because lockups expire, it is impossible to know the total days to redemption for any of the funds in the vendor data; it is likewise impossible to infer the length of individual components from Form PF's aggregated redemption horizons.

[29]This statistic applies to Qualifying Hedge Funds only, since we restrict to only funds that report investor share liquidity measures, and only Qualifying Hedge Funds report that information.

[30]This information is sourced from Form ADV.

we include controls for fund size, strategy, fees, leverage, fund domicile, and share restrictions to better determine whether any observed differences are the result of observables or the endogenous decision to list in a vendor database.

3 The Size of the Hedge Fund Industry

3.1 Net Assets Under Management

Our estimate of the size of the hedge fund industry is the aggregation of AUM reported to vendor databases plus AUM reported to Form PF, while avoiding double-counting. We include, in the vendor-listed AUM, funds that report AUM to at least one data vendor. The vendor-listed AUM therefore comprises funds that report only to one or more vendor databases, as well as funds that report to both vendor databases and file Form PF. Funds that report returns to a data vendor, but not AUM, are included in the non-listed-only sample if they also report to Form PF (which requires disclosure of AUM).

The top panel of Figure 2 plots the time-series of AUM for vendor-listed and non-listed funds. Funds reporting to a data vendor that also file Form PF are included in the vendor-listed set. The total AUM of funds reporting to at least one public vendor database is just over \$2.84 trillion as of December 2019. The total AUM of non-listed funds is \$3.20 trillion, more than \$360 billion more than the AUM reported across the *entirety* of the vendor data.

The time-series of total AUM (vendor-listed plus non-listed) is shown in the bottom panel of Figure 2. Aggregate industry AUM is \$6.04 trillion as of the end of 2019. This is about 46% larger than the next highest estimate of \$4.14 trillion as of December 2019, which is provided by the SEC using (only) data derived from Form PF.[31] We note that our estimate is not larger due to a new estimation technique or imputation procedure; rather, our estimates are larger simply because our merged data contain a larger cross-section of funds than has been previously tabulated in any other academic, industry, or regulatory analyses. The bottom row of Panel A in Table 2 reports the values of listed and non-listed AUM, as well as their totals, as of year-end over the period 2013–2019. Note that total non-listed AUM is substantial during every year of this period, which indicates a long-term potential for biases in studies that include only vendor-listed hedge funds.

Further, while the AUM reported to commercial data vendors has declined since the end of 2017, the AUM reported only to Form PF has grown steadily over the sample period. From the end of the first quarter of 2013 to the end of the fourth quarter of 2019, vendor-listed AUM has grown from \$2.19 to \$2.84 trillion (a 30% increase), while non-listed AUM has grown from \$1.79 to \$3.20 trillion (a 79% increase). This further highlights the importance of

[31] See SEC Private Fund Statistics: https://www.sec.gov/divisions/investment/private-funds-statistics/private-funds-statistics-2021-q3.pdf.

regulatory data to build a more precise estimate of the size and growth of the industry.

We also note that while the regulatory data are an important source of information not available in the vendor data, they are also incomplete. Figure 3 shows the time-series of total AUM managed by listed-funds that have no corresponding SEC identifier; that is, for funds with no requirement to file Form PF or Form ADV. As of 2019, such funds managed more than $1.5 trillion in net AUM, which demonstrates a significant data gap in U.S. regulatory collections of global hedge fund data. While U.S. regulatory agencies have no direct oversight over such funds, these funds participate in U.S. financial markets and may engage with U.S. counterparties. This fact illustrates the importance of our merged vendor databases in providing a more complete picture of the industry beyond our regulatory data.

It is useful to compare the incremental increase in AUM due to the inclusion of Form PF data to that which would be achieved through the inclusion of "mega funds" from other sources. Edelman, Fung, and Hsieh (2013) show that estimates of AUM for some of the largest hedge funds are available through various public sources, including the HFM Absolute Return Billion Dollar Club list, the Top 100 Hedge Fund list, and the Top 50 European Hedge Fund list. Does Form PF, which captures data on larger funds, contribute any additional information about industry size beyond what could be obtained through these publications alone? The answer is yes. The inclusion of data from these sources would add an additional $1.2 trillion to our existing vendor data estimate of $2.5 trillion as of the end of 2016, bringing the total to roughly $3.7 trillion. In sharp contrast, the Form PF data contribute an additional $2.7 trillion at the end of 2016, $1.5 trillion more in AUM than would be captured from these additional sources. Thus, our tabulation of hedge fund assets greatly exceeds even the most earnest prior attempts to classify the true size of the hedge fund industry.

We note that we are not able to capture data for funds that neither report on Form PF nor list in any of the major vendor database. This will include small hedge funds that do not meet the minimum size threshold for regulatory reporting and choose not to list, or large funds that are not required to register with the SEC (for example, funds with no U.S. investors). The SEC's Form ADV, which is a publicly available filing containing information on private fund advisers, does provide information on regulatory assets under management for a wider set of hedge funds.[32] However, Form ADV AUM comprises both equity capital and borrowing, so cannot be used to infer the net assets of U.S. registered hedge funds. Yet, these "omitted assets" are unlikely to be substantial. At the end of 2019, we

[32]Form ADV exempts investment advisers that *only* advise private funds and have total regulatory assets under management below $150 million. That means advisers that advise both public and private funds will be required by file Form ADV and report certain information about the hedge funds they advise. Form PF, on the other hand, requires reporting only for advisers with at least $150 million in private fund regulatory assets under management.

find that funds with AUM less than $100M constitute only around 4.2% of total AUM in our combined vendor data. Further, our data include the vast majority of the "mega funds" examined in Edelman, Fung, and Hsieh (2013).[33] Thus, while our estimates may not capture the full amount of AUM in the hedge fund industry, the assets we miss are unlikely to be material.

3.2 Gross Assets Under Management

A well-known feature of hedge funds is the use of balance sheet leverage. Balance sheet leverage arises from the investment of borrowed funds — generally through collateralized borrowing (e.g., securities lending or repurchase agreement borrowing) or through direct borrowing from the fund's prime brokers — and simultaneously increases the assets and liabilities of a fund. All else equal, leverage increases the magnitudes of gains and losses of an investment strategy relative to the payoff of the strategy funded strictly through investor equity capital. Because hedge fund investments are funded by both investor capital and borrowing, the *gross assets* of the fund are a better measure of funds' economic exposures and potential systemic risk to various sectors of financial markets than net assets, which represent only the equity capital of the fund.[34]

Form PF explicitly collects the gross AUM of reporting hedge funds in addition to their AUM, giving us a direct and reliable measure of funds' balance sheet leverage.[35] Unfortunately, the vendor databases only collect this information for a subset of funds, as fund leverage is a potentially sensitive part of a fund's strategy. For vendor-listed funds, we calculate funds' gross asset values (GAV) by scaling the AUM reported to the vendor database by the fund's reported leverage ratio. If the leverage ratio is not reported, we impute it as the AUM-weighted mean leverage ratio within the same quarter and investment style of vendor-listed funds. If the fund reports multiple leverage ratios that differ across databases, we use the median.

While our approach for determining leverage in vendor-listed funds is likely to suffer from estimation error, the average leverage we estimate is highly sensible. We estimate average leverage of 1.86 at the end of 2019 for vendor-listed funds, virtually identical to the average level of 1.89 estimated from non-listed funds.[36] An alternative approach to estimating gross AUM in listed funds that simply multiplies AUM by average leverage (in total, or by

[33] We verify this by checking if the funds captured by the EFH methodology appear in our merged vendor data, and if not, if they appear in Form ADV.

[34] The so-called "Quant Crisis" of August 2007 illustrated that deleveraging by some very large hedge funds very likely was responsible, in part, for the dislocation in financial markets (Khandani and Lo (2011)). Another example is the panic created by the specter of a large "wind-down" by Long-Term Capital Management in September 1998. More recently, hedge funds contributed to the large sell-off in U.S. Treasury markets during the March 2020 instability episode through their unwinding of the cash-futures basis trade (Schrimpf, Shin, and Sushko (2020); Barth and Kahn (2021), Kruttli, Monin, Petrasek, and Watugala (2022)).

[35] Balance sheet leverage is simply gross AUM divided by net equity AUM.

[36] We exclude Form PF funds that report a leverage ratio greater than 75-to-1.

strategy) estimated from Form PF would therefore produce very similar results.

The top panel of Figure 4 shows the value of gross AUM separately for listed and non-listed funds, and the bottom panel shows their total. As before, funds reporting to both a data vendor and that file Form PF are included in the vendor-listed set. Our estimates indicate that gross assets in vendor-listed funds grew from just over \$4.1 trillion in March 2013 to \$5.3 trillion in December 2019. Over the same period, gross assets in non-listed funds grew from \$3.1 trillion to \$6.1 trillion. Total gross AUM in the hedge fund industry exceeds \$11.3 trillion at the end of 2019. Note, again, the higher growth rate of total gross AUM of the non-listed funds over the sample period, as compared to vendor-listed hedge funds.

We make one additional modification to our estimates of hedge fund gross AUM using Form ADV. Because investment advisers must report to Form ADV if their total regulatory assets under management (effectively gross assets) exceed \$150 million, while advisers report only to Form PF if their *private* regulatory assets under management exceed \$150 million, there are some hedge funds that are reported on Form ADV but not on Form PF. For these hedge funds (that also do not list in vendor datasets), we add the value of their regulatory assets under management to our measure of total gross AUM from Form PF.

We once again exercise an abundance of caution by excluding funds that report either as a feeder fund on Form ADV, that categorize themselves as a fund-of-funds, or that are a parallel fund in a parallel fund structure. The addition of these funds increases our estimates of total gross assets in the hedge fund industry to \$12.1 trillion as of December 2019. The time-series of gross AUM that include data Form ADV is shown in Figure 5.

Our estimates of hedge fund net and gross AUM suggests that the hedge fund industry is substantially larger than previously thought. The much larger scale of the industry has implications for the potential amount of arbitrage capacity in financial markets, and the extent to which traditional financial intermediaries may be connected to less-regulated financial actors. In the next two sections, we decompose AUM by fund strategy and domicile to better understand where hedge fund assets reside.

3.3 Net and Gross Assets by Fund Strategy

Panel A of Table 2 decomposes the size of the industry by funds' investment strategies. Our strategy classifications are derived from Form PF, which asks funds to report the fraction of their assets that fall into each of 22 pre-selected strategy categories. These 22 categories fall under eight broader categories: Equity, Relative Value, Macro, Event Driven, Managed Futures, Credit, Invests in Other Funds, and Other. We classify funds that have less than 75% of AUM in any particular category as Multi-strategy funds.

The first observation from Table 2 is that researchers using vendor data, even if it is aggregated across all major vendors, miss a large fraction of funds within each strategy category. This calls into question the robustness of the conclusions found in the numerous previous hedge fund studies that rely exclusively on data from commercial vendors.[37] This issue is particularly salient given the non-random nature of reporting to vendor databases.

Table 2 shows that in 2013 the most under-represented strategies are the Multi-strategy and Other categories, for which vendor databases include only 33% and 12% of reported AUM, respectively.[38] By the end of 2019, Equity-strategy funds reporting to vendor datasets also comprised less than half of all Equity-style hedge fund assets, and two other strategies, Relative Value and Event Driven, had just more than half of assets listed with public data vendors, comprising 52% and 54% of assets, respectively.

One confounding factor in this analysis is the size of the "Other" category in the Form PF data. Vendor-listed funds, which presumably intend to raise additional capital, are incentivized to provide prospective investors with a clear strategy mandate. No such incentive exists in Form PF data, and funds are allowed to indicate that no broad strategy category fits their specific investment objective. In this case, funds are allowed to select the strategy "Other" and write in a self-reported strategy description. This may suggest that assets in the "Other" category in Form PF data are artificially inflated, and the over-representation of vendor-listed funds in certain strategy categories simply results from too many Form PF funds being categorized as "Other" when a standard category is suitable.

However, an alternative explanation is that funds with particularly unique or bespoke strategies are the most harmed by or susceptible to reverse-engineering or front-running, and such funds would face the highest costs of reporting to public vendor databases (see Glode and Green (2011)). We should then expect that the "Other" category in Form PF would be significantly larger than the "Other" category in the vendor data, because managers of rightly categorized "Other" strategies endogenously choose not to report to vendors to maintain a greater degree of secrecy. In this case, the results in Table 2 would suggest an important data gap arising from analyses based only on vendor data: many hedge funds with non-standard or difficult-to-classify strategies my be substantially under-represented.

Table 3 shows similar conclusions based on the distribution of gross assets under management by strategy. However, some important differences emerge. Based on gross assets in December 2019, Relative Value assets are much more concentrated in vendor-listed funds, whereas Event Driven funds are disproportionately contained in the non-listed sample. Non-listed Equity-style assets comprise just under half of all gross assets, while the significant

[37]For example, as documented in EFH, numerous very large hedge funds do not report to vendor databases. As discussed earlier, our approach captures a much larger set of these non-listed funds than would be captured by the EFH methodology.

[38]This finding is consistent with the below-noted tendency of non-listed funds to be vague about their strategies when they file Form PF with the SEC, presumably because they wish to be opaque about their strategies; thus, the large amount of AUM in "Other" in the non-listed database.

differences in Other, Multi-strategy, and Macro strategies remain.

3.4 Net and Gross Assets by Fund Domicile

Table 4 decomposes net and gross assets under management by fund domicile. Fund domicile is a standard characteristic that funds report to the vendor databases because it often associates with considerations relevant to prospective investors, such as tax implications. Information on non-listed fund domicile is provided on Form ADV. Based on a mapping of Form PF funds to Form ADV we are able to determine fund domiciles for funds reporting to Form PF as well.

The results in Table 4 show that Caribbean-domiciled hedge funds have the largest amount of total AUM missing from the vendor databases. The fraction of assets in Caribbean-domiciled funds missing from the vendor data has also grown over time. Roughly 51% of AUM is missing from the vendor data at the end of 2013, while more than 63% is missing by the end of 2019. Nearly 59% of North American domiciled assets are missing from the vendor data at the end of 2019, although this is unsurprising given that the Form PF collection is based on private funds advised by investment advisers that are registered with SEC, and therefore is expected to be skewed toward U.S.-domiciled funds.

Meanwhile, European-domiciled hedge funds are dramatically over-represented in the vendor data. Only 9% of European-domiciled assets are included in the non-listed sample. This may in part reflect UCITS hedge funds, a European regulatory designation that imposes restrictions on various fund activities but permits marketing the fund to a wider range of investors.[39] Although UCITS hedge funds could be marketed in the U.S., only a very tiny fraction of UCITS funds are targeted to U.S. investors. Therefore, information on UCITS hedge funds will be found only in vendor data. Nonetheless, the vendor data likely captures a substantial portion of UCITS hedge funds given their wider marketability (Joenväärä and Kosowski (2021)).

The remaining domiciles appear to contain relatively little of the total net and assets of the global hedge fund industry. Similar patterns by domicile arise for total gross AUM, which are shown in Table 5.

4 Performance

> *The immediate question might be how so many hedge fund managers failed to do better...A bigger question, however, is why the investment performance of hedge funds has been so poor for so long.*
>
> – Dan McCrum, Financial Times, January 5th, 2017

[39] UCITS stands for “Undertakings for Collective Investment in Transferable Securities”.

The dramatic underperformance of hedge funds is pretty amazing considering the survivorship and backfill biases in the index data that skew hedge fund returns upwards by 3% to 5% per year.

– Peter Lazaroff, Enterprising Investor blog at CFAInstitute.org, February 24, 2016

4.1 Returns

Since the 2007 – 2009 financial crisis, the financial press and industry observers have lamented the relatively poor performance of hedge funds. While the nature of hedge funds, at least in their original conception, presupposes that funds should underperform a broad stock index during bull markets and outperform during bear markets, the *degree* of recent under-performance has received scrutiny. Using only data on vendor-listed funds, Bollen, Joenväärä, and Kauppila (2021) confirm this anecdotal evidence by showing that hedge funds have underperformed their benchmarks since 2008.

Most hedge fund studies focus on the returns provided to investors or the alphas generated from various asset-pricing models.[40] The vast majority of these studies rely solely on returns provided by data vendors, and industry observers are likewise limited to data available from industry service providers or private, and likely incomplete, collections.[41] However, without a comprehensive accounting of fund returns, it is difficult to know the extent to which the performance of funds with publicly available data is representative of the experience of the industry as a whole.

The top panel of Figure 6 shows the AUM-weighted average net-of-fee rate of return earned by vendor-listed and non-listed funds over the sample period. In order to highlight the importance of the endogenous reporting decision, returns for vendor-listed funds are not corrected for backfill bias, although value-weighting returns already dampens the size of the bias considerably (see Joenväärä, Kauppila, Kosowski, and Tolonen (2021)). The top panel of Figure 6 indicates that the returns to the vendor-listed and non-listed funds are highly correlated; the time-series correlation of the value-weighted return series is 88% over the sample period. However, while these returns move together, economic magnitudes differ; vendor-listed funds have consistently worse performance. This finding is consistent with the mean performance statistics reported in Table 1.

Differences in the level of monthly rates of return produce large differences in cumulative returns, particularly if non-listed funds exhibit lower return volatility, as indicated in the figure. The bottom panel of Figure 6 plots the growth of one dollar invested at the end of January 2013 in the value-weighted portfolios of funds in the listed and

[40]Getmansky, Lee, and Lo (2015) and Agarwal, Mullally, and Naik (2015) provide excellent surveys of studies of hedge fund returns and alphas.

[41]Exceptions include EFH and Brown, Goetzmann, Liang, and Schwarz (2008).

non-listed databases.[42] That is, the bottom panel of Figure 6 reports the cumulative, total return earned by the value-weighted portfolio of listed and non-listed funds, separately. A dollar invested in vendor-listed funds would be worth \$1.20 at the end of 2019; the same investment in non-listed funds would be worth \$1.63. That is, the cumulative performance for non-listed funds is more than triple that of vendor-listed funds over the sample.

4.2 Performance by Strategy

Figure 7 shows that the outperformance of non-listed funds is consistent across nearly every hedge fund strategy. One dollar invested in the AUM-weighted portfolio of vendor-listed Equity-style funds would be worth \$1.34 by December 2019; one dollar invested in the portfolio of non-listed Equity-style funds would be worth \$1.78, a cumulative performance difference of 44 percentage points. Similar outperformance of the non-listed funds is found in Relative Value strategies (16% total return vs. 53%), Credit strategies (10% vs. 45%), Event Driven strategies (31% vs. 103%), Multi-strategy (11% vs. 50%), Managed Futures (8% vs. 19%), and Other strategies (13% vs. 66%). The only strategy for which the non-listed funds did not outperform the vendor-listed funds is the Macro strategy (16% total return for vendor-listed funds vs. 8% for non-listed funds).

4.3 Positive Flows and Funds in Both Data Sets

There may be different reasons why hedge funds choose not to list in vendor databases, many of which have been previously addressed in the literature, and which would naturally lead to average return differences between listed and non-listed funds. First, as documented by EFH, very large and successful hedge funds — those most likely to have managers who possess active-management skill — may be reticent to list in vendor databases due to the potential for reverse-engineering or otherwise intensified attention to their strategies. Consistent with this interpretation, Agarwal, Fos, and Jiang (2013) and Aragon, Hertzel, and Shi (2013) show that hedge funds' "confidential holdings" exhibit superior performance. Large and successful funds are also more likely to experience diseconomies-of-scale in the capacity of their investment style.[43] Second, unsuccessful hedge funds may strategically choose not to publicize their returns in order to minimize the publicity that their poor performance generates (e.g., through word-of-mouth or through the media), and to stem outflows from existing investors or to promote inflows from new investors.

In section 4.5.4, we formally examine whether a variety of fund characteristics, such as size, investor share liquidity, leverage, fund domicile, or fees, help to explain the difference in performance between vendor-listed and non-listed funds. Here, we aim to better understand performance differences by examining the returns of two specific

[42] We begin the cumulative return series at the beginning of February 2013 because value-weighting requires non-missing AUM data in period $t-1$.

[43] We note, here, that large and successful funds that are unlisted may choose to be more anonymous due to their higher expected diseconomies-of-scale, as compared with smaller and less-successful funds (Yin (2016)).

types of funds: those that report to a data vendor *and* to Form PF, and those that have positive cumulative flows.

In the top panel of Figure 8, we repeat the aggregate performance series of vendor-listed and non-listed (only) funds from Figure 6, but also include the performance of funds that report to at least one vendor and to Form PF (labeled "Funds in Both").[44] Again, we weight returns by beginning-of-period net AUM. Funds that reside in both data sets exhibit performance that is higher than those that report to one or more data vendors, but also exhibit lower performance than those that report only to Form PF. The cumulative return for funds listed with vendors and filing Form PF is 37% over the sample period, compared to 63% for non-listed (only) funds and 20% for listed funds.

These results provide some insight into the forms of selection bias in vendor reporting. The comparison between funds that report only to Form PF, and those that report to both a vendor and to Form PF, is an apples-to-apples comparison of the selection bias induced by voluntary reporting. Both sets of funds are required to report to Form PF, but only the latter set of funds also *choose* to list with a vendor. Meanwhile, the comparison between vendor-listed funds and funds in both data sets isolates the selection bias arising from SEC registration, which comprises both the choice to offer the fund to U.S. investors and the ascension above the minimum size threshold for mandatory Form PF filing. The top panel of Figure 8 offers evidence that both sources of potential selection bias are material; among funds that list with a vendor, funds that also report to Form PF — as a result of choosing to cater to U.S. investors and being above a minimum size — outperform funds that do not file Form PF. And, among funds that file Form PF, funds that also list with a vendor underperform those that do not.[45] Both types of selection-bias suggest that the returns of funds that report to data vendors represent downward-biased estimates of true industry performance.

In the bottom panel of Figure 8, we replace the "funds in both" from the top panel with non-listed funds that have positive cumulative flows as of that date. We do so because one source of performance difference could be that vendor-listed funds are generally open to new investors, whereas funds that have reached their efficient scales might generally be closed to new investors and will not list with any vendor. If more highly skilled managers are more likely to reach efficient scale due to the positive relationship between flows and performance, then funds that are closed to new investors might exhibit superior performance, and might be mechanically correlated with the decision to not list.

We test this by restricting the non-listed aggregate performance measure to include only funds that, at date t, have positive cumulative flows between the date they are first observed in the Form PF data and date t. We restrict

[44] Note that the vendor-listed series in Figure 8 still includes funds that report to Form PF as well; that is, the "Funds in Both" series is a strict subset of the "Vendor Listed" series. We do this to maintain comparability with the bottom panel of Figure 6.

[45] In unreported results, we find some differences in investor-type (e.g. pensions, sovereign wealth funds, and high net-worth individuals) between non-listed and dual-reporting funds, but we are unable to compare investor type for funds that only list with a vendor and do not file Form PF.

to positive *cumulative* flows because funds that are closed to new investors may still have positive flows in quarters following investor redemptions in order to return the fund to its efficient size through additional contributions by incumbent investors. We continue to weight returns by net AUM at the beginning of each month. As shown in the bottom panel of Figure 8, this restriction only slightly decreases the aggregate returns earned by non-listed funds, and remains well above the aggregate returns earned by listed funds. As of December 2019, the value-weighted cumulative return for non-listed funds with positive cumulative flows is 56%, compared to 63% for the unrestricted non-listed sample. Similarly, in unreported results, we find no evidence that the degree of "inside ownership"—the percentage of the fund owned by fund managers—is materially related to returns for non-listed funds.

Prior literature indicates that returns to hedge funds, in general, are likely *worse* than those listing with vendors due to backfill and survivorship biases and, potentially, the desire for hedge fund advisers to selectively list their funds when past performance is favorable. Our findings suggest that the bias associated with the endogenous decision to list is strongly *negative* for returns, and much larger than the positive biases due to survivorship, backfilling returns, or opportunistic strategic listing. We return to this interpretation at the end of the section.

4.4 Systematic Risk

There are two possible sources of difference between the performance of vendor-listed and non-listed funds: exposures to systematic risk factors and risk-adjusted returns (alpha). To decompose the sources of difference in total returns, we estimate a standard factor model of the form:

$$R_{i,t} = \alpha_i + \beta_i' F_t + \varepsilon_{i,t}, \tag{1}$$

where $R_{i,t}$ is fund i's return in period t, α_i is the average return not captured by systematic risk, β_i is a vector of sensitivities to systematic risk factors, F_t is the time t vector of systematic risk factors, and $\varepsilon_{i,t}$ is the residual. We use risk factors from the "Global 7" factor model developed in Joenväärä, Kauppila, Kosowski, and Tolonen (2021). The Global 7 model has been shown to have greater explanatory power for hedge fund returns than the traditional Fung and Hsieh (2004) model, and has the added benefit of comprising true asset pricing factors (returns to long-short portfolios sorted on characteristics) rather than the asset-based style portfolios of Fung and Hsieh (2004). The Global 7 model includes: the global equity market excess return (MKT-RF) and size factor (SMB) from Fama and French (2012); the global value (HML) and cross-sectional momentum (CS MOM) factors from Asness, Moskowitz, and Pedersen (2013); the time-series momentum factor (TS MOM) from Moskowitz, Ooi, and Pedersen (2012); the betting against beta factor (BAB) from Frazzini and Pedersen (2014); and, the traded liquidity factor (PS LIQ) from

Pastor and Stambaugh (2003). Each of the Global 7 factors have been shown to be important sources of systematic risk in hedge fund portfolios.

Figure 9 plots the densities of each $\widehat{\beta}_i$, estimated from gross-of-fee returns, separately for listed and non-listed funds. We focus on gross-of-fee returns to evaluate the risk exposures of hedge funds' invested capital, rather than the exposures ultimately faced by investors. Factor exposures are roughly similar between listed and non-listed funds, with listed funds demonstrating larger exposures to market, size, value, and betting-against-beta factors, and non-listed funds having slightly greater exposure to liquidity risk. For instance, the mean market beta for vendor-listed funds is 0.50 with a standard deviation of 0.59. For non-listed funds, the mean market beta is lower at 0.37, with a standard deviation of 0.50. Betas on the other factors are comparably similar. Across all factors, non-listed funds demonstrate tighter exposures to systematic risks; that is, the tails of the distributions for vendor-listed funds are fatter.

While there are observable differences in factor exposures between listed and non-listed funds, these differences are insufficient to explain the heterogeneous performance observed in the previous section. In fact, the generally lower exposure to systematic risk factors suggests non-listed funds would have lower expected returns all else equal. This points toward alpha as the primary source of differences in performance.

4.5 Alphas

In this section, we examine the extent to which alphas differ between listed and non-listed funds. Anecdotes of large and successful funds that do not publicly disclose performance information are plentiful. James Simon's Medallion fund has produced apparently superior alphas with negative risk factor betas and has not reported to a vendor database since 2003 (Cornell (2020)). Elliot Advisers, Two Sigma, Millennium, DE Shaw, and Citadel are just a few more prominent examples of advisers that list few if any of their hedge funds.

We evaluate alphas through three separate approaches: the intercept from the first stage regression in equation (1) (Jensen's alpha), a bootstrap approach that incorporates the sampling distribution of fund returns, and a Fama-MacBeth approach that allows us to control for fund characteristics. We also estimate the value-added by funds' investment performance as in Berk and van Binsbergen (2015). Lastly, we evaluate both the level and persistence of alphas in listed and non-listed funds.

4.5.1 Jensen's Alpha

Figure 10 plots the densities of $\widehat{\alpha}_i$ estimated from equation (1) for listed and non-listed funds and gross-of-fee returns. As expected from the evidence in sections 4.1 and 4.4, estimated alphas in non-listed funds are substantially larger and less varied. Monthly alphas for vendor-listed funds have a mean of -0.063%, a median of -0.032%, and a standard deviation of 4.116%, while for non-listed funds the mean alpha is 0.410%, the median is 0.201%, and the standard deviation is 2.226%. This corresponds to an implied annual difference in median alpha of 2.83 percentage points, and an even larger difference in mean annual alpha of 5.83 percentage points.

Figure 10 also shows that while alphas in non-listed funds are larger throughout much of the distribution, the largest difference in alphas occurs at the lower quantiles. The 25$^{\text{th}}$ percentile of monthly $\widehat{\alpha}_i$ for vendor-listed funds is -0.67%, whereas for non-listed funds the 25$^{\text{th}}$ percentile is -0.17%, a difference of 0.50% per month. The 75$^{\text{th}}$ and 90$^{\text{th}}$ percentile values of monthly $\widehat{\alpha}_i$ for vendor-listed funds are 0.50% and 1.00%, and for non-listed funds are 0.61% and 1.22%. Further, the left tail of the alpha distribution is significantly longer for vendor-listed funds, indicating that large negative estimates of alpha are more likely for vendor-listed funds. The long left tail for vendor-listed funds is also the source of the larger differences in mean alphas compared to median alphas.

4.5.2 Berk and van Binsbergen Value Added

While returns comparisons are useful for understanding differences in performance between listed and non-listed funds, they may not be the appropriate measure for assessing manager *skill*. Berk and van Binsbergen (2015) develop a measure of skill in the mutual fund industry that is based on *value* rather than returns. Specifically, the authors demonstrate that appropriately measured skill is determined by the product of gross-of-fee, excess (risk-adjusted) returns and assets under management:

$$V_{i,t} = \widetilde{R}_{i,t} \times AUM_{i,t-1}. \tag{2}$$

In equation (2), $V_{i,t}$ is the dollar value of what fund i adds above the “benchmark” return in period t, $\widetilde{R}_{i,t}$ is the gross-of-fee return net of the benchmark return, and $AUM_{i,t}$ is the net assets under management.

As demonstrated above, non-listed funds generate substantially higher Jensen’s alpha than vendor-listed funds, despite managing a roughly similar level of assets. This suggests that aggregate value-added by non-listed funds could also be significantly higher. In Figure 11, we plot the aggregate, cumulative value of $V_{i,t}$ for vendor-listed and non-listed funds, separately. We use the portion of gross-of-fee returns from equation (1) that are not explained by systematic risk exposures as the measure of excess returns: $\widetilde{R}_{i,t} = \alpha_i + \varepsilon_{i,t}$. Between the beginning of 2013 and the end of 2019, non-listed funds generated a total of more than $600 billion in value above the risk-adjusted benchmark. In stark contrast, vendor-listed funds have lost nearly $200 billion in value. The $800 billion dollar difference in value-added between listed and non-listed funds over the sample period highlights the substantial economic differences that arise from the differences in performance that we estimate.

4.5.3 Bootstrap Distributions

The differences in the distributions of $\widehat{\alpha}_i$ in Figure 10 are significant, but they don't lend themselves easily to hypothesis testing without parametric assumptions. In this section, we offer evidence that follows from the bootstrap methodology developed in Fama and French (2010). For each fund i, we generate a vector of empirical returns under the counterfactual that $\alpha_i = 0$. That is, we construct $\widetilde{R}_{i,t} = R_{i,t} - \widehat{\alpha}_i$, where $\widehat{\alpha}_i$ is estimated from equation (1). The set of returns $\widetilde{R}_{i,t}$ maintain all of the empirical characteristics of the actual realized returns, except that alphas for each fund are zero by construction. Using this zero-alpha empirical distribution of returns, we then bootstrap 10,000 simulated samples of the same size as our empirical data (84 months). In the Fama and French (2010) approach, a simulated sample is generated by randomly sampling dates with replacement. However, hedge fund returns are known to have a strong autocorrelation structures (Getmansky, Lo, and Makarov (2004)), and sampling dates at random would destroy this structure. Instead, we employ the stationary bootstrap of Politis and Romano (1994), which samples an initial date $t = 1$ at random from the empirical data (a date between January 2013 and December 2019). Then, for each remaining date $t = 2, ..., T$, we set $s_t^b = 1 + s_{t-1}^b$ with probability $1 - \frac{1}{L}$, and sample s_t^b uniformly otherwise. If t equals December 2019, we set $t+1$ to January 2013 if $t+1$ is not randomly drawn. This ensures that bootstrapped dates are sampled in continuous "blocks" to preserve the autocorrelation structure. Following Ledoit and Wolf (2008), we set $L = 6$ in our analyses.

Finally, for each of the 10,000 bootstrapped (zero-alpha) samples, we re-estimate the factor model specified in equation (1), and calculate the t-statistic of the estimated alpha for each fund, $t(\widehat{\alpha}_i^b)$, where b indexes the bootstrap samples. As in Fama and French (2010), we use the t-statistics of alphas rather than the level of alphas to account for differences in estimation precision that arise from differences in the number of months the fund is observed in the data. The interpretation of $t(\widehat{\alpha}_i^b)$ is similar to an information ratio or appraisal ratio — it is the estimated level of alpha divided by its standard error. This gives us 10,000 cross-sectional samples of alpha t-statistics from a sample with alphas equal to zero by construction. This comprises the empirical zero-alpha counterfactual distribution to which we can compare the actual observed distribution of hedge fund alpha t-statistics.

The benefit of this bootstrap approach is that it allows for the empirical distributions of listed and non-listed funds to differ; we do not need to impose that the only difference in the distributions are due to differences in means. Further, the empirical distributions of fund returns may not be well-approximated by ex-ante specified parametric distributions, and the non-parametric bootstrap approach avoids incorrect inferences that may result from assumptions about functional forms.

Figure 12 plots the probability density functions (PDFs) of alpha t-statistics from the actual hedge fund data against the PDFs obtained by taking the average t-statistics across all 10,000 bootstrapped samples at each percentile. The top-left panel plots the simulated and actual PDFs of gross-of-fee alpha t-statistics for vendor-listed funds, while the right-panel plots PDFs for non-listed funds. Estimated alpha t-statistics in vendor-listed funds are near zero on average. Vendor-listed funds have slightly larger alphas in the upper end of the distribution than the zero-alpha sample, but slightly smaller alphas in the lower end. That is, actual alphas appear to have fatter tails compared with the simulated data for vendor-listed funds.

Conversely, the top-right panel of Figure 12 shows that for non-listed funds, the actual gross-of-fee alphas are centered well to the right of zero. Further, the distribution is similar at very low return values, but the right-tail is much fatter for the actual data compared to the simulated data. This suggests that, unlike vendor-listed funds, alpha t-statistics in non-listed funds are significantly larger than implied by the zero-alpha sample.

The bottom row of Figure 12 plots PDFs based on net-of-fee returns. The results are largely unchanged. Actual vendor-listed alphas lie slightly to the left of zero, with fatter tails than the zero alpha distribution. For non-listed funds, alphas remain centered to the right of zero, with a thicker right tail. This suggests the positive alphas in non-listed funds are not entirely captured by managers in the form of higher fees (as may be predicted in a Berk and Green (2004) style model, for instance), but instead are partially passed through to fund investors.

Table 6 offers an alternative way to view the distributions of alpha t-statistics. The first two columns of each set of comparisons — listed and non-listed and gross- and net-of-fee — repeats the information in Figure 12, which shows the average value of alpha t-statistics across the 10,000 bootstrapped samples at various percentiles ("Sim") and the value of the actual alpha t-statistics at those same percentiles ("Act"). The third column shows the percentage of bootstrapped samples for which the simulated value of the alpha t-statistic at that percentile is below the actual value at that percentile, which is equivalent to a p-value in a one-sided hypothesis test under the null that alpha is zero.

Table 6 shows that the statistical significance of alpha t-statistics from non-listed funds is much larger than vendor-listed funds. For vendor-listed funds and gross-of-fee returns, at the 50^{th} percentile just over 50% of the bootstrapped, zero-alpha samples had alpha t-statistics below the actual value. The fraction of bootstrapped samples that are below the actual value for vendor-listed funds doesn't surpass 95% until between the 70^{th} and 80^{th} percentiles. Alternatively, for non-listed funds, by the 30^{th} percentile 96.1% of bootstrapped samples generate alpha t-statistics below the actual observed alpha t-statistic, and this percentage remains above 99% for each displayed percentile after. That is, one can reject the null hypothesis at the 1% level that actual alphas in non-listed funds are equivalent to those in the zero-alpha sample at each percentile above the 30$^{\text{th}}$.

These key takeaways are similar for net-of-fee returns as well. For vendor-listed funds, net-of-fee alpha does not eclipse the 95% threshold until after the 97$^{\text{th}}$ percentile. For non-listed funds, this threshold is surpassed between the 40$^{\text{th}}$ and 50$^{\text{th}}$ percentiles, and for every displayed percentile after the null of zero-alpha can be rejected at the 2%-level or better. Table 6 offers evidence that alphas in vendor-listed funds between 2013 and 2019 are statistically greater than zero only in the upper end of the distribution, while they are economically large and significantly different from zero for non-listed funds throughout a majority of the distribution.

4.5.4 Fama-MacBeth Regressions

The previous sections show a significant difference in the distributions of estimated alphas between listed and non-listed funds. In this section, we examine whether this difference can be explained by observable fund characteristics. Here, we restrict the analysis to include only non-listed funds that file Form PF quarterly (these are almost exclusively *Qualifying Hedge Funds*), because the covariates we include below are updated quarterly only for these funds; we note that *Qualifying Hedge Funds* constitute more than 80% of gross AUM reported on Form PF.[46]

We estimate two-stage Fama and MacBeth (1973) regressions. The first-stage is given by equation (1), which estimates standard time-series factor regressions for each fund i. In the second stage, we estimate cross-sectional regressions for each period $t \in \{1,2,...,T\}$ of $\widehat{\alpha}_{i,t}(=\alpha_i+\varepsilon_{i,t})$ from the first stage on an indicator for whether the fund lists with a vendor (*List*) and other controls (X):

$$\widehat{\alpha}_{i,t} = \beta_{0,\alpha,t} + \beta_{\ell,\alpha,t} \times List_i + \beta_{X,t} X_{i,t} + \varepsilon_{i,t} \tag{3}$$

Final coefficient estimates are then obtained as the time-series averages of the T second stage cross-sectional regressions, with heteroskedastic and autocorrelated Newey and West (1987) standard errors. We include alphas from various factor models on the left hand side, including the G7 model described above, the standard Fung and Hsieh (2004) seven factor model, and the Fung-Hsieh model augmented by an emerging market factor from Edelman, Fung, Hsieh, and Naik (2012) and the out of the money option factors from Agarwal and Naik (2004). We also include a specification with gross-of-fee alphas from the G7 model, and net-of-fee alphas from the G7 model based on unsmoothed returns as in Getmansky, Lo, and Makarov (2004).

[46] See the SEC's Private Fund Statistics here: https://www.sec.gov/files/investment/private-funds-statistics-2022-q3.pdf.

The Fama-MacBeth approach is useful because a number of potentially important characteristics can be included in the control set, X. Because non-listed funds are substantially larger, and previous research suggests better managers will endogenously acquire more capital and reach larger efficient sizes (e.g., Berk and Green (2004) among open-end mutual funds), size may be an important source of the performance gap between listed and non-listed funds.[47] Performance differences could also result from differences in the liquidity of fund portfolios. Aragon (2007) and Barth and Monin (2020) show that portfolio illiquidity and investor share restrictions can explain a significant portion of hedge fund alpha; if non-listed funds have more illiquid portfolios or tighter share restrictions, this could explain their higher returns. We proxy for portfolio illiquidity using the one-period lagged excess return, as previous work demonstrates that return autocorrelations are related to the illiquidity of the assets. We also construct indicator variables for funds with highly liquid and highly illiquid shares. We define highly liquid funds as those for which all investors are contractually permitted to redeem their capital within at least seven days, including lockups, notice periods, and redemption frequencies. Highly illiquid funds are defined as those for which the most restricted investors are unable to redeem capital within one year or longer. We include the levels of management and incentive fees (Ackermann, McEnally, and Ravenscraft (1999); Agarwal, Daniel, and Naik (2009)). Form PF does not explicitly report management and incentive fees separately, so we estimate them following the method of Barth and Monin (2020). We include indicator variables for whether the fund employs leverage in their investment strategy and whether the fund is domiciled in the Caribbean ("offshore"). Finally, to account for differences in performance associated with various investment objectives, we include controls for the same broad strategy classifications described in section 3.3.[48]

Such analyses are difficult due to the sharing restrictions and confidentiality associated with the respective commercial vendor and Form PF data sets. Performing regressions while respecting the confidentiality of our data sets is nonetheless possible with some simple linear algebra. Let X_1 denote the design matrix and y_1 the observations for the vendor-listed funds, and X_2 and y_2 the same for the non-listed funds. The combined design matrix and observation vector are defined as:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{4}$$

[47] Due to the difficulty in scaling various hedge fund strategies, fund size and performance may also be related. See Teo (2009) and Ramadorai (2013) as two of many examples.

[48] In the Form PF data, returns are provided at a monthly frequency but the other control variables are reported quarterly. We fill forward missing values in non-quarter-end months based on the values at the end of the previous quarter.

The combined OLS regression coefficient estimates are then obtained as

$$\widehat{\beta} = \left(X'X\right)^{-1}X'y = \left(X_1'X_1 + X_2'X_2\right)^{-1}\left(X_1'y_1 + X_2'y_2\right) \tag{5}$$

That is, we can estimate regressions on the full sample of data by estimating $X_1'X_1$, $X_2'X_2$, $X_1'y_1$, and $X_2'y_2$ separately, by two sets of researchers who have different access to data, and then combining them after.[49]

Table 7 reports coefficient estimates. Column (1) shows results for net-of-fee excess returns, whereas columns (2)–(7) show results for alphas estimated from various factor models. In each case, the coefficient on the vendor-listed indicator is negative and statistically significant, consistent with the results in Table 6. Column (1) shows that non-listed funds earn an additional 25 basis points per month in net-of-fee excess returns. In column (2), we find that gross-of-fee alphas from the G7 model are also 25 basis points higher per month on average for non-listed funds.

The remaining columns focus on net-of-fee performance, although gross performance yields similar findings. Column (3) shows that net-of-fee alpha from the G7 model is 35 basis points per month greater on average for non-listed funds. To evaluate whether differences in alphas result from return smoothing, column (4) uses alphas derived from returns that have been filtered by an MA(2) process as in Getmansky, Lo, and Makarov (2004). The coefficient on the vendor-listed indicator is -0.38, highly similar to the coefficient in column (3). This is unsurprising given that the lagged excess return is already included as a control to partially account for smooth returns. In columns (5)–(7), we use alphas from the Fung and Hsieh (2004) style benchmarks, augmented by an emerging market factor from Edelman, Fung, Hsieh, and Naik (2012) and the out of the money option factors from Agarwal and Naik (2004). Coefficient estimates are similar to those estimated for the G7 model. although somewhat smaller. In each specification, t-statistics are no smaller than 2.45, and exceed 5.50 under the G7 factor models, indicating the results are strongly statistically significant.

The significantly larger alphas for non-listed funds persist despite the inclusion of various controls potentially related to outperformance. These controls represent virtually all fund attributes available in the commercial vendor data — that is, if fund attributes other than the endogenous listing decision are the ultimate source of performance differences, such attributes are not available in the vendor data. Nonetheless, observable characteristics are indeed related to performance. Table 7 shows that highly liquid funds earn significantly lower alphas on average, whereas highly illiquid funds earn significantly larger alphas. The lagged net excess return, intended to capture the effect on alpha from smoothed returns, is also strongly related to performance, foreshadowing the persistence results we

[49]The matrices $X_1'X_1$ and $X_1'y_1$ are transformations of the vendor data that would be impossible to reverse engineer into their constituent raw data series. Similarly, $X_2'X_2$ and $X_2'y_2$ cannot be used to "reverse-engineer" Form PF data.

document in the following section. Yet, while certain characteristics are associated with performance, none of the observables in Table 7 can explain the large differences in alpha between listed and non-listed funds. Instead, the endogenous decision to report continues to be a strong driving factor for the primary underlying source of these differences.

4.5.5 Performance Persistence

Tables 6 and 7 document a substantial difference between the alphas earned by listed and non-listed funds. This could arise either because non-listed funds are better able to generate superior risk-adjusted performance, or because they take additional risks that are poorly captured by standard factor models. An alternative possibility is that predicted returns for non-listed funds are estimated with greater noise, for instance if risk-factor exposures are less stable over time.

One approach to assessing whether the alphas documented in Tables 6 and 7 arise from estimation error is to examine performance persistence. If alphas are precisely estimated – that is, if they capture real information about the return generating process – then we would expect, on average, that high-alpha funds in one period would earn high alphas in future periods. In this sense, performance persistence serves as an out-of-sample analysis. Persistence in hedge fund returns has also been of interest to researchers and practitioners generally (see Kosowski, Naik, and Teo (2007), Jagannathan, Malakhov, and Novikov (2010), Glode and Green (2011), Boyson (2008), and Fung, Hsieh, Naik, and Ramadorai (2008) as few examples). However, if listed and non-listed funds display different levels of persistence, previous findings may be incomplete.

To examine whether performance persistence differs between listed and non-listed funds, we perform Fama-MacBeth regressions as in Busse, Goyal, and Wahal (2010). For each monthly cross-section, we regress the performance over a future horizon on the performance over a past horizon, plus controls for fund size and strategy. Performance is measured as the average monthly return minus the return to the fund's equal-weighted style index (calculated separately for listed and non-listed funds).[50] Using style indices instead of risk factors avoids potential benchmark misspecification, to the extent that funds within a style are subject to similar risk factors (see Jagannathan, Malakhov, and Novikov (2010) for further discussion). In the analysis, we use three past horizons: three months, six months, and one year; and four future horizons: three months, six months, first year ahead, and second

[50]We use style portfolios rather than a particular factor model because the estimation and prediction horizons are short, and coefficient estimates would be imprecisely estimated using a time-series regression with factors.

year ahead. Notice that the first and second years ahead are non-overlapping. Specifically, we estimate:

$$\widetilde{r}_{i,t+k_1:t+k_2} = \beta_{0,t} + \beta_{1,t}\widetilde{r}_{i,t-k:t} + \beta_{2,t}Z_{i,t} + \varepsilon_{i,t}, \tag{6}$$

where $t+k_1$ to $t+k_2$ defines the horizon over which we estimate future returns, $t-k$ to k defines the horizon of past returns, and $Z_{i,t}$ is a vector of controls. Standard errors are adjusted for autocorrelation, which may arise from overlapping estimation windows.

We estimate equation (6) separately for the sample of listed and non-listed funds. To assess whether $\beta_{1,t}$ differs systematically between listed and non-listed funds, we also estimate the model for the full sample of funds and interact an indicator for whether the fund lists in a vendor database ($List_{i,t}$):

$$\widetilde{r}_{i,t+k_1:t+k_2} = \beta_{0,t} + \beta_{p,t}List_{i,t} + \beta_{1,t}\widetilde{r}_{i,t-k:t} + \gamma_i \times List_i \times \widetilde{r}_{i,t-k:t} + \beta_{2,t}Z_{i,t} + \varepsilon_{i,t}, \tag{7}$$

where $\gamma_{i,t}$ captures the performance persistence difference between listed and non-listed funds.

We estimate two specifications. The first contains no corrections for smoothed returns. In the second, to avoid estimates being contaminated by short-term autocorrelation due to illiquid strategies or return smoothing, we apply an MA(2) filter as in Getmansky, Lo, and Makarov (2004) to both fund returns and style indices.[51]

Table 8 reports results. For each past and future horizon, we report the coefficient on past performance estimated separately for listed and non-listed funds, as well as the coefficient on the interaction term $\widehat{\gamma}_t$. Coefficients are time-series averages across all monthly cross-sections, and all returns are converted to monthly equivalents to ensure comparability of coefficients across different horizons. Panel A includes controls for fund size and strategy but no correction for smoothing. For vendor-listed funds, coefficients range between 0.054 to 0.179, and are statistically insignificant for prediction horizons greater than six months. Conversely, for non-listed funds, coefficients are much larger, ranging from 0.118 to 0.321, and are statistically significant across all past and future performance horizons, except for the 1-year estimation/2nd-year prediction specification, with t-statistics that exceed 2.6 in each case, and that are almost always greater than 3.0. The coefficient on the interaction term shows the magnitude of differences in average persistence. It ranges from -0.079 to -0.219. The statistical significance of $\widehat{\gamma}_t$ generally follows the statistical significance of listed funds; while non-listed funds generally have strongly statistically significant persistence estimates, listed funds only show statistical significance at shorter prediction horizons, and this discrepancy carries

[51]To increase the power of our tests we do not correct returns for backfill bias in the vendor data. Jorion and Schwarz (2019) show that persistence results are not sensitive to the choice of backfill bias correction method.

through to the interaction term. In panel B, we include corrections for return smoothing. These corrections have little effect on our findings.

The results in Table 8 offer evidence of substantially larger performance persistence of non-listed funds compared to listed funds. For funds that list, we find relatively weak persistence and only for shorter prediction horizons. Non-listed funds display much larger coefficient estimates with statistical significance at longer prediction horizons. These findings offer some assurance that estimated differences in alphas are not the result of estimation error, since historical outperformance is a strong predictor of future outperformance for non-listed funds. These results are consistent with Glode and Green (2011), who demonstrate that persistence in hedge fund returns may arise endogenously as compensation to investors for maintaining secrecy surrounding a novel but replicable profit-generating strategy. This theory is also consistent with our finding that funds which operate non-standard strategies are much less likely to report publicly.

In total, the results in sections 4.1 – 4.5 are consistent with hedge fund manager skill informing the decision to list with a hedge fund data vendor. Managers with unproven skill that seek to raise capital will turn to data vendors to garner attention from prospective investors. Because skill is scarce, managers in the vendor data produce zero alpha on average and demonstrate little performance persistence. Conversely, managers with more established skill are already known to investors and are able to raise capital without reporting to data vendors. For non-listed funds, alpha is present for both gross-of-fee and net-of-fee returns because investors bear some of the costs of the alpha generating strategy (through informal expectations of capital permanence, for instance). For listed funds, gross-of-fee alpha is zero on average which implies net-of-fee alpha must also be zero on average. However, for listed funds net-of-fee alpha turns positive in the upper tail of the distribution just as does gross-of-fee alpha. This indicates that the existence of net-of-fee alpha in the Form PF data but not in the vendor data is due to the existence of gross-of-fee alpha, rather than vendor-listed funds diluting alpha in the form of higher fees.

The sign of selection bias in hedge fund returns has been examined previously, but those studies have reached different conclusions. Edelman, Fung, and Hsieh (2013) analyze a set of large non-reporting funds and find that vendor performance data reasonably well-represent the broader industry. In contrast, Aiken, Clifford, and Ellis (2013) and Agarwal, Fos, and Jiang (2013) conclude that selection bias in vendor data is positive, and that vendor data exaggerate the average skill in the industry. Our findings, based directly on high-quality information about non-listed funds, show the opposite. Our results indicate that vendor-listing selection bias is both large and negative, and that performance estimated from vendor data substantially *understates* the true performance of the industry.[52]

[52] We have replicated the Agarwal, Fos, and Jiang (2013) and Aiken, Clifford, and Ellis (2013) methodologies and found that their result of

Other well-known biases in hedge fund vendor data, such as backfill bias, which results from funds reporting return histories after a stretch of good performance, and delisting bias, where funds with poor performance stop reporting to a vendor database before their ultimate demise, would imply a positive performance bias in vendor data.[53] Our results suggest that these biases are small compared to the negative selection bias induced by non-listed funds.

Our interpretation of these empirical findings has additional predictions for fund flows. For listed funds, investor flows should be more sensitive to performance as investors update their uncertain beliefs about manager skills. Funds that do not list in the vendor data have less uncertainty around true manager skill, and the relationship between performance and flows should be weaker, as short periods of returns have less effect on investors' beliefs about true manager ability. We examine these predictions, as well as their consequence for the flow-performance relationship, in the following section.

5 Flows

5.1 Aggregate Flows

The larger growth of assets under management in non-listed funds, demonstrated in Figures 2 and 4, could result from higher returns or higher net inflows of investor capital. The previous section showed that non-listed funds have outperformed funds that list with vendor databases. In this section, we examine whether flows to non-listed funds also differ. Several hedge fund studies examine the relationship between investor flows and past performance.[54] These studies seek to determine whether hedge fund investors infer manager skill from the fund's past performance record. However, the selection bias demonstrated in the vendor data brings into question whether these past studies are similarly biased in their inference of the flow-performance relation.

If managers with less certain or harder to forecast skills report to vendor databases to attract prospective investors, we may expect an exaggerated response of investor flows to observed past performance as investors update their beliefs about manager ability. Alternatively, if the performance of a non-listed fund is more indicative of true manager skill, we may expect the flow-performance relationship to be steeper for non-listed funds, as performance carries a more precise signal of manager ability. Which of these two effects will dominate is an empirical question.

Due to the delisting issues associated with hedge fund data — listing and delisting that arises from strategic and voluntary reporting in vendor databases, and with size-based reporting thresholds in regulatory data — calculating

an upward bias in returns remains in our newer 2013-2019 period. This indicates that our opposite finding of a downward bias is unlikely to stem from a change in the direction of the bias, but rather because our data are more comprehensive.

[53] On the importance of backfill bias and related discussion, see Jorion and Schwarz (2019) and Bhardwaj, Gorton, and Rouwenhorst (2014); for a nice discussion of delisting bias see Edelman, Fung, and Hsieh (2013).

[54] See Christoffersen, Musto, and Wermers (2014) for a survey of studies of the flow-performance relation.

aggregate net investor flows is not a trivial task. Further, neither the vendor databases nor Form PF collect data on subscriptions and redemptions explicitly, which means that flows must be approximated from AUM and fund returns. One approach would be to calculate the total AUM in the listed and non-listed funds in quarter t and $t-1$, and use the AUM-weighted average returns to each to infer the net flows in quarter t. However, this method will conflate the net flows to funds that report to a database (or Form PF) in both t and $t-1$, the increase in AUM due to newly reporting funds that did not report in $t-1$, and the decrease in AUM due to funds that report in $t-1$ but exit the data prior to reporting assets in period t. For this reason, it is also not possible to calculate credible flow estimates based on aggregated asset and performance data.

Instead, our approach is to calculate flows for each fund that reports in both quarters $t-1$ and t, and then to calculate average flows by value-weighting individual fund flows by AUM in $t-1$:

$$Flow_{i,t} = \frac{AUM_{i,t} - AUM_{i,t-1}(1+r_{i,t})}{AUM_{i,t-1}} \tag{8}$$

$$Flow_t = \sum_i \frac{AUM_{i,t-1}}{AUM_{t-1}} \times Flow_{i,t}, \;\; AUM_{t-1} = \sum_i AUM_{i,t-1} \tag{9}$$

For each fund i we calculate quarterly flows, $Flow_{i,t}$, using the standard approach in the literature, which subtracts AUM in period $t-1$, multiplied by the fund's quarterly return, from AUM in period t and scaled by AUM in $t-1$. Aggregate flows, $Flow_t$, are then simply the AUM-weighted average of individual fund flows (weighted by AUM in quarter $t-1$). Similarly to returns, we calculate total flows, $Flow_t$, separately for listed versus non-listed funds as well as for strategy.

Figure 13 shows the value-weighted, quarterly percentage flows for listed and non-listed funds. The empirical patterns differ substantially from the return results shown in Figure 6; in virtually all quarters, the weighted-average percentage flows to vendor-listed funds exceed those to non-listed funds. Further, for non-listed funds, flows are negative in almost all quarters, whereas flows are positive in many quarters for vendor-listed funds. Over the full sample, the mean quarterly value-weighted percentage flow is 0.26% for listed funds and -1.12% for non-listed funds.

Figure 14 shows that, similar to the return results reported earlier, the flow results are largely consistent across fund strategy. Within nearly every strategy, the vendor-listed funds experience greater percentage flows than non-listed funds. For many strategies, the size of these flow differences is substantial. Vendor-listed Equity funds had an average quarterly percentage flow of 0.44%, versus -0.58% for non-listed funds. Similar disparities exist for Relative Value funds (0.05% versus -0.58%), Macro funds (-0.99% vs. -2.20%), Credit funds (1.26% versus -1.53%), Event

Driven funds (-0.06% versus -3.09%), Multi-strategy funds (0.81% versus -1.46%), and Other strategy funds (1.70% versus -1.04%). Managed Futures is the only fund strategy where the vendor-listed funds have lower average flows than non-listed funds (-0.07% vs. 0.89%).

The results in Figures 13 and 14 are not necessarily surprising. Funds that report to vendor databases likely do so to increase investor awareness and raise additional capital, and one would expect such funds to have greater net flows than funds that are not actively marketing to new prospective investors. Further, successful funds may actively return capital to investors in order to preserve an efficient scale of investments. The significantly larger alphas of non-listed funds documented in section 4.5 would in such cases be consistent with the net *negative* investor flows of non-listed funds documented in this section. As one dramatic example, Renaissance Technologies' famed Medallion fund maintained a $9.5 – $10 billion net asset value between 2010 and 2018, despite the *lowest* annual return over this period being 29.01% (see Cornell (2020)). This suggests a staggering $30 billion of capital was returned to Medallion fund investors over the period while maintaining the same level of net AUM. Still, the flow results offer an interesting contrast with the return results; while the vendor-listed funds perform considerably worse than non-listed funds, they nonetheless have raised substantially more capital as a percentage of AUM than their better-performing, non-listed counterparts. This dichotomy suggests that the association between performance and flows is likely to be notably different between listed and non-listed funds. We explicitly study the flow-performance relationship in the next section.

5.2 The Flow-Performance Relationship

A large body of literature investigates the relationship between performance and investor flows in delegated asset management.[55] The combination of results in Sections 4 and 5.1 has important implications for this relationship. Funds reporting to the vendor data simultaneously exhibit (i) poorer performance and (ii) greater net percentage flows. This suggests that the true flow-performance relationship in hedge funds may differ significantly from that estimated in the existing literature, which almost exclusively relies on publicly available vendor data.

In Table 9, we report results from Fama-MacBeth regressions similar to those described in Section 4.5.4, but with quarterly flows as the dependent variable. In column (1), we include only an indicator variable for whether the fund lists with a data vendor, and confirm the results in Figure 13. Without additional controls, vendor-listed funds experience a 4.01 percentage point larger flow *per quarter* than non-listed funds. Column (2) shows this coefficient changes only slightly with the addition of strategy controls.

[55] The literature is far to large to cite exhaustively. See Fung, Hsieh, Naik, and Ramadorai (2008), Agarwal, Green, and Ren (2018), Getmansky (2012), and Liang, Schwarz, Sherman, and Wermers (2019) as a few of many examples.

In columns (3)–(6), we investigate the flow-performance relationship. Columns (3) and (4) include the fund's percentile rank in the net excess return distribution for each period (lagged by one quarter). Previous research has found that the performance rank may be a better predictor of flows than performance levels (Sirri and Tufano (1998), Liang, Schwarz, Sherman, and Wermers (2019)). Due to data sharing restrictions and the confidentiality of Form PF data, performance distributions are calculated separately for listed and non-listed funds.[56] Column (3) shows that, consistent with previous studies, the performance rank of a fund is strongly related to subsequent investor flows. An increase in rank of 10 percentiles (going from the 50^{th} to the 60^{th} percentile, for instance) is associated with an 80 basis point higher expected quarterly flow. The coefficient has a t-statistic greater than seventeen. Flows are also strongly positively related to the flow in the previous quarter, and negatively related to fund size. Each is consistent with funds raising capital until they reach an efficient scale.

Once performance-rank and other controls are included, vendor-listed funds no longer demonstrate greater flows, on average, than non-listed funds. This suggests it may be differences in fund characteristics, or the slope of the flow-performance relationship, that explains the difference in flows between listed and non-listed funds. In column (4), we interact performance rank with an indicator for whether the fund lists with a data vendor. The coefficient on the interaction term is economically large and statistically significant. Vendor-listed funds expect an 80 basis points greater flow per quarter for a 10 percentile-rank increase in performance than non-listed funds. The t-statistic is 12. This indicates that flows are substantially more sensitive to performance for vendor-listed funds. The specification in column (4) includes controls for fund size, lagged flow, liquidity, fees, leverage, and fund domicile, suggesting again that the endogenous decision to list with a data vendor is linked to observed differences in flows and performance, rather than only observable fund characteristics.

Because separately ranking performance among listed and non-listed funds may be problematic (e.g., for investors who have access to both listed and non-listed funds), in columns (5) and (6) we instead include the level of the lagged net excess return as the measure of performance. The conclusions are the same as those found in columns (3) and (4). Column (5) shows that performance is strongly positively related to flows; a one percentage point higher quarterly return in the previous quarter is associated with 20 basis point greater flow. Again, the result is highly statistically significant. In column (6), the coefficient on the interaction term shows that vendor-listed funds expect to receive 27 basis points higher flows, per quarter, for a one percentage point higher lagged quarterly return than do non-listed funds. The t-statistic of 5.15 again suggests the finding is strongly significant. While neither the separate

[56] That is, a vendor-listed fund with median (50^{th} percentile rank) performance in a given quarter may have a different return than the median non-listed fund in the same quarter.

rankings of performance in columns (3) and (4) or the level of performance in columns (5) and (6) are the ideal specification, the similarity and strength of the results across specifications suggests the flow-performance relationship is dramatically different between listed and non-listed funds.

Lastly, in column (7), we examine whether the stronger flow-performance association among listed funds is asymmetric. Previous research has shown that the flow-performance relationship is convex in equity mutual funds but concave in bond funds.[57] In column (7), we examine whether the difference in the flow-performance relationship between listed and non-listed funds arises from differences in curvature. To do so, we decompose performance into its positive and negative segments, and interact each with the listed-fund indicator variable.

Column (7) suggests that the flow performance relationship is steeper for listed funds in both the positive and negative portions of the return distribution, though the difference in the segment of positive returns is much stronger. For positive lagged quarterly returns, net investor flows are 31 basis points higher per percentage point of additional return for listed funds compared to non-listed funds, but only 19 basis points higher for a similar increase in negative returns. Further, the coefficient on positive returns is highly significant, with a t-statistic grater than four, whereas the t-statistic is less than two for negative returns. This suggests the flow-performance relationship is significantly more convex for listed funds than for non-listed funds.

Our results suggest that vendor data dramatically overstate the sensitivity of flows to performance in hedge funds. The sensitivity of flows to performance is substantially smaller for non-listed funds. Because our sample includes many more vendor-listed funds than non-listed funds, the sensitivity to performance in the full sample is highly similar to the sensitivity for vendor-listed funds in the specifications with interactions. However, Figure 2 shows that total amount of assets managed by listed and non-listed funds are roughly equal, which would suggest that full-sample regressions weighted by size rather than fund count would produce a coefficient on performance that is dramatically smaller than the coefficient estimated only from vendor data. In addition to the selection bias associated with vendor-listing, the flow-performance relation estimated from vendor data may be affected by delisting due to poor performance or new listings due to strong performance. We believe this may be an interesting opportunity for future research.

Our flow-sensitivity results have important implications for one dimension of systemic risk in hedge funds. While we show the hedge fund industry is substantially larger than previously thought, we also find that the "unreported" (non-listed) capital is likely more permanent, earning significantly higher returns and being dramatically

[57] See Christoffersen, Musto, and Wermers (2014) for a summary of the literature on equity mutual fund flows, and Goldstein, Jiang, and Ng (2017) for evidence of a concave flow-performance relation in fixed income mutual funds.

less sensitive to performance. Previous research has highlighted the fragility associated with more volatile and performance-sensitive flows (Chen, Goldstein, and Ji (2010), Goldstein, Jiang, and Ng (2017), Aragon, Nanda, and Zhao (2020)). Our results imply aggregate hedge fund capital may be less fragile than implied by publicly available data. To the extent that "runs on hedge funds" are a potential threat to financial stability, we find the threat may not be so dire.

6 Conclusions

We use data from a merged set of six commercial hedge fund data vendors, in combination with filings from the first systematic U.S. regulatory collection on large hedge funds, to form the most comprehensive hedge fund data set to date. We estimate worldwide net assets under management of $6.0 trillion and gross assets under management that exceeds $11.3 trillion, both of which are more than 40% greater than the largest prevailing estimates. These findings indicate a much bigger "footprint" of hedge funds in financial markets than previously believed, and highlight the importance of regulatory data collections for filling data gaps and assessing the potential systemic risks associated with hedge fund activities.

Our substantially more comprehensive sample of hedge funds results because neither the vendor data nor regulators have a complete view of the total hedge fund industry. Vendor data services collect information only on a voluntary basis, and regulatory data is collected only for funds within the appropriate jurisdictions' regulatory perimeters. While U.S. regulators have no direct oversight over private funds with no U.S. investors, such funds may nonetheless be significant participants in U.S. financial markets or have relationships with U.S. counterparties. Our estimates of the size of missing non-U.S. fund assets should therefore be of interest to U.S. regulators with little visibility into such funds.

Decomposing AUM by strategy and fund domicile shows that the AUM of all strategies are significantly understated by publicly available vendor data. Many, including Equity, Event Driven, Relative Value, Multi-strategy, and Other, roughly 50% or less of assets accounted for by the vendor data. Further, vendor data accounts for less than 41% of AUM of funds domiciled in either the Caribbean or United States (although, importantly, does have good coverage of European UCITS funds).

Our analysis then turns to an examination of fund performance. Our findings suggest that non-listed funds dramatically outperform those that list with vendor databases, both in aggregate and within nearly every fund strategy category. We find that this outperformance is driven almost exclusively by larger alphas, rather than greater exposures to systematic risk factors. Risk factor sensitivities are largely similar between listed and non-listed funds, while non-listed funds have significantly larger alphas. The differences in estimated alphas survive various statistical tests and are not explained by observable characteristics, including fund size, strategy, investor share illiquidity, fees, leverage, or fund domicile. This superior performance led non-listed funds to add $800 billion more in value than vendor-listed funds as measured by the Berk and van Binsbergen (2015) method. Additionally, we document that risk-adjusted returns are substantially more persistent for non-listed funds, providing further evidence that persistent fund characteristics such as manager skill is a likely source of return differences.

Finally, given the vast literature on the hedge fund and mutual fund flow-performance relation, we estimate the average flows (as a percentage of AUM) from investors to hedge funds in the vendor and regulatory data. *A priori*, it is not clear whether hedge funds that choose to list in vendor datasets should have a greater ability to gather assets (due to the increased publicity) or a lesser ability (due to the worse performance documented in this paper). We find that the average net flow is significantly *higher* for vendor-listed funds compared with non-listed funds, despite these funds having significantly worse performance. We then show that the flow-performance relationship for non-listed funds is substantially weaker than for vendor-listed funds, suggesting estimates from only vendor data likely overstate the sensitivity of investor flows to performance.

Our results demonstrate a multitude of potential biases associated with commercial hedge fund vendor data. First, the total AUM of funds that list with a vendor has grown by 30% over the period 2013-2019, whereas the growth of assets in non-listed funds has grown by 79%, suggesting the vendor data may be getting less representative over time. Second, the existing literature (e.g., Aiken, Clifford, and Ellis (2013)) suggests that the performance of vendor-listed funds is likely to *overestimate* the true performance of the industry, due to backfill, delisting, and survivorship biases. We find that (negative) selection bias dominates these effects, so that performance estimates based on vendor-listed funds are instead biased significantly *downward*. Selection appears to be largely related to skill; differences in performance between vendor-listed and non-listed funds is exclusively due to alpha. Third, the combination of better performance and lower net flows associated with non-listed funds results in a flow-performance relationship that is much flatter than is estimated from vendor data alone.

Our findings suggest many avenues for future research. First, while our work demonstrates that the endogenous decision to list publicly may be associated with unobserved manager skill, we stop short of a full examination of this choice. A better understanding of the microfoundations that produce this association are an important next step. Our results also suggest that due to stickier capital, non-listed funds may face fewer limits for arbitrage. This could have implications for the study of market efficiency as well as the role of arbitrage capital in broader systemic risks. Finally, our work suggests that alternative data sources may be increasingly important in the study of private, nonbank financial institutions. As in Harris, Jenkinson, and Kaplan (2014) and Ang, Gorovyy, and van Inwegen (2011), data obtained from capital allocators may provide a path forward.

References

ACKERMANN, C., R. MCENALLY, AND D. RAVENSCRAFT (1999): "The Performance of Hedge Funds: Risk, Return, and Incentives," *The Journal of Finance*, 54(3), 833–874.

AGARWAL, V., N. D. DANIEL, AND N. Y. NAIK (2009): "Role of Managerial Incentives and Discretion in Hedge Fund Performance," *The Journal of Finance*, 64(5), 2221–2256.

AGARWAL, V., V. FOS, AND W. JIANG (2013): "Inferring Reporting-Related Biases in Hedge Fund Databases from Hedge Fund Equity Holdings," *Management Science*, 59(6), 1271–1289.

AGARWAL, V., T. C. GREEN, AND H. REN (2018): "Alpha or beta in the eye of the beholder: What drives hedge fund flows?," *Journal of Financial Economics*, 127(3), 417 – 434.

AGARWAL, V., K. A. MULLALLY, AND N. Y. NAIK (2015): "The Economics and Finance of Hedge Funds: A Review of the Academic Literature," *Foundations and Trends in Finance*, 10(1), 1–111.

AGARWAL, V., AND N. Y. NAIK (2004): "Risks and Portfolio Decisions Involving Hedge Funds," *The Review of Financial Studies*, 17(1), 63–98.

AGGARWAL, R. K., AND P. JORION (2010): "The performance of emerging hedge funds and managers," *Journal of Financial Economics*, 96(2), 238 – 256.

AIKEN, A. L., C. P. CLIFFORD, AND J. ELLIS (2013): "Out of the Dark: Hedge Fund Reporting Biases and Commercial Databases," *The Review of Financial Studies*, 26(1), 208–243.

AMIN, G. S., AND H. M. KAT (2003): "Welcome to the Dark Side," *The Journal of Alternative Investments*, 6(1), 57–73.

ANG, A., S. GOROVYY, AND G. B. VAN INWEGEN (2011): "Hedge fund leverage," *Journal of Financial Economics*, 102(1), 102–126.

ARAGON, G. O. (2007): "Share restrictions and asset pricing: Evidence from the hedge fund industry," *Journal of Financial Economics*, 83(1), 33 – 58.

ARAGON, G. O., M. HERTZEL, AND Z. SHI (2013): "Why Do Hedge Funds Avoid Disclosure? Evidence from Confidential 13F Filings," *Journal of Financial and Quantitative Analysis*, 48(5), 14991518.

ARAGON, G. O., V. NANDA, AND H. ZHAO (2020): "Investor Protection and Capital Fragility: Evidence from Hedge Funds around the World," *The Review of Financial Studies*, 34(3), 1368–1407.

ARAGON, G. O., AND P. E. STRAHAN (2012): "Hedge funds as liquidity providers: Evidence from the Lehman bankruptcy," *Journal of Financial Economics*, 103(3), 570 – 587.

ASNESS, C. S., T. J. MOSKOWITZ, AND L. H. PEDERSEN (2013): "Value and Momentum Everywhere," *The Journal of Finance*, 68(3), 929–985.

BARTH, D., AND J. KAHN (2021): "Hedge Funds and the Treasury Cash Futures Disconnect," Working paper.

BARTH, D., AND P. MONIN (2020): "Illiquidity in Intermediary Portfolios: Evidence from Large Hedge Funds," Working paper.

BERK, J. B., AND R. C. GREEN (2004): "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy*, 112(6), 1269–1295.

BERK, J. B., AND J. H. VAN BINSBERGEN (2015): "Measuring skill in the mutual fund industry," *Journal of Financial Economics*, 118(1), 1–20.

BHARDWAJ, G., G. B. GORTON, AND K. G. ROUWENHORST (2014): "Fooling Some of the People All of the Time: The Inefficient Performance and Persistence of Commodity Trading Advisors," *The Review of Financial Studies*, 27(11), 3099–3132.

BOLLEN, N. P., J. JOENVÄÄRÄ, AND M. KAUPPILA (2021): "Hedge Fund Performance: End of an Era?," *Financial Analysts Journal*, (Forthcoming).

BOYSON, N. M. (2008): "Hedge Fund Performance Persistence: A New Approach," *Financial Analysts Journal*, 64(6), 27–44.

BOYSON, N. M., C. W. STAHEL, AND R. M. STULZ (2010): "Hedge Fund Contagion and Liquidity Shocks," *The Journal of Finance*, 65(5), 1789–1816.

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., W. GOETZMANN, B. LIANG, AND C. SCHWARZ (2008): "Mandatory Disclosure and Operational Risk: Evidence from Hedge Fund Registration," *The Journal of Finance*, 63(6), 2785–2815.

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

BRUNNERMEIER, M. K., AND L. H. PEDERSEN (2008): "Market Liquidity and Funding Liquidity," *The Review of Financial Studies*, 22(6), 2201–2238.

BUSSE, J. A., A. GOYAL, AND S. WAHAL (2010): "Performance and Persistence in Institutional Investment Management," *The Journal of Finance*, 65(2), 765–790.

CHAN, N., M. G. SHERMAN, S. M. HASS, AND A. W. LO (2006): "Do Hedge Funds Increase Systemic Risk," 91(4).

CHEN, Q., I. GOLDSTEIN, AND W. JI (2010): "Payoff complementarities and financial fragility: Evidence from mutual fund outflows," *Journal of Financial Economics*, 97(2), 239–262.

CHRISTOFFERSEN, S. E., D. K. MUSTO, AND R. WERMERS (2014): "Investor Flows to Asset Managers: Causes and Consequences," *Annual Review of Financial Economics*, 6(1), 289–310.

CORNELL, B. (2020): "Medallion Fund: The Ultimate Counterexample?," *The Journal of Portfolio Management*, 46(4), 156–159.

DIXON, L., N. CLANCY, AND K. B. KUMAR (2012): *Hedge Funds and Systemic Risk*. RAND Corporation, Santa Monica, CA.

EDELMAN, D., W. FUNG, AND D. A. HSIEH (2013): "Exploring uncharted territories of the hedge fund Industry: Empirical characteristics of mega hedge fund firms," *Journal of Financial Economics*, 109(3), 734 – 758.

EDELMAN, D., W. FUNG, D. A. HSIEH, AND N. Y. NAIK (2012): "Funds of hedge funds: performance, risk and capital formation 2005 to 2010," *Financial Markets and Portfolio Management*, 26(1), 87–108.

FAMA, E. F., AND K. R. FRENCH (2010): "Luck versus Skill in the Cross-Section of Mutual Fund Returns," *The Journal of Finance*, 65(5), 1915–1947.

——— (2012): "Size, value, and momentum in international stock returns," *Journal of Financial Economics*, 105(3), 457–472.

FAMA, E. F., AND J. D. MACBETH (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81(3), 607–636.

FLOOD, M., AND P. MONIN (2016): "Form PF and the Systemic Risk of Hedge Funds: Risk-measurement precision for option portfolios," *Journal of Alternative Investments*, 18(4), 125–147.

FLOOD, M., P. MONIN, AND M. BANDYOPADHYAY (2015): "Gauging Form PF: Data Tolerances in Regulatory Reporting on Hedge Fund Risk Exposures," Working Paper No. 15-13, Office of Financial Research.

FRAZZINI, A., AND L. H. PEDERSEN (2014): "Betting against beta," *Journal of Financial Economics*, 111(1), 1 – 25.

FUNG, W., AND D. A. HSIEH (2000): "Performance Characteristics of Hedge Funds and Commodity Funds: Natural vs. Spurious Biases," *Journal of Financial and Quantitative Analysis*, 35(3), 291–307.

——— (2002): "Hedge-Fund Benchmarks: Information Content and Biases," *Financial Analysts Journal*, 58(1), 22–34.

——— (2004): "Hedge Fund Benchmarks: A Risk-Based Approach," *Financial Analysts Journal*, 60(5), 65–80.

FUNG, W., D. A. HSIEH, N. Y. NAIK, AND T. RAMADORAI (2008): "Hedge Funds: Performance, Risk, and Capital Formation," *The Journal of Finance*, 63(4), 1777–1803.

GABAIX, X., A. KRISHNAMURTHY, AND O. VIGNERON (2007): "Limits of Arbitrage: Theory and Evidence from the Mortgage-Backed Securities Market," *The Journal of Finance*, 62(2), 557–595.

GETMANSKY, M. (2012): "The Life Cycle of Hedge Funds: Fund Flows, Size, Competition, and Performance," *Quarterly Journal of Finance*, 02(01), 1250003.

GETMANSKY, M., P. A. LEE, AND A. W. LO (2015): "Hedge Funds: A Dynamic Industry in Transition," *Annual Review of Financial Economics*, 7(1), 483–577.

GETMANSKY, M., A. LO, AND I. MAKAROV (2004): "An econometric model of serial correlation and illiquidity in hedge fund returns," *Journal of Financial Economics*, 74(3), 529–609.

GLODE, V., AND R. C. GREEN (2011): "Information spillovers and performance persistence for hedge funds," *Journal of Financial Economics*, 101(1), 1 – 17.

GOLDSTEIN, I., H. JIANG, AND D. T. NG (2017): "Investor flows and fragility in corporate bond funds," *Journal of Financial Economics*, 126(3), 592–613.

HARRIS, R. S., T. JENKINSON, AND S. N. KAPLAN (2014): "Private Equity Performance: What Do We Know?," *The Journal of Finance*, 69(5), 1851–1882.

HOMBERT, J., AND D. THESMAR (2014): "Overcoming limits of arbitrage: Theory and evidence," *Journal of Financial Economics*, 111(1), 26–44.

JAGANNATHAN, R., A. MALAKHOV, AND D. NOVIKOV (2010): "Do Hot Hands Exist among Hedge Fund Managers? An Empirical Evaluation," *The Journal of Finance*, 65(1), 217–255.

JOENVÄÄRÄ, J., M. KAUPPILA, R. KOSOWSKI, AND P. TOLONEN (2021): "Hedge Fund Performance: Are Stylized Facts Sensitive to Which Database One Uses," *Critical Finance Review*, (Forthcoming).

JOENVÄÄRÄ, J., AND R. KOSOWSKI (2021): "The Effect of Regulatory Constraints on Fund Performance: New Evidence from UCITS Hedge Funds," *Review of Finance*, 25(1), 189 – 233.

JORION, P., AND C. SCHWARZ (2019): "The Fix Is In: Properly Backing out Backfill Bias," *The Review of Financial Studies*, 32(12), 5048–5099.

KHANDANI, A., AND A. LO (2011): "What happened to the quants in August 2007? Evidence from factors and transactions data," *Journal of Financial Markets*, 14(1), 1 – 46.

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.

KRUTTLI, M., P. MONIN, L. PETRASEK, AND S. WATUGALA (2022): “LTCM Redux? Hedge Fund Treasury Trading and Funding Fragility,” Working paper.

LEDOIT, O., AND M. WOLF (2008): “Robust performance hypothesis testing with the Sharpe ratio,” *Journal of Empirical Finance*, 15(5), 850 – 859.

LIANG, B. (2000): “Hedge Funds: The Living and the Dead,” 35(3), 309–326.

LIANG, B., C. SCHWARZ, M. G. SHERMAN, AND R. WERMERS (2019): “Share Restrictions and Investor Flows in the Hedge Fund Industry,” Working paper.

LIU, X., AND A. S. MELLO (2011): “The fragile capital structure of hedge funds and the limits to arbitrage,” *Journal of Financial Economics*, 102(3), 491–506.

MALKIEL, B. G., AND A. SAHA (2005): “Hedge Funds: Risk and Return,” *Financial Analysts Journal*, 61(6), 80–88.

MITCHELL, M., AND T. PULVINO (2012): “Arbitrage crashes and the speed of capital,” *Journal of Financial Economics*, 104(3), 469–490.

MOSKOWITZ, T. J., Y. H. OOI, AND L. H. PEDERSEN (2012): “Time series momentum,” *Journal of Financial Economics*, 104(2), 228 – 250, Special Issue on Investor Sentiment.

NEWEY, W. K., AND K. D. WEST (1987): “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55(3), 703–708.

PASTOR, L., AND R. F. STAMBAUGH (2003): “Liquidity Risk and Expected Stock Returns,” *Journal of Political Economy*, 111(3), 642–685.

POLITIS, D. N., AND J. P. ROMANO (1994): “The Stationary Bootstrap,” *Journal of the American Statistical Association*, 89(428), 1303–1313.

POSTHUMA, N., AND P. J. VAN DER SLUIS (2003): “A Reality Check on Hedge Fund Returns,” Working paper.

RAMADORAI, T. (2013): “Capacity constraints, investor information, and hedge fund returns,” *Journal of Financial Economics*, 107(2), 401 – 416.

SCHRIMPF, A., H. S. SHIN, AND V. SUSHKO (2020): “Leverage and margin spirals in fixed income markets during the Covid-19 crisis,” BIS Bulletins 2, Bank for International Settlements.

SIRRI, E. R., AND P. TUFANO (1998): “Costly Search and Mutual Fund Flows,” *The Journal of Finance*, 53(5), 1589–1622.

STULZ, R. M. (2007): “Hedge Funds: Past, Present, and Future,” *Journal of Economic Perspectives*, 21(2), 175–194.

TEO, M. (2009): “Does Size Matter in the Hedge Fund Industry?,” Working paper.

YIN, C. (2016): “The Optimal Size of Hedge Funds: Conflict between Investors and Fund Managers,” *The Journal of Finance*, 71(4), 1857–1894.

7 Tables and Figures

7.1 Tables

Table 1: Summary Statistics

	Vendor-Listed Funds						**Non-listed Only Funds**					
	Mean	**S.D.**	**25th**	**50th**	**75th**	# Obs.	**Mean**	**S.D.**	**25th**	**50th**	**75th**	# Obs.
AUM ($ billions)	0.37	1.82	0.01	0.05	0.21	650,971	0.61	1.80	0.03	0.13	0.46	107,864
Net Return (%)	0.35	6.08	-1.49	0.36	2.17	841,417	0.55	4.54	-0.71	0.40	1.70	301,302
Gross Return (%)	0.56	5.66	-1.34	0.54	2.41	431,133	0.72	4.88	-0.66	0.48	1.89	301,302
Management Fee (%)	1.39	0.57	1.00	1.50	2.00	513,236	1.51	2.06	0.00	0.96	1.92	57,083
Performance Fee (%)	15.33	7.96	10.00	20.00	20.00	513,236	18.67	11.94	8.78	19.04	28.06	57,083
Highly Liquid	27.57					513,236	15.04					57,083
Highly Illiquid	16.37					513,236	40.39					57,083
Leverage	50.77					513,236	78.49					57,083
Offshore	27.10					513,236	49.92					57,083
Flow (% quarterly)	3.78	29.24	-4.84	-0.07	5.05	165,884	-0.10	13.59	-3.98	-0.15	2.10	18,891

Table 1 reports summary statistics for funds that list with at least one hedge fund data vendor and non-listed funds that report only on Form PF. For vendor-listed funds, all variables are reported at a monthly frequency. For non-listed funds, monthly returns as well as (estimated) fees are reported monthly, but the remainder are reported at a quarterly frequency. AUM represents net assets under management, which is equivalent to the equity value of the fund (and excludes leverage). Net Return is the nominal monthly return net of fees, and Gross Return is the monthly return excluding fees. Management Fee is the fee charged based on assets under management. Performance Fee is the fee applied to returns earned above the high-water mark and hurdle rate.

Table 2: Net Assets Under Management by Strategy ($ Billions)

	2013			2014			2015			2016		
Strategy	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total
Equity	687	818	1,505	762	877	1,639	794	892	1,686	769	852	1,621
Other	479	65	545	554	67	621	599	63	663	718	71	789
Multi-strategy	398	196	593	426	227	654	392	264	656	389	255	643
Relative Value	112	281	393	113	280	393	114	253	367	132	256	388
Macro	190	518	708	221	506	728	219	479	698	214	457	671
Credit	93	198	292	125	225	350	123	348	472	123	344	467
Event Driven	131	178	309	121	201	323	148	173	321	145	161	306
Managed Futures	14	250	263	21	250	271	24	300	324	36	289	325
Total	2,104	2,504	4,608	2,343	2,634	4,977	2,413	2,772	5,186	2,526	2,684	5,210

	2017			2018			2019		
Strategy	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total
Equity	926	1,082	2,008	880	970	1,851	1,048	983	2,030
Other	816	91	907	872	86	958	964	98	1,062
Multi-strategy	456	262	719	396	217	614	398	187	585
Relative Value	163	305	468	175	268	443	215	252	467
Macro	216	498	714	198	444	642	209	440	650
Credit	101	419	520	159	450	610	180	481	661
Event Driven	145	172	317	154	162	316	145	159	304
Managed Futures	47	310	357	42	257	299	42	243	285
Total	2,869	3,141	6,010	2,877	2,855	5,731	3,201	2,843	6,044

Table 2 reports total net and gross assets under management for vendor-listed and non-listed funds, in aggregate and by strategy. Strategy categories are determined by Form PF. Values are reported as of the end of the calender year.

Table 3: Gross Assets Under Management by Strategy ($ Billions)

	2013			2014			2015			2016		
Strategy	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total
Equity	1,046	1,299	2,345	1,151	1,398	2,549	1,185	1,495	2,681	1,100	1,411	2,511
Other	631	87	717	741	93	833	771	84	854	1,044	98	1,143
Multi-strategy	904	381	1,285	961	457	1,418	981	556	1,537	1,000	585	1,585
Relative Value	158	582	740	154	581	736	159	545	704	194	571	765
Macro	254	1,242	1,496	290	1,192	1,482	292	1,076	1,369	278	1,024	1,302
Credit	170	437	607	252	455	707	272	666	938	232	638	870
Event Driven	289	271	560	310	317	628	323	283	606	381	276	658
Managed Futures	16	316	332	25	312	337	28	371	400	43	351	394
Total	3,467	4,616	8,083	3,885	4,805	8,689	4,012	5,077	9,089	4,273	4,954	9,227

	2017			2018			2019		
Strategy	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total
Equity	1,342	1,736	3,078	1,324	1,620	2,944	1,618	1,662	3,280
Other	1,108	136	1,244	1,184	135	1,319	1,338	161	1,499
Multi-strategy	1,257	622	1,879	1,162	552	1,714	1,171	450	1,621
Relative Value	236	701	937	284	638	922	320	628	947
Macro	277	1,108	1,384	250	954	1,204	266	957	1,222
Credit	240	762	1,002	425	798	1,223	736	863	1,599
Event Driven	397	292	689	462	274	736	571	274	845
Managed Futures	60	357	417	57	298	355	62	283	346
Total	4,917	5,713	10,630	5,148	5,268	10,416	6,082	5,277	11,359

Table 3 reports total net and gross assets under management for vendor-listed and non-listed funds, in aggregate and by strategy. Strategy categories are determined by Form PF. Values are reported as of the end of the calender year.

Table 4: Net Assets by Geographic Region ($ Billions)

Region	2013 Non-Listed Only	2013 Listed	2013 Total	2014 Non-Listed Only	2014 Listed	2014 Total	2015 Non-Listed Only	2015 Listed	2015 Total	2016 Non-Listed Only	2016 Listed	2016 Total
Caribbean	1,043	1,010	2,053	1,176	1,043	2,219	1,206	1,026	2,233	1,218	957	2,175
North America	776	566	1,341	847	616	1,463	861	740	1,601	967	731	1,698
Europe	101	828	930	101	883	983	94	923	1,017	93	900	993
Others	184	99	284	219	93	312	252	83	335	248	96	344
Total	2,104	2,504	4,608	2,343	2,634	4,977	2,413	2,772	5,186	2,526	2,684	5,210

Region	2017 Non-Listed Only	2017 Listed	2017 Total	2018 Non-Listed Only	2018 Listed	2018 Total	2019 Non-Listed Only	2019 Listed	2019 Total
Caribbean	1,348	1,016	2,364	1,345	910	2,256	1,523	892	2,415
North America	1,117	855	1,972	1,115	858	1,973	1,252	877	2,129
Europe	123	1,147	1,270	148	969	1,118	160	961	1,121
Others	281	123	404	268	117	385	266	113	379
Total	2,869	3,141	6,010	2,877	2,855	5,731	3,201	2,843	6,044

Table 4 reports total net assets under management for vendor-listed and non-listed funds, in aggregate and by fund domicile. For non-listed funds, domicile is determined from Form ADV. Values are reported as of the end of the calender year.

Table 5: Gross Assets by Geographic Region ($ Billions)

	2013			2014			2015			2016		
Region	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total
Caribbean	1,936	1,962	3,898	2,191	2,057	4,248	2,256	2,029	4,285	2,354	1,902	4,255
North America	1,109	878	1,987	1,206	959	2,165	1,219	1,092	2,311	1,384	1,075	2,459
Europe	126	1,625	1,751	121	1,650	1,771	109	1,839	1,948	105	1,842	1,947
Others	296	151	447	366	139	505	428	117	545	430	136	567
Total	3,467	4,616	8,083	3,885	4,805	8,689	4,012	5,077	9,089	4,273	4,954	9,227

	2017			2018			2019		
Region	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total	Non-Listed Only	Listed	Total
Caribbean	2,695	1,977	4,672	2,851	1,816	4,666	3,663	1,784	5,447
North America	1,564	1,259	2,823	1,586	1,307	2,893	1,804	1,380	3,184
Europe	135	2,300	2,435	166	1,974	2,140	182	1,944	2,125
Others	523	177	700	545	171	717	434	169	603
Total	4,917	5,713	10,630	5,148	5,268	10,416	6,082	5,277	11,359

Table 5 reports total gross assets under management for vendor-listed and non-listed funds, in aggregate and by fund domicile. For non-listed funds, domicile is determined from Form ADV. Values are reported as of the end of the calender year.

Table 6: Distribution of Alphas

	Gross of Fee Returns						Net of Fee Returns					
	Vendor-Listed			Non-Listed Only			Vendor-Listed			Non-Listed Only		
Pct	Sim	Act	% < Act	Sim	Act	% < Act	Sim	Act	% < Act	Sim	Act	% < Act
1	-2.96	-3.93	12.04	-3.01	-3.20	33.67	-2.95	-4.51	3.82	-2.92	-3.40	21.40
2	-2.52	-3.52	9.15	-2.42	-2.57	34.59	-2.52	-4.00	3.18	-2.38	-2.80	20.23
3	-2.26	-3.21	8.72	-2.14	-2.27	34.94	-2.27	-3.71	2.74	-2.11	-2.48	20.41
4	-2.08	-3.04	7.62	-1.95	-2.07	35.06	-2.08	-3.48	2.48	-1.92	-2.30	18.98
5	-1.94	-2.84	7.66	-1.80	-1.94	33.20	-1.94	-3.24	2.57	-1.78	-2.13	18.82
10	-1.46	-2.24	6.11	-1.34	-1.38	40.96	-1.47	-2.67	1.34	-1.33	-1.59	20.41
20	-0.92	-1.48	5.82	-0.85	-0.71	68.37	-0.94	-1.93	0.62	-0.84	-0.90	36.21
30	-0.57	-0.91	8.44	-0.51	-0.22	96.10	-0.58	-1.38	0.46	-0.51	-0.40	68.62
40	-0.28	-0.45	16.62	-0.24	0.23	99.93	-0.29	-0.90	0.54	-0.24	0.00	93.15
50	-0.01	-0.01	50.71	0.01	0.64	99.98	-0.02	-0.45	0.97	0.01	0.40	98.66
60	0.25	0.45	86.83	0.26	1.03	99.95	0.25	-0.01	5.18	0.26	0.78	99.01
70	0.54	0.88	93.98	0.54	1.50	99.93	0.54	0.48	41.22	0.54	1.21	99.09
80	0.90	1.46	97.18	0.88	2.08	99.89	0.90	1.01	68.86	0.88	1.79	99.29
90	1.43	2.39	98.75	1.39	3.04	99.90	1.44	1.83	85.25	1.40	2.68	99.43
95	1.94	3.26	98.98	1.87	4.04	99.91	1.95	2.66	91.27	1.88	3.64	99.61
96	2.11	3.64	99.26	2.03	4.56	99.94	2.11	2.92	91.91	2.04	3.98	99.65
97	2.33	4.24	99.68	2.23	5.13	99.93	2.33	3.37	94.31	2.24	4.54	99.79
98	2.65	5.24	99.87	2.52	6.17	99.93	2.64	4.24	97.46	2.54	5.43	99.83
99	3.25	8.41	99.99	3.08	9.57	99.98	3.25	7.11	99.95	3.11	8.97	99.94

Table 6 shows various percentile values for alpha t-statistics from the actual data, and the values calculated by averaging (at each percentile) across the 10,000 bootstrapped samples. Pct is the percentile, Sim is average alpha t-statistic at that percentile across the 10,000 bootstrapped samples, Act is the alpha t-statistic value in the actual data at that percentile, and % $<$ Act is the percentage of the 10,000 bootstrapped samples that have alpha t-statistics that are lower than the actual alpha t-statistic at that percentile. Values are shown for gross-of-fee and net-of-fee returns, and for listed and non-listed funds.

Table 7: Fama MacBeth Performance Regressions

Dep. Var.	Net Excess Return	Gross G7 Alpha	Net G7 Alpha	Net G7 (GLM Adj.) Alpha	Net FH Alpha	Net FH + Em Mkt Alpha	Net FH + Option Alpha
List	-0.25	-0.25	-0.35	-0.38	-0.20	-0.16	-0.13
	-3.55	-5.54	-7.26	-7.59	-3.49	-2.82	-2.45
log(Size)	-0.04	-0.04	-0.03	-0.03	-0.03	-0.04	-0.02
	-3.58	-4.62	-3.97	-4.26	-3.89	-4.82	-2.33
Net excess return (% pq)	0.04	0.04	0.03	0.02	0.03	0.03	0.03
	3.55	4.58	4.19	2.99	3.47	3.53	3.45
Management fee (%)	0.00	0.00	-0.03	-0.02	-0.01	0.00	-0.01
	0.05	-0.31	-2.46	-2.27	-1.11	0.13	-0.91
Incentive fee (%)	0.00	0.01	0.01	0.01	0.01	0.01	0.01
	0.51	6.57	5.12	4.41	6.16	6.46	5.75
Highly liquid	-0.07	-0.27	-0.26	-0.26	-0.16	-0.11	-0.11
	-1.57	-6.54	-7.23	-7.27	-4.17	-2.74	-3.13
Highly illiquid	0.11	0.27	0.26	0.25	0.17	0.11	0.12
	1.96	4.69	4.41	4.18	3.62	2.45	2.64
Leverage Indicator	-0.01	0.04	0.02	0.03	0.04	0.00	0.05
	-0.24	1.63	0.86	1.16	1.40	0.15	1.73
Offshore Indicator	0.01	0.12	0.06	0.03	0.07	0.07	0.10
	0.31	2.34	1.35	0.73	1.62	1.72	2.28
Strategy Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	527,601	353,023	505,538	505,501	505,538	505,538	505,538
Number of Cross-sections	81	81	81	81	81	81	81

Table 7 shows regressions of monthly percentage risk-adjusted returns, defined as $\widehat{\alpha}_{i,t} = R_{i,t} - \beta_i' F_t$, on various controls (lagged by one month). List is an indicator equal to one if the fund lists with a public data vendor; log(Size) is the log value of net assets; Net excess return is the net-of-fee quarterly return minus the risk-free rate, Management fee is the fee rate charged on assets under management, Incentive fee is the variable fee charged as a proportion of returns, Highly liquid is an indicator variable for funds where all investors can redeem within seven days, and Highly illiquid is an indicator for funds where the longest amount of time needed to redeem capital exceeds 365 days. Strategy controls indicate whether broad strategy indicator variables are included in the regression. t-statistics are reported below coefficients with Newey-West standard errors.

Table 8: Performance Persistence

Panel A: Standard Controls

α Estimation Horizon		3 Months			6 Months			1 Year		
α Prediction Horizon	**Reported Value**	**Non-Listed**	**Listed**	$\widehat{\gamma}_t$	**Non-Listed**	**Listed**	$\widehat{\gamma}_t$	**Non-Listed**	**Listed**	$\widehat{\gamma}_t$
3 Months	Estimate	0.229	0.102	-0.133	0.195	0.082	-0.119	0.126	0.054	-0.079
	t-statistic	4.481	2.195	-2.379	4.267	2.009	-1.885	2.602	1.796	-1.282
6 Months	Estimate	0.296	0.134	-0.172	0.244	0.108	-0.146	0.149	0.069	-0.092
	t-statistic	4.584	1.898	-1.777	4.292	1.572	-1.228	3.282	1.223	-0.696
1st Year	Estimate	0.313	0.109	-0.219	0.244	0.097	-0.167	0.147	0.089	-0.085
	t-statistic	3.870	0.882	-1.271	3.594	0.712	-0.921	3.297	0.685	-0.666
2nd Year	Estimate	0.321	0.179	-0.202	0.239	0.156	-0.146	0.118	0.072	-0.103
	t-statistic	3.223	0.887	-1.209	3.125	0.634	-1.190	0.863	0.634	-2.242

Panel B: GLM (2004) Corrections

α Estimation Horizon		3 Months			6 Months			1 Year		
α Prediction Horizon	**Reported Value**	**Non-Listed**	**Listed**	$\widehat{\gamma}_t$	**Non-Listed**	**Listed**	$\widehat{\gamma}_t$	**Non-Listed**	**Listed**	$\widehat{\gamma}_t$
3 Months	Estimate	0.213	0.079	-0.138	0.184	0.072	-0.116	0.121	0.048	-0.079
	t-statistic	4.438	1.676	-2.600	4.493	1.834	-1.929	3.355	1.653	-1.932
6 Months	Estimate	0.286	0.118	-0.175	0.242	0.100	-0.151	0.150	0.062	-0.098
	t-statistic	4.628	1.825	-2.004	4.456	1.569	-1.408	3.656	1.240	-0.861
1st Year	Estimate	0.309	0.097	-0.226	0.249	0.090	-0.177	0.153	0.084	-0.096
	t-statistic	3.371	0.800	-1.473	3.758	0.696	-1.027	4.156	0.722	-1.018
2nd Year	Estimate	0.323	0.171	-0.207	0.244	0.155	-0.150	0.116	0.070	-0.106
	t-statistic	2.724	0.809	-1.161	3.447	0.660	-1.166	0.966	0.668	-2.062

Table 8 shows regressions of future performance, measured as returns minus the return to the broad strategy index, on historical performance. All returns are converted to monthly equivalents to ensure comparability of coefficients across different horizons. Regressions are estimated on the full sample as well as separately for listed and non-listed funds. $\widehat{\gamma}_t$ is the coefficient on the interaction between an indicator for whether the fund lists with a data vendor and the estimation-horizon return. Controls include fund size and strategy. Getmansky, Lo, and Makarov (2004) corrections use returns filtered from an MA(2) process. Note that the 1st year and 2nd year prediction horizons are non-overlapping. t-statistics are reported below coefficients with Newey-West standard errors.

Table 9: Fama MacBeth Flow Regressions

Dep. Var.	Flow (% Qtr)	Flow (% Qtr)	Flow (% Qtr)	Flow (% Qtr)	Flow (% Qtr)	Flow (% Qtr)	Flow (% Qtr)
List	4.01	3.89	-0.33	-4.28	-0.09	-0.38	-0.47
	10.16	8.24	-0.85	-6.38	-0.27	-0.84	-0.95
List x Performance Rank				0.08			
				12.01			
Performance Rank (percentile)			0.08	0.01			
			17.38	2.87			
List x Net Excess Return						0.27	
						5.15	
Net Excess Return (%, pq)					0.20	-0.04	
					10.11	-0.81	
List x Max(0, Ret) (% pq)							0.31
							4.02
List x Min(0, Ret) (% pq)							0.19
							1.89
Max(0, Ret) (% pq)							-0.14
							-2.07
Min(0, Ret) (% pq)							0.20
							2.79
log(Size)			-1.44	-1.44	-1.43	-1.43	-1.46
			-28.33	-28.48	-27.47	-27.29	-27.28
Flow_{t-1} (%, pq)			0.09	0.09	0.09	0.09	0.09
			5.11	5.08	5.10	4.87	4.88
Management fee (%)			-0.41	-0.38	-0.40	-0.40	-0.38
			-8.21	-8.08	-8.29	-8.47	-8.82
Incentive fee (%)			-0.03	-0.03	-0.03	-0.03	-0.03
			-3.42	-3.68	-3.58	-3.61	-3.33
Highly liquid			0.97	0.96	0.84	0.86	0.84
			2.32	2.33	2.03	2.11	2.03
Highly illiquid			-0.15	-0.14	0.01	0.01	0.01
			-1.25	-1.10	0.06	0.07	0.05
Strategy Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Leverage Controls	No	No	Yes	Yes	Yes	Yes	Yes
Offshore Indicator	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	207434	207434	162891	162891	162891	162891	162891
Number of Cross-sections	27	27	26	26	26	26	26

Table 9 shows regressions of quarterly investor flows on various controls. List is an indicator equal to one if the fund lists with a hedge fund data vendor; Performance rank is the percentile value of fund net-of-fee returns in the previous quarter within the listed/non-listed set and the fund's strategy category, Net excess return is the net-of-fee quarterly return minus the risk-free rate; log(Size) is the log value of net assets, Management fee is the fee rate charged on assets under management, Incentive fee is the variable fee charged as a proportion of returns, Highly liquid is an indicator variable for funds where all investors can redeem within seven days, and Highly illiquid is an indicator for funds where the longest amount of time needed to redeem capital exceeds 365 days. Strategy controls indicate whether broad strategy indicator variables are included in the regression. t-statistics are reported below coefficients with Newey-West standard errors.

7.2 Figures

Figure 1: Counts of Vendor-Listed and Non-Listed Hedge Funds

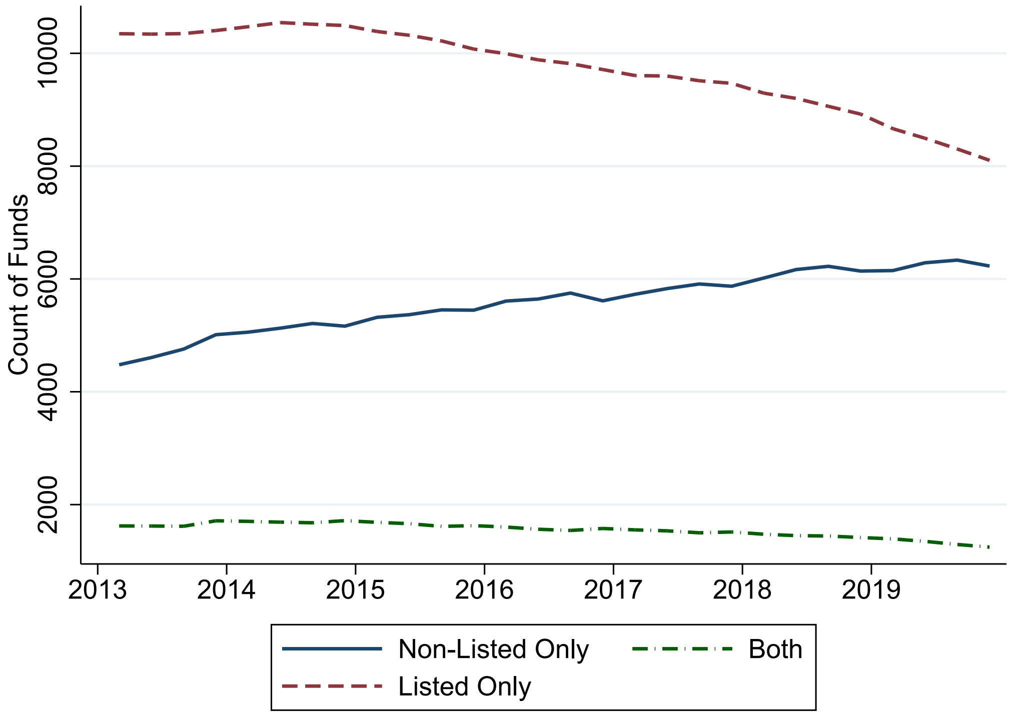
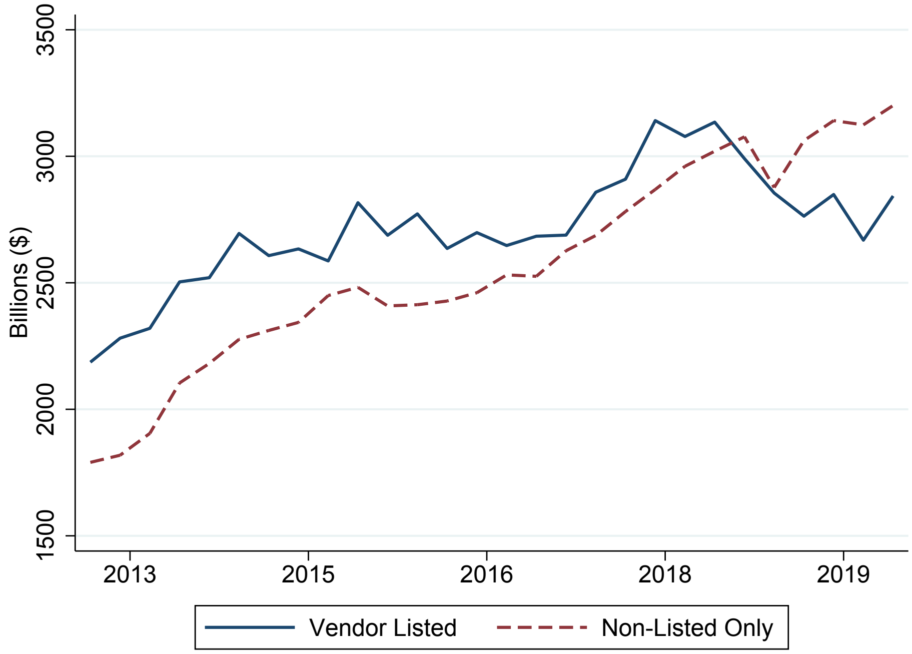


Figure 1 presents the number of funds that fall within each of three categories: those that report only to a vendor database, those that report only to Form PF, and those that report to both Form PF and at least one vendor database.

Figure 2: Hedge Fund Industry Net Assets



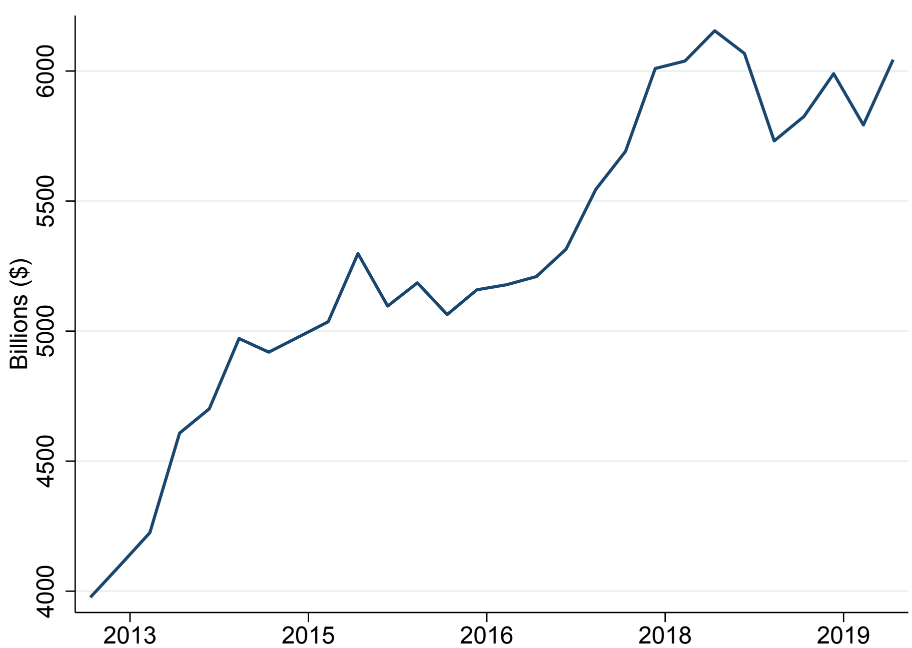


Figure 2 plots net assets under management for vendor-listed and non-listed only funds, separately, as well as their total. Values are reported as of the quarter-end of each year, beginning with March 2013 and ending December 2019. Form PF annual filers have net AUM values filled forward until the next observation, up to 11 months ahead.

Figure 3: Net and Gross Assets for Listed-Only Funds

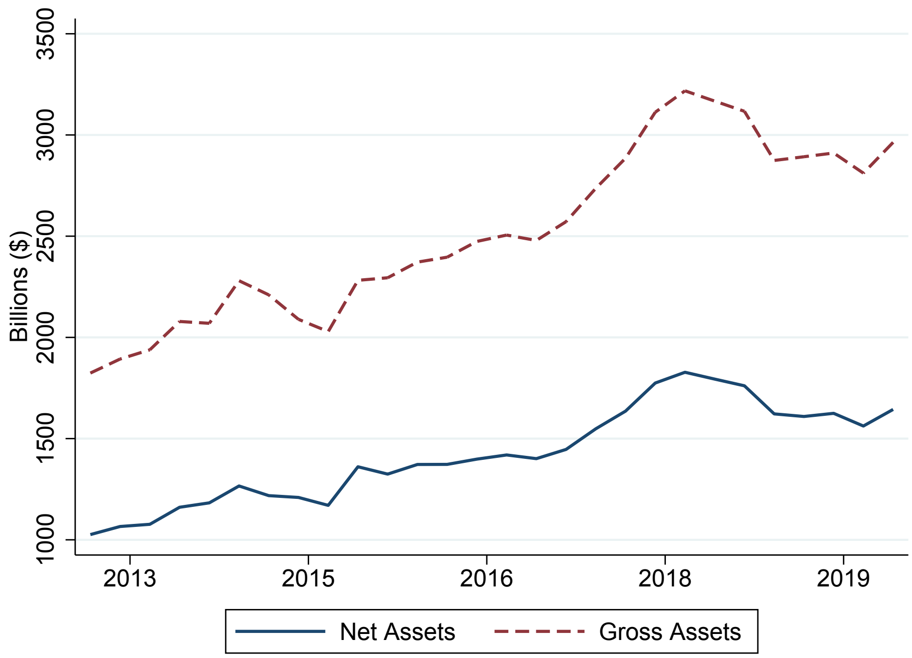


Figure 3 plots net and gross assets under management for funds that report to at least one vendor database but do not register with SEC as an investment adviser. Such funds do not report on Form PF or Form ADV. Values are reported as of the quarter-end of each year, beginning with March 2013 and ending December 2019.

Figure 4: Hedge Fund Industry Gross Assets

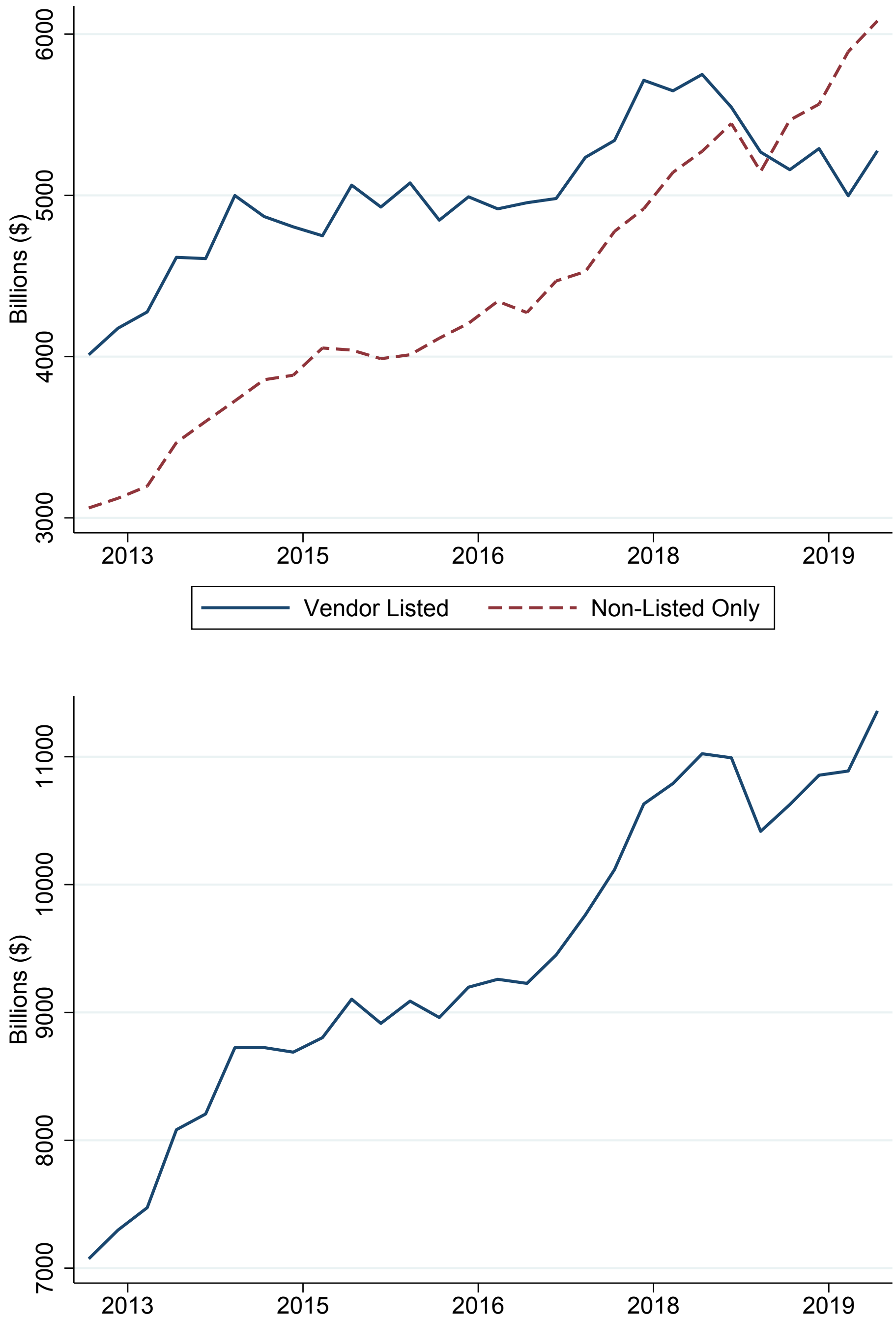
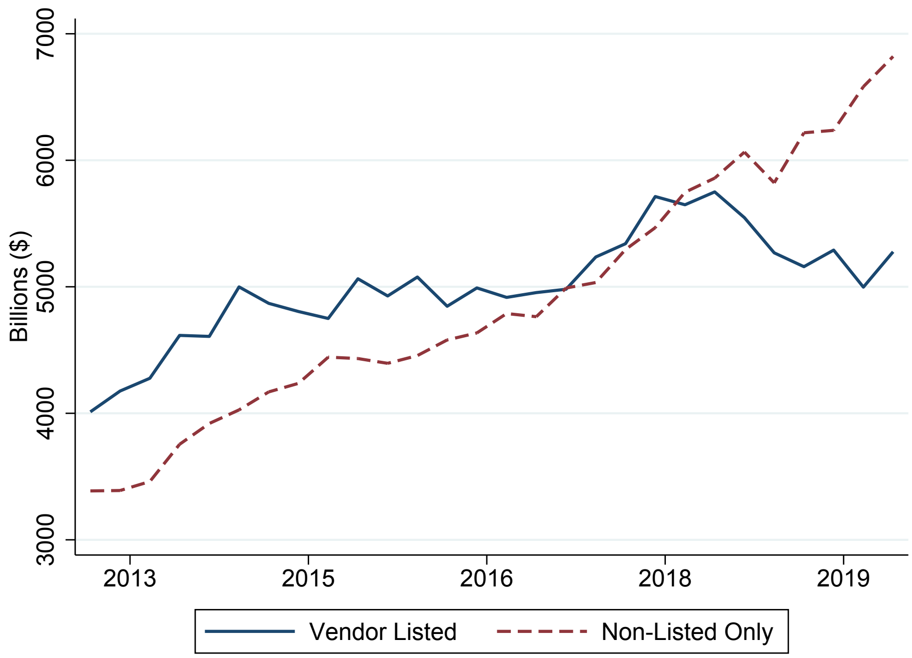


Figure 4 plots gross assets under management for vendor-listed and non-listed only funds, separately, as well as their total. Values are reported as of the quarter-end of each year, beginning with March 2013 and ending December 2019. Form PF annual filers have gross AUM values filled forward until the next observation, up to 11 months ahead.

Figure 5: Hedge Fund Industry Gross Assets (Form PF + Form ADV)



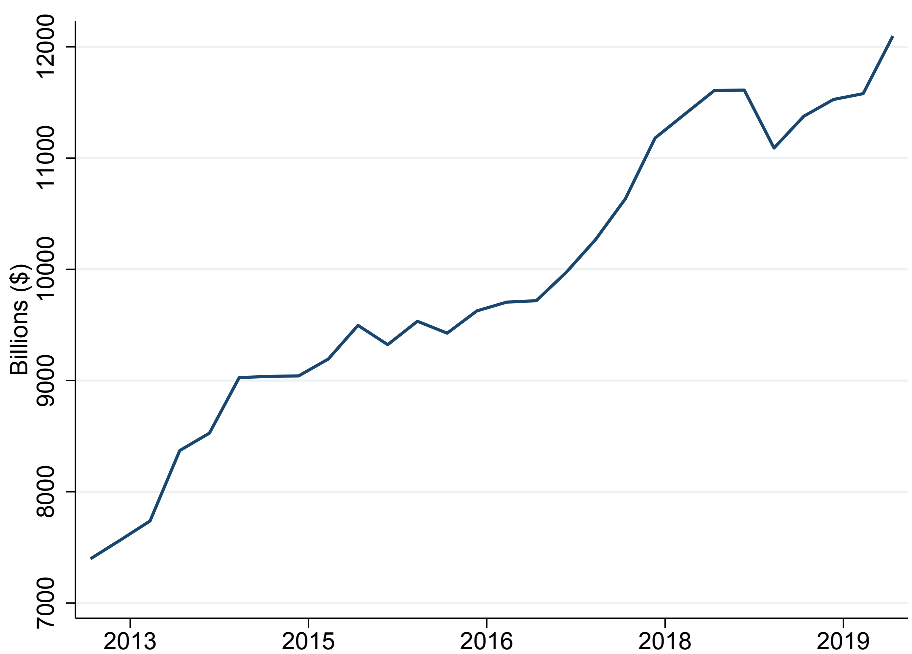
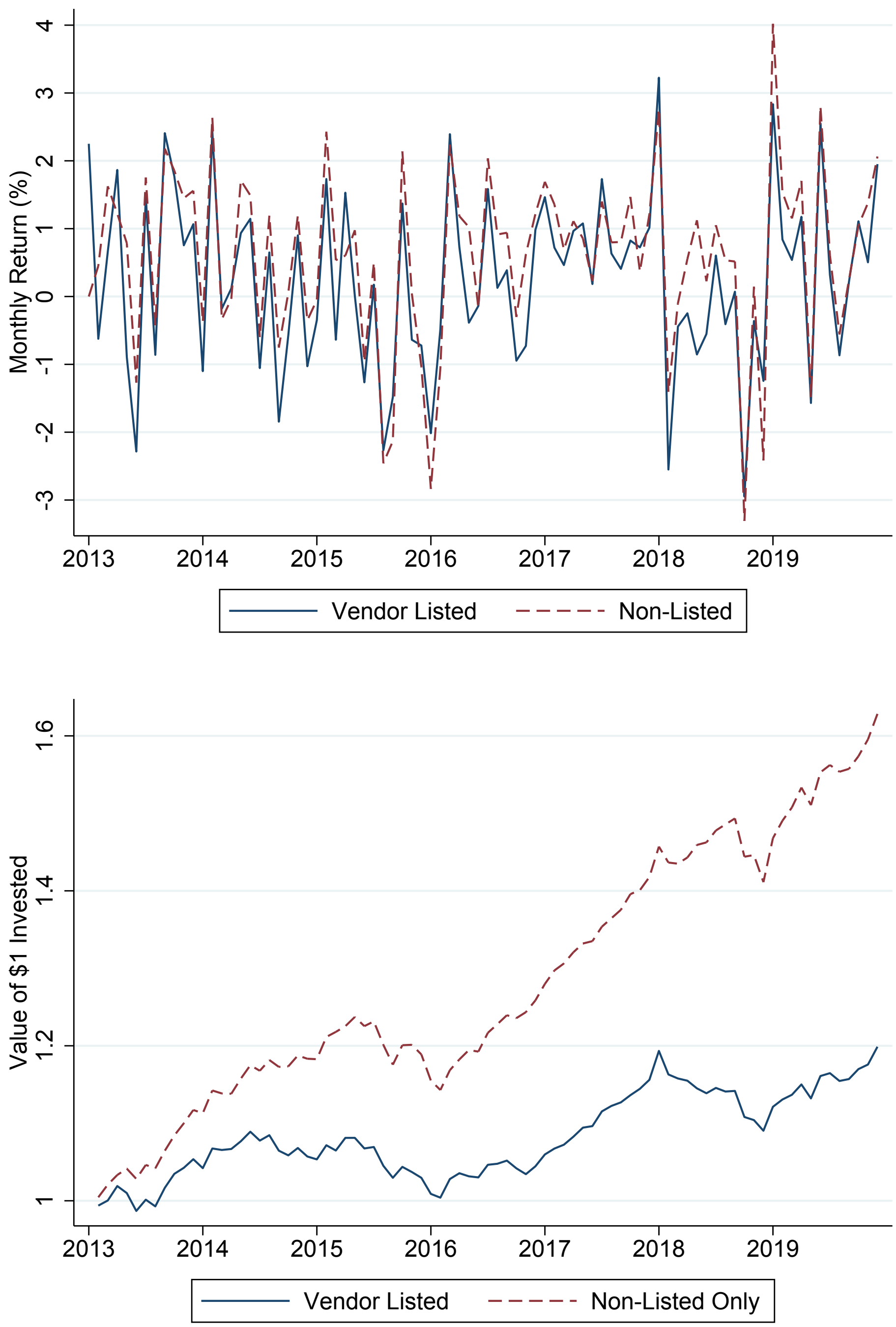


Figure 5 plots gross assets under management for vendor-listed and non-listed funds, separately, as well as their total. In this figure, non-listed gross assets include gross assets as reported on Form PF, or as reported on Form ADV for funds that do not report gross assets to Form PF (or to any vendor database). Values are reported as of the quarter-end of each year, beginning with March 2013 and ending December 2019. Form PF annual filers have gross AUM values filled forward until the next observation, up to 11 months ahead.

Figure 6: Cumulative Returns



The top panel of Figure 6 plots the monthly, weighted-average rate of return, net of fees, for funds that report to at least one vendor database and those that report only on Form PF. Returns are value-weighted by funds' net assets. The bottom panel plots the cumulative, total return over the full sample period, based on the rate of return values reported in the top-panel.

Figure 7: Cumulative Returns by Strategy

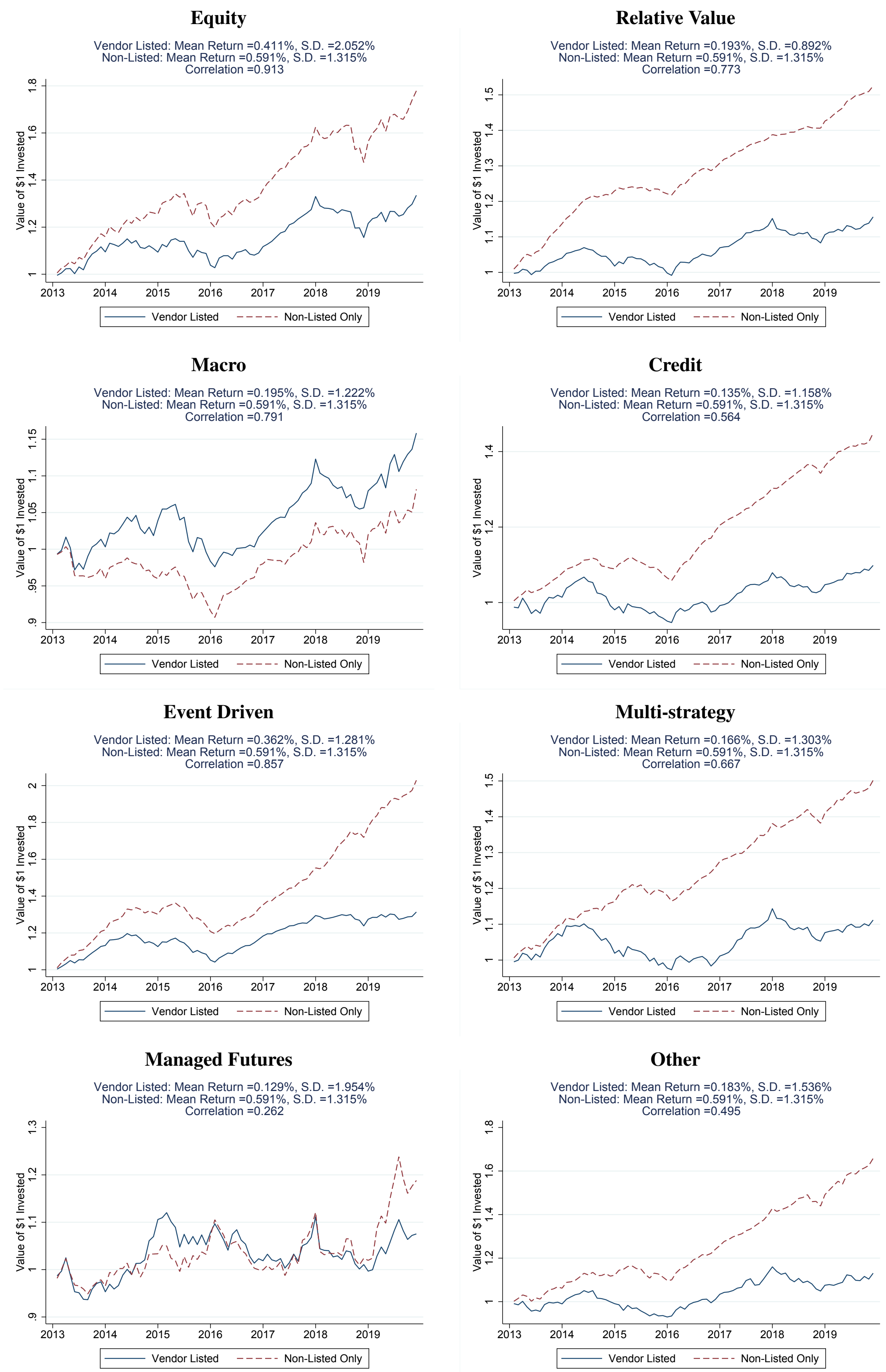
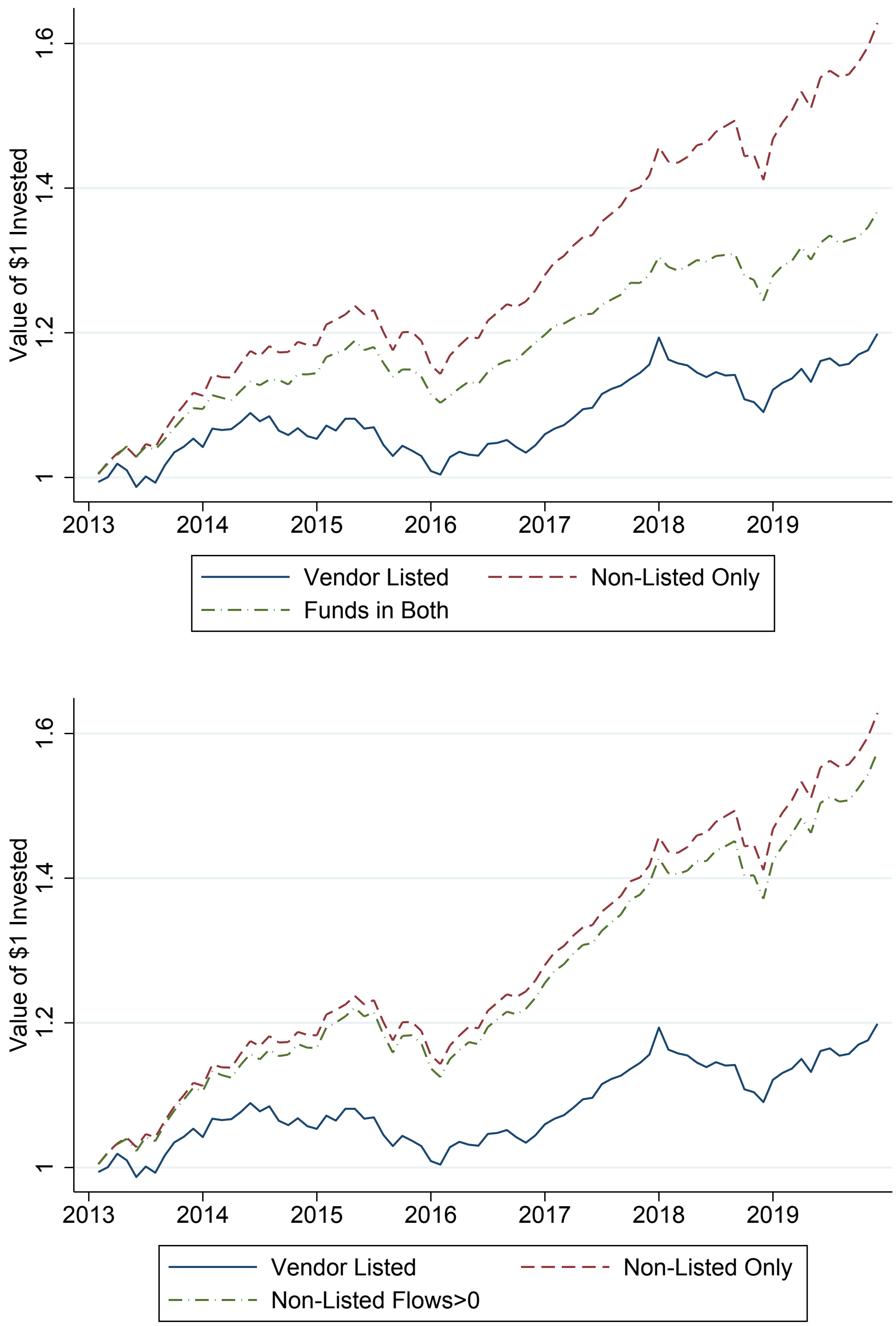


Figure 7 plots the monthly, weighted-average, cumulative total return over the full sample period for listed and non-listed funds, separately for each strategy category. Rates of return are weighted by funds' net assets within listed/non-listed, strategy, month and year.

Figure 8: Cumulative Returns: Alternative Explanations



The top panel of Figure 8 plots the cumulative, total return, net of fees, for funds that list with at least one vendor database, those that report only on Form PF, and those that list with a vendor and report on Form PF. Returns are value-weighted by funds' net assets. The bottom panel plots the same return series for listed (only) funds, non-listed funds, and non-listed funds whose cumulative investor flows as of that month and year were positive. Note that the vendor-listed series in both panels includes funds that report to Form PF as well; that is, the "Funds in Both" series in the top panel is a strict subset of the "Vendor Listed" series. We do this to maintain comparability with the bottom panel of Figure 6.

Figure 9: Betas for Vendor-Listed and Non-Listed Funds

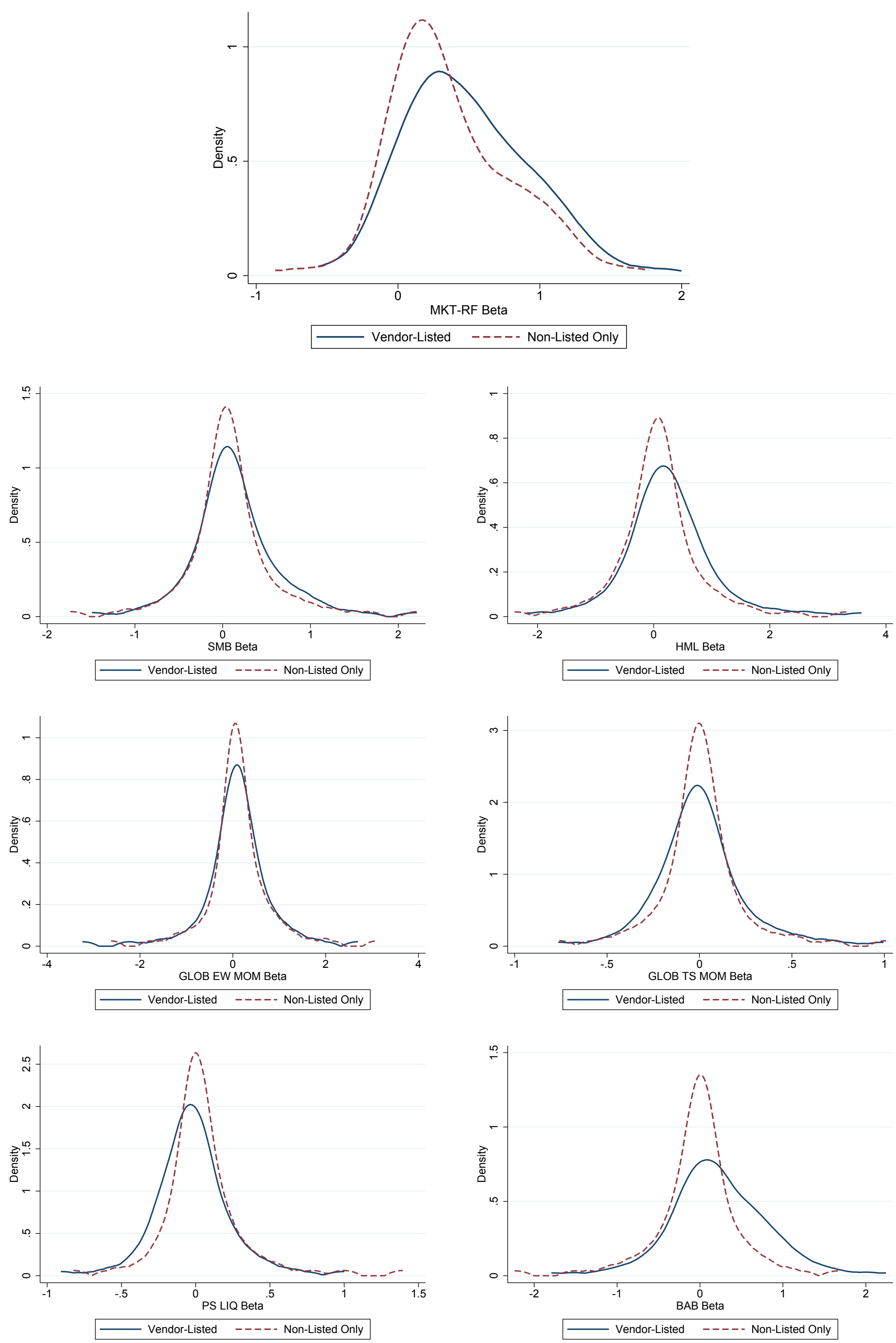


Figure 9 plots the empirical PDF of factor betas estimated from equation (1) and based on the risk factors included in the Global 7 factor model. For non-listed funds, the PDFs are constructed as smoothed density plots through the values between (and including) the 1st and 99th percentiles.

Figure 10: Jensen's Alphas

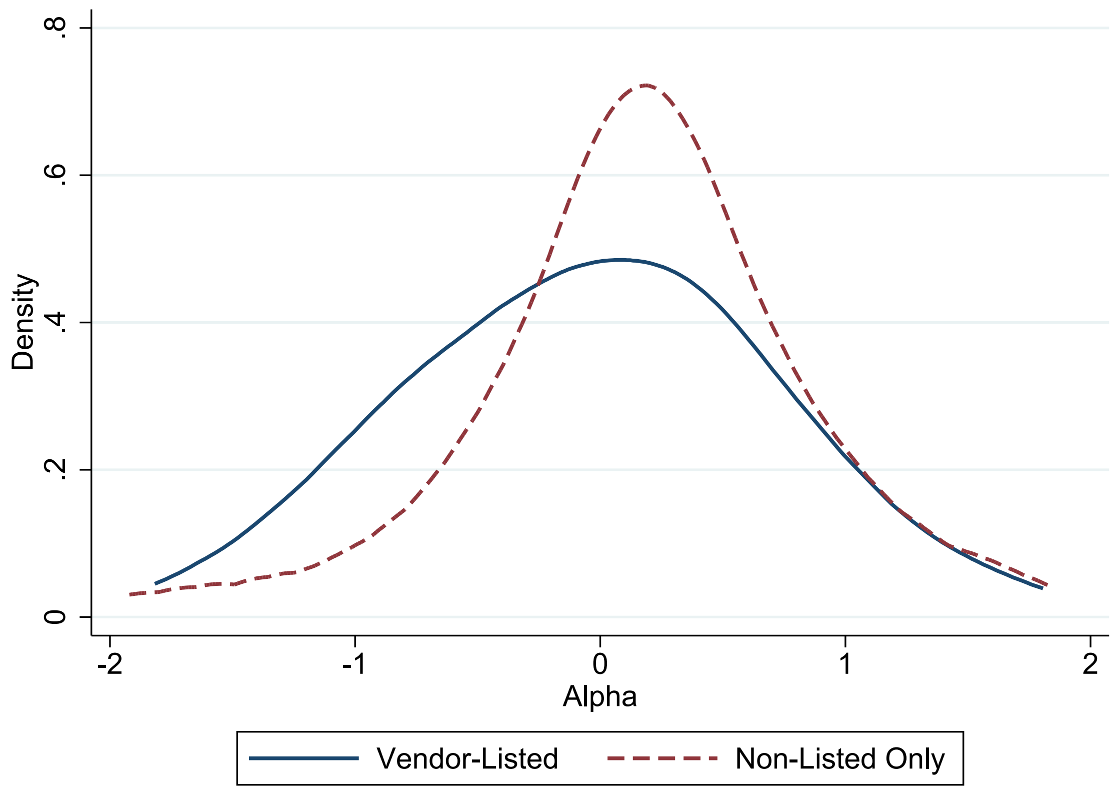


Figure 10 plots the empirical PDF of Jensen's alpha, the fund-level intercept estimated from equation (1), for listed and non-listed funds separately. the PDFs are constructed as smoothed density plots through the values between (and including) the 1^{st} and 99^{th} percentiles.

Figure 11: Berk and van Binsbergen (2015) Value Added

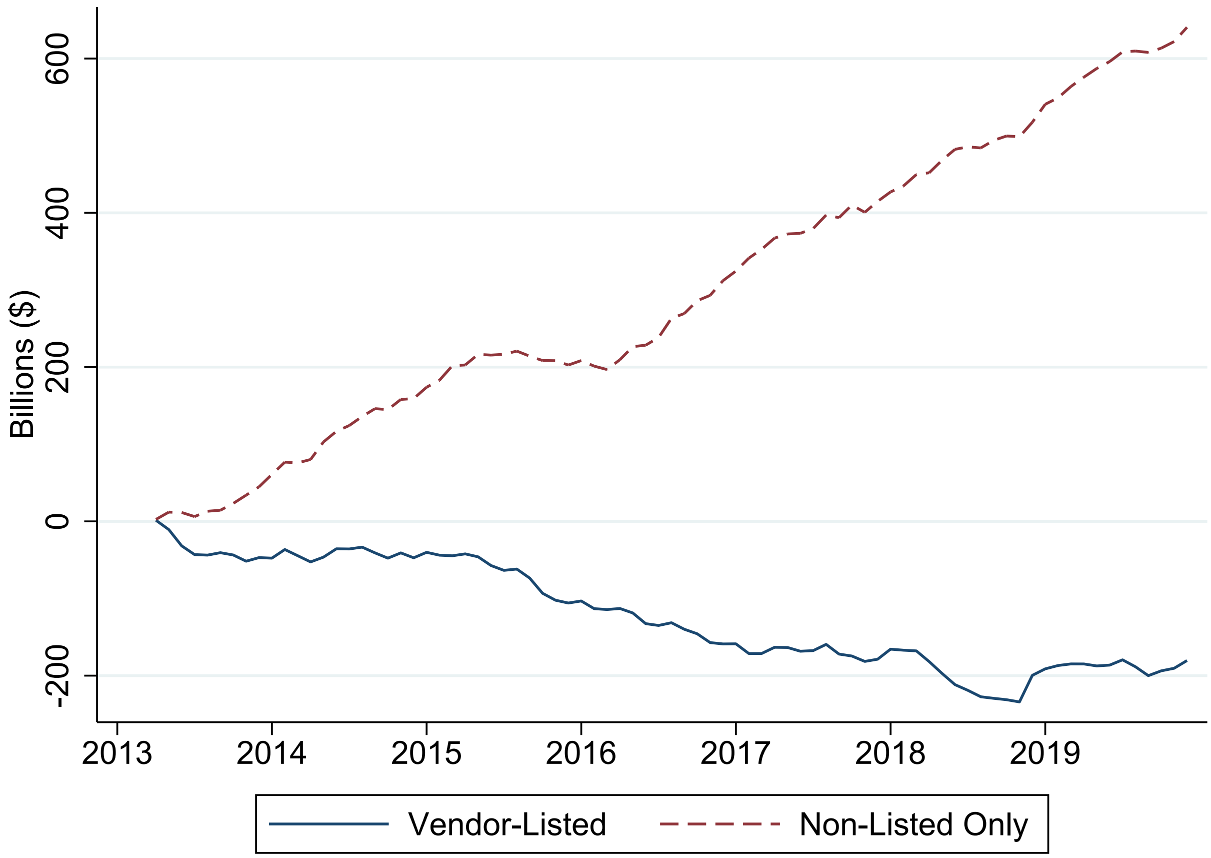


Figure 11 plots the cumulative value-added by vendor-listed and non-listed funds, separately, based on the methodology of Berk and van Binsbergen (2015).

Figure 12: PDFs of Estimated Alphas

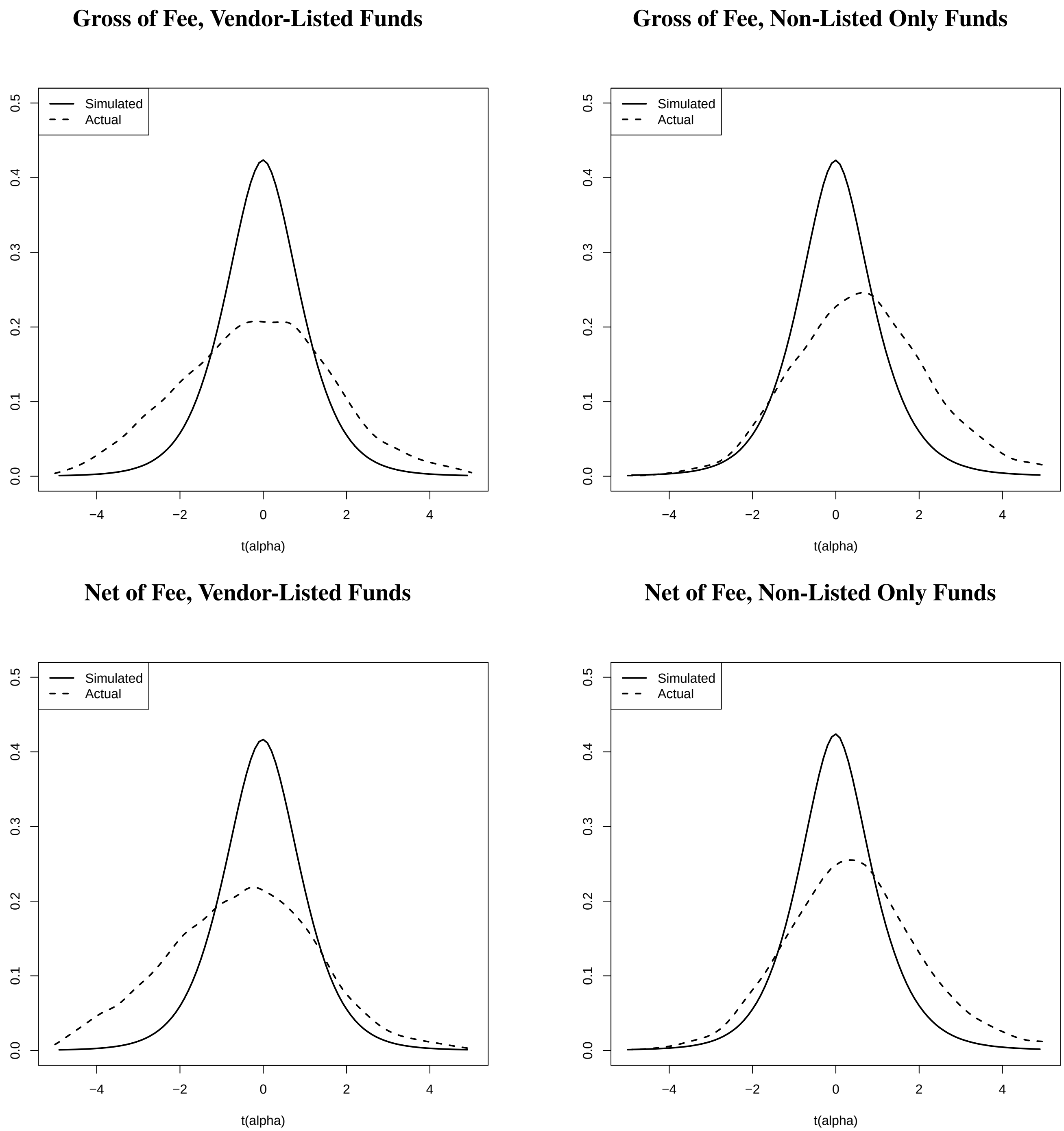


Figure 12 plots the empirical PDFs of alpha t-statistics from the actual data against the PDFs implied by the average values of alpha t-statistics at each quantile over the 10,000 bootstrapped samples. PDFs are calculated using kernel density estimation. For data using Form PF, the kernel density estimation is based on the 99 integer-valued percentiles of the distribution of the alphas. All plots include data from January 2013 – December 2019.

Figure 13: Aggregate Flows

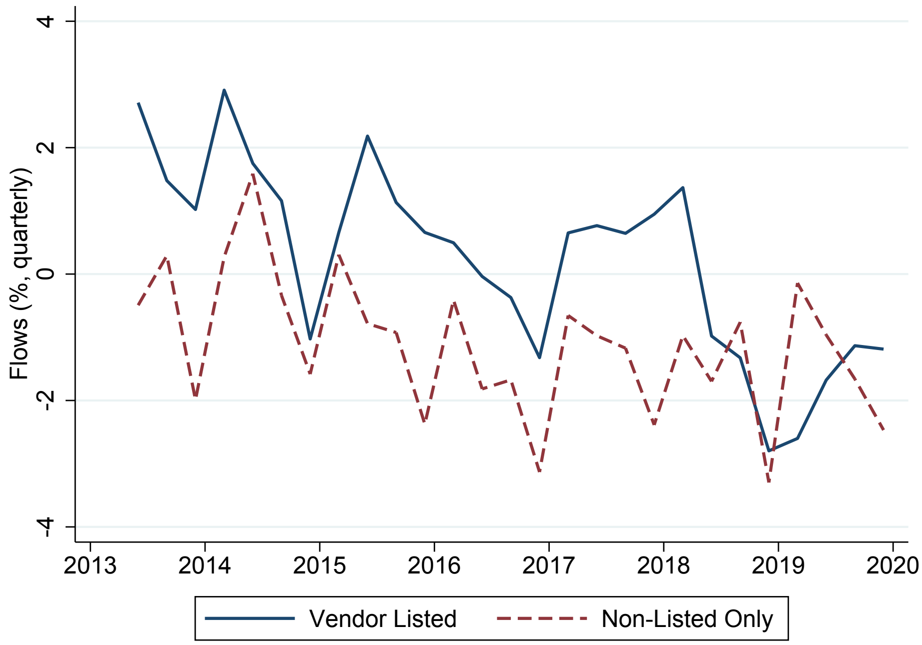


Figure 13 plots the quarterly, weighted-average (net) investor flows for funds that report to at least one vendor database and those that report only on Form PF. Net flows are value-weighted by funds' net assets. The bottom panel plots the cumulative, total flow over the full sample period based on the net flows reported in the top-panel.

Figure 14: Aggregate Flows by Strategy

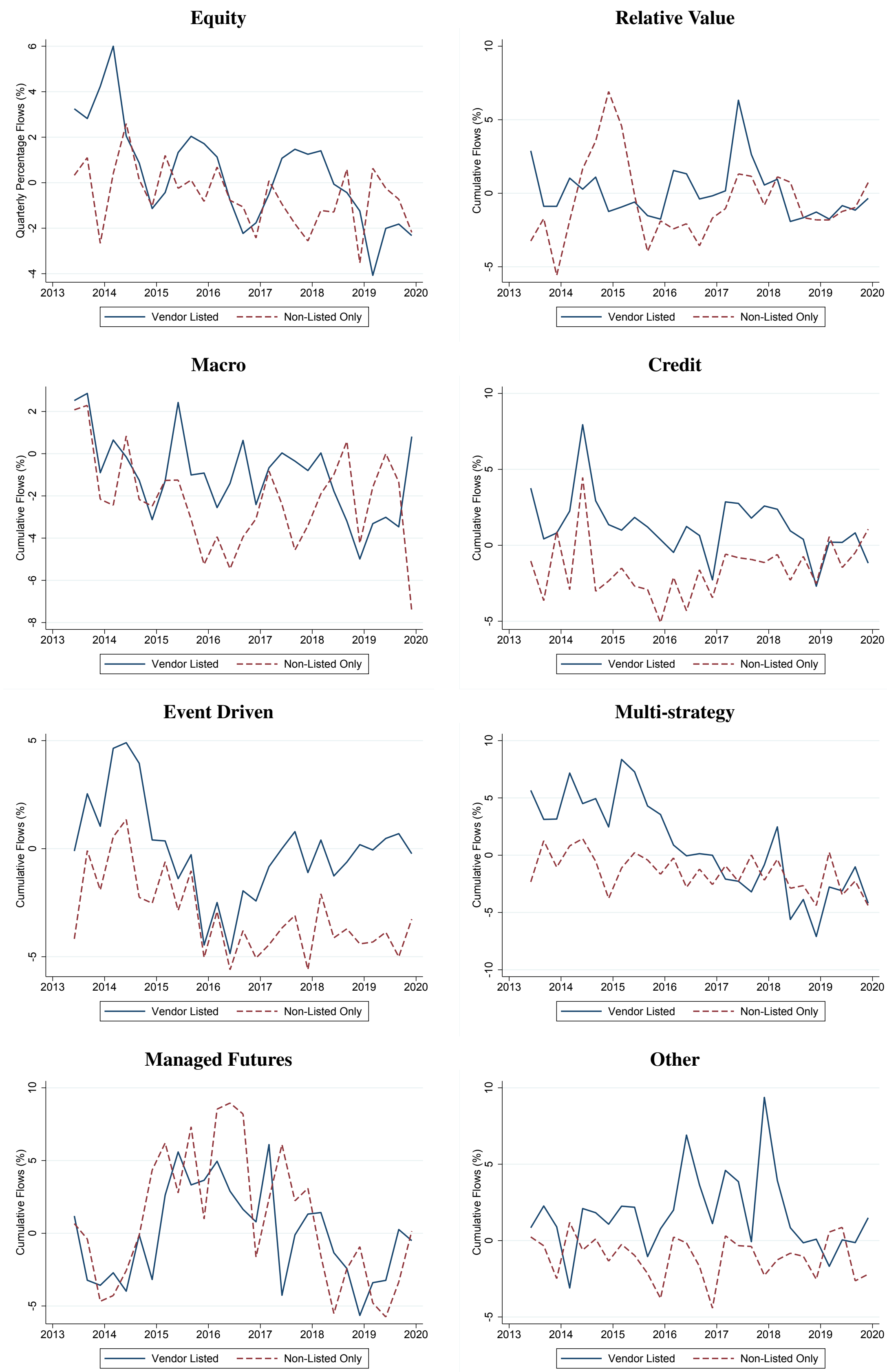


Figure 14 plots the quarterly, weighted-average, cumulative total (net) investor flows over the full sample period for listed and non-listed funds, separately for each strategy category. Net flows are weighted by funds' net assets within listed/non-listed, strategy, quarter and year.