Hedge Funds: Fees, Return Revisions, and Asset Disclosure

Michael P. Streatfield

Jesus College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2012

This thesis is a collection of three essays on hedge funds with contributions to the empirical understanding of their fees, and their voluntary disclosure of returns and assets under management, using a large consolidation of widely-employed publicly available hedge fund databases.

First, time-series variation in reported fees is analysed using fund launches within hedge fund management companies, and conditioning fees at launch on fund family characteristics. Larger and better performing fund families launch high fee funds. Funds with high management fees at launch do not perform any differently from low fee funds, though funds with high incentive fees marginally outperform. An interval regression technique is proposed to overcome the discrete nature of reported fees.

Secondly, the reliability of voluntary disclosures of financial information is analysed with a different measure of time-variation — tracking changes to statements of historical performance recorded at different points in time. This uncovers evidence that historical returns are routinely revised. These revisions are not merely random or corrections of earlier mistakes; they are partly forecastable by fund characteristics. Moreover, funds that revise their performance histories, significantly and predictably underperform those that have never revised.

Finally, the availability, and timing, of the selective disclosure of assets under management by funds is examined. More than a third of funds have asset records falling short of returns published. There is evidence of strategic disclosure by funds — asset reporting drying up after times of fund stress, such as poor performance or outflows. Furthermore, investors should take heed of the greater propensity for shortfall funds to trigger fraud performance flags.

These results suggest that unreliable disclosures: constitute a valuable source of information for current and potential investors; have implications for researchers; and, exhort market regulators to include assets, not just returns, in the debate around mandatory disclosure by financial institutions.

Hedge Funds: Fees, Return Revisions, and Asset Disclosure

Michael P. Streatfield
Jesus College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2012

To my wife, Nikki, and, The Mexican and Didi for their unending support and love

Acknowledgements

My deepest gratitude goes to all who have supported my doctoral research. I appreciate those who have seen less of me as I have delved into my research. Inordinate thanks goes to Nikki. To my family, your patience and quiet support speaks volumes. To my friends from afar that have not forgotten me — Phil and James — I look forward to catching up sometime soon over a game of cards. To my mentors, Leon Camphor and Raymond Goldblatt, you inspired my curiosity for research.

I would like to thank Professor Tarun Ramadorai for his encouragement and guidance as my supervisor. His econometric mind, academic insights, and approach to writing were invaluable contributions to shaping my doctoral research experience. I will never forget these moments. Thank you too for the rewarding teaching opportunities. Thanks also to Andrew Patton for his teaching, and the pleasure of working with him as a co-author of the third chapter of this thesis.

I have an intellectual debt to the Faculty of Saïd Business School. Professors Alan Morrison, Jose Martinez, Tom Noe, Han Ozsoylev, Ludovic Phillapou, Tim Jenkinson, and Mungo Wilson, and on the econometrics side, Kevin Sheppard, I appreciate your brilliant teaching and input into my research. I am truly grateful to Dimitrios Tsomocos for his guidance, and championing me in my MFE year. I would also like to remember Tom Karkinsky, who is no longer with us.

I would like to especially thank the Oxford-Man Institute of Quantitative Finance for not only providing the hedge fund data and financial support, but also furnishing a wonderful working environment for these ideas to flourish.

My graduate studies have been enriched by the interactions with my fellow students. I learnt a great deal working alongside the duo of Alex Vardoularkis and Udara Peiris, the grandfathers of the doctoral program. Special thanks to Ken Okamura for your wisdom and rigour not only on finance, but also with that fine alternative asset, wine. Thank you too, Ken, for a very helpful review of my fourth chapter. To Mario, thank you for being such a good friend. Thanks also to Andrea, Bastiaan, Cavit, George, Lizzy, Pancho, Richard, and Sumudu. To Li Lin zǎo.

My time at Oxford has been home to marvellous memories at Jesus College. To my best friend, Stephanie Dobrowolski, all I can say is TIO. I will treasure stimulating conversations and relaxing times with Emily, Jon, Martin, Matt, Rex, Sarah, Garth, and Yannick on the water.

Annabel, a huge thank you for being an excellent proof reading guru.

This doctoral journey that led me to Oxford was inspired by fine friends and colleagues who have walked this path in their own fields — Jessica Tiffin, Cara Winterbottom, Richard Saldanha, Terence Moll, and in particular, the irrepressible Michael Power.

I am deeply grateful to Investec Asset Management for their valuable financial support. Hendrik du Toit, your love for Oxbridge made this possible.

Abstract

This thesis is a collection of three essays on hedge funds with contributions to the empirical understanding of their fees, and their voluntary disclosure of returns and assets under management, using a large consolidation of widely-employed publicly available hedge fund databases.

First, time-series variation in reported fees is analysed using fund launches within hedge fund management companies, and conditioning fees at launch on fund family characteristics. Larger and better performing fund families launch high fee funds. Funds with high management fees at launch do not perform any differently from low fee funds, though funds with high incentive fees marginally outperform. An interval regression technique is proposed to overcome the discrete nature of reported fees.

Secondly, the reliability of voluntary disclosures of financial information is analysed with a different measure of time-variation — tracking changes to statements of historical performance recorded at different points in time. This uncovers evidence that historical returns are routinely revised. These revisions are not merely random or corrections of earlier mistakes; they are partly forecastable by fund characteristics. Moreover, funds that revise their performance histories, significantly and predictably underperform those that have never revised.

Finally, the availability, and timing, of the selective disclosure of assets under management by funds is examined. More than a third of funds have asset records falling short of returns published. There is evidence of strategic disclosure by funds — asset reporting drying up after times of fund stress, such as poor performance or outflows. Furthermore, investors should take heed of the greater propensity for shortfall funds to trigger fraud performance flags.

These results suggest that unreliable disclosures: constitute a valuable source of information for current and potential investors; have implications for researchers; and, exhort market regulators to include assets, not just returns, in the debate around mandatory disclosure by financial institutions.

Contents

List of Figures

List of Tables

"And I will show you something different from either
Your shadow at morning striding behind you
Or your shadow at evening rising to meet you;
I will show you fear in a handful of dust."
— T.S. Eliot *The Wasteland*

1 Introduction

1.1 Hedge funds as an asset class

Hedge funds are well-known for their secrecy — creating mystique and safeguarding their proprietary investment processes. However, hedge funds regularly reveal their fortunes through publication of their performance to investment databases. This thesis explores information disclosed by hedge funds to unpack time variation in fees, and to challenge the reliability of these voluntary submissions.

Hedge funds are lightly regulated investment vehicles, with a stark contrast of immense investment freedom, alongside limited access to the public.[1] Hedge funds provide a fascinating backdrop for study. This is because the nature of the investment vehicle combines diverse opportunities for investment managers to exercise talent, with performance fee structures strongly incentivising the hedge fund manager, and yet, investors in hedge funds face asymmetric information as reporting on positions and strategies is limited.

[1] Stulz (2007) provides a fine introduction to those less familiar with this asset class.

This thesis unearths a number of findings from hedge fund public information. Firstly, using fund launches by hedge fund management companies as a unit of analysis — it reveals the time variation in fees and the consequences of rising management fees. Secondly, using repeated snapshots of the hedge fund databases — it uncovers extensive revisions of prior hedge funds returns. Thirdly, comparing return and asset information production — it exposes the extent and risk of non-disclosure of asset data.

As a point of clarification, I am using the term 'hedge fund' loosely in this introduction to encompass a range of hedge fund types. A distinct type of hedge fund investment is a fund-of-funds, a fund that invests in hedge funds. Another category of hedge fund investment is a Commodity Trading Adviser (CTA), which I classify under the Managed Futures hedge fund strategy.[2] The scope of analysis is this wider group of funds-of-funds, CTAs, and hedge funds.

The next two sections of this introduction will explain the rationale and research questions that led to these findings, and outline the thesis contribution. I will then close with a briefing on how the thesis is structured.

1.2 Hedge fund fees and incentives

Hedge fund remuneration structures are distinguished by fund revenue being highly geared to performance through an incentive fee. Hedge fund incentive fees are a variable component that is charged as a fixed percentage share of the outperformance of a fund over a threshold return or hurdle rate (this hurdle rate can be zero). These variable incentive fees, generally collected on an annual basis, are in addition to a typically fixed annual management fee charged by the fund.

[2]CTAs are trend following funds, which trade in futures contracts, and are registered with the Commodity Futures Trading Commission (for more detail, see Fung and Hsieh, 2001).

Hedge fund incentive fees are uni-directional, in that they are charged only on outperformance. There is no explicit penalty on underperformance. Hedge fund managers thus only share in the performance upside. This is not the practice in U.S. mutual funds where performance-related fees have to be symmetric (Elton, Gruber, and Blake, 2003).

Investors in hedge funds use other mechanisms to balance the playing field and control downside performance exposure. Hurdles set above zero ensure investors earn a minimum return before incentive fees are payable. More effective are high-water marks (HWM), which condition the payment of future performance fees to be payable only when the fund exceeds the maximum share value of the investor, i.e. be paid only on 'new' outperformance. Goetzmann, Ingersoll, and Ross (2003) assess and find value in high-water marks for both investors and managers. However, there is no escaping the importance of fee levels, as high-water marks have limitations. For example, they cannot prevent underperforming funds, now deep 'underwater,' from simply closing, returning capital to investors, and then reopening.

A stylised fact exists of hedge fund fees being "2 and 20", namely, 2% p.a. fixed management fee and a performance-related incentive fee of 20% of the fund's outperformance of a hurdle. The first hedge fund, by Alfred W. Jones in 1976, charged a 20% incentive fee and there is substantial anchoring by hedge funds to this level. However, this assertion for management fees is challenged in Chapter 2, where the mean fee reported in a large consolidated hedge fund data-set is calculated to be 1.511% p.a. (median 1.5%).

Fees matter as they directly impact and lower the return that investors receive. In Chapter 2, three research questions are posed regarding hedge fund fees. Firstly, how do hedge fund fees vary across time and in the cross-section of funds? Although hedge fund databases seem static when considering the most recent version of the data presented, there is one source of time-variation in hedge fund fees, namely the fees at fund launches. Using a large consolidation of hedge fund databases, over the period 1995 to 2009, the fees for newly launched funds are examined.

Secondly, what are the drivers of fund fees? Using the panel of fund launches, this is assessed by conditioning management and incentive fees on the administrative characteristics, past performance, size, and location of the management fees that launched them. Although the use of management companies is prevalent in the mutual fund literature, this approach to analysing hedge funds is nascent.

Theoretical models of hedge fund fees expect managers to focus on maximising fee income whilst avoiding liquidation risk (Goetzmann et al., 2003; Lan, Wang, and Yang, 2012). Better performing hedge fund managers, are expected to attract flows. This lowers liquidation risk, allowing for greater leverage, and potentially better future incentive fees. Lan et al.'s model would anticipate a trade-off in management fees for higher incentive fees for better performers, however this neglects realities such as capacity constraints, industry structure, and the managers' future expectations of their strategy and the industry.

And thirdly, given the findings regarding the characteristics of the funds that launch higher-fee funds, what is the impact on investors? Do higher fees motivate higher performance, leaving investors better off on a net-of-fee basis? This is tested by conditioning subsequent post-launch performance on the fee levels and management company characteristics, like size and past performance.[3]

This thesis contributes to the empirical understanding of hedge fund fees through the following findings. New insights into the causes and consequences of time-variation in hedge fund fees are presented by analysing fees set, and reported, by management companies launching new funds. Tracking these launchers over the period 1995 to 2009 reveals reported management fees have broadly been rising since 2000, while incentive fees have remained constant, or have declined slightly. Using a panel of launching funds, management companies that are larger, and those with better recent past performance, launch funds with higher management fees. High recent past performing

[3] This approach assumes the investment approach is shaped by the existing management company. Another potential line of enquiry could be to look within the management companies to the identities of the managers themselves. In the case where the investment talent switches to another company for a fund launch, these individuals could be tracked across companies. This data is not comprehensibly available for such cases, and so the past characteristics of the launching management company is used instead.

management companies, and those with high-water marks in the family, also launch with higher incentive fees. These results are robust to the use of interval regression techniques (see Stewart, 1983), to get closer to the underlying continuous distribution of fees. This conditioning of fees to be bounded within an interval, provided the actual fee is close to the fee reported, ensures that despite the discrete nature of reported fees, and the potential of rebating to some investors, the drivers of the variation in fees can be confirmed. Given that better performing management companies are launching higher fee funds, the post-launch performance of these funds is tested. Net-of-fee performance for high and low management fee funds is indistinguishable, so it seems to investors that these higher management fees at launch are 'money for nothing.' There is some evidence that higher incentive fees at launch signal higher future raw returns. These results have implications for regulators and prospective investors as to the influence and role of management companies in the provision of value for money in the hedge fund industry. The next focus of this thesis regards return and asset disclosure.

1.3 Voluntary disclosure

1.3.1 Hedge fund databases

To gain an appreciation of how returns are disclosed by hedge funds, I first need to position the availability of hedge fund data relative to more well-known financial assets, such as mutual funds. For mutual funds, the provision of information is tightly legislated and readily available. Listings of mutual funds are available in the press and standard information is encouraged to be available online.[4] Hedge fund data differs in that submission is voluntary.[5] Hedge fund managers can choose when, where, and

[4]See, for example, Securities Exchange Commission (SEC) press release 2008-275, available at http://www.sec.gov/news/press.shtml.

[5] The focus of this thesis is on *voluntary* disclosure to public hedge fund databases. Note, there are some mandatory reporting requirements to the Securities Exchange Commission (SEC), and the Commodity Futures Trading Commission (CFTC), affecting some U.S. hedge funds. For example, the SEC requires larger funds, with publicly traded assets over U.S.$100 m, to make Form 13-F filings of their long asset positions (For details, see Agarwal, Jiang, Tang, and Yang, 2011b). The SEC and CFTC are jointly seeking greater confidential reporting requirements (see SEC press releases 2011-23 and 2011-226, available at http://www.sec.gov/news/press.shtml).

what to report. Access to this data is via costly subscription to hedge fund database providers.

The consequence of self-reporting is that hedge funds can report a fund to more than one database. This leaves hedge fund data highly fragmented with no single database capturing a lion share of the industry. This is demonstrated in Figure 1.1, sourced from funds in the consolidated hedge fund database. This shows the percentage of funds covered by a database individually and shared across combinations of databases. Looking at the edges of the Venn diagram shows the funds uniquely captured by a single database. This ranges from 12.6% in Tremont Advisory Shareholder Services (TASS) to 8.3% in Morningstar. Although there is substantial overlap between databases, relying on a single database yields a narrow view of the industry.[6]

The best attempt to surmount this fragmentation is to cast the net wide across multiple databases to capture as many hedge funds as possible.[7] This approach inevitably leads to duplication of funds owing to their entry into multiple databases, and, the existence of multiple share classes and currency classes. This repetition needs to be corrected and the best representative fund chosen. Such a process is followed across five databases. The end result is a large, comprehensive set of more than 18,000 hedge funds.[8]

Hedge fund researchers still face other known difficulties. Some of these issues are listed. There are known data biases like survivorship bias (where defunct funds drop out of databases), and backfill bias (historic data added retrospectively when a fund first enters a database) (see Fung and Hsieh, 2009, for a recent update). Self-reporting leads to positive selection biases (Agarwal, Fos, and Jiang, 2012; Aiken, Clifford, and Ellis, 2012). There are quality differences between databases (for example, see survivorship differentials by Joenväärä, Kosowski, and Tolonen, 2012). Hedge fund reporting can be strategically delayed (Aragon and Nanda, 2011). There is statistical evidence casting doubts over what is published, such as smoothing of returns

[6] For example, the widely-used TASS neglects more than half (56.9%) of the industry's funds.

[7] Naturally, this still excludes hedge fund managers who chose not to report to any database.

[8] Appendix A documents the consolidation procedure used to gather, de-duplicate and select this set of hedge fund data across the TASS, HFR, CISDM, BarclayHedge, and Morningstar databases.

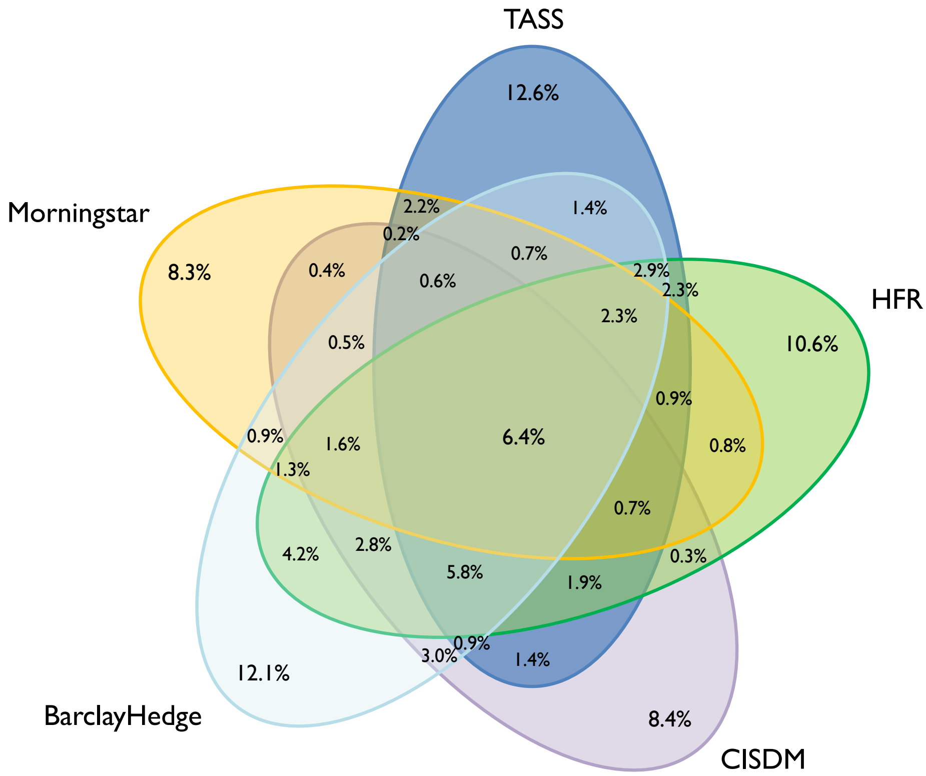


Figure 1.1: *Illustration of the Fragmentation of Hedge Fund Data*
Consolidation of merger of five hedge fund databases from BarclayHedge, the Center for International Securities and Derivative Markets (CISDM), Hedge Fund Research (HFR), Morningstar, and Tremont Advisory Shareholder Services (TASS) (now Lipper). The percentage of funds covered by all potential combinations of database entry is shown.

(Getmansky, Lo, and Makarov, 2004), or intra-year increases in December (Agarwal, Daniel, and Naik, 2011a).

The public hedge fund databases provide a valuable outlet for hedge funds to strategically advertise their products, particularly given the past SEC controls on soliciting the public (Jorion and Schwarz, 2010). Discrepancies in reporting in, and between, these databases do occur, and the penalties do not seem material (Fung and Hsieh 2009; Liang 2000 discuss persisting data discrepancies like differing inception dates, and fund open statuses between databases, Liang 2003 notes the role of auditing). However on the mandatory side, hedge funds face significant penalities for non-reporting, or false reporting, to the SEC. The new Form PF disclosure penalties were initially proposed as perjury, albeit downgraded in subsequent drafts. Recent cases by the SEC in their preemptive investigation of "suspicious investment returns" outlines their enforcement actions, and interest, in this area of misreporting of SEC data.[9] For example, firms were charged with fraud, and individuals were barred from the financial services industry. In other cases, hedge funds face enforcement actions of return gains and pay penalties; freeze and return of offshore assets; and, permanent injunctions.[10] This thesis does not source or examine data from the mandatory side. However, this interaction of voluntary and mandatory disclosure is a growing research area (See Agarwal et al., 2012; Schwarz and Potter, 2012). As the SEC's "Aberrational Performance Inquiry" tools develop, one might anticipate better alignment between forms of disclosure.

This thesis makes a further contribution to the hedge fund literature by documenting direct, and not statistical, evidence of misreporting in Chapter 3 — the extent to which *voluntary* historical hedge fund data is revised across time, and the predictable underperformance in these funds that revise.

1.3.2 Database vintages and changes in historic data

In Chapter 3, the research question is whether the information voluntarily disclosed by hedge funds is reliable. This is gauged by tracking changes in statements of performance in these databases across different points in time, between 2007 and 2011. In each "vintage" of these databases, hedge funds provide information on their entire performance track record, from when they began reporting to the most recent period. Since these updates are not incremental, this affords the opportunity for funds to change their histories. By comparing successive vintages, the extent of changes in performance data is established.

[9] See SEC press release 2011-252, http://www.sec.gov/news/press/2011/2011-252.htm.
[10] See SEC press releases 2011-194, 2011-135, and 2011-66, respectively.

Furthermore, these changes are conditioned on fund characteristics to determine the nature of those who revise returns, dubbed revisers. The next stage of analysis is to see if these revisions hold any information content for investors. At each vintage, portfolios of funds are formed of those that revised their histories at least once (revisers) versus those where no return revisions are observed (non-revisers). Differences in returns between these two portfolios are tested on a risk-adjusted basis, and subject to robustness tests.

The findings in this thesis highlight the extent of revisions in the historical performance records of hedge funds. By tracking snapshots of publically available hedge fund databases, and comparing successive vintages, direct evidence of changes to return records is exposed. These changes are widespread, with nearly 40% of the 18,382 funds revising at least one prior return. These revisions are not isolated to recent returns and are not merely random reporting errors. Conditioning on fund characteristics and past performance, finds the likelihood of revisions higher in funds that are larger, more volatile, and less liquid. [11] There is also variation by strategy, with Funds-of-Funds more likely to make changes, and Managed Futures funds less likely to revise. Comparing initial to final vintages of reported fund performance finds that funds are initially positioning themselves optimistically, then performance is on average being revised downwards.[12] There is a cost to revision. Detecting a fund has revised its past returns, and then subsequently tracking this reviser's future performance, helps us to predict that it will underperform funds which do not revise their returns. This performance differential is significant on a risk-adjusted basis. This result is robust to extreme values, fund type, database, and, recency and size of the revision.

[11] Note larger hedge funds are not necessarily more liquid, unlike what might be expected for mutual funds.

[12] Although one might consider original valuations when tracking revisions, the working assumption is the updated information in the latest vintage is the final expression of all prior changes. Performance comparisons use this most recent return information. In addition, I chose this as it is the only information set available to prospective investors when using the databases to make investing decisions.

1.3.3 Hedge fund asset disclosure

Hedge fund databases provide not only returns, but also the funds' assets under management. Asset information is valuable to hedge fund researchers as both asset and return information is required to assess hedge fund flows. The assets of a fund need to be valued before returns are calculated. However, in an environment of self-reporting, the availability of a value is no guarantee of its publication. There are marketing incentives for hedge fund managers to report their asset growth, but there can be negative consequences of this disclosure when managers want to hide outflows, or are over smoothing returns.

In Chapter 4, the research question concerns the extent and nature of the disclosure of assets under management. Asset disclosure is measured relative to returns for those funds willing to share some asset information. Firstly, is there a shortfall between the publication of assets and returns by funds? Secondly, are there drivers of fund asset disclosure, and, when assets go 'missing', what forces are at work? I estimate the propensity of funds having a shortfall in asset disclosure relative to returns, conditioned on fund characteristics, like total restrictions, strategy and past performance.

Finally, I ask is the removal of information benign, or does weakened disclosure herald an increased risk of fraud for investors? Fraud is always a fear for investors, particularly where information is limited. The large losses by investors in Bernard Madoff's Ponzi scheme, illustrates that the fear of this risk is warranted. Drawing on the auditing (Nigrini and Mittermaier, 1997), database quality (Straumann, 2009), and performance flag literature (Bollen and Pool, 2012), I estimate the frequency of which fraud flags are triggered. I consider funds with a shortfall in disclosure as the "problem" funds.[13] Since the reliability of voluntary disclosure of returns is questioned in Chapter 3, I test to ensure that the shortfall in asset disclosure is not merely a shadow recast from this issue.

[13]This is the parlance of Brown, Goetzmann, Liang, and Schwarz (2008) who checked mandatory Form ADV disclosures by hedge funds for disclosed fraud and SEC violations.

The findings of this thesis regarding the level of asset disclosure by hedge funds are as follows. Funds reporting returns are not as forthcoming with asset information. A third of funds are missing up to a year of asset values. This shortfall in selective asset disclosure is concentrated in funds that are diversified, have smooth return profiles, and lower restrictions. There is strategic disclosure by hedge fund managers — when funds go through a bad patch, i.e. outflows or weak relative performance, then the shortfall propensity increases. Weakening asset disclosure is a negative sign for investors, as the incidence triggering fraud performance flags is greater for these funds. Funds with a shortfall in asset disclosure have a different pattern of fraud flags to return revisers, and their characteristics differ from revising funds, which were more volatile and had greater fund restrictions. These findings thus exhort that assets, and not only returns, play a role in the debate on disclosure.

In summary, this work is important because it brings insights into hedge fund fees, return misreporting, and asset disclosure to researchers, regulators, and investors. This research identifies performance, size, and fee-setting behaviours of hedge fund management companies. Stylised facts regarding the level ("2 and 20") and direction (i.e. falling) of fees reported by hedge funds is not born out in the data. Voluntary disclosure is shown to be unreliable. Historic returns in many funds, and not just returns in recent months, are revised on average downwards. There are economic costs to enhancing disclosure, but these findings indicate the benefit for investors — since funds that revise returns subsequently underperform, and funds in shortfall carry more fraud risk.

1.4 Thesis structure

The main contributions to this thesis are in the respective Chapters: 2, 3, and 4. Each chapter is presented in the format of a discrete research 'paper.' Each chapter has its own brief introduction with relevant literature outlined, followed by data, methodology, and result sections; and, ends with a chapter conclusion.

Figures have been displayed within the text. However, the tables, given their volume and size on the page, have been managed in the following manner. The core tables for each chapter are in an appendix at the end of that chapter. Supplementary tables can be found at the back in Appendix F. A full list of tables can be found in the preamble on page vi.

Finally, Chapter 5 concludes the thesis and makes policy recommendations for appropriate voluntary disclosure.

"But if '10 is the new 15' when it comes to returns, as some investors claim, then 15 may be the new 20 when it comes to fees. Investors may be less willing to pay the typical boom-era fees of 2% of assets and 20% of returns, if those returns are subdued."

— Hedge funds: bigger, safer but duller
The Economist, 28 August 2010.

2 Understanding Variation in Reported Hedge Fund Fees

In this chapter, we[14] analyse the determinants of hedge fund management and incentive fees in a large consolidated hedge fund dataset. We detect time-series variation in fees by concentrating our attention on fund launches, and conditioning fees at launch on fund family characteristics. Larger and better performing fund families launch high fee funds. Funds with high management fees at launch do not perform any differently from low management fee funds, though funds with high incentive fees marginally outperform. Our results are robust to the use of an interval regression technique to uncover the underlying continuous distribution of fees from the discrete reported fees.

2.1 Introduction

Hedge fund fees have been the subject of significant popular controversy at least since Alfred W. Jones launched the first hedge fund, with the by now well-recognized

[14] My co-author and I have a working paper under the title "Money for Nothing? Understanding Variation in Reported Hedge Fund Fees" (with T. Ramadorai).

"2 and 20" structure. Since the advent of organized data collection on hedge funds in the early 1990s, researchers have found that hedge fund fees have had a significant impact on the investment industry, overturning the trend in falling investment costs. A sense of the magnitude of these fees can be obtained from French (2008), who finds that the growth in hedge funds and other high-fee charging asset managers, has absorbed more than two-thirds of the decline in trading costs (arising from technological improvements and consequently, better liquidity) experienced by the active portfolio management industry.

This chapter is concerned with the issue of how hedge fund fees vary across time and in the cross-section of funds. The evidence on hedge funds in the academic literature seems to indicate that their fees remain stubbornly high, despite the fact that hedge fund assets under management have grown exponentially since the early 1990s, and hedge fund returns have been falling for more than a decade (Hsieh, 2006). This appears to go against the standard intuition, that supernormal profits can only persist for so long in the presence of entry into an economic activity — in hedge funds, for some reasons, fees appear to be resilient to large-scale entry to the activity.

One problem in analyzing hedge fund fees, is that the databases report snapshots of administrative characteristics of these funds, including fees. This makes it difficult to capture the dynamics of hedge fund fees to understand whether they have been falling, rising, or remaining constant over the past decade and a half since the advent of large scale data capture on hedge funds. We attempt to surmount this obstacle, by looking at one source of time-variation in hedge fund fees that can be uncovered from the seemingly static hedge fund databases, namely the management and incentive fees at fund launch.[15]Using one of the largest available consolidated databases of hedge funds, we find that over the period from 1995 to 2009, there have been several broad trends in hedge fund fees. The most distinct of these is that management fees for newly launched funds have been rising, especially between 2000 and 2009, and

[15]In Chapter 3 when investigating changes in returns, we present another approach to unearth time-variation, by comparing database snapshots gathered at different points in time. However, the data in this chapter are from a single vintage as at June 2009.

that incentive fees at launch seem to be falling slightly over the same period. These trends are also reflected in the average management and incentive fees of all live funds reporting to databases over the sample period.

We then move to the panel of fund launches, and condition fund management and incentive fees at launch on the administrative characteristics, past performance, location, and asset gathering behaviour of the management companies that launch them. This use of management companies to analyse trends in the hedge fund industry is relatively nascent,[16] although it has been extensively employed in the literature on mutual funds. We find that large management companies, and those that have recently performed well relative to their launching peers, all launch funds with higher management fees than their counterparts. High return management companies also launch funds with higher incentive fees. Furthermore, as Khorana, Servaes, and Tufano (2009) find for mutual funds, we detect significant geographical variation and variation by strategy in hedge fund and fund-of-fund fees.

One possible criticism of our analysis is that hedge funds self-report to databases, generating important biases in their reported performance and administrative characteristics. In the case of fees, the specific concern is the reported fees 'on the tin,' as it were, do not accurately capture the range of underlying variation of true hedge fund fees, that could be subject to various rebates and side notes.[17] We attempt to tackle this issue by employing a technique that is more frequently used in the labour economics literature — we treat reported hedge fund fees as bounded in an underlying interval, and then employ the interval regression approach of Stewart (1983) to back out the determinants of the underlying continuous distribution of fees. The working assumption here is that while the reported fee might not be the true fee, it lies close to the reported amount. We consider the case of these fees lying in a region

[16]Boyson (2009) has a paper looking at performance characteristics of hedge fund families limited to the TASS database. Kolokolova (2010) considers family behaviour in hedge fund liquidation in Altvest funds. A new important study on the larger, often non-reporting, hedge fund families is by Edelman, Fung, and Hsieh (2012).

[17]The use of side letters is wide spread enough to lead the UK's Financial Services Authority to publish guidance on their disclosure (FSA 06/02 discussion paper). This guidance does not require preferential fee terms to be disclosed, increasing the need for approaches like the interval regression technique that we employ.

around typical reported fees, as well as a "rounding-up" scenario in which the true fee lies strictly below typically reported levels. Using this methodology, we are able to confirm all of our results from the analysis of reported discrete hedge fund fees.

Finally, we come to the analysis at the core of this chapter. If recently successful management companies launch high-fee funds, is this fee paid by investors 'money for nothing?' Or do these higher-fee funds achieve higher gross returns, so in equilibrium, net of fee returns are equivalent? To answer this question, we condition the subsequent net-of-fee performance of launchers on their management and incentive fee levels, as well as on other characteristics of the management companies such as past performance and size. We find that high management fee launches have no better (or slightly lower) performance than low management fee launchers when considering returns or alphas, suggesting that these fees are indeed money for nothing. Another way to state this result is that if fees are in equilibrium, our findings are consistent with bargaining power residing with the management company. On the other hand, there is some evidence that funds with high incentive fees at launch have slightly higher raw returns than funds that have lower incentive fees at launch, which suggests that there might still be some money on the table for savvy investors.

The organization of the chapter is as follows. Section 2.2 discusses related hedge fund and mutual fund literature. Section 2.3 presents the methodology that we employ. Section 2.4 describes the data. Section 2.5 describes the results from estimating the specifications, and Section 2.6 concludes. The core tables can be found in Section 2.7 and Appendix F contains additional analyses.

2.2 Related literature

Our research contributes to the literature on actively managed investments, comprehensively documenting hedge fund compensation and adding a hedge fund perspective to what is known for mutual fund fees.[18] Our use of the family, or management company as the main unit of analysis on the right-hand side of our specifications is gaining ground in the literature on hedge funds. This helps us detect the determinants of reported fees, while analyses that have focused only on funds have been thus far unable to detect such relationships (see Goetzmann et al., 2003).[19]

Unlike hedge funds, the significant majority of mutual funds are in mutual fund families. Nanda, Wang, and Zheng (2004) note that 80% of mutual funds in 2006 were in families, and Gaspar, Massa, and Matos (2006) indicate 98% of U.S. equity fund assets, and 90% of funds, are in complexes. In the mutual fund literature, authors such as Massa (2003) explore the structural benefits of fund families, and find that families profit from investor heterogeneity, by providing services such as costless switching within diverse in-house funds, thus removing the drive for funds to engage solely in performance maximisation. Gaspar et al. (2006) find that mutual fund families strategically employ cross-fund subsidisation (via reverse trades and favourable IPO allocations) to boost the performance of high value funds in their families. Family affiliations can also provide liquidity insurance between funds, and prevent fire-sales (Bhattacharya, Lee, and Pool, 2012). Given the convexity of fund flows based on past performance, there are positive gains to be had from this performance shifting activity — Sirri and Tufano (1998) note this relationship, and Gallaher, Kaniel, and Starks (2006) confirm that the convexity also operates at the fund family level. Nanda et al. (2004) argue star performers in a family can attract attention to its other funds, thus boosting market share. Continuing along the same lines, it seems that there are costs to mutual fund investors of being beholden to a single family through increased correlation within family funds. Elton, Gruber, and Green (2007) find this

[18]On top of the traditional annual management fee, hedge funds also charge an incentive fee. This fee is generally earned on any positive performance over a threshold or hurdle rate (which can be just zero) and there is no penalty for underperforming the hurdle rate. This contrasts sharply with mutual funds: in the U.S., mutual fund performance fees are required to be symmetric (with the fees earned by managers declining when performance is negative) and are thus less common in practice for mutual funds. See Elton et al. (2003) for a discussion on 'fulcrum' fees.

[19]Two recent exceptions in the hedge fund literature apply other innovative techniques to get around this problem. Schwarz (2007) employs snapshots of hedge fund databases, and Aragon and Nanda (2011) employ proprietary data from one database (Lipper TASS) on the time-series of fee changes in funds.

correlation is largely owing to common holdings, but also warn of families clustering around high or low risk type funds creating further divergences for focused investors. In consonance with this literature, we find that larger fund families, and those that have had high recent performance, charge higher fees to outside investors on newly launched funds, imposing additional costs on these investors.

This thesis also contributes to the literature on hedge fund fees. Hedge fund researchers often make use of fees as an explanatory variable (see, for example Agarwal, Daniel, and Naik, 2009; Liang, 2000; Ramadorai, 2012). It would appear, therefore, that a deeper understanding of fees and their drivers is essential indirectly as well as directly. Fee analyses in earlier hedge fund papers, have tended to concentrate on descriptive statistics of the static administrative fields. Fung and Hsieh (2000) provide an early estimate of the level of hedge fund and fund-of-fund fees using the TASS database. They find that for fixed fees, the mode is 1–2%, and the median is 1.5%.[20] For incentive fees, the mode is 0%, and the median is 10%. Fung and Hsieh (2006, p. 8) in an update mention that current "fees are roughly the same as they were in 1999." This conclusion that fund fees do not change much is a frequent theme in the literature. Liang (2000) finds hedge funds do not change their fees in practice, documenting only 1% of fund's fees change over a period of one year. In addition, Ackermann, McEnally, and Ravenscraft (1999) confirm that incentive fees are set in the offer documentation and rarely change over a fund's lifetime. Employing a larger dataset, a longer time period, and identification using new fund launches, we find that mean hedge fund fees have risen over this period.

This rise in management fees is intriguing given the increased institutionalisation of the hedge fund space. This might be a consequence of hedge fund managers having a preference for switching their revenue mix from uncertain incentive fee revenue to regular management fee income. This switch will be dependent on the hedge fund

[20] Despite this finding, reference to the stylised fact of 2 and 20% lingers. For example, persisting in the parameter calibration in theoretical papers: Hodder and Jackwerth (2007, p. 814), Foster and Young (2010, p. 1440), Lan et al. (2012, p. 12). Likewise, in the recent press, *Financial Times*, "How to avoid the average hedge fund", 4 November 2012, or, *Wall Street Journal*, "Caxton to cut its fees", 15 October 2012.

manager's view of the level of future performance anticipated. The incentive fee can remain unchanged, although the manager might expect future incentive revenue to fall. This can be influenced by the state of the fund's high-water marks. Buraschi, Kosowski, and Sritrakul (2012) note the optionality of future incentive fee revenue is affected by a fund's distance from the high-water mark, leading to endogenous leverage decisions by the hedge fund managers. They illustrate, through a structural estimation approach, how isolating this risk shifting can give investors a clearer picture of manager skill. In this chapter from a fee-setting perspective, newly launching funds circumvent this high-water mark effect by starting afresh. However, the rise in management fees has theoretical support. Managers will balance future fee income with the threat of liquidation of assets (Goetzmann et al., 2003). Lan, Wang, and Yang (2012)'s theoretical hedge fund fee model anticipates the manager's selection of leverage is dependent on liquidation risk, leading to deleveraging and reliance on management fees when prudency is called for.[21] These fund level models neglect the buttressing a management company can provide to limit liquidation (seen earlier in Bhattacharya et al., 2012). Nevertheless, deleveraging, capacity constraints, or weaker performance expectations through crowded strategies, might be behind this management fee increase.

Our use of interval regressions to back out the underlying continuous distribution of fees is important in light of rebating behaviour that has been detected in other asset classes (and that exists in the form of side notes in hedge funds, from anecdotal evidence).[22] Christoffersen (2001) documents the extensive behaviour of fee waiving in institutional mutual funds, for both low and high performing funds. In these funds,

[21]The Lan et al. (2012) model has heroic assumptions about constant alpha, does not have decreasing returns to scale to alpha on asset size, has an unrealistic high-water mark construction that is more akin to a compound hurdle rate, than a cumulative loss account driven by incentive fees paid (which leads the model to have few incentive fee payouts), and has a common high-water mark regardless of flows. In practice, funds have a spread of high-water mark levels based on the timing of individual investor flows that are dealt with by equalisation reserves, or multi-unit accounting with multiple share classes.

[22]Going somewhat against this, is the industry practice for investors in hedge funds to use most favoured nation clauses, to prevent being prejudiced by preferential fee discounts offered to other investors. See, for example, "Top tips for negotiating fund manager fees," *Financial Times*, 6 December 2010. This could limit the extent of fee rebates.

fee waivers are a quick way to change net performance figures and boost positioning.

Fees have also been analysed in the mutual fund literature. Khorana et al. (2009) in a comprehensive global study of mutual fund fees find that there is substantial variation in fees. Fees obviously vary by investment objective and type of fund (for example, they are lower for index funds). In addition, they find that mutual fund fees vary by domicile nation. They find that fees are lower for larger funds and fund complexes, which is different from our finding for hedge funds that large families launch higher fee funds. Hedge fund results on management fees and size might differ with Khorana et al. (2009)'s findings for three reasons. Firstly, hedge fund revenue is a combination of management and incentive fees, so comparing solely on management fee is potentially invalid. Secondly, hedge funds face capacity constraints, so as firms get larger the expected revenue from incentive fees might diminish, leading to a preference for higher management fees. Whereas the mutual fund industry has open-ended funds, and has a focus on market share. Thirdly, the mutual fund industry is more mature, and does not have the extensive lock-up provisions hedge funds have to protect withdrawals of capital. This influences pricing models. Large firms in a maturing industry might lower fees to protect their 'cash cow' asset bases. (Contrast this with Guidotti and Nagy (2011)'s extension of Berk and Green (2004)'s model for hedge funds where a manager's response is to raise management fees as they grow.) Khorana et al. also find that fees are higher for funds domiciled in offshore locations, whereas for hedge funds, we find that offshore funds tend to have lower incentive fees.

Finally, our analysis of fund performance conditional on fees at launch has parallels in several papers. Brown, Goetzmann, and Park (2001) find that incentive fees do not encourage additional risk taking by managers when they are behind their high-water marks (although managers might shut down or spawn additional currency classes, or new classes off a clean track record). Theory by Panageas and Westerfield (2009) also suggests that high water marks do not generate risk-shifting effects, though Guasoni and Obloj (2011) suggest that this conclusion depends somewhat on the risk aversion of the manager.

2.3 Methodology

While returns and assets under management information are reported as time-series in hedge fund databases, one limitation that has prevented researchers from analysing time-variation in the administrative characteristics of hedge funds (such as their fees), is the static nature of the reporting of these fund attributes. As a consequence, most statements about hedge fund fees have tended to be about their cross-sectional distribution (see, for example, Fung and Hsieh, 2000, 2006).[23] However, the seemingly static nature of the administrative characteristics conceals two sources of time variation in hedge fund and fund-of-fund fees. The first is the fact that new funds have been launched every month since the inception of the databases, allowing us to track how fees have varied across various vintages of launches. Second, while the same hedge fund or fund-of-fund does not launch repeatedly, multiple funds-of-funds and hedge funds belong to the same management company. This allows us to track how fees vary conditional on performance, capital flows, size and other dynamic attributes, with our unit of analysis the management company rather than the individual fund. Such management company-level analysis has been undertaken before in research on mutual funds (as reviewed in the previous section), but somewhat rarely in hedge funds thus far. In addition to discussing the time-varying nature of hedge fund fees using the approaches outlined below, we also provide statistics pertaining to the cross-sectional distribution of fees, to compare results from our broader database over a more extended time period with the estimates provided in earlier literature.

2.3.1 Empirical model

Writing $MgmtFee_{i,k,t}$ ($IncFee_{i,k,t}$) for the management fee (incentive fee) for fund i launched by management company k in quarter t, we empirically model these fees as a function of variables proposed by theory and used in prior empirical work:

$$MgmtFee_{i,k,t} = f(Performance_{k,t-1}, Restrictions_{k,t-1}, Flows_{k,t-1}) + u_{i,k,t} \quad (2.1)$$

[23]Two notable exceptions, mentioned above, are Schwarz (2007) and Agarwal and Ray (2011).

$$IncFee_{i,k,t} = g(Performance_{k,t-1}, Restrictions_{k,t-1}, Flows_{k,t-1}) + v_{i,k,t} \quad (2.2)$$

The function is linear in parameters and variables. The right-hand side consists of pure cross-sectional variables, as well as those that vary across both funds and time. Consequently, the data for all transactions are stacked, and specifications are estimated using pooled OLS. We also employ strategy-specific fixed effects, time dummies (for three broad periods of five years each) and regional dummies (akin to Khorana et al. (2009)'s use of country dummies) in estimation, to ensure we are capturing variation in fees arising from all possible sources. Since monthly data on the dependent variable is sometimes sparse or highly volatile on account of months in which there are very few launches, our regressions employ quarterly data. To account for the residual correlations, we construct cross-sectional correlation and autocorrelation consistent bootstrap standard errors (Politis and Romano, 1994), the details of which are in Appendix E. We also verify these standard errors using a Rogers (1983, 1993) covariance matrix, as well as standard errors corrected for hetereoskedasticity using White (1980)'s method.

2.3.2 Interval regression analysis

One important issue that we face when estimating the specifications in (2.1) and (2.2), is that there is significant clustering of fees at typical breakpoints, such as 1.5% and 2% for management fees, despite the considerable heterogeneity of funds' investment objectives and geographies. Figure 2.2 visually shows that this source of discreteness appears to be very important in the reported hedge fund data. Our concern is that estimates of equation (2.1) and (2.2) are likely to yield somewhat crude estimates given this form of data censoring during the reporting process.

Labour economists have been faced with a parallel problem in which income, one important variable of interest, is often collected in broad bands. The underlying income levels then need to be estimated. Stewart (1983) extends the concept of a censored sample to one in which the dependent data is grouped. Stewart notes that this provides more consistent estimates in general than simple OLS based on, say, the mid point of the interval. We therefore use this interval regression approach, to estimate underlying fees after allocating the reported fee to an interval. These interval regression estimates complement our estimates from (2.1) and (2.2) — Appendix D provides details about the interval regression procedure.

2.3.3 Regressors

The variables we consider as determinants of hedge fund fees can be broadly categorized into four categories, namely, performance, restrictions, flows, and regional variables. This subsection briefly explains the variables employed in each category.

Performance

The three performance measures employed in this chapter, are the alpha of the management company's returns on the Fung-Hsieh seven-factor model measured in the 24 months prior to fund launch, the size of the management company, and the age of the management company. The first two of these variables are measured using tercile dummies which show ranks (high, medium, low) relative to all other management companies launching funds in the quarter, and lagged one quarter to avoid any mechanical association. (Age is simply a rank standardised from 0 to 1). The use of these terciles also accounts for possible non-linearities in the fee-performance and fee-size relationships, a possibility given the well-established convexity of the flow-performance relationship (see Chevalier and Ellison, 1999, and Agarwal, Daniel, and Naik, 2005).[24] As Berk and Green (2004) suggest, if managers hold all the bargaining power in their negotiations with investors, management groups with strong performance histories will not drop fees as they will be in demand by investors. This is slightly more complicated for hedge funds than in mutual funds, since there are two types of fees. We expect that managers will raise management fees until weaker net

[24] There is also some evidence of concavity in the size-fee relationship from mutual funds. Khorana et al. (2009) use logs for their size measures for mutual fund fees analysis noting: "We employ a log specification for the two scale measures because we expect their marginal effects to decline, consistent with the previous literature analyzing fees (e.g., Baumol et al., 1980)."

performance has an impact on total (management plus incentive) fees at the margin.[25]

Management company size is employed on account of the extensive evidence on capacity constraints in hedge fund strategies, which shows that larger hedge funds, or hedge funds which have experienced high capital flows in the past, have lower expected future performance (see Ramadorai, 2011, Fung, Hsieh, Naik, and Ramadorai, 2008, Zhong, 2008 and Kosowski, Naik, and Teo, 2007 for hedge funds, and Pollet and Wilson, 2008 among others for mutual funds). Analogous to fund size, a group of funds in a management company might have the flexibility to charge less if the overall business is operating at scale. Rational management companies can respond by changing their incentive/management fee allocation. We anticipate that large managers faced with the threat of diminishing returns will switch to higher management fees (given higher demand and established distribution) and be willing to lower incentive fees. Goetzmann et al. (2003) explore this in their theoretical model and also note that large managers take on less new business reflecting limits on capacity. Furthermore, for mutual fund fees, Khorana et al. (2009) observe that larger funds and fund complexes charge lower fees.[26] Finally, age is also employed as a performance measure, following the work of Berk and Stanton (2007), which implies that expected excess NAV returns should be negatively related to the tenure of the investment manager, proxied by management company age.[27]

[25] Guidotti and Nagy (2011) confirm this in their extension of the Berk and Green model for hedge funds. Unlike mutual funds, the decreasing returns to scale impact hedge fund incentive fees. Their model results in remuneration maximised before investors would stop allocating flows, and promotes either hedge fund soft closures, or management fee increases.

[26] Fees affect the asset levels of funds, so there might be a concern of correlation between fees and size within the same fund. In the regressions, a potential endogeneity problem is limited because the asset variable used predates the launch of the fund, and, secondly, the asset value is of the total management company assets (and excludes the launch fund by construction).

[27] Given the nature of voluntary disclosure, coupled with hedge fund failure, attrition of funds will affect the sample data. This bias is mitigated in an number of ways. Using a large comprehensive sample that spans more than one database, this prevents the disappearance of funds that switch data publishers. Secondly, by using management company data, as opposed to a single fund, the age is not dependent on a single fund stopping reporting. Thirdly, ranks, and not absolute ages, are used in regressions so that the measure is not distorted by periods of poor market conditions for launches.

Restrictions

We employ two measures of restrictions for a management company. The first is the cross-fund average of total restrictions (redemption notice period + lockup period + redemption frequency) across all funds in the management company. This is then ranked across all launching management companies (the rank is standardised between 0 and 1). The second measure looks at the existence of any high-water mark and hurdle rate provisions, amongst the other funds in the management company in the period prior to launch. The indicator is set to one if any are in place. The reason we include these measures of restrictions is twofold. First, we expect that fees would change with the type of clientele in the fund, and the level of total restrictions might proxy for the investor base. For example, management companies primarily catering to funds-of-funds might have significantly more liquid funds, and potentially lower fees, depending on whether managers or capital providers have greater bargaining power. Another issue illustrated by Chen, Goldstein, and Jiang (2010) for mutual funds, which also pertains to the size variable described as a performance variable above, is that larger investors might expect discounts in management fees since large fund investors flows are less reactive to short term performance. Second, there is a liquidity-fee trade-off that has been detected by authors such as Aragon (2007), who shows that investments in funds with high lockups have higher expected returns to compensate investors for the illiquidity risk inherent in such investments. This compensation could also show up in fees.

Flows

The research on capacity constraints in hedge fund strategies shows that larger hedge funds, or hedge funds which have experienced high capital flows in the past, have lower expected future alphas. One critical assumption that is responsible for driving alpha down in models such as Berk and Green (2004), is that managers have higher bargaining power than outside investors, yielding high fee income, but zero alpha in equilibrium. In models such as Goetzmann et al. (2003), strong flows improve survival rates as the threat of withdrawal is diminished and incentive fees can be earned. Consequently, we might expect that strong flows to star performers will provide additional pricing power to raise management fees. For these reasons, we include past flows to the management company as an explanatory variable in our specifications.

The Berk and Green model is silent about entry into a highly profitable business. It might be, that incorporating entry would generate pressure for new launches to be priced at a discount to get attention in a crowded space. Empirically, therefore, it is an open question whether high past flows will translate into higher or lower fees at launch, depending on the number of other simultaneous launches.

Region

Khorana et al. (2009) show that the location of a mutual fund has an important impact on its fees. Teo (2009b) finds that a fund's proximity to investments can enhance hedge fund returns. Furthermore, several authors show that offshore funds are able to charge slightly more management fees in exchange for optimising taxation.[28] We therefore include dummy variables for the headquarters of the management company launching the fund, although in contrast to the mutual fund literature, we use more broad regional rather than country-specific dummy variables. We suspect the country to have less influence on hedge funds, in contrast to mutual funds, since hedge funds are largely unregulated (no regulated fee classes or required disclosures). Hedge funds are also more global in nature, and set up in common offshore jurisdictions, creating less opportunity for differential country effects, but the possibility of broader regional effects. The dummy variables are constructed to distinguish four regions, namely, North America, Europe, Offshore, and all other regions.

[28]See Khorana et al. (2009) for mutual funds, Brown, Goetzmann, and Ibbotson (1999) for hedge funds. Liang and Park (2008) also find that hedge fund offshore investors collect higher illiquidity premiums from these vehicles and not just the assets themselves.

2.3.4 Money for nothing? Fees and future performance

Once we understand the determinants of fees at launch, we turn to the question of whether the captured variation in fees turns up in future fund performance. Is fee variation just 'money for nothing' for the launching funds with higher fees? We relate the post launch performance of the fund to the fee at launch and the prior management company factors. In particular, we model the funds' post-launch performance to their fee levels at launch, taking into account similar control factors, strategy, time and regional headquarter, to equation (2.1) and (2.2) using a specification in which the initial fee dummies represent above median fees at launch. In these specifications, we measure future performance first on a raw return basis, then on a risk-adjusted basis using Fung-Hsieh alphas and ultimately on an information ratio basis using the t-statistic of alpha. (Kosowski et al. (2007) discuss this last measure and show that performance persistence is more accurately estimated when using the t-statistic of alpha). For a fund i launched by management company k at time t, we estimate:

$$\begin{aligned} FutureFundPerf_{i,k,t+1:t+24} = g(&HiMgtFee_{k,t-1}, HiIncFee_{k,t-1}, Perf_{k,t-1:t-24}, \\ &Restrictions_{k,t-1}, Flows_{k,t-1:t-24}) + Controls_{i,k,t-1} + u_{i,k,t} \quad (2.3) \end{aligned}$$

Hedge funds report returns net of both types of fees. Our hypothesis is, that if managers with ability are setting fees higher in anticipation of claiming the surplus for themselves, then we shall expect coefficients for the higher management fee dummies to be negatively signed, given future management fee drag. Although incentive fees do not show as much variability, managers expecting future higher performance would not lower incentive fees, so we expect β_2 to be slightly positively correlated with future performance. One way to look at these specifications is through the lens of Berk and Green (2004), in which fees tell us about the relative bargaining power of managers and fund investors. In a rational equilibrium, fee changes should convey no signal about future fund performance. However, in a situation in which managers are overcharging investors, high fees should signal declines in future performance. Similarly, high fees might signal increased future performance in a situation in which investors have not invested up to fund capacity, or managers have not raised fees enough to eliminate impacts on future alpha.

The risk of endogeneity needs to be broached as it is an inescapable problem in empirical financial economics. This is a serious problem as it can bias estimates towards zero and provide less precision **??**. We use two performance measures on both sides of equation (2.3), however we dealt with this risk in two ways. Firstly, by considering performance over distinct, and sufficiently large, time periods, and, secondly, between different entities. The methodology behind the database construction also contributes to mitigate this risk.

Recall, the right hand side contains $Perf_{k,t-1:t-24}$, the relative two-year pre-launch performance of the management company k calculated as an average of all funds already in the management company. This measure excludes the launching fund's i performance by construction. The left hand side has $FutureFundPerf_{i,t+1:t+24}$, the post-launch two-year performance of the fund i that management company k launches.

There is a residual risk that the performance of these entities can be strongly related (putting aside the periods are not contemporaneous). Consider the case if the management company funds are not diverse, and the newly launching fund is simply a copy of an existing fund, then the returns generated could be highly correlated. (This can certainly happen in practice. Management companies do launch different share classes of the same fund, such as adding a Euro class to a U.S. dollar fund offering. Although in this dataset this possibility is remote, as we are evaluating distinct fund representatives. Highly correlated shareclasses are removed in the deduplication process outlined in Appendix A.) The construction of this sample thus lowers the probability of this new launch being a 'copycat' fund. It is more likely that this new fund is a different product offering, such as a new strategy, which would lower correlation between the return measures.

Furthermore, the time periods are not overlapping. The choice of time period of two years, for both the past and future horizons, leaves the average difference between the two measures of being two years. Hedge fund persistence is typically found to be operating strongly at the quarterly horizon (Agarwal and Naik, 2000; Baquero, Ter Horst, and Verbeek, 2005), with evidence this also exists at the annual level (Kosowski et al., 2007); although seen by others as statistically weak and varying by flows (Baquero and Verbeek, 2009). Therefore with the wide gap in time period lowering performance persistence, and as the entities are distinct; correlations are expected to be low. We believe these elements diminish the concern of endogeneity in the performance measures.

From the earlier specifications, (2.1) and (2.2), to (2.3), we switch from explaining fee levels using management company variables *before* fund i is launched, to using relative fee levels as an indicator variable to test if the high fee cohort signals future declines in performance in fund i post launch. We are assuming the fees will not be changing once the fund has launched, this lessens the issue of casuality. In addition, although performance is reduced by fees, some of this impact is absorbed by considering relative rankings, not absolute measures. An alternative is to instrument the fee variable by finding a variable that is related to fees, but not related to the performance variable.

Another endogeneity problem not raised, is the correlation between performance, flows, and asset levels as these variables are inter-related by construction. A parallel project, undertaken but not documented in this thesis, tested these relationships and found no undue excessive correlations that could skew regression results. This was explored through extensive simulation to test the potential range of correlations, given the observed management fund structures and hedge fund performance levels.

2.4 Data

2.4.1 Consolidated hedge fund and fund-of-fund data

We use a large cross-section of hedge funds and funds-of-funds over the period from January 1994 to June 2009, which is consolidated from data in the TASS/Lipper, HFR, CISDM, Morningstar, and BarclayHedge databases. Appendix A contains details of the process followed to consolidate these data. The funds in the combined database come from a broad range of vendor-classified strategies, which are consolidated into nine main strategy groups: Security Selection, Global Macro, Relative Value, Directional Traders, Funds of Funds, Multi-Process, Emerging Markets, Fixed Income, and Other. The set contains both live and dead funds, the percentage of the funds in the data that are live and dead is reported in Table F.4 in Appendix F. The distribution of live versus defunct funds is roughly similar across the larger databases, and the total percentage of defunct funds is 46%, which is comparable to the ratio reported in Agarwal et al. (2009) of 44%, although their sample period ends in 2002.

We filter the data in a few standard ways. First, we remove funds with no return or assets under management histories, and missing fund level data (such as no inception date, management fee, incentive fee, minimum investment size, strategy or management company recorded). Second, to counteract known hedge fund biases, like selection and backfill biases, we remove funds that never exceed U.S.$5 million in assets over their lifetimes. This also allows us to control incubation bias (we note that Evans (2010) recommends a different type of correction, which is easier to employ for mutual funds than for hedge funds).

In this environment of voluntary disclosure, this work is reliant on the quality of data provided by managers. The opportunity for misreporting certainly exists. We address this in three ways. Firstly, we did some spot checks on the fee data, particularly for outliers, where other sources of data could be found (such as in factsheets). Secondly, we mitigate single manager errors by increasing the power through a larger sample, and drawing on different database sources. Thirdly, the interval regression technique is an attempt, albeit limited to small sized reporting errors, to test the validity of the results. Alternative approaches of dealing with potential measurement error are using instrumental variables, providing a sufficiently strong instrument could be found (Hausman, 2001), or appropriate structural models.

Table 2.2 shows that this leaves us with a total of 10,647 funds, of which 7,008 are hedge funds, 963 are CTAs or CPOs, and 2,676 are funds-of-funds (roughly 25% of the total set of funds). Approximately 40% of these funds are offshore domiciled, according to the databases, with the remainder predominantly located in North America and Europe. The investor interest in Europe for funds-of-funds over hedge fund vehicles, is reflected in its relatively higher proportion versus other regions.

2.4.2 Management and incentive fee data

Table 2.2 shows the summary statistics of the fees for the funds and their major fund group. The first interesting finding from this table is that the conventional wisdom of hedge funds' mainly issuing "2 and 20%" contracts is inaccurate. Management fees in the hedge fund space are lower than 2%, averaging 1.511% (median 1.5%). (This is not a recent development, see Figure 2.1). However, the second part of the conventional wisdom does hold true, i.e., hedge funds in the consolidated database are clustered around an incentive fee of 20%. Table 2.2 also highlights clearly the differences between fund-of-fund and hedge fund fees. What might not be appreciated is that the management fees of some funds-of-funds can be very high, (the median management fee for funds-of-funds is the same as that of hedge funds, although the average is 1.35%).

The other interesting feature of Figure 2.1 is that it reveals broad trends that can be detected from the analysis of launching funds. In particular, it does seem as though management fees of the new launchers have been rising over time, especially between 2000 and 2008, whereas incentive fees seem to have been static or slightly decreasing over the period. Later in the chapter, we confirm the existence of these broad trends after controlling for other determinants of fees.

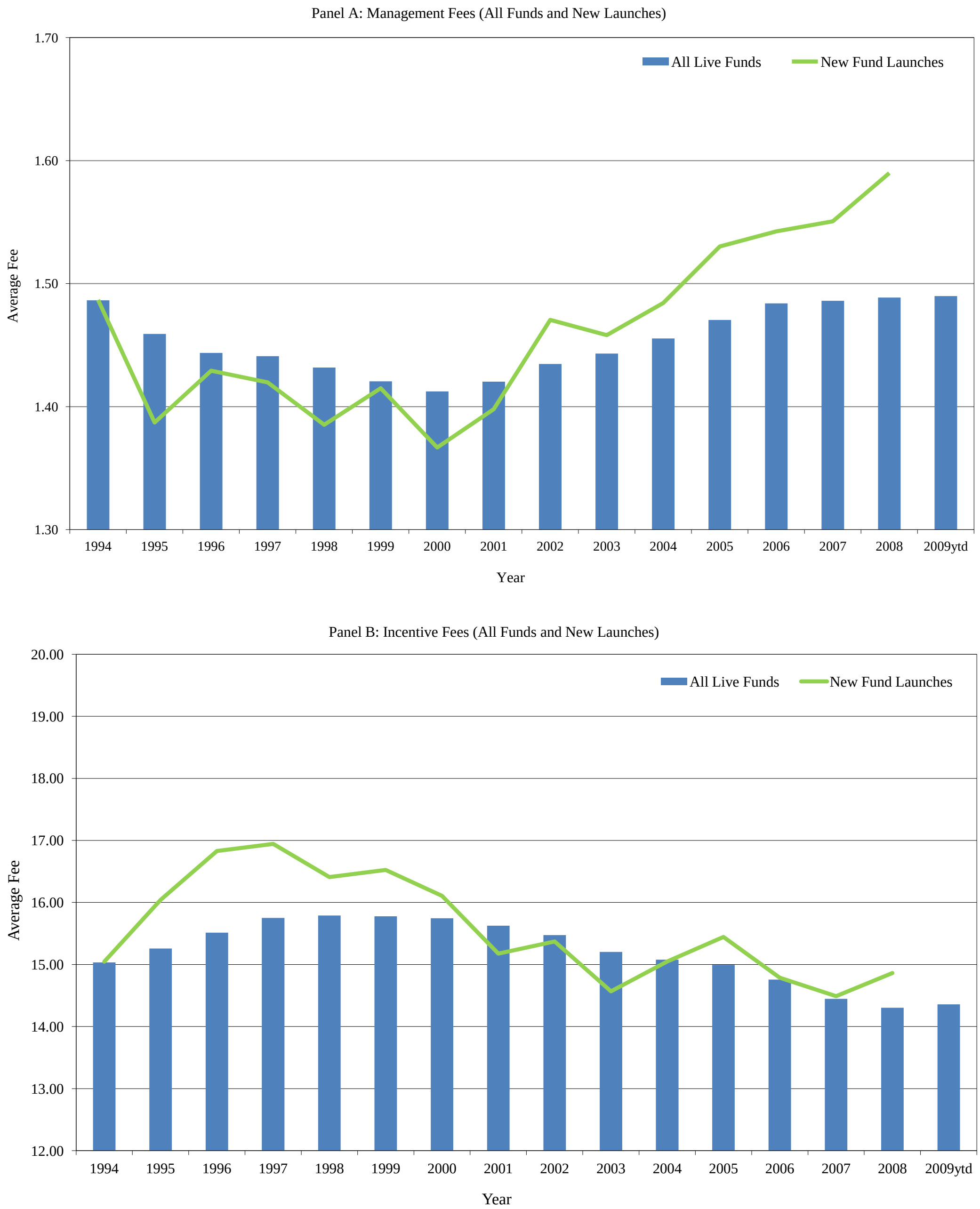


Figure 2.1: *Fees of All Funds Across Time*

The figures show the average fee values (blue bars) for all types of funds across time. All live funds are all funds currently reporting returns at the end of each year. New fund launches (green line) refers to funds launching in that calendar year, and gives an indication of the current fee setting trend. Panel A plots management fees, and Panel B plots incentive fees.

Figure 2.2 Panels A and B show the distribution of fee values and their discrete nature. Although fees could be set across a continuum, the diagrams show how fees charged are anchored to set values. There are a few key fee points chosen such as 1%, 1.5% and 2% for management fees and 10%, 15% and predominantly 20% for incentive fees. This suggests two important points. The first is that a specification such as the interval regression analysis might be necessary to extract the determinants of the underlying continuous variation in hedge fund fees. Second, in contrast to the conventional wisdom, there does seem to be some variation in the cross-section of fees. Our goal is to detect the causes of this cross-sectional variation, as well as the causes of time-series variation in reported hedge fund fees.

2.4.3 Management companies

Our primary unit of analysis in this chapter is newly launching funds, but we condition the fees of these new launchers on attributes of the management company to which the fund belongs (we refer to the management company interchangeably as the family of the fund). Management companies contain a number of different funds which could be fund of hedge funds, hedge funds, or a combination of both. Appendix A contains details of our consolidation of management company names from the raw databases.

To ensure that we are able to condition fees at launch on management company characteristics, we restrict our analysis to those management companies in each period, which have at least one fund with a sufficient history of two year of returns. Table 2.3 describes these data: the number of funds in this launch group is 4,657 funds from 1,383 management companies. The largest number of launches from management companies with pre-existing funds is in the 2000 to 2004 period (over 2,000 launches take place in this period), closely followed by the 2005 to 2008 period. Mechanically, the average age of these management companies is increasing over time. The average size of these management companies is also increasing over time, at a pace consistent with the enormous growth in hedge fund assets from 1995 to 2007. However, the management companies' returns have been falling over time, again reflecting

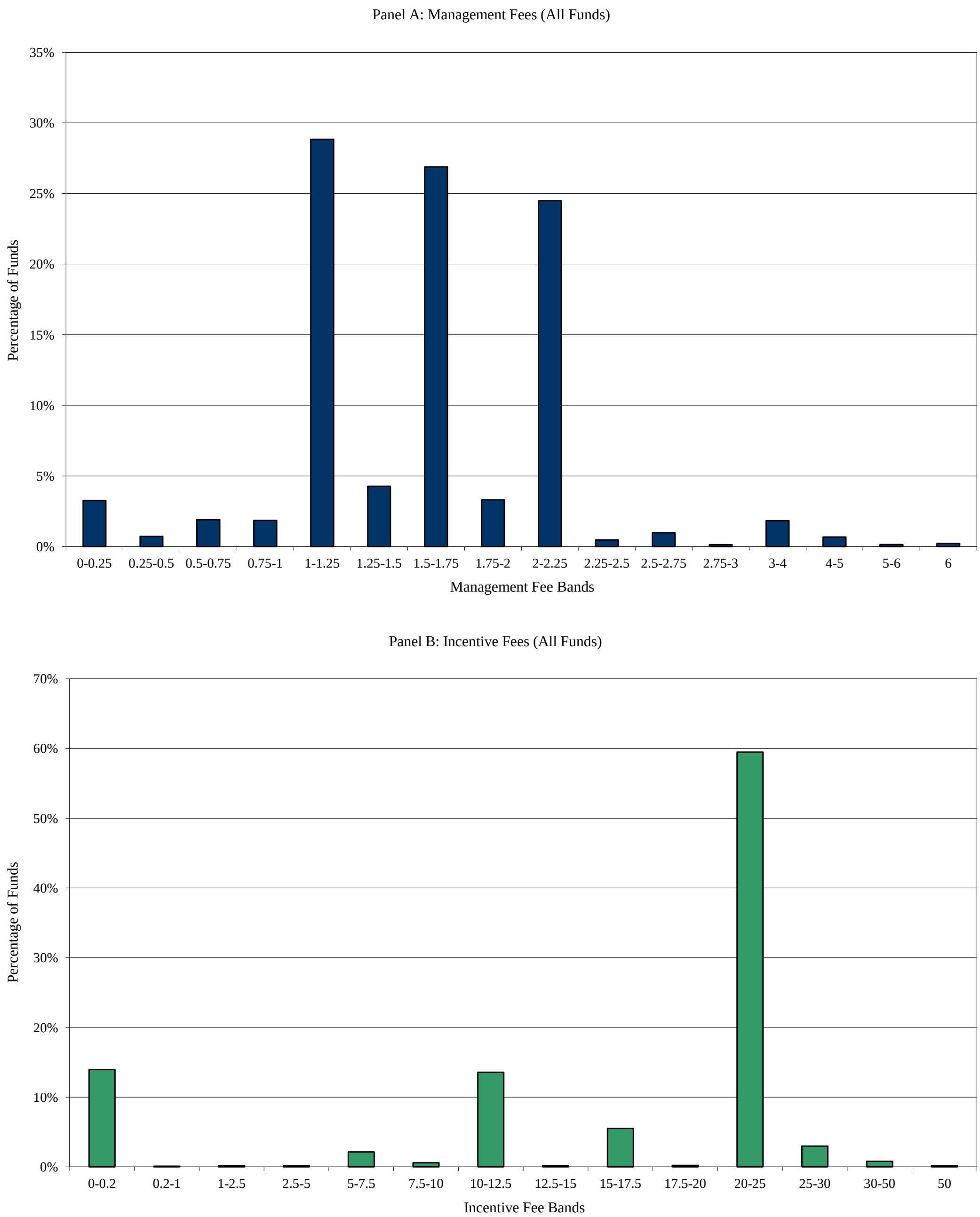


Figure 2.2: *Distribution of Fees of All Funds*
The figures show the distribution of fee values for all types of funds, as a percentage of funds with fees that fall within the selected fee bands. Bands include the lowest value and are up to but excluding the highest value. Panel A plots management fees, and Panel B plots incentive fees.

the declining performance of the hedge fund industry that several authors have attributed to capacity constraints (Fung et al., 2008; Naik, Ramadorai, and Stromqvist, 2007). Average flows have first risen, then fallen across the three periods.

Measurement of management company flows and performance

We use quarterly data on returns and assets under management (AUM) of individual funds, and aggregate these data to the management company level for use in our regression specifications. Fund AUM data are winsorised at the outset at the 1st and 99th percentile to mitigate the influence of outliers.[29] We calculate aggregate value-weighted and equal-weighted returns for management company k as:

$$R^{ew}_{k,t} = \frac{\sum_{i=1}^{N(k,t)} R_{it}}{N(k,t)}, R^{vw}_{k,t} = \frac{\sum_{i=1}^{N(k,t)} AUM_{i,t-1} R_{it}}{\sum_{i=1}^{N(k,t)} AUM_{i,t-1}} \tag{2.4}$$

Where $R^{ew,vw}_{it}$ is the net (equal or value-weighted) return for fund i in month t, and $N(k,t)$ is the number of funds in management company k in month t. Returns for longer periods (quarters or years) are constructed as arithmetic averages of the monthly returns.

To make appropriate risk adjustments in analysing performance, we calculate alphas and t-alphas using the Fung and Hsieh seven-factor model for hedge fund returns (Fung and Hsieh, 2001). These factors have been shown to have considerable explanatory power for hedge fund and fund-of-fund returns. The factors are an equity market factor (S&P 500); equity size factor (Russell 2000 less S&P 500); bond market factor using a constant-maturity adjusted 10-year Treasury bond yield; bond credit spread factor, using change in Moody's BAA credit spread over a constant-maturity adjusted 10-year Treasury bond yield; and three trend-following strategy factors, formed from excess returns on portfolios of lookback straddle options for currencies, commodities and bonds.[30]

[29] The results are robust to not doing so.

[30] Data for the trend following factors can be found on David Hsieh's website (http://faculty.fuqua.duke.edu/~dah7/HFRFData.htm). Datastream and the Federal Reserve website are sources for the equity and bond factors respectively.

We compute money flows for fund i for the quarterly period, following the methodology in Fung et al. (2008), as:[31]

$$F_{i,t} = AUM_{i,t} - AUM_{i,t-1} \times (1 + R_{i,t}) \tag{2.5}$$

To aggregate these flows to the management company level, we sum the family fund dollar flows for the funds that reported assets as the start of the quarter and then divide by beginning of period AUM, i.e., $F_{k,t} = \frac{\sum_{i \in k,t-1} F_{i,t}}{\sum_{i \in k,t-1} AUM_{i,t-1}}$, for each management company k at time t.

2.5 Results

Tables 2.4 and 2.5 show the results of estimating equation (2.1) for management fees, and (2.2) for incentive fees respectively. For management fees, funds launched by management companies with high recent past performance, have significantly higher fees than their less fortunate counterparts. Large management companies, and older management companies, also clearly charge higher fees than smaller funds (since there are time and strategy dummies in the specification, this is being driven by cross-sectional variation across funds at launch, rather than the broader trends in size and age that we saw in Table 2.3). Hedge funds with high total restrictions tend to charge lower management fees and higher incentive fees and display the liquidity-fee trade-off that has been observed in the literature (Aragon, 2007), although this is not significant for funds-of-funds. There also does seem to be significant regional variation in fees, with management companies domiciled in North America charging lower management fees than their Offshore and European counterparts. The strategy-dummies also show that managed futures funds and macro funds charge higher management fees than the other strategies.

[31] This formulation makes the assumption that any cash flows come into the fund at the end of the period after returns have been accrued.

Turning to Table 2.5, which looks at the determinants of incentive fees, it appears that better performing families launch funds with higher incentive fees. However, for size, the picture is more mixed. High-water marks within the management company allow for higher incentive fees which suggests an institutional clientele effect. In terms of regional variation, it seems that Offshore and particularly European funds charge lower incentive fees at launch than their North American counterparts.

The next section presents results from the interval regression technique, that we employ in order to derive efficient estimates in the presence of discretized reported fees.

2.5.1 Interval regressions

We observe significant clustering of fees at typical breakpoints, like 1.5% and 2% for management fees, despite the considerable heterogeneity of funds' investment objectives and geographies. These fees are in a sense being "rounded" when reported. Using these reported fees, our interest is to understand the determinants of the underlying continuous distribution of fees, given fund characteristics and industry drivers.

Labour economists have faced a parallel problem in which income, the variable of interest, is collected in broad bands. To surmount this obstacle, Stewart (1983) extends the concept of a censored sample to that of one in which the dependent data is grouped, and notes that this provides more consistent estimates in general than simple OLS based on, say, the mid point of the interval.

We believe the interval regression technique is preferable to an ordered probit. Ordered probits assume a natural ordering of categories, for example, levels of schooling that one progresses through. For fees, we are making no such assumption as the fees are set in one shot at launch. The assumption driving the interval regression is that of a normally distributed latent variable that we observe only in intervals, which we believe accurately characterizes the nature of reported hedge fund fees.

Our approach is to employ the interval regression technique to estimate underlying fees. This requires allocating the reported fees to an interval; we begin by constructing bands based on the observed reported fees. From Figure 2.2 Panel A and B, we know that the fees are focused at certain discrete break points. Appendix F, Table F.3 shows the fee bands constructed to capture the key fee points for both management and incentive fee. We were parsimonious with the bands to minimise the number of parameters and to ensure sufficient data in each category, but we keep the highest band to isolate the impact of the extreme fee values. To ensure robustness, we test formulations where endpoints are focused on reported fees (as if observed fees were rounded up), and formulations in which the probability mass is considered to reside in the centre of the fee band. Results for both band choices are shown in Table 2.6.

The details of the interval regressions can be found in Table 2.6 for the different fee types for all funds. The interval regression is in general consistent with the significance of the OLS coefficients in Table 2.4 and 2.5, and in many cases provides stronger significance. We find no reversal in signs for significant OLS coefficients discovered above. Minor changes do occur, namely that one of the country factors is less significant than the OLS approach for incentive fees; and the total restriction rank is now significant for management fees. In summary, we infer that even with the practice of reported fees being clustered, we are able to confirm the drivers on the unobserved actual fee using the interval regression approach.

2.5.2 Fees and future performance

The previous sections have described how fees at launch vary conditional on the past performance and other characteristics of the management companies that launch them. We now turn to analysing whether high fund fees at launch signal high future performance for the funds being launched. If fee-setting and performance are in a rational equilibrium, then high fee funds should not generate higher net-of-fee performance than low fee funds.

Table 2.7 shows the analysis for launched funds dependent on classifying them given above and below median management and incentive fees in each launch year. The table also conditions future fund performance on past management company performance measures as well as fees (to account for performance persistence in management company returns), on the size of the management company at launch, and the flows into the management company prior to launch. Past and future performance are measured over a period of 24 months.

The focus of attention in this table is the high management and high incentive fee dummies. Irrespective of performance measure, the management fee coefficient for higher fee funds is negative, although only statistically significant for the t-alpha performance measure. The table seems to indicate that funds which launch with relatively higher management fees, show no higher future returns than funds with lower management fees, and there is mild evidence of some negative performance for high-performing funds. It does appear that higher management fees are 'money for nothing,' suggesting that the bargaining power between fund investors and fund managers if anything lies on the side of fund managers.

Turning to the other determinants of future fund performance, we find that it is only significantly related to past management company performance when we consider the t-statistic of alpha, echoing the earlier work of Fung et al. (2008) and persistence findings of Kosowski et al. (2007). Management company size does seem to be a drag on future fund performance, which could point to funds launching new funds to circumvent capacity constraints. Higher management company restrictions positively forecast future fund performance, reaffirming that giving the hedge fund manager more flexibility to manage liquidity is beneficial. Taken together, these findings suggest that capacity constraints operate at the family level, and not just at the fund level.

Focusing now on incentive fees in Table 2.7, it seems there are somewhat more encouraging results from the perspective of fund investors. High incentive fees are positively associated with future expected raw returns, regardless of the controls employed. The magnitude of the outperformance is statistically significant in all cases, albeit small — relatively high incentive fees at launch translate into a 12 basis points per quarter higher raw return, or close to 0.5% per annum for the high incentive fee-funds. However, these results do not translate into statistically significant higher alpha. It also seems smaller management companies, particularly those charging higher incentive fees, have a performance advantage. This echoes the findings at the fund level on capacity constraints (Fung et al., 2008; Naik et al., 2007).

2.5.3 Robustness checks

We also implemented several robustness checks. We changed the prior performance period and subsequent horizon periods (to 12 months); estimated performance using both raw returns and alphas using a market factor model; grouped performance ranks using different percentiles such as quintiles versus terciles; augmented the set of controls; used unwinsorized data; and weighted funds with management company by size rather than equal-weighting.[32] Despite these changes, the results of our analysis remained broadly the same.

2.6 Conclusions

Employing data on the fees that are set by management companies for their newly launching funds, this chapter presents new insights about the causes and consequences of time-variation in hedge fund fees. Overall, we find broad trends in management fees — they appear to have been rising since 2000. Incentive fees for newly launching funds have remained constant, or slightly declined. In the panel, we find that larger management companies, and those with high recent past performance, are more prone

[32]Since many management companies have only one other fund at launch the equal and asset weighted returns are equal, so the impact is less significant in practice. We also find that equal and asset weighted management company returns were within 0.5% of each other in 85% of the cases.

to launching new funds with higher management fees. High recent past performance management companies also launch new funds with higher incentive fees. These results are robust to the use of interval regression techniques to get closer to the underlying fee distribution.

We then turn to the consequences of launching high-fee funds, conditioning future net-of-fee performance of newly launched funds on their management fee, and incentive fee levels. We find that higher management fees at launch appear to be 'money for nothing,' in the sense that there is no detectable difference between the future performance of high and low-fee funds. There is some evidence that higher incentive fees at launch do appear to signal higher future raw returns.

Our results have implications for regulators and for prospective investors in hedge funds, and suggest that paying careful attention to the performance, size, and fee-setting behaviour of management companies might yield useful insights about value for money in the hedge fund industry.

2.7 Appendix: Tables

Table 2.1: *Summary Statistics, Overall Dataset*

This table shows the breakdown of the database by fund type, strategy, and region headquarters, for all funds with fee information up to June 2009, including defunct funds. Hedge Fund Group (HFG) indicates the group of hedge funds, CTAs and CPOs aggregated together.

	Hedge Fund Group (HFG)				
	Hedge Funds	**CTAs / CPOs**	**Total HFG**	**Funds-of-Funds**	**Total Funds**
Funds	7,008	963	7,971	2,676	10,647
% of Total Funds	65.82%	9.04%	74.87%	25.13%	100.00%
Strategy % by Fund Type					
Security Selection	34.36%	-	30.21%	-	22.62%
Macro	9.76%	-	8.58%	-	6.42%
Relative Value	1.76%	-	1.54%	-	1.16%
Directional Traders	19.53%	-	17.17%	-	12.86%
Funds-of-Funds	-	-	-	100.00%	25.13%
Multi-Process	19.15%	-	16.84%	-	12.60%
Emerging	5.64%	-	4.96%	-	3.71%
Fixed Income	9.80%	-	8.62%	-	6.45%
Managed Futures	-	100.00%	12.08%	-	9.04%
	100.00%	100.00%	100.00%	100.00%	100.00%
Region % by Fund Type					
North America	44.14%	41.23%	43.78%	28.21%	39.87%
Offshore	39.50%	30.01%	38.35%	42.38%	39.36%
Europe	12.26%	16.82%	12.81%	27.32%	16.46%
Other	4.11%	11.94%	5.06%	2.09%	4.31%
	100.00%	100.00%	100.00%	100.00%	100.00%

Table 2.2: *Fee Summary Statistics by Major Fund Group*

This table shows summary statistics for the fees of all funds, also split by major fund group. Hedge Fund Group indicates the group of hedge funds, CTAs and CPOs aggregated together. Perc indicates percentile. Means are equal weighted.

	N	**Average**	**Std. Dev**	**25^{th} Perc**	**Median**	**75^{th} Perc**
			Management Fees			
Hedge Fund Group	7,971	1.511	0.662	1.000	1.500	2.000
Funds-of-Funds	2,676	1.345	0.561	1.000	1.500	1.500
All Funds	10,647	1.470	0.642	1.000	1.500	2.000
			Incentive Fees			
Hedge Fund Group	7,971	17.806	6.428	20.000	20.000	20.000
Funds-of-Funds	2,676	8.204	6.848	0.000	10.000	10.000
All Funds	10,647	15.393	7.750	10.000	20.000	20.000

Table 2.3: *Characteristics of Management Companies Launching Subsequent Funds*

These are descriptive statistics of the management companies (mgtcos) that have launched more than one fund over the period, and where sufficient assets and return information was available. Age of management company until launch is stated in years and measured from the first month returns are reported since January 1994, to the fund launch. Averages and standard deviations are shown as annual measures, though calculated from monthly data for returns, and quarterly for flows. Averages are equal weighted.

	Total Period	**1994 to 1999**	**2000 to 2004**	**2005 to 2008**
Funds launched over period	4,657	847	2,014	1,796
Unique mgtcos launching funds	1,383	429	745	799
New funds launched per mgtco				
Average	3.367	1.974	2.703	2.248
Median	2.000	1.000	2.000	1.000
Age of mgtco at launch (years)				
Average	5.678	2.928	5.560	7.106
Median	5.000	2.750	5.750	6.417
Mgtco assets managed prior to launch (U.S.$ millions)				
Average	857.4m	211.9m	650.6m	1,393.8m
Median	184.5m	69.6m	164.1m	403.9m
Std. Dev.	1,877.0m	398.0m	1,333.4m	2,556.8m
Mgtco returns 12 months prior launch				
Average	12.010	16.326	11.401	10.658
Median	9.043	13.900	7.954	8.673
Std. Dev.	4.478	5.541	4.569	3.635
Mgtco flows 12 months prior launch				
Average	17.60%	16.04%	20.94%	14.59%
Median	6.90%	5.80%	8.92%	5.28%
Std. Dev.	22.29%	21.17%	24.12%	20.48%

Table 2.4: *Explaining Management Fee*

These are coefficients from a regression of management fee for all funds combined and both fund type groups. This is explained by management company (mgtco) variables prior to launch, such as family size, prior average performance and flow statistics, and average fund characteristics within the funds mgtco (computed then ranked across all management companies); plus relevant dummies. Size, return variables and flows are allocated into terciles based on other launching mgtcos that were reporting in the period. Return variables are alphas estimated using the Fung-Hsieh seven-factor model. Regressors are described in the text. Return and flow variables look at environment eight periods prior to fund inception date, others are one period prior. t-statistics, shown in parentheses, have standard errors estimated using a cross-correlation and autocorrelation consistent bootstrap estimator. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

Panel A: Management Fee Regression Results

Management Company Variables		**All Funds**		**Hedge Fund Group**		**Funds-of-Funds**	
Relative Performance	High 3	0.111	***	0.013		-0.122	**
		(5.283)		(0.409)		(-2.280)	
	2	0.089	***	0.059	**	0.049	
		(4.012)		(2.314)		(0.929)	
	Low 1	base					
Size Prior Launch	High 3	0.087	***	0.128	***	0.048	
		(2.873)		(4.067)		(0.792)	
	2	0.039		0.054		-0.078	*
		(1.144)		(1.492)		(-1.910)	
	Low 1	base					
Past Flows	High 3	0.006		0.035		-0.093	***
		(0.164)		(1.011)		(-2.675)	
	2	-0.041		-0.027		0.045	
		(-1.223)		(-0.564)		(1.016)	
	Low 1	base					
Total Restrictions		-0.049		-0.049		0.079	
		(-0.868)		(-0.801)		(1.510)	
HWM Hurdle		0.025		0.088	*	-0.105	**
		(0.680)		(1.659)		(-2.347)	
Launch Age		0.090	***	0.003		0.139	*
		(2.583)		(0.072)		(1.789)	
N		3,881		2,744		1,137	
Adjusted R^2		0.108		0.091		0.078	

Panel B: Management Fee Regression Controls						
	All Funds		Hedge Fund Group		Funds-of-Funds	
Strategy Fixed Effects						
Security Selection	1.358	***	1.379	***		
	(10.423)		(9.149)			
Macro	1.671	***	1.684	***		
	(11.994)		(10.998)			
Relative Value	1.587	***	1.601	***		
	(11.706)		(10.868)			
Directional Traders	1.471	***	1.489	***		
	(11.038)		(9.925)			
Funds-of-Funds	1.244	***			1.413	***
	(9.712)				(13.021)	
Multi-Process	1.399	***	1.417	***		
	(11.929)		(10.336)			
Emerging	1.563	***	1.588	***		
	(10.875)		(9.939)			
Fixed Income	1.332	***	1.353	***		
	(9.640)		(8.546)			
Managed Futures	1.900	***	1.928	***		
	(19.914)		(20.554)			
Region Fixed Effects						
North America	-0.181	***	-0.197	***	-0.186	***
	(-2.923)		(-2.719)		(-2.648)	
Offshore	-0.042		-0.086		0.059	
	(-0.635)		(-1.110)		(0.820)	
Europe	-0.049		-0.137		0.081	
	(-0.687)		(-1.613)		(1.152)	
Period Fixed Effects						
2000–2004	0.040		0.060		-0.042	
	(1.175)		(1.415)		(-0.619)	
2005–	0.073	***	0.113	***	-0.065	
	(2.807)		(3.808)		(-0.880)	

Table 2.5: *Explaining Incentive Fee*

These are coefficients from a regression of incentive fee for all funds combined and both fund type groups. This is explained by management company (mgtco) variables prior to launch, such as family size, prior average performance and flow statistics, and average fund characteristics within the funds mgtco (computed then ranked across all management companies); plus relevant dummies. Size, return variables and flows are allocated into terciles based on other launching mgtcos that were reporting in the period. For terciles, 2 and 3 reflect differences from Low base case 1. Return variables are alphas estimated using the Fung-Hsieh seven-factor model. Regressors are described in the text. Return and flow variables look at environment eight periods prior to fund inception date, others are one period prior. t-statistics, shown in parentheses, have standard errors estimated using a cross-correlation and autocorrelation consistent bootstrap estimator. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

Panel A: Incentive Fee Regression Results

Management Company Variables		**All Funds**		**Hedge Fund Group**		**Funds-of-Funds**	
Relative Performance	High 3	0.714	**	0.190		0.631	
		(2.261)		(0.746)		(1.011)	
	2	0.252		0.283		0.936	
		(0.957)		(0.686)		(1.337)	
	Low 1	base					
Size Prior Launch	High 3	-0.285		0.028		-0.431	
		(-1.045)		(0.077)		(-0.772)	
	2	0.163		0.280		0.010	
		(0.652)		(0.784)		(0.024)	
	Low 1	base					
Past Flows	High 3	0.359		0.567	*	-0.272	
		(1.206)		(1.669)		(-0.503)	
	2	-0.524	*	-0.020		-0.012	
		(-1.690)		(-0.060)		(-0.016)	
	Low 1	base					
Total Restrictions		0.391		0.854	*	0.501	
		(0.990)		(1.677)		(0.603)	
HWM Hurdle		2.821	***	2.029	***	4.090	***
		(7.400)		(3.600)		(5.975)	
Launch Age		-0.861	**	-1.477	***	-0.857	
		(-2.066)		(-3.008)		(-1.179)	
N		3,881		2,744		1,137	
Adjusted R^2		0.377		0.077		0.082	

Panel B: Incentive Fee Regression Controls						
	All Funds		Hedge Fund Group		Funds-of-Funds	
Strategy Fixed Effects						
Security Selection	15.429	***	15.794	***		
	(20.677)		(21.162)			
Macro	17.518	***	17.921	***		
	(23.566)		(22.332)			
Relative Value	18.692	***	18.983	***		
	(23.279)		(24.057)			
Directional Traders	16.496	***	16.792	***		
	(21.270)		(22.484)			
Funds-of-Funds	7.019	***			7.631	***
	(9.135)				(4.479)	
Multi-Process	15.288	***	15.592	***		
	(21.731)		(19.330)			
Emerging	14.424	***	15.035	***		
	(15.817)		(16.592)			
Fixed Income	16.373	***	16.506	***		
	(21.988)		(21.998)			
Managed Futures	18.547	***	18.820	***		
	(29.189)		(26.718)			
Region Fixed Effects						
North America	-0.085		1.065	**	-4.574	***
	(-0.183)		(2.521)		(-3.395)	
Offshore	-0.540		-0.112		-3.056	**
	(-1.078)		(-0.225)		(-2.110)	
Europe	-2.241	***	-1.843	***	-5.185	***
	(-3.919)		(-3.032)		(-4.057)	
Period Fixed Effects						
2000–2004	-0.219		-0.382		1.045	
	(-0.841)		(-1.569)		(1.204)	
2005–	-0.558	*	-1.179	***	1.669	*
	(-1.649)		(-3.231)		(1.835)	

Table 2.6: *Explaining Fees — Interval Regression*

These are coefficients from an interval regression of fees, for all launching funds against relative management company (mgtco) variable terciles indicators, strategies as a constant, with region and period fixed effects. t-statistics are shown in parentheses using standard errors clustered by launch quarter. Two alternative band choices are shown I) Management Fee bands 0–0.75%, 0.75%–1.25%, etc. and Incentive Fee bands 0–7.5%, 7.5%–12.5%, etc.; and II) Management Fee Bands 0–0.5%, 0.5%–1%, etc., and Incentive Fee bands 0–5%, 5%–10%, etc. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

Panel A: Fee Interval Regression Results

		Management Fee				**Incentive Fee**			
Mgtco Variables		**Band I**		**Band II**		**Band I**		**Band II**	
Relative Performance									
	High 3	0.103	***	0.088	***	0.694	***	0.690	***
		(4.870)		(3.880)		(2.590)		(2.680)	
	2	0.083	***	0.074	***	0.168		0.189	
		(4.120)		(3.960)		(0.710)		(0.850)	
	Low 1	base							
Size Prior Launch									
	High 3	0.076	***	0.084	***	-0.103		-0.135	
		(3.350)		(3.720)		(-0.460)		(-0.630)	
	2	0.025		0.021		0.164		0.148	
		(1.190)		(0.990)		(0.800)		(0.740)	
	Low 1	base							
Past Flows									
	High 3	-0.007		-0.012		0.271		0.234	
		(-0.310)		(-0.520)		(1.110)		(0.960)	
	2	-0.040	*	-0.041	**	-0.470	*	-0.471	*
		(-1.830)		(-2.020)		(-1.760)		(-1.850)	
	Low 1	base							
Total Restrictions		-0.121	***	-0.131	***	0.082		0.097	
		(-3.040)		(-3.520)		(0.210)		(0.260)	
HWM Hurdle		0.004		0.001		1.895	***	1.797	***
		(0.160)		(0.040)		(5.360)		(5.200)	
Launch Age		0.078	***	0.059	*	-0.710	**	-0.681	**
		(2.600)		(1.910)		(-2.030)		(-2.000)	
N		3,881		3,881		3,881		3,881	
McFadden's Adj R^2		0.93%		3.54%		16.07%		15.91%	
Log Pseudolikelihood		-5210.8		-5183.6		-5230.5		-5305.1	

Panel B: Fee Interval Regression Controls								
	Management Fee				**Incentive Fee**			
	Band I		**Band II**		**Band I**		**Band II**	
Strategy Fixed Effects								
Security Selection	1.372	***	1.178	***	16.475	***	14.235	***
	(23.670)		(21.110)		(27.600)		(24.620)	
Macro	1.609	***	1.396	***	18.101	***	15.826	***
	(23.230)		(20.750)		(32.010)		(27.690)	
Relative Value	1.572	***	1.361	***	19.285	***	17.017	***
	(18.240)		(16.930)		(27.110)		(24.720)	
Directional Traders	1.474	***	1.277	***	17.269	***	15.071	***
	(22.530)		(20.840)		(28.880)		(25.580)	
Funds-of-Funds	1.241	***	1.096	***	8.727	***	6.809	***
	(19.650)		17.46		(14.660)		(11.870)	
Multi-Process	1.414	***	1.207	***	16.338	***	14.087	***
	(24.330)		(20.770)		(27.770)		(24.930)	
Emerging	1.550	***	1.373	***	15.549	***	13.333	***
	(21.780)		(20.250)		(19.480)		(17.330)	
Fixed Income	1.352	***	1.156	***	17.186	***	14.938	***
	(19.220)		(16.060)		(26.590)		(23.570)	
Managed Futures	1.729	***	1.514	***	18.867	***	16.679	***
	(28.040)		(24.970)		(35.150)		(32.260)	
Region Fixed Effects								
North America	-0.109	**	-0.106	***	0.219		0.033	
	(-2.520)		(-2.600)		(0.500)		(0.070)	
Offshore	-0.008		0.000		-0.232		-0.365	
	(-0.200)		(0.000)		(-0.530)		(-0.810)	
Europe	-0.032		0.020		-1.582	***	-1.667	***
	(-0.760)		(0.470)		(-3.180)		(-3.360)	
Period Fixed Effects								
2000–2004	0.046	*	0.044		-0.293		-0.282	
	(1.720)		(1.570)		(-1.370)		(-1.340)	
2005–	0.105	***	0.105	***	-0.524	**	-0.510	**
	(4.130)		(4.200)		(-2.130)		(-2.070)	

Table 2.7: *Explaining Future Performance*

These are coefficients from a regression of future performance of launching funds, for all funds combining both fund type groups. Regressors are described in the text. Future horizon is two years and alphas are estimated using the Fung-Hsieh seven-factor model. Management company(mgtco) performance and flow variables look at environment eight quarterly periods prior to fund inception date, others are one quarterly period prior to fund inception date. t-statistics, shown in parentheses, have standard errors estimated using a cross-correlation and autocorrelation consistent bootstrap estimator. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

Panel A: Return Predictability Regression Results

		All Funds					
Dependent Variable		**Future Returns**		**Future fund α (FH7 model)**		**Future fund α t-stat (FH7 model)**	
		Past Returns		**Past α**		**Past α t-stats**	
High Management Fee		-0.026		-0.027		-0.260	*
		(-0.500)		(-0.493)		(-1.913)	
High Incentive Fee		0.123	**	0.090		0.060	
		(2.193)		(1.230)		(0.372)	
Management Company Variables							
Performance Value		0.000		-0.031		0.129	***
		(-0.004)		(-0.636)		(4.290)	
Size Prior Launch	High 3	-0.176	***	-0.152	**	-0.357	**
		(-3.971)		(-2.424)		(-2.170)	
	2	-0.125	**	-0.094	*	-0.235	
		(-2.477)		(-1.792)		(-1.631)	
	Low 1	base		base		base	
Past Flows	High 3	-0.034		0.051		0.020	
		(-0.532)		(0.800)		(0.159)	
	2	-0.043		0.037		-0.169	*
		(-0.794)		(0.684)		(-1.848)	
	Low 1	base		base		base	
Total Restrictions		0.054		0.930	***	0.600	***
		(0.641)		(3.302)		(3.408)	
N		2,874		2,864		2,864	
Adjusted R^2		0.074		0.036		0.117	

Panel B: Return Predictability Regression Controls

	All Funds					
	Past Returns		**Past α**		**Past α t-stats**	
Strategy Fixed Effects						
Security Selection	1.153	***	0.930	***	1.908	***
	(4.570)		(3.302)		(5.404)	
Macro	1.274	***	0.980	***	2.002	***
	(4.852)		(3.720)		(5.907)	
Relative Value	1.209	***	0.990	***	1.999	***
	(3.544)		(3.238)		(3.271)	
Directional Traders	1.268	***	0.930	***	1.550	***
	(4.807)		(3.223)		(4.200)	
Funds-of-Funds	1.019	***	0.722	***	2.412	***
	(4.427)		(2.718)		(6.884)	
Multi-Process	1.156	***	0.880	***	2.524	***
	(5.050)		(3.396)		(6.198)	
Emerging	1.265	***	1.204	***	2.733	***
	(3.972)		(3.860)		(5.950)	
Fixed Income	0.978	***	0.825	***	3.831	***
	(4.572)		(3.357)		(7.275)	
Managed Futures	1.235	***	0.733	***	1.543	***
	(5.188)		(3.425)		(4.742)	
Region Fixed Effects						
North America	0.042		0.056		0.276	
	(0.350)		(0.507)		(1.382)	
Offshore	-0.085		-0.034		-0.078	
	(-0.830)		(-0.291)		(-0.351)	
Europe	-0.255	*	-0.202		-0.193	
	(-1.913)		(-1.457)		(-0.920)	
Period Fixed Effects						
2000–2004	-0.254		-0.305		-0.780	*
	(-1.363)		(-1.401)		(-1.828)	
2005–	-0.661	**	-0.071		-0.767	**
	(-1.985)		(-0.380)		(-2.036)	

"When the facts change, I change my mind. What do you do, sir?"

— attributed to John Maynard Keynes

3 Hedge Fund Return Revisions

In this chapter, we[33] analyse the reliability of voluntary disclosures of financial information, focusing on widely-employed publicly available hedge fund databases. Tracking changes to statements of historical performance recorded at different points in time between 2007 and 2011, we find that historical returns are routinely revised. These revisions are not merely random or corrections of earlier mistakes; they are partly forecastable by fund characteristics. Moreover, funds that revise their performance histories significantly and predictably underperform those that have never revised, suggesting that unreliable disclosures constitute a valuable source of information for current and potential investors. These results speak to current debates about mandatory disclosures by financial institutions to market regulators.

3.1 Introduction

In January 2011 the Securities and Exchange Commission proposed a rule requiring U.S.-based hedge funds to provide regular reports on their performance, trading positions, and counterparties, to a new financial stability panel established under the

[33]My co-authors and I have a working paper under the title "Change You Can Believe In? Hedge Fund Data Revisions" (with A. J. Patton and T. Ramadorai).

Dodd-Frank Act. A modified version of this proposal was voted for adoption in October 2011, and will be phased in starting late 2012. The proposal requires detailed quarterly reports (using new Form PF) for 200 or so large hedge funds — those managing over U.S.$1.5 billion — which collectively account for over 80% of total hedge fund assets under management; and for smaller hedge funds, these reports will be less detailed, and required only annually. The proposal states clearly that the reports would only be available to the regulator, with no provisions in the proposal regarding reporting to funds' investors. Nevertheless, hedge funds argued against the proposal, citing concerns that the government regulator responsible for collecting the reports, could not guarantee that their contents would not eventually be made public.[34]

The economic theory literature almost uniformly predicts that providing more information to consumers is welfare enhancing (an early example is Stigler, 1961, also see Jin and Leslie, 2003, 2009, and references therein). Hedge funds, however, are notoriously protective of their proprietary trading models and positions, and generally disclose only limited information, even to their own investors. One important piece of information that many hedge funds do offer to a wider audience, is their monthly investment performance. This information (as well as information on fund characteristics and assets under management)[35] is self-reported by thousands of individual hedge funds to one or more publicly available databases. These databases are widely used by researchers, current and prospective investors, and the media. As SEC rules preclude advertising by hedge funds, disclosing past performance and fund size to these publicly available databases, is thought to be one of the few channels that hedge funds can use to market themselves to potential new investors (see Jorion and Schwarz, 2010, for example).

[34] See SEC press releases 2011–23 and 2011–226, available at http://www.sec.gov/news/press.shtml. For response from the hedge fund industry, see "Hedge Funds Gird to Fight Proposals on Disclosure", *Wall Street Journal*, February 3 2011.

[35] Note that the information provided does not include the holdings or trading strategies of the fund.

In this chapter we closely examine hedge fund disclosures to these publicly available databases, with the goal of providing empirical evidence to underpin the current debate on hedge fund disclosure regulation. We are particularly interested in whether these voluntary disclosures by hedge funds are reliable guides to their past performance, and we attempt to answer this question by tracking changes to statements of performance in these databases recorded at different points in time, between 2007 and 2011. In each "vintage"of these databases,[36] hedge funds provide information on their performance from the time they began reporting to the database, until the most recent period. We find evidence that in successive vintages of these databases, older performance records (pertaining to periods as far back as fifteen years) of hedge funds are routinely revised. This behavior is widespread: nearly 40% of the 18,382 hedge funds in our sample have revised their previous returns by at least 0.01% at least once, over 20% of funds have revised a previous monthly return by at least 0.5%, and over 15% by at least 1%. These are very substantial changes, comparable to, or exceeding the average monthly return in our sample period of 0.64%.

While positive revisions are also commonplace, negative revisions are more likely and larger when they occur, i.e., on average, initially provided returns present a more rosy picture of hedge fund performance than finally revised performance. This suggests the danger of prospective investors being wooed into making decisions based on initially reported histories which are then subsequently revised. Moreover, these revisions are not random, indeed, we employ information on the characteristics and past performance of hedge funds to predict them. For example, Funds-of-Funds and hedge funds in the Emerging Markets style are significantly more likely to have revised their histories of returns than Managed Futures funds. Larger funds, more volatile funds, and less liquid funds, are also more likely to revise.

[36]This has links with the "real time data" literature in macroeconomics, see Croushore (2011) for a recent survey.

To provide an example of the sort of episode to which we refer, consider the (anonymised but true) case of Hedge Fund X, which was incorporated in the early 1990s. Four months later the fund began reporting to a database, and a year after inception it reported assets under management (AUM) in the top quintile of all funds. In the mid 2000s, the fund experienced a troubled quarter and saw its AUM halve in value. It then ceased reporting AUM figures. The fund's performance recovered, and during the last quarter of 2008 it reported a particularly good double digit return, putting it in the top decile of funds. However, a few months later this high return was revised downward significantly, into a large negative return. A similar pattern emerged later that year, when a previously reported high month return was substantially adjusted downward in a later vintage, along with two other past returns altered. A further sequence of poor returns was then revealed, and the fund was finally reported as closed in 2009.

The example provided above suggests that these revisions should be interpreted as negative signals by investors, that is, that they are manifestations of the asymmetric information problem embedded in voluntary disclosures of financial information. However, it is entirely possible that revisions are innocuous despite being systematically associated with particular fund characteristics. For example, they might simply be corrections of earlier mistakes, and therefore contain no information about future fund performance — although such corrections would have to be substantive, as simple errors such as digit transpositions and decimal point errors make up only a negligible fraction of the revisions observed in our sample.

To better understand the information content of revisions, at each vintage of data we categorize hedge funds into those that have revised their return histories at least once (revisers) and the remainder (non-revisers). We find that on average, revising funds significantly underperform non-revising funds, and that there is a far greater risk of experiencing a large negative return when investing in a revising fund. In short, this method reveals in real time, that funds with unreliable reported returns are likely to underperform in the future. The finding is virtually unchanged by risk-adjustment using various models, not greatly affected by varying the size threshold for detecting significant revisions, stronger for revisions pertaining to periods far back in time, stronger for funds with higher levels of asset illiquidity, and robust to various other changes in parameter values. The results from these robustness checks also provide some evidence that performance differentials between revisers and non-revisers, are higher for more illiquid funds, but they are by no means restricted to these funds.

Our analysis suggests that mandatory, audited disclosures by hedge funds, such as those proposed by the SEC in 2011, could be beneficial to investors and not just regulators, and contributes to a growing list of examples highlighting the benefits of an independent auditor or regulator for financial institutions. For example, Danielsson, Embrechts, Goodhart, Keating, Muennich, Renault, Shin, et al. (2001) note that under Basel II European banks were given the choice of either using a standardized model to measure their risk exposures (used in setting their capital requirements), or using their own in-house models. These in-house models were subject to audit by the banking regulator, but owing to the complexity of each bank's models it is questionable whether it was possible or feasible for the regulator to properly monitor their effectiveness. After the financial crisis, it was noted in the press and in the finance literature that these models appear to have under-estimated the true risk of many banks' positions.

The remainder of the chapter is structured as follows. Section 3.2 reviews related literature. In Section 3.3, we describe the data and introduce how we determine revisions. Section 3.4 outlines our methodology. We present our main empirical results in Section 3.5, and some robustness checks in Section 3.6. Section 3.7 concludes. The core tables can be found in Section 3.8 and Appendix F contains additional analyses.

3.2 Related literature

Several previous authors have noted problems with self-reported hedge fund returns. The fact that hedge fund managers voluntarily disclose returns to hedge fund databases, means that they are able to choose if and when to start reporting, and when to stop reporting. This leads to substantial data biases not seen in traditional data sets, such as listed equities or registered mutual funds. Ackermann et al. (1999), Fung and Hsieh (2000, 2009) and Liang (2000) provide an overview of these biases such as survivorship, self-selection and backfill.

Self-reporting also leads to the possibility of using different models to value assets, as well as the possibility of earnings smoothing. For example, Getmansky et al. (2004) document high serial correlation in reported hedge fund returns relative to other financial asset returns, and consider various reasons such as underlying asset illiquidity to explain this. Asness, Krail, and Liew (2001) note that the presence of serial correlation leads reported returns to appear less risky and less correlated with other assets than they truly are, thus providing an incentive for hedge fund managers to intentionally "smooth" their reported returns, a form of earnings management for the hedge fund industry. Cassar and Gerakos (2011) match due diligence reports with smoothing measures, and find that smoother returns are associated with managers, who have greater discretion in sourcing the prices used to value the fund's investment positions. Bollen and Pool (2008) extend Getmansky et al. (2004) to consider autocorrelation patterns that change with the *sign* of the return on the fund, with the hypothesis being that hedge fund managers have a greater incentive to smooth losses than gains, and they find evidence of this in their analysis. This finding is reinforced using a different approach in Bollen and Pool (2009), who document that there are substantially fewer reported monthly returns that are small and negative than one might expect. When aggregating to bimonthly returns no such problem arises, suggesting that the relative lack of small negative returns in the data is caused by temporarily overstated returns. Jylha (2011) extends Bollen and Pool (2009) work on misreporting, by conditioning the search for pooled distribution discontinuities on various fund attributes.

Agarwal et al. (2011a) find evidence that hedge funds tend to underreport returns during the calendar year, leading to a spike in reported returns in December, that cannot be explained using risk-based factors (a similar result for quarter-end returns for mutual funds can be found in Carhart, Carpenter, Lynch, and Musto (2002)). The motivation for doing so is that hedge funds are paid incentive fees once a year based on annual performance. At higher frequencies, Patton and Ramadorai (2012) find that estimated hedge fund risk exposures appear to be highest at the beginning of the month, and lowest just prior to end of month reporting periods.

Others have looked at 13-F filings by hedge funds to uncover evidence of unreliable voluntary disclosure, such as Cici, Kempf, and Puetz (2011) who find evidence that these filings often appear to be valued at prices different from prevailing closing prices in CRSP, Ben-David, Franzoni, Landier, and Moussawi (2012b) who present evidence that hedge funds appear to increase holdings of illiquid stocks at critical reporting valuation dates, and Agarwal et al. (2011b) who find that hedge funds are the greatest users of confidentiality provisions to delay reporting of sensitive positions in 13-F filings.[37] While this thesis is related to this stream of research, the new empirical phenomenon we document might be better labeled "history management" — with closer parallels to earnings restatements rather than to earnings management (see Dechow, Ge, and Schrand (2010) for a comprehensive review of the accounting literature on the subject).

The literature on hedge funds has also considered the role of mandatory disclosures for hedge funds. For a unique, and brief, period in 2006 before the rule was vacated, the SEC required hedge funds to disclose a variety of information such as potential conflicts of interest, and past legal and regulatory problems. These Form ADV disclosures were designed to deter fraud, or control operational risk more generally. Brown, Goetzmann, Liang, and Schwarz (2012); Brown et al. (2008) report evidence that these mandatory disclosures of information related to operational risk were beneficial to investors. The authors find that the information in these disclosures enabled investors to select managers that went on to have better performance, and that conflicts identified in the Form ADV filings were correlated with other flags for operational risks.

[37] Along the same lines, Aragon and Nanda (2011) examine the timing issues surrounding short-run history management. While they do not examine return revisions, they find that the reporting of bad news by hedge funds is strategically delayed until weak performance reverses.

Our analysis of changes in the reported histories of hedge fund returns is also related to Ljungqvist, Malloy, and Marston (2009), who study changes in the I/B/E/S database of analysts' stock recommendations. These authors document that up to 20% of matched observations are altered from one database to the next, using annual vintages of the IBES database from 2001 to 2007. Like us, they find that these revisions are not random: recommendations that were further from the consensus, or from "all star" analysts, were more likely to be revised than others, and undoing these changes reduces the persistence in the performance of analyst recommendations. While the focus of these authors was primarily to illuminate problems of replicability in academic research, our concerns run deeper on account of the environment of limited disclosure for hedge funds. This environment generates a greater reliance on self-reported hedge fund data. We demonstrate that hedge fund return revisions could skew allocations by investors reliant on the initial return presented. Moreover, the significantly lower future returns and greater downside risks in troubled times experienced by funds with unreliable disclosures, suggests that the issue we identify represents a source of risk to hedge fund investors, and quite possibly a broader systemic risk.

Finally, it is worth noting here that information on the trading strategies and positions of hedge funds also has implications for how they are compensated. Foster and Young (2010) show theoretically, the difficulty of devising a performance-based compensation contract for hedge fund managers that rewards skilled managers, but not unskilled managers. With only returns histories made available for performance evaluation, unskilled managers can mimic skilled managers arbitrarily well, simply by taking on an investment with a small probability of a large crash. Foster and Young (2010) argue that transparency of positions, not just performance, is needed to separate skilled managers from unskilled managers.

3.3 Data

3.3.1 Consolidated hedge fund and fund-of-fund data

We employ a large cross-section of hedge funds and funds-of-funds over the period from January 1994 to May 2011, which is consolidated from data in the TASS, HFR, CISDM, Morningstar, and BarclayHedge databases. Appendix A contains details of the process followed to consolidate these data. The funds in the combined database come from a broad range of vendor-classified strategies, which are consolidated into 10 main strategy groups: Security Selection, Macro, Relative Value, Directional Traders, Funds-of-Funds, Multi-Process, Emerging Markets, Fixed Income, Managed Futures, and Other (a catch-all category for the remaining funds).[38] The set contains both live and dead funds. Returns and assets under management (AUM) are reported monthly, and returns are net of management and incentive fees.

3.3.2 Hedge fund database vintages

Hedge fund data update their databases from time to time. These updates not only include the incremental changes since the previously published version, but also the entire history of returns for each fund including incremental changes. This allows us to compare reported histories across vintages of these databases at various points in time. We compare a total of 40 vintages of the different databases between July 2007 and May 2011.[39] At each of these vintages $v \in \{1, 2, \ldots 40\}$, we track changes to returns for all available databases. Not every database is updated with the same periodicity, and in those cases the newer vintage is simply set to the previous one, thus forcing zero detected changes.

[38] The mapping between these broad strategies and the detailed strategies provided in the databases is reported in Appendix B.

[39] Vintages were collected in July 2007, and then monthly from January 2008 to May 2011, with February and November 2009 omitted owing to data download errors.

We apply some standard filters to the data before analysis. First, we remove 82 funds with very large or small returns to eliminate a possible source of error (truncating between monthly return limits of −90%, and +200%).[40] Second, we remove 186 funds that report data only quarterly. Third, we remove funds with insufficient return histories (less than 12 months) and missing fund level data (such as no "Strategy" or "Offshore" indicators recorded). Fourth, as less than one-third of Morningstar funds passed these quality filters, we remove the remaining 832 Morningstar funds to ensure sufficient depth by database. The final cleaned dataset contains 18,382 unique hedge funds.

Table 3.1 shows some characteristics of the sample. On average, funds report for five and a half years, have U.S.$104 MM in assets, and generate returns of approximately 0.64% per month. Slightly over a quarter of them are Funds-of-Funds, with Security Selection and Managed Futures being the predominant hedge fund strategies represented in the data. Approximately one-third of the funds are from the TASS database, with the CISDM database accounting for the smallest share of the four databases represented in our final sample, at just under 10% of funds.

3.3.3 Changes: Revisions, deletions, and additions

We compare return histories across successive vintages and group changes into three categories, namely, additions, deletions, and revisions. To help elucidate these categories, consider $Ret_{i,t,v}$, the return for fund i at time t reported in vintage v of the database. We drop i and t for ease of exposition, and let $v-1$ indicate the previously available vintage for the database in which the fund's data was reported (this might not necessarily be immediately one vintage prior as not all databases update simultaneously). An addition implies that a 'new' return appears in a later vintage, i.e., Ret_{v-1} was not in the database, but Ret_v is present. Clearly there are legitimate circumstances in which this would happen, such as when a new fund launches, or when new return updates are provided for months between the dates at which the

[40] Although −100 would be the natural choice, we used −90 to specifically remove cases in which data providers use large negative returns as placeholders for missing observations.

two vintages were captured. In order to rule these cases out, when counting additions we exclude all fund launches (in which there is no return for the entire fund in the preceding vintage) and exclude return months within 12 months from the vintage $v-1$ (to avoid picking up late reporting).[41] A deletion implies that a return goes missing between vintages, i.e., Ret_{v-1} was reported but Ret_v was not. We define as revisions cases in which both Ret_{v-1} and Ret_v are available but are not equal to each other. As mentioned above, we filter out small changes (less than 1 basis point) that might be attributable to rounding.

Table 3.3 shows the prevalence of these three different types of changes to funds' return histories. Over 40% of funds have one of the three types of changes described above ("Any Change"). Of these, revisions of pre-existing data are the most frequent, at 38%, followed by deletions at 6%, and additions at 2%. (Some funds have multiple types of changes, and so the sum of the individual categories is greater than the "Any Change" proportion.) This large percentage of funds with revisions demonstrates that this is a widespread problem: funds that have had at least one change in their reported history manage around 46% of the average total assets under management (this number peaks at U.S.$1.8 trillion in June 2008).

Panels B and C of Table 3.2 report summary statistics on the size of revisions observed in our sample. We observe that 38% (6,906 funds) of funds revise their returns at least once by at least 1 basis point, and 22% of funds revise at least once by at least 50 basis points. Panel C reveals that the mean absolute revision is 82 basis points. To provide an appropriate comparison, the mean monthly return across hedge funds is 64 basis points, as reported in Table 3.1, i.e., lower than the mean absolute monthly revision. The revisions that we detect are therefore substantial.

[41] For example, consider the case in which vintage $v-1$ for a fund was captured in June 2009, and this vintage shows fund histories up to February 2009. The next vintage v, is captured in August 2009 and this vintage shows fund histories up to July 2009. We would disregard any additions of data occurring after the month of June 2008 when computing the additions for this fund. So for example, if March 2009 and April 2009 returns are missing in $v-1$ but present in v, these months would not be counted as additions, to ensure that we do not capture late updates of returns by the fund's manager to the database provider. Our focus for additions is backfilling of past history rather than short-term lags in fund reporting.

Panel D of Table 3.2 reports on the "recency" of the revisions that we detect in our data, defined as the difference between the date of the return and the date at which a revision was detected. Each of the columns of Panel D shows the proportion of revising funds remaining once we exclude revisions near the vintage date (for example, when $k > 3$ we ignore revisions of returns that occur within three months of the date of the return). As we increase k, the proportion of funds that are flagged as having revised their returns declines, from 37.6% in total, down to 18.7% when we ignore any revision within a year of the return date. This reveals that almost one half of the return revisions in our sample relate to returns that are more than 12 months in the past.[42] Presaging results from later in this chapter, it seems unlikely that these revisions are merely corrections of data entry errors or a simple consequence of illiquid positions being marked-to-market.

Panel E of Table 3.2 investigates whether the revisions that we find in our data are mainly attributable to common data entry errors. We consider three such errors: sign changes (where the revised return is identical to the original return except for the sign), decimal place errors (where the revised return differs from the original return by exactly a factor of 0.01, 0.1, 10, or 100), and transposition errors (where adjacent digits in the original return are transposed in the revised return).[43] We find that these contribute only a negligible fraction of the observed revisions — only 3.3% of funds have one of these types of errors, compared with the 37.6% of funds that have revised their returns at least once. Thus these common types of data entry errors do *not* appear to be the primary source of the return revisions that we uncover in our data.

[42] A reviewer suggested some asset classes, like private equity, could have long standing revisions occurring after accounting audits have been concluded. We find even beyond two years there are substantial older revisions. Revisers as % of funds (revisions) for longer periods of k are: 2 years – 13%(48%), 3 years – 10% (38%), and, 5 years – 6% (22%).

[43] Verhoeff (1969) gives statistics of the frequency of errors, and estimates transposition errors can be 10% to 20% of errors.

Table 3.3 shows the prevalence of return revisions by strategy, and reveals the degree of heterogeneity across hedge fund strategies. The style with the highest proportion of revising funds is the Fund of Funds category (52.5%), followed by the Relative Value (43.6%) and Emerging Markets (42.9%) categories. The style categories with the lowest proportions of revising funds are Managed Futures (28.4%), Security Selection (28.7%) and Macro (30.4%). This ordering is in line with previous studies of hedge fund liquidity, see Getmansky et al. (2004) for example, and suggests that funds in less liquid styles are more likely to revise their reported returns. It also seems to be the case that conditioning on the minimum size of revisions does not greatly affect these relative proportions across strategies. We study the determinants of revising behaviour, including strategy affiliations, in more detail using a probit model described in the next section.

3.3.4 Hedge fund return factors

To make appropriate risk adjustments in analyzing portfolio performance for the revising and non-revising funds, we calculate alphas via the widely-used Fung and Hsieh seven-factor model for hedge fund returns (Fung and Hsieh, 2001). The Fung-Hsieh factors have been shown to have considerable explanatory power for hedge fund and fund-of-fund returns. They comprise four market related factors: an equity market factor (S&P 500); equity size factor (Russell 2000 less S&P 500); bond market factor using a constant-maturity adjusted ten-year Treasury bond yield; bond credit spread factor, using change in Moody's BAA credit spread over a constant-maturity adjusted ten-year Treasury bond yield; and three trend-following strategy factors formed from excess returns on portfolios of lookback straddle options for bonds (PTFSBD), currencies (PTFSFX), and commodities (PTFSCOM).[44] In robustness checks, we also add an eighth factor to the Fung-Hsieh set, namely, MSCI Emerging Market index returns; and employ the Fama-French-Carhart and Pastor-Stambaugh models as al-

[44]Data for the trend following factors can be found on David Hsieh's website (http://faculty.fuqua.duke.edu/~dah7/HFRFData.htm). Datastream and the Federal Reserve website are sources for the equity and bond factors respectively.

ternative risk-adjustment models.[45]

3.4 Methodology

We begin by documenting the characteristics of funds that are prone to return history changes, focusing our analysis mainly on the most prevalent category of changes, namely revisions. We then go on to analyse the determinants of the size and sign of revisions, documenting the differences between initially perceived and final histories. This enables a better understanding of how an investor using the database would see different pictures of hedge fund performance, if he or she had employed different vintages of the data. Finally, we form portfolios of reviser and non-reviser funds to ascertain the information content of revisions for future performance and shortfalls.

3.4.1 Which funds revise?

Our first step is to assign a '1' to any fund which experiences a revision of returns across any two vintages of data. Assigning a '0' to all other funds, we then estimate a cross-sectional fund level probit regression, conditioning this variable (which we label Rev_i for fund i) on various fund characteristics, which are described below, and denoted by the vector X_i:

$$Rev_i = \alpha + X_i'\beta + u_i. \tag{3.1}$$

In the above regression, the right-hand side comprises pure cross-sectional variables, including lifetime summaries of variables such as fund performance and fund size. While it is a useful summary of the fund characteristics associated with revisions, it does not permit us to make any causal statements. We therefore also estimate a version of the above specification in which the right-hand side variables are computed at vintage $v-1$, in order to explain revisions occurring between vintages $v-1$and v. This also allows us to understand whether revisions are autocorrelated across vintages, i.e., whether funds that have revised returns in the past are likely to do so

[45] An anonymous referee suggested that tradeable bond portfolios could be used instead of the Fung-Hsieh bond market and credit spread factors. Some indicative work (not shown) confirms the findings are robust to this, and actually strengthens the results.

again in the future.[46]

$$Rev_{i,v} = \alpha + \gamma Rev_{i,v-1}, X'_{i,v-1}\beta + u_{i,v} \tag{3.2}$$

In these specifications, we employ a range of conditioning variables on the right-hand side. We employ assets under management (AUM) to study whether changes are more likely to occur for larger or smaller hedge funds, ranking funds by their lifetime average AUM, computed using data at the final vintage available for the fund (or vintage $v-1$ in (3.2)). We also use the average of lifetime returns for each fund, again computed using data at the final vintage (or vintage $v-1$ in (3.2)). This is to capture the possibility that weaker performing funds might resort to changes to recast their histories. Third, we use the standard deviation of lifetime returns, to capture the fact that funds with more volatile returns might experience pressure to delete or recast disappointing performance. Finally, we use a measure of return smoothing suggested by Getmansky et al. (2004), namely the first-order autocorrelation coefficient of lifetime returns. In all cases in which we employ ranks, they are standardized between 0 and 1.

In addition to these variables computed from return and AUM histories, we also consider a variety of fund characteristics as explanatory variables. We include strategy fixed effects in our specifications to control for the possibility that differences in volatility and liquidity occasioned by the use of these different strategies, as well as differential access to information about these strategies (for example, underlying returns for obscure investments by Emerging Markets funds might be difficult to independently verify) might lead to differences in the propensity to alter data. We also include database fixed effects as controls, such as verification of returns pre-loading, implemented by each database vendor might vary, thus influencing the propensity for changes. We employ an indicator for whether the fund is offshore or onshore, as

[46]Standard errors are clustered by database in equation (3.1) and by vintage in equation (3.2). The former is to control for the possibility that errors across funds are correlated according to the database, as there might be unobserved determinants controlling revisions in specific databases. The latter is to control for the possibility that there are certain periods in which unexplained revisions are more likely to be prevalent. Appendix F also presents results which explain the prevalence of additions, deletions and 'any change,' a catch-all category encompassing all three types of changes.

funds in offshore jurisdictions might be subject to less scrutiny, and condition on the lockup restrictions imposed by the fund on its investors — these restrictions provide liquidity safeguards for the fund manager, but also might allow managers to hide from the reputational consequences of changing data within the period of the lockup. We also include an indicator for whether a fund has a high-water mark or a hurdle rate, as well as an indicator for whether the fund has any audit information available in the database.[47]

Finally, we include a variable which computes the number of returns in fund i's lifetime history up to vintage v. This is to control for the purely mechanical possibility that if there is a small fixed chance of data capture error, then a longer return history provides more exposure to return revisions. Of course, this is also a measure of the age of a fund, so this variable has multiple interpretations. Table F.6 in Appendix F contains descriptive statistics for several of these variables.

3.4.2 Determinants of the size and direction of revisions

Having determined which funds revise, we turn next to understanding the impact of revising history on the historical performance record of funds. We do so by comparing the initially reported return for fund i in month t, with the same fund-month return as seen in the last database vintage in which it appears. This analysis attempts to answer the following question: if an investor only looked at a return expressed by the fund's portfolio manager the first time it was made public, how does this differ from what the investor might see in the database at the last available vintage?

Our next step is to condition the return differences occasioned by revisions on various fund characteristics and period fixed effects. The dependent variable in these regressions is the average difference, for all years in which a fund experienced return revisions, between the final set of annual returns provided by a fund and the first set of annual returns provided by the same fund for the same year. For example, if a

[47] Underlying databases differ in the types and level of information they provide, with some providing the date of last audit, other providing annual audit flags, and yet others providing auditor names. Our indicator takes the value '1' if *any* audit information is available for the fund, and zero otherwise.

fund Y initially reported 6% average annual return for year t, and at the final vintage this average stood at 4%, then the return difference variable would be -2%.

In these specifications, we only include periods in which the fund had at least six months of return observations, to reduce the noise in the dependent variable. We explain both the absolute value of all such differences as well as the signed revisions on the independent variables. Period dummies include crisis dummies for the 1998–1999 period, the 2000–2001 period, and the 2008–2009 period. Several of the remaining regressors have been described earlier, with three new additions, namely the rank of flows experienced by the fund relative to all other funds in the same year; the management fee; and the incentive fee of the fund.

3.4.3 Are revisions informative about future performance?

Our final question is whether knowing that a fund has revised its past performance constitutes useful information about its future performance. The null hypothesis here is that these revisions are innocuous and provide no information about future returns. One alternative is that they are an indicator of either poor operational controls or of dishonesty, both of which provide negative information about revising funds (as in Brown et al., 2008). A third possibility is that revisions are a sign of honesty, in the sense that revisers 'fess up' to past mistakes. In this case, we might expect performance to be higher for revisers than non-revisers.

To consider these hypotheses rigorously, we employ two methods to determine the performance differentials between revising and non-revising funds. Our first approach is to form portfolios of the returns of funds based on their revising behaviour, allocating funds to one of two groups, "reviser" funds that have revised at least once, and "non-reviser" funds that have had no revisions up until a given vintage. At the first vintage, by definition, all funds are non-revisers. At each subsequent vintage, once we observe revising behaviour, we allocate funds into these two groups, moving several funds from the non-reviser portfolio to the reviser portfolio at each step. Once a fund is categorized as a reviser, we track all its subsequent returns in the reviser portfolio.

Note that this is a real-time strategy: consider the example of a fund making its first ever return revision, say of its previously reported January 2007 return, in the August 2008 database vintage. Once we detect this historical return revision, we immediately classify the fund as a reviser. The reviser portfolio will then include the fund's returns from September 2008 until the end of our sample period, and the non-reviser portfolio will no longer track its returns from September 2008 onwards. Thus, at each time period, the non-reviser portfolio contains funds that have never revised data in any previous vintages, although it could contain funds that are yet to be identified as revisers. Within each portfolio, we weight all monthly returns of funds equally, computing a time-series of portfolio returns.[48] We can then look at whether there are differences in the returns of reviser and non-reviser portfolios, and risk-adjust these return differences in various ways.

3.5 Results

3.5.1 Which funds revise?

Table 3.4 shows the results of estimating the probit regression equation (3.1) for revisions. (The results for other change types, including whether a fund made any one of the three different types of changes, can be found in Appendix F.) These regressions present the marginal effects of each continuous right-hand side variable, that is, the change in probability in the dependent variable that results from an infinitesimal change in each of these variables. For dummy variables, such as offshore, the effect is captured for the discrete change of the variable from 0 to 1.

Table 3.4 reveals that asset size and return autocorrelation are positive and significant determinants of a fund's propensity to report a change in history.[49] The number of returns present for a fund has a significant effect on the propensity to make a revision, although this could be simply a mechanical effect as described above. Turning to

[48] In Section 3.6.4 we consider using the median of the returns on the reviser and non-reviser funds to address concerns about outliers driving the results, and show this is not an issue in our sample.

[49] Although these marginal effects are focused on the median rank, we confirm in Appendix F that these effects are present when considering other quantiles.

the strategy indicators, Funds-of-Funds show the highest chance of reporting changes, which is perhaps unsurprising, as Fund-of-Fund performance numbers are a function of underlying hedge fund performance numbers, suggesting that their revisions might simply be a function of revisions in the hedge funds that they hold.[50]

An increase in the total restrictions (lockup plus redemption notice period) on removing capital from the fund has a positive and significant effect on the propensity to report changes in histories. This might be correlated with greater asset illiquidity, as suggested by Aragon (2007), or constitute evidence that having a "longer period in which to hide" prior to withdrawals by investors shields funds from the adverse consequences of revisions. The presence of a high-water mark or hurdle rate provision in the fund is also associated with a higher propensity to revise, a finding to which we return a little later in the chapter. The presence of audit information, reflected in the audit flag, has a large positive and marginally significant coefficient. At first glance this seems counter-intuitive, as one might expect that funds not subject to audits would have more latitude to change returns. However, it might be the case that auditing could trigger corrections in returns – alternatively frequent changes in returns might prompt investors to press for funds to undergo audits. Finally, fund performance rank is negatively correlated with the propensity to make any changes at the 10% significance level, suggesting that poorer performing funds are associated with revisions. However, the direction of causality is unclear from this analysis — as revisions might presage poorer performance — and we investigate it in greater detail below.

In Table 3.5, we find that average returns forecast revisions with a positive coefficient, i.e., funds with better past performance are more likely to revise returns. Taken together with our finding from Table 3.4, this foreshadows a result from the next section, namely that reviser funds tend to perform poorly in periods *subsequent* to revisions. Using information from the previous vintage also enables us to include

[50] In Tables F.14 and F.15 in Appendix F, we present results corresponding to Table 3.4 and 3.5 but with Funds-of-Funds removed from the sample. The results are very similar and all of the main conclusions hold.

an indicator variable, for whether the fund revised returns in the previous vintage. We find that the coefficient on this indicator variable is highly significant, revealing that some funds are regular revisers of their returns.

3.5.2 Determinants of the size and direction of revisions

We now turn to explaining the size and direction of revisions. As a first step, we take all 6,906 reviser funds and construct a portfolio using their reported returns, and report these returns using two different sets of data, namely the very first vintage of returns for each fund, and the final 'last' vintage available for these funds,[51] once the impact of all revisions has been incorporated. We plot the returns on this portfolio in Figure 3.1. While the first vintage appears in July 2007, revisions occur across the entire possible range of return history from 1994 to 2011, hence this figure plots these two alternative reported histories.

The figure shows clearly that the cumulative difference between final and initial returns has a significant negative trend. What a prospective investor infers about fund performance depends on when he or she sees it, apparently, and (especially in periods of stress, as we shall see later) this last performance is significantly lower than initially reported performance. This suggests the danger of prospective investors being wooed into making decisions based on initially reported histories which are then subsequently revised.

While it is tempting to infer a great deal from this plot, it is certainly consistent with multiple possibilities. The first is dishonesty – that is, performance is reported to be higher than actual in order to increase commitments to funds, and subsequently revised back once many years have elapsed. Another possibility is that valuation errors of both types might occur, but fund managers might have greater incentives to correct them downwards rather than upwards. That is, acknowledging overestimation of past returns might allow managers to push historical high-water marks down, thus allowing the earlier collection of incentive fees — an interpretation that gains support

[51]'Last' in the sense of last observed. The fund might still be alive, and, continue reporting in the future.

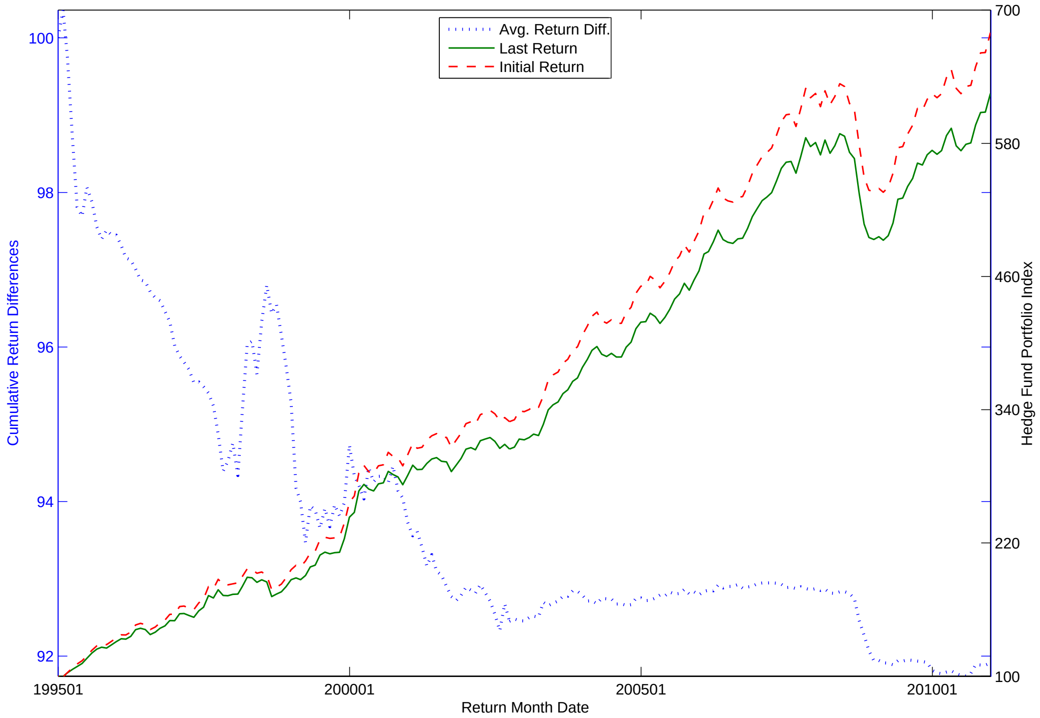


Figure 3.1: *Cumulative Differences between Last and Initial Returns*
The figure shows the cumulative average return differences between the last expression of the return at the most recent available vintage (denoted 'Last') and the first time the return is expressed in a database (denoted Initial) for reviser funds. The picture shows the performance histories that would have been seen initially, versus that seen once the impact of all revisions has been taken into account. The index is based to 100 at the time of the second year of the return data, 31 December 1994.

from the higher propensity of funds with high-water mark provisions to revise. Conversely, acknowledging underestimation of past returns requires payments to investors (without even accounting for high-water marks), hence there may be few incentives to do so.[52] A third is that changes in management or auditors cause re-evaluations of accounting techniques and past reported performance figures, generating significant revisions to previously optimistic assessments in the future. Fourth, fee revisions may cause a chain of NAV re-valuations with consequences for older performance numbers, a possibility for which we attempt to control a little later in the chapter. Finally, illiquidity and the consequent possibility of original estimates being revised upon finally realised valuations is also a possibility. However, it is important to keep in mind that the revisions pertain to periods many years in the past — in some cases, up to 15

[52] We thank Istvan Nagy for suggesting this explanation.

years, making it harder to explain all revisions as consequences of later marking to market, and even if the illiquidity explanation is the proximate cause, there is clearly a significant positive bias in initial estimates. In our predictive analysis of revisions, we explore the possibility of many of these explanations for revisions.

Our next step, as described in the methodology section, is to construct calendar-year returns for any fund/year that contained at least one revised return using both initial and finally reported data, and explain the difference between the two, i.e., final less initial, using a number of variables. Panel A of Table 3.6, which analyses the absolute value of these differences, shows that return revisions are on average large. Moreover, these revisions are larger in absolute value during crises, with all three of the crisis dummy variables having significantly positive coefficients. Of these, the very largest revisions pertain to the 1998 to 1999 crisis period, adding 1.75% to the already large baseline revision. This is followed by the 2000 to 2001 NASDAQ crisis period with roughly 80 bp per annum, and the most recent crisis, with 60 bp per annum.

Turning to the fund characteristics, it appears that offshore funds have larger absolute revisions, in line with our conjecture, that potentially weaker enforcement in such jurisdictions might lead to more important revisions. Perhaps surprisingly, funds with audit information appear to be associated with revisions that are larger in absolute value, suggesting that at least some revisions might be occasioned by the enhanced scrutiny generated by recent audits or the appointment of a new auditor. In keeping with this result, Jylha (2011) finds that funds with prominent auditors have more misreporting discontinuities, although Liang (2003) finds no such evidence in his earlier study of the auditing of TASS returns. Finally, the table shows that smaller funds, and those with high incentive fees have larger revisions, which is consistent with greater incentives for dishonesty, as well as with the possibility of larger revaluations when fee-structures change.

Panel B of Table 3.6 explains return differences, rather than their absolute values, and finds that during crisis periods, in particular the 2000–01 and 2008–09 periods, revisions are significantly negative, meaning that the initially reported return tends to be revised downwards in subsequent vintages of the database, as seen earlier. The table also shows that large funds with high management fees tend to make upward revisions.

We now turn to evaluating the predictive content of revisions, constructing portfolios of revisers and non-revisers as successive vintages reveal their identities.

3.5.3 The future performance of revisers and non-revisers

Figure 3.2 plots the cumulative performance of the reviser and non-reviser portfolios constructed as described in Section 3.4.3. Panel A shows that the returns of the revisers are appreciably lower than those of non-revisers. This difference is economically substantial with a cumulative difference of 11.2% emerging after just over three years.[53] This substantial return difference between the two portfolios, at first glance suggests that our classification of funds into revisers and non-revisers has substantial predictive content. However, in order to better understand these differences, and to ensure that they are not simply driven by differences in the risk-loadings or characteristics of funds, we need to risk-adjust (and potentially characteristic-adjust) these returns.

[53] Note that even in the early periods of the out-of-sample period, we still have a substantial number of firms in the “reviser” portfolio, growing from 274 revising firms detected in the first month.

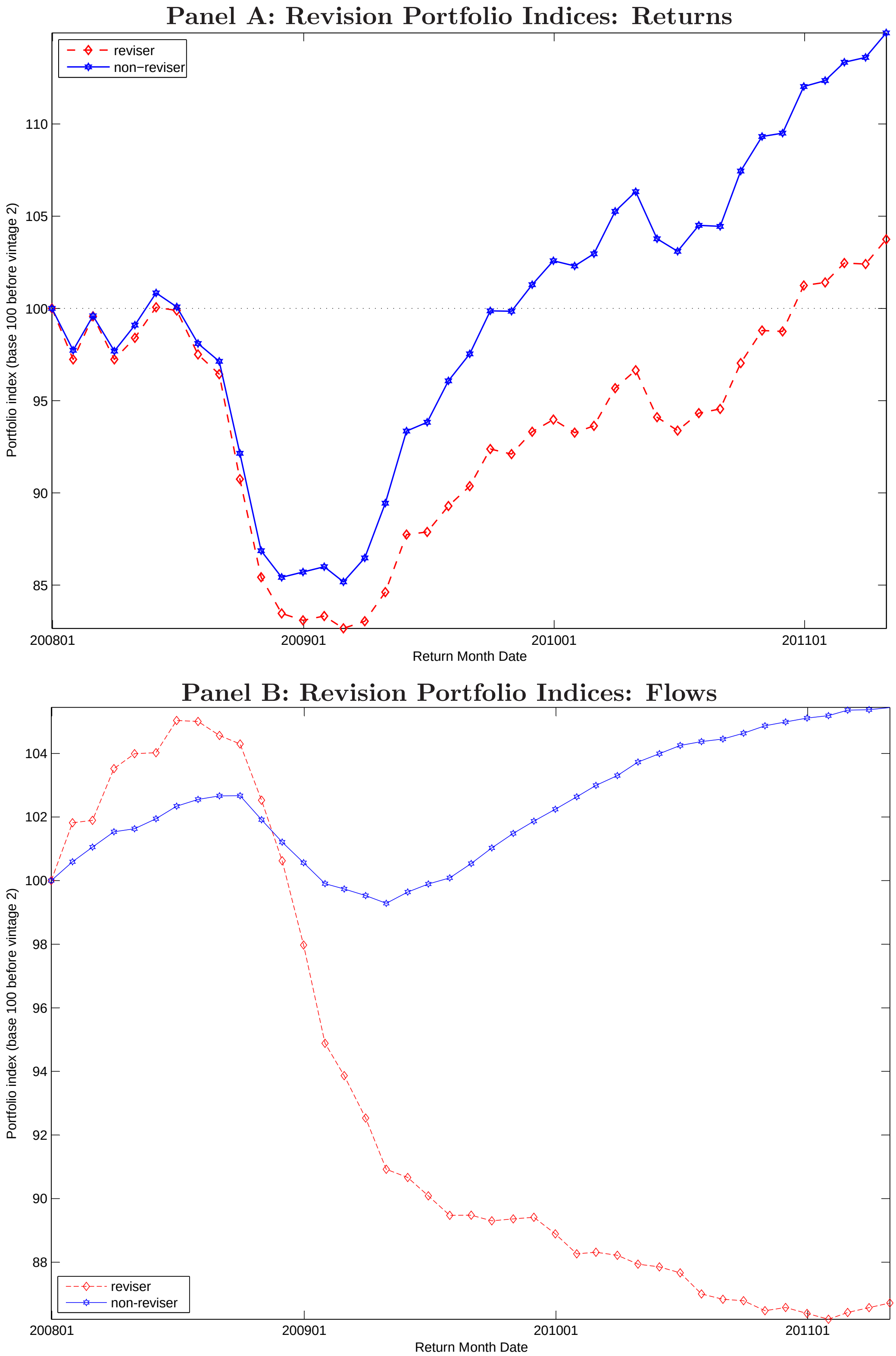


Figure 3.2: *Portfolio Performance — Revisers and Non-revisers*
Figures show the cumulative performance of reviser and non-reviser portfolios. The non-reviser portfolio holds performance of funds that never revise between vintages, plus the early records of funds before they become revisers. The index is based to 100 at 31 December 2007, just before the second vintage starts. Returns are equally weighted in portfolios. Flow calculations employ average assets reported across all vintages.

Before moving to risk-adjusted returns, Figure 3.2 Panel B plots cumulative flows (as usual, flows are computed as changes in AUM unaccounted for by returns, i.e., $F_{it} = \frac{AUM_{it} - AUM_{it-1}(1+R_{it})}{AUM_{it-1}}$) for both reviser and non-reviser portfolios, using data from the final vintage. The plot shows that the reviser portfolio experiences very significant outflows beginning in August–September 2008, during the Lehman collapse. The impact of big outflows and subsequent fire sales of fund assets, might be one potential reason for the poor performance of the reviser portfolio (see Coval and Stafford (2007), and Jotikasthira, Lundblad, and Ramadorai (2012) for evidence of the importance of this mechanism). The flows might also simply be responding to poor performance, à la de Long, Shleifer, Summers, and Waldmann (1990).

Table 3.7 presents results from a variety of models for risk adjusting the return difference between the reviser and non-reviser portfolios, and shows that the findings are very robust to this choice. The alpha of the non-reviser-reviser difference from the Fung-Hsieh seven-factor model is 0.23% per month, or 2.8% per annum net of all fees and costs. We plot cumulative alpha (i.e., $\alpha + \varepsilon_t$ for each time-series portfolio regression) estimated using the Fung-Hsieh seven-factor model in Figure 3.3, and find that it resembles the plot of raw returns: the non-revisers consistently outperform the revisers. We also consider risk adjustment using the Fama-French three factor model, as well as augmented variants that include momentum and liquidity factors, and find that the future poor performance of the "reviser" portfolio is not explained by these alternative models.[54]

Having established that the reviser/non-reviser return differential is not explained by differences in exposure to risk factors, we next consider several possibilities for drivers of this result. One inference is to consider revisions as a sign of dishonesty or poor operational controls within the fund. If either of these were the case, we might also expect to see differences in the tail risk of revisers relative to non-revisers — the dramatic outflows from the reviser portfolio suggest that these differences might

[54]Table F.16 in Appendix F presents results that correspond to Table 3.7, but with funds-of-funds excluded. The risk-adjusted excess performance is smaller for non-FOFs, around 0.17% per month compared with 0.23% per month, however the difference in performance is still strongly statistically significant across all risk adjustment models.

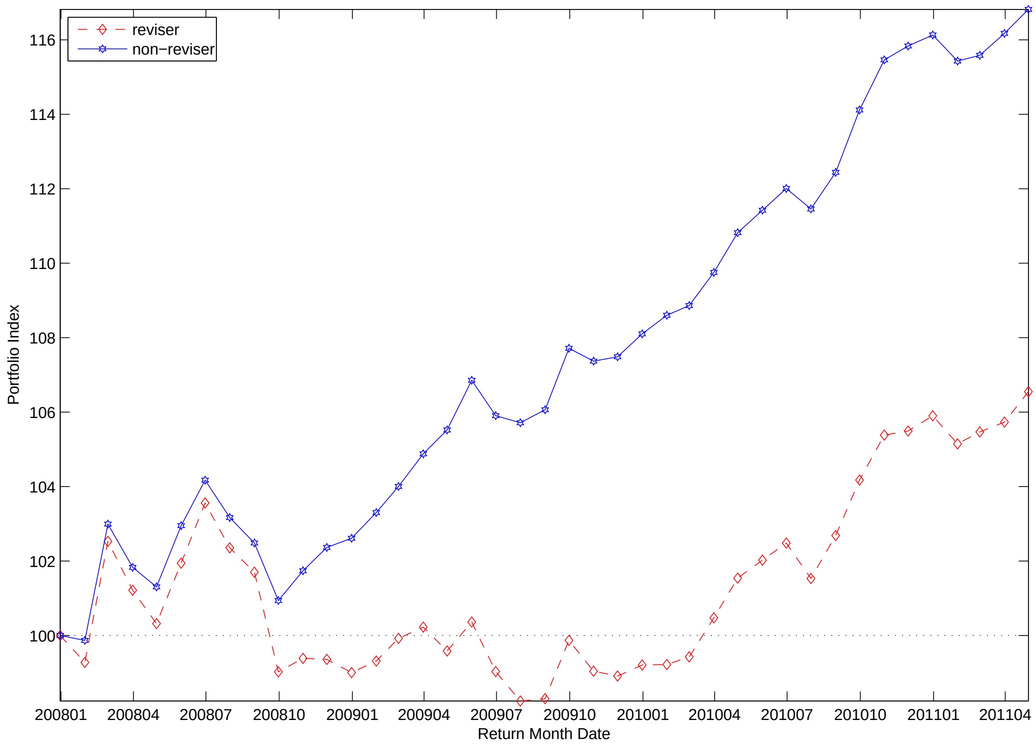


Figure 3.3: *Cumulative Alpha*
The figure plots cumulative $\alpha + \epsilon_t$ using the Fung-Hsieh seven-factor model for the reviser portfolio and non-reviser portfolios. The index is based to 100 at 31 December 2007, just before the second vintage starts.

be stark. To verify this, we employ the historical simulation method, in which we estimate the bottom decile of performance from all returns seen from the beginning of the reviser portfolio up until each date, moving through time (this is done at the individual fund level within each of the portfolios). We also average the returns falling below these empirically computed decile thresholds to arrive at an expected shortfall measure.

Figure 3.4 plots these measures for the cross section of underlying funds of the respective portfolios. We caution here that we have a relatively small sample of data, implying that our estimates of tail quantities are somewhat imprecise, and these plots should be taken as suggestive rather than definitive. Nevertheless, the figures show that the empirical bottom decile and the expected shortfall of the reviser portfolio is virtually always below the non-reviser portfolio over the entire period for both portfolio and cross-sectional measures. There is a dramatic divergence during the crisis with the empirical percentile and the expected shortfall collapsing in the months of October and November 2008. While the tail risk of the revisers at the fund level recovers and seems quite similar to that of the non-revisers in the more recent periods, this could be attributed to the weakest funds having been eliminated from the portfolio during the crisis period. Overall, it appears from this analysis that investors are at greater downside risk when investing in funds that revise their returns. We also checked the results using lower percentile thresholds, and the conclusions are similar.

[*Rest of page intentionally blank as Figure 3.4 is on the next page.*]

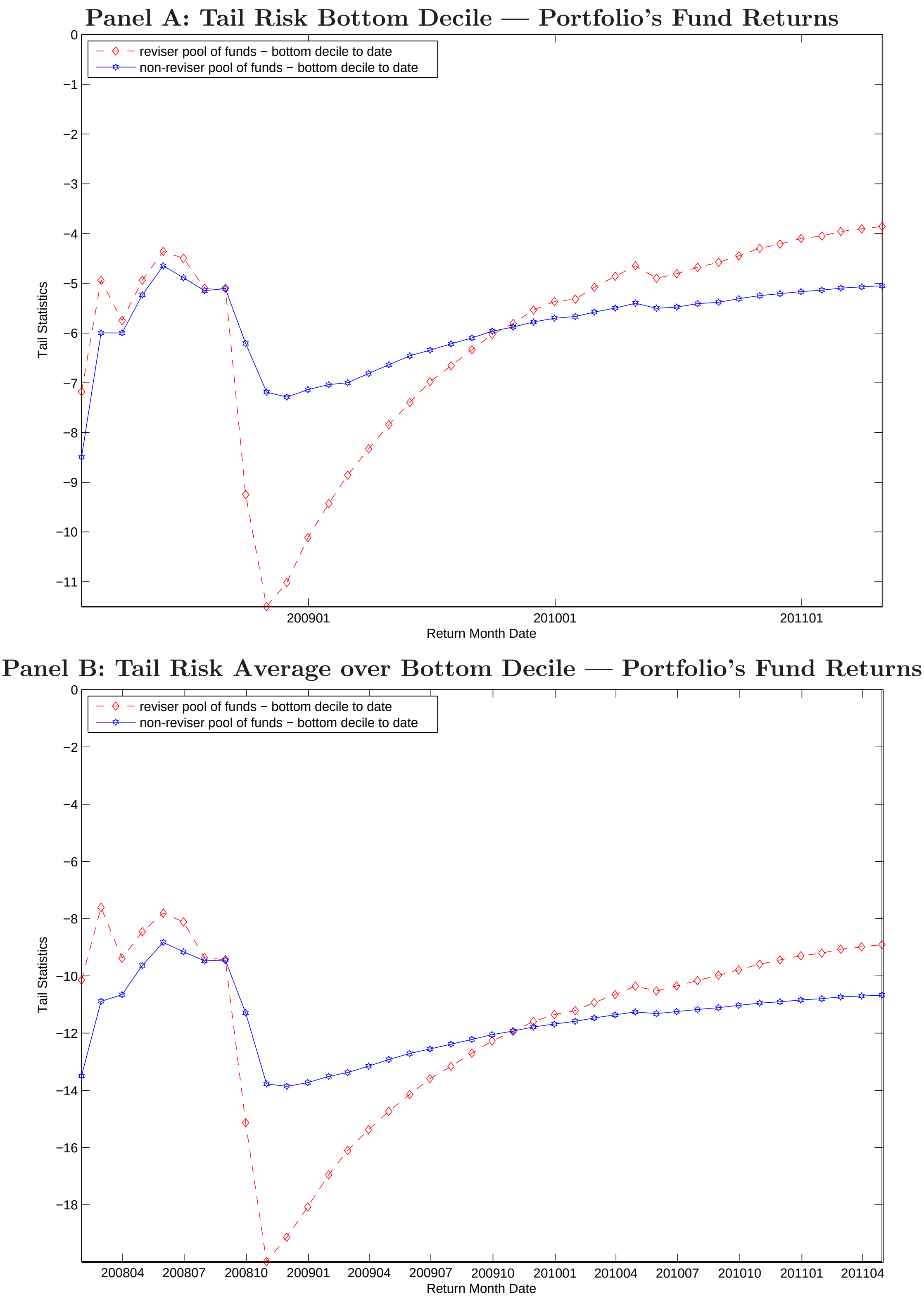


Figure 3.4: *Tail Risk Percentiles for Reviser and Non-reviser Portfolios*
The figures show the bottom decile tail statistics for the reviser portfolio and non-reviser portfolio. Panel A shows the empirical bottom decile for the portfolio fund returns, using historical simulation. Panel B shows the average return of those portfolio fund returns in this bottom decile, as a measure of expected shortfall.

The recovery of the tail risk, in the reviser portfolio towards the end of the sample period that we consider, does suggest that these funds might hold more illiquid assets in their portfolios, which simultaneously drives revisions, sharp falls in asset values, and subsequent recoveries. In this sense, we might simply be picking up differences in asset holdings. The next section explores this and other potential determinants of our findings.

3.6 Robustness checks

In this section we present the results of a battery of robustness checks of our main empirical findings. We vary several of the parameters in our analysis, double-sort the funds by various characteristics as well as by revisers/non-revisers, consider the impact of extreme returns, exclude funds of funds, and examine results for individual hedge fund databases separately. Appendix F presents additional robustness checks and analyses.

3.6.1 Varying the minimum size of the revision

The first parameter that we vary, is the minimum size of a change for it to be labelled a "revision." This is one way to control for the possibility, that our results might be driven by the initial marking to market of illiquid assets. It also allows us to see if we can obtain stronger predictability signals by conditioning on larger revisions. Our main analysis uses a 1 basis point threshold for identifying revisions. We increase this threshold to 10, 50, and 100 basis points as alternatives, in each case only classifying as revisions changes in returns across successive vintages, that are greater than the threshold.

Panel A of Table 3.8 reveals that the return differences reported in Table 3.7 persist, with the estimated monthly alphas across these thresholds ranging from 0.23% to 0.26%. Indeed, our results appear slightly stronger when we only consider funds with larger revisions in our set of revisers.

3.6.2 Varying the minimum age of the revision

Our next robustness check is to give a "free pass" to revisions that occur close to the vintage date. We define the "recency", k, of a revision as the number of months between the date of the return date and the date of the vintage in which the revision was observed. For example, if the return for the month of January 2008 was revised between the December 2008 and January 2009 vintages of data, then this revision would have $k = 12$ months. The parameter k is useful for evaluating various different hypotheses. By setting k to be large, we can evaluate only those funds that revise "ancient history." Moreover, using a large k eliminates the incorporation of funds into the reviser portfolio that relatively quickly revised returns. In other words, we can give a free pass to such small k revisers, to allow for the possibility that funds might employ estimated returns for recent time periods, which could be revised on account of accounting procedures, or because of the re-valuation of illiquid securities in light of more accurate information. The larger we set k, the less likely that we are picking up such revaluation revisions. In our baseline results, we set $k \geq 1$, i.e., we include all revisions. We employ $k > 3,\ 6$, and 12, to identify revisions older than one quarter, six months, and one year.

Panel B of Table 3.8, shows that our results become slightly *stronger* as k increases, peaking at $k > 6$, and descending slightly for $k > 12$, but still higher than unrestricted $k \geq 1$. It is worth noting here that we take additional care with two cases: First, for each k, we ensure that funds revising returns more recent than the threshold k are *not* included in the *non-reviser* portfolio — that is, they do not appear in our plots — to ensure that we compare "true" non-revisers with high-k revisers. Second, in any given vintage, we do not include funds in both reviser and non-reviser portfolios if they simultaneously conduct low- *and* high-k revisions.[55] This is to allow for the possibility of a benign AUM or valuation error found months ago that could, in some cases, cause a cascade of revisions. For example, an incorrectly processed share corporate event could trigger off such a case. Despite these exclusions, high-k revisions are associated

[55]Of course, if they only conducted a high-k revision in a subsequent vintage, they would then be included in the reviser portfolio.

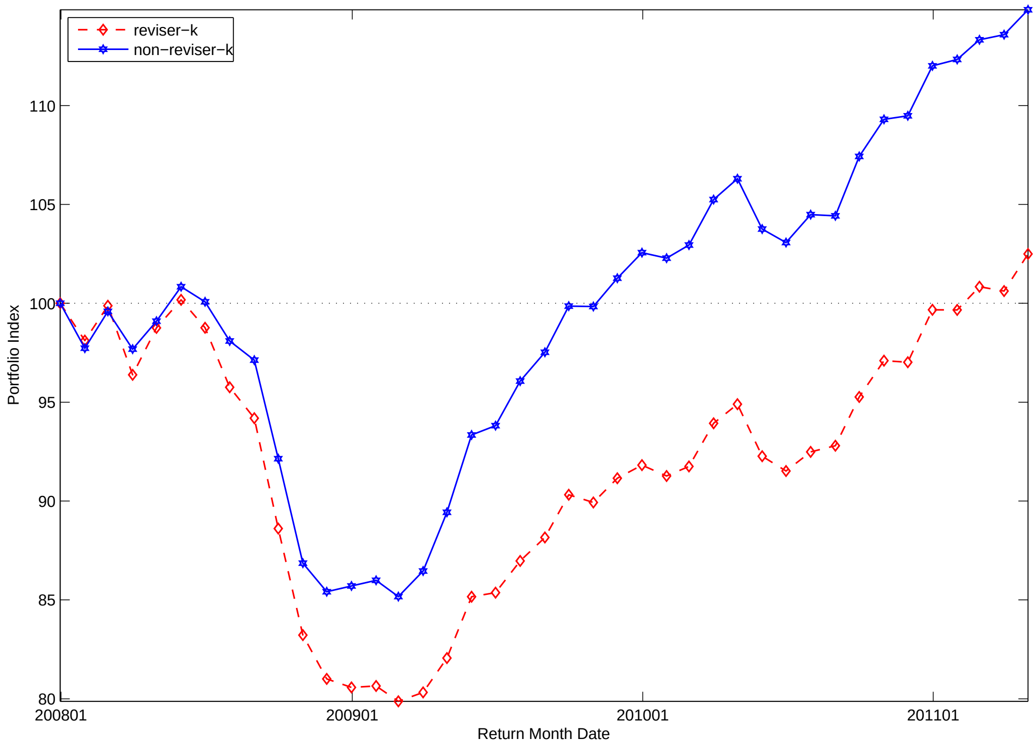


Figure 3.5: *Portfolio Performance — Conditioning on Recency*
The figure shows the cumulative performance of the reviser and non-reviser portfolios. The index is based to 100 at 31 December 2007, just before the second vintage starts. The plot excludes recent revisions near the vintage date. That is, at each date, only funds which revise returns over a year old are included in the construction of the reviser portfolio.

with significant return differentials between revisers and non-revisers — Figure 3.5 plots the return difference between revisers and non-revisers for $k > 12$.

3.6.3 Two-way sorts on fund characteristics

In our earlier probit analysis, we found that reviser and non-reviser funds have different characteristics.[56] While the factor loadings of the return difference between these groups should capture such differences, we perform an additional test to check that our results are not driven by such characteristic-based differences. To do so, we double-sort by these characteristics and the reviser/non-reviser classification. We consider five such fund characteristics, three of which have been identified in the liter-

[56]Table F.13 of Appendix F presents a formal comparison of some key characteristics of reviser and non-reviser funds. This sub-section presents two-way sorts for all variables that are found to be significantly different across revisers and non-revisers.

ature as relevant for expected returns, namely, the first autocorrelation of fund returns (a measure of the smoothness of the fund's returns à la Getmansky et al. (2004), the total lockup period imposed by the fund (see Aragon, 2007) and the size of the fund, to control for the impact of capacity constraints (see Fung et al., 2008). Additionally, we double-sort by the fund's total return volatility and the history length (a measure of age) of the fund, as these were found (see Table F.13) to be significantly different across reviser and non-reviser funds.

Given the nature of the fund characteristics that we employ for these double sorts, this analysis also allows us to investigate whether fund asset illiquidity (correlated with both the GLM measure, and lockup periods, according to the extant literature) helps explain the reviser-non-reviser difference. Specifically, if this were the case, we would expect to see no differences between revisers and non-revisers within each portfolio of funds (independently) double-sorted by illiquidity proxies (autocorrelation, lockup, fund size), but pronounced differences across these illiquidity-sorted groups. If, however, we continue to see variation in reviser and non-reviser portfolio returns within these groups, this would suggest that the revisions provide orthogonal information to underlying asset illiquidity.[57]

The alphas of the return differences between reviser and non-reviser funds of these double-sorted portfolios are reported in Table 3.9, and are all statistically significant, at varying degrees. There are several interesting features in this table. First, we find that reviser-non-reviser differences are particularly stark among funds that have high return autocorrelation. Getmansky et al. (2004), for example, highlight that their measure of return smoothness could be either on account of true asset illiquidity or deliberate return-smoothing among funds. Our result that smooth-return-revisers have worse performance than smooth-return-non-revisers, suggests that our measure might allow investors to discriminate between these two possibilities for observed return smoothness. Second, small funds, young funds, and funds with volatile returns

[57] Of course if these proxies for illiquidity are not as good a measure of underlying asset illiquidity as our revisions measure, it is possible that the explanation might still apply. In that case, the interpretation is that we have found a better measure of asset illiquidity — although the other robustness checks (especially varying k) militate against this explanation.

all show stark differences between reviser and non-reviser portfolio returns. This suggests that when revising behaviour is detected in funds with relatively higher incentives to establish their reputations, it might well be construed as a particularly negative signal about their future return prospects.

3.6.4 Controlling for extreme returns

One might worry, that the poor future performance of hedge funds that have revised their returns is attributable to a few extreme returns. In this section, we consider the reviser/non-reviser performance differential using the *median* return for each of these groups rather than the mean return. Of course, the median return cannot be interpreted as the return on a portfolio of hedge funds, unlike the mean return, but it does allow us to investigate the sensitivity of our results to rare, large returns.

Table 3.10 presents results that correspond to Table 3.8, using the same set of models to estimate risk adjusted returns. We find that the risk-adjusted median return is slightly smaller than the risk-adjusted mean return (around 0.21% per month compared with 0.23%), but is strongly significant across all risk adjustment models. Thus the negative future performance of revising funds is *not* attributable to the extreme poor performance of a few revising funds or, conversely, to the extreme high performance of a few non-revising funds.

3.6.5 Excluding funds-of-funds

The returns reported by funds of hedge funds (FOFs) are of course a function of the returns earned by the individual hedge funds in which the FOF is invested. If an individual fund revises past returns then, unless it is offset by a revision in the opposite direction by another hedge fund, the FOF will have to revise its past returns. This leads to worries of double counting, and to whether our results are robust to the removal of FOFs from the analysis.

Tables F.14, F.15 and F.16 in Appendix F replicate the results presented in Tables 3.4, 3.5, and 3.7. The former two tables refer to the results from probit regressions on the types of funds that revise their returns, and are largely unchanged following the exclusion of FOFs. The latter table presents results on the future performance differential between revisers and non-revisers. We find that the risk-adjusted average return on the difference portfolio is slightly lower when FOFs are excluded (0.17% per month compared with 0.23%), but it remains strongly significant across all risk adjustment models. Thus revising returns remains a significant predictor of poor future performance for both individual funds and funds of hedge funds.

3.6.6 Empirical results for single databases

In addition to tracking vintages of hedge fund databases over the period July 2007 to May 2011, this project also involves the consolidation of the four largest hedge fund databases (TASS, HFR, BarclayHedge, and CISDM). Part of this consolidation process, described in detail in Appendix A, involves the identification of funds that appear in more than one database. To avoid labeling as a "revision" a return that differs across two databases, we associate each fund with a *single* database (choosing the database with the longest history for that fund, if more than one database is available). Nevertheless, to address any concerns that the revisions we detect are owing to the computationally-intensive tasks associated with merging and tracking vintages of multiple hedge fund databases, we also present results separately using just a single database at a time.

Tables F.17 and F.18 in Appendix F replicate the probit model results presented in Tables 3.4 and 3.5. We see from these tables that the parameter estimates and significance levels are quite consistent across databases. The only noteworthy difference is that the "prior revision indicator" variable, capturing autocorrelation in a fund's tendency to revise returns, is significant for TASS, HFR and BarclayHedge, but not for CISDM. This is likely attributable to the fact that the CISDM database is updated less frequently than the other three databases.

In Table F.19 in Appendix F, we present results on the reviser/non-reviser performance differential separately for each database, using the Fung-Hsieh seven-factor model to risk adjust the returns. For the CISDM database we have too few updates in the out-of-sample period, to include it separately in this analysis. The results for the other three databases are in line with the main results: the reviser portfolio underperforms the non-reviser portfolio. The degree of under-performance is weakest in the TASS database (0.16% per month) and greatest in the BarclayHedge database (0.53%) per month, and in all three cases this difference is statistically significant. Thus these results are not driven by our use of a consolidated hedge fund database.

3.7 Conclusions

This chapter examines the reliability of voluntary disclosures of performance information by hedge funds. We do so, by tracking revisions to historical performance records by hedge funds, in several publicly available hedge fund databases. We find evidence that in successive vintages of these databases, older performance records (pertaining to periods as far back as 15 years) of hedge funds are routinely revised. These revisions are widespread, with nearly 40% of the 18,382 hedge funds in our sample (managing around 45% of average total assets) having revised their historical returns at least once. These revisions are not merely random reporting errors: they are partly predictable using information on the characteristics and past performance of hedge funds, with larger, more volatile, and less liquid funds more likely to revise their returns. Initially reported performance track records present a far rosier picture of historical performance, than track records that include all changes made in subsequent data vintages. Perhaps most interestingly, detecting that a fund has revised one of its past returns, helps us to predict that it will subsequently underperform funds that have never revised their returns.

Recent policy debates on the pros and cons of imposing stricter reporting requirements on hedge funds have raised various arguments. The benefits of disclosures include market regulators having a better view on the systemic risks in financial markets, and investors and regulators being able to better determine the true, risk-adjusted, performance of the fund. The costs include the administrative burden of preparing such reports, and the risk of leakage of valuable proprietary information, in the form of trading strategies and portfolio holdings. This chapter's analysis suggests that mandatory, audited disclosures by hedge funds, such as those proposed by the SEC last year and due to be implemented in 2012, would be beneficial to regulators. We believe that it would also be worth considering how these reporting guidelines, which currently only apply to funds' disclosures to regulators, could also apply to disclosures to prospective and current investors, so as to help them make more informed investment decisions.

3.8 Appendix: Tables

Table 3.1: *Summary Statistics, Overall Dataset*

This table shows summary statistics on funds that we employ in our analysis, with time-series statistics in Panel A computed only using the May 2011 (final) vintage of the 40 vintages of data that we capture. AUM refers to assets under management. Panel A shows broad statistics on returns and AUM, Panel B shows the strategies into which the funds are classified, and Panel C shows the databases from which the funds are sourced.

Panel A: Fund Summary Statistics

	Num. Funds	Average Fund AUM U.S.$ MM	Average Fund Return	Average Fund History Length (years)
	18,382	104.19 m	0.640	5.535

Panel B: Fund Strategies

	Fund Count	Count%
Security Selection	3,009	16.37%
Macro	1,201	6.53%
Relative Value	250	1.36%
Directional Traders	2,358	12.83%
Funds-of-Funds	4,846	26.36%
Multi-Process	1,877	10.21%
Emerging	821	4.47%
Fixed Income	957	5.21%
Other	174	0.95%
Managed Futures	2,889	15.72%
	18,382	100.00%

Panel C: Funds by Database

	Fund Count	Count%
TASS	6,604	35.93%
HFR	4,742	25.80%
CISDM	1,698	9.24%
BarclayHedge	5,338	29.04%
	18,382	100.00%

Table 3.2: *Summary Statistics of Return Changes across Vintages*

This table shows summary statistics of changes in returns between successive vintages. If $Ret_{i,t,v}$ is the return for a fund i, for month t, at vintage v, then Deletion (Del) means $Ret_{i,t,v-1}$ was available but $Ret_{i,t,v}$ is not available. Addition (Add) means that $Ret_{i,t,v-1}$ was missing but $Ret_{i,t,v}$ is available. (Add excludes fund launches; the first time a return appears for a fund; and additions within 12 months of the vintage $v-1$ date so as to avoid picking up late reporting.) Revision (Rev) means that $Ret_{i,t,v-1}$ and $Ret_{i,t,v}$ are both available but are not equal to one another. (Rev excludes revisions that are less than 1 bp in absolute value to avoid picking up spurious changes in significant digits in reporting e.g. from 2 to 4 decimal places.) Any Change means the fund experienced at least one of the change types (Del, Add, Rev) in the period of analysis. Panel A shows counts of these types of changes, Panel B shows the proportion of revising funds with at least one revision that is greater than or equal to (in absolute value) various size thresholds, Panel C shows the proportion of revising funds after excluding revisions near the vintage date (for example, when $k > 6$ only funds with revisions that occur six months prior to the date of the vintage are considered revisers), Panel D shows various percentiles of revisions, their absolute value, and separately for positive and negative revisions, and finally, Panel E explores potential reasons for innocuous revisions: sign changes are revisions where the return is identical across vintages, except for their sign, decimal errors are revisions that differ between vintages by 1 or 2 decimal places (e.g., 1.75 and 0.175), and digit transpositions identify revisions which are re-orderings of adjacent digits (e.g., 1.75 and 1.57).

Panel A: Changes Breakdown at Fund Level

	Fund Count	Any Change Count	Deletions Count	Additions Count	Revisions Count
Funds	18,382	7,421	1,078	370	6,906
% of Total		40.4%	5.9%	2.0%	37.6%

Panel B: Size of Revisions

	Fund Count	Revisions Count			
		at least 0.01%	at least 0.1%	at least 0.5%	at least 1%
Funds	18,382	6,906	5,803	3,972	2,973
% of Total		37.6%	31.6%	21.6%	16.2%

Panel C: Summary Statistics for the Distribution of Revisions

	Revisions	Absolute Revisions	Positive Revisions	Negative Revisions
Count	87,504	87,504	42,815	44,689
Mean	-0.044	0.815	0.788	-0.841
Median	-0.020	0.130	0.130	-0.134
99th perc	6.449	10.700	10.454	-0.014
95th perc	1.585	3.386	3.240	-0.020
75th perc	0.128	0.480	0.470	-0.050
25th perc	-0.140	0.050	0.050	-0.486
5th perc	-1.770	0.020	0.020	-3.500
1st perc	-7.190	0.013	0.013	-10.942

Panel D: Recency of Revisions					
		Minimum Recency of Revisions Count			
	Fund Count	**1 or more months**	**more than 3 months**	**more than 6 months**	**more than 12 months**
Funds	18,382	6,906	5,461	4,355	3,436
% of Total	100.0%	37.6%	29.7%	23.7%	18.7%
Revisions		87,504	63,834	51,469	43,233
% of Total		100.00%	72.95%	58.82%	49.41%

Panel E: Statistics for Potential Types of Revisions					
	Reviser Count	**Sign Change**	**Decimal Place**	**Digit Transposition**	**Sign *or* Decimal *or* Transpose**
Funds	6,906	246	74	340	604
% of Total	37.6%	1.34%	0.40%	1.85%	3.29%
Revisions	87,504	280	424	389	1,093
% of Total	100%	0.32%	0.48%	0.44%	1.25%

Table 3.3: *Summary Statistics of Revisions by Strategy*

This table shows the percentage of funds in each strategy with absolute value revisions of at least 1 bp, 10 bp, 50 bp, or 100 bp. For example, of the 3,009 Security Selection funds, 28.7% have past history which is revised by at least 1 bp, 22.9% by at least 10 bp, 16% by at least 50 bp, and 12.1% by at least 1%.

		Revisions as % of Funds in Strategy			
Strategy	**Fund Count**	**at least 0.01%**	**at least 0.1%**	**at least 0.5%**	**at least 1%**
Security Selection	3,009	28.7%	22.9%	16.0%	12.1%
Macro	1,201	30.4%	24.5%	15.4%	11.2%
Relative Value	250	43.6%	32.8%	20.8%	15.6%
Directional Traders	2,358	31.5%	24.4%	16.5%	12.4%
Funds-of-Funds	4,846	52.5%	47.4%	33.5%	25.0%
Multi-Process	1,877	37.8%	31.4%	19.8%	15.0%
Emerging	821	42.9%	35.6%	27.4%	22.7%
Fixed Income	957	34.6%	27.6%	18.2%	12.9%
Other	174	40.8%	35.1%	27.0%	20.7%
Managed Futures	2,889	28.4%	22.8%	14.6%	10.5%
All Funds	18,382	37.6%	31.6%	21.6%	16.2%

Table 3.4: *Probit Regression for Revisions*

The table shows the marginal effects from a probit regression. The dependent variable takes the value of 1 if a fund had revised data over any of the 40 vintages that we capture, and 0 otherwise. The independent variables are lifetime average returns, lifetime average AUM, standard deviation of returns, and the autocorrelation of returns, all measured as ranks relative to the other funds in the data; and the number of return observations in the return history of the fund. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located Offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods) and a flag which takes the value of 1 if there is any information pertaining to audits available in any of the databases. We also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Mean**	**Z-stat**	
Lifetime Avg. AUM (Rank)	0.238	0.500	3.99	***
Lifetime Avg. Return (Rank)	-0.084	0.500	-1.92	*
Lifetime Ret. Std. (Rank)	0.077	0.500	1.93	*
Return Autocorrelation (Rank)	0.116	0.500	7.75	***
Return History Length	0.024	5.535	5.55	***
Offshore	-0.043	0.501	-21.44	***
Total Restrictions	0.009	1.829	2.26	**
Audit	0.137	0.712	1.48	
High-Water Mark or Hurdle	0.185	0.604	5.66	***
Database Fixed Effects				
HFR	-0.064	0.258	-3.75	***
CISDM	-0.039	0.092	-0.52	
BarclayHedge	0.091	0.290	4.75	***
Strategy Fixed Effects				
Macro	0.087	0.065	25.5	***
Relative Value	0.180	0.014	3.18	***
Directional Traders	0.007	0.128	0.58	
Funds-of-Funds	0.252	0.264	17.66	***
Multi-Process	0.098	0.102	4.62	***
Emerging	0.117	0.045	20.32	***
Fixed Income	0.038	0.052	0.90	
Other	0.138	0.009	1.47	
Managed Futures	0.179	0.157	5.05	***
N	18,382			
Pseudo R^2	0.135			

Table 3.5: *Probit Regression for Revisions at Vintage Level*

This table runs essentially the same specification as in Table 3.4, the difference is that we employ the panel structure of the data, and the fund-vintage is now our unit of analysis. The dependent variable takes the value of 1 if a fund revised data between the last available vintage $v-1$ and the current vintage v. The ranks of the lifetime variables are therefore now measured using data in vintage $v-1$ on assets under management, and returns. We also add an independent variable that takes the value of 1 if the fund experienced a data revision in the prior vintage, and 0 otherwise. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods) and a flag which takes the value of 1 for the fund if there is any information pertaining to audits available in any of the databases. We also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by vintage. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Mean**	**Z-stat**	
Lifetime Avg. AUM (Rank) (v-1)	0.030	0.500	15.04	***
Lifetime Avg. Return (Rank) (v-1)	0.013	0.500	3.76	***
Lifetime Ret. Std. (Rank) (v-1)	0.002	0.500	0.81	
Return Autocorrelation (Rank) (v-1)	0.008	0.500	3.52	***
Return History Length (v-1)	0.001	5.310	5.63	***
Prior Vintage Revision Indicator	0.227	0.068	16.61	***
Offshore	-0.004	0.502	-3.57	***
Total Restrictions	0.001	1.902	4.86	***
Audit	0.019	0.691	8.61	***
High-Water Mark or Hurdle	0.014	0.606	9.39	***
Database Fixed Effects				
HFR	-0.001	0.256	-0.350	
CISDM	-0.024	0.098	-1.900	*
BarclayHedge	0.014	0.285	2.320	**
Strategy Fixed Effects				
Macro	0.015	0.065	8.17	***
Relative Value	0.014	0.013	5.57	***
Directional Traders	-0.001	0.128	-1.12	
Funds-of-Funds	0.039	0.262	20.56	***
Multi-Process	0.009	0.093	4.43	***
Emerging	0.010	0.043	5.61	***
Fixed Income	0.005	0.051	3.99	***
Other	0.015	0.009	6.73	***
Managed Futures	0.027	0.160	13.69	***
N	560,428			
Pseudo R^2	0.219			

Table 3.6: *Explaining Revision Return Differences*

This table conditions the return differences occasioned by revisions on various fund characteristics and period fixed effects. The dependent variable is the average difference, for all years in which a fund experienced return revisions, between the final set of annual returns provided by a fund and the first set of annual returns provided by the same fund for the same year. For example, if fund Y initially reported 4% average annual return for year t, and at the final vintage, this average stood at 6%, then the return difference variable would be 2%. We only include periods in which the fund had at least 6 months of return observations, to reduce the noise in the dependent variable. Panel A takes the absolute value of all such differences as the dependent variable, and Panel B conditions the signed revisions on the independent variables. Period dummies include crisis dummies for the 1998-1999 period, the 2000-2001 period, and the 2008-2009 period. The remaining regressors have been described earlier in these tables, with three new additions, namely the rank of prior flows and returns experienced by the fund relative to all other funds in the same year; the Management fee and the Incentive fee of the fund. t-statistics, shown in parentheses, are robust to heteroskedasticity and clustered at the fund-level. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

Panel A: Absolute Value of Differences						
	Coef.	***t*-stat**		**Coef.**	***t*-stat**	
Constant	1.099	-24.976	***	1.153	-6.167	***
Crisis dummy1: 1998–99	1.75	-3.232	***	1.753	-3.276	***
Crisis dummy2: 2000–01	0.784	-2.601	***	0.765	-2.551	**
Crisis dummy3: 2008–09	0.579	-9.188	***	0.574	-9.094	***
Offshore				0.233	-2.732	***
Total Restrictions				-0.011	(-0.818)	
High-Water Mark or Hurdle				-0.158	(-1.584)	
Audit				0.278	-2.309	**
Management Fee				0.008	-0.113	
Incentive Fee				0.02	-3.44	***
Asset t-1 rank				-0.904	(-5.853)	***
Return prior year t-1 rank				-0.153	(-1.220)	
Flow prior year t-1 rank				-0.08	(-0.753)	
N	10,004			10,004		
Adjusted R^2	0.012			0.023		

Panel B: Return Differences						
	Coef.	***t*-stat**		**Coef.**	***t*-stat**	
Constant	-0.005	(-0.126)		-0.172	(-1.013)	
Crisis dummy1: 1998–99	-0.39	(-0.616)		-0.41	(-0.647)	
Crisis dummy2: 2000–01	-0.783	(-2.475)	**	-0.795	(-2.524)	**
Crisis dummy3: 2008–09	-0.348	(-5.250)	***	-0.345	(-5.204)	***
Offshore				-0.096	(-1.360)	
Total Restrictions				0.015	-1.079	
High-Water Mark or Hurdle				-0.1	(-1.098)	
Audit				-0.063	(-0.574)	
Management Fee				0.129	-2.027	**
Incentive Fee				0.002	-0.378	
Asset t-1 rank				0.21	-1.82	*
Return prior year t-1 rank				0.133	-1.023	
Flow prior year t-1 rank				-0.17	(-1.399)	
N	10,004			10,004		
Adjusted R^2	0.003			0.004		

Table 3.7: *Do Revisions Predict Future Returns?*

This table regresses the difference in returns between the reviser and non-reviser portfolios over the 40 months from January 2008 to the end of the sample period, May 2011, on several different sets of factors. Panel A employs subsets, followed by the full set, of factors from the Fung-Hsieh model. Panel B employs the Fama-French 3 factor model, adds a momentum factor, and finally adds the Pastor-Stambaugh Liquidity factor (P-S factors are only available until December 2010). Newey-West heteroskedasticity and autocorrelation robust standard errors (with three lags) are employed to assess statistical significance. Regression betas are shown with t-statistics shown in parentheses beneath coefficients. The significance of the alpha is denoted by stars at the 10% (*), 5% (**) and 1% (***) levels respectively.

Panel A: Return differences (Fung-Hsieh Model)					
Factors	**Constant**	**Market**	**FH 4**	**FH 7**	**FH 8**
Constant	0.256***	0.252***	0.235***	0.229***	0.228***
	(-3.388)	(-4.202)	(-2.993)	(-2.877)	(-2.922)
SP500	-	0.019	0.017	0.018	0.015
	-	(-1.631)	(-1.166)	(-1.422)	(-0.952)
SMB	-	-	0.028	0.025	0.026
	-	-	(-2.025)	(-1.532)	(-1.464)
BOND10YR	-	-	-0.163	-0.288	-0.280
	-	-	(-0.930)	(-1.049)	(-0.993)
CREDSPR	-	-	0.043	-0.026	-0.007
	-	-	(-0.244)	(-0.107)	(-0.026)
PTFSBD	-	-	-	-0.288	-0.288
	-	-	-	(-0.439)	(-0.439)
PTFSFX	-	-	-	0.950	0.944
	-	-	-	(-1.763)	(-1.785)
PTFSCOM	-	-	-	-1.471	-1.457
	-	-	-	(-2.147)	(-2.133)
EMERGING	-	-	-	-	0.003
	-	-	-	-	(-0.200)
N	40	40	40	40	40
Adjusted R^2		0.061	0.027	0.074	0.045

Panel B: Return differences (Fama-French 3 factors + Momentum + Pastor-Stambaugh Liquidity Model)			
Factors	**FF3**	**FF3 + Mom**	**FF3 + Mom + Liquidity**
Constant	0.246***	0.213***	0.244***
	(-3.152)	(-3.963)	(-4.982)
MKTRF	0.755	-0.604	0.241
	(-0.582)	(-0.648)	(-0.304)
SMB	1.186	1.848	2.209
	(-0.722)	(-1.093)	(-1.354)
HML	3.112	0.467	-2.649
	(-2.083)	(-0.385)	(-1.509)
UMD	-	-3.660	-3.420
	-	(-9.312)	(-9.288)
PSLIQ	-	-	-2.339
	-	-	(-2.663)
N	40	40	36
Adjusted R^2	0.152	0.095	0.09

Table 3.8: *Robustness Checks: Size and Recency*

This table conditions the results in Table 3.7 on the size and recency of revisions. Panel A shows the impact of using different size thresholds for considering revisions as important. For example, the first column (1 bp) of Panel A reproduces the results from Panel A of Table 3.7, and '10 bp' only includes funds with revisions which are greater than 10 bp in absolute value in the construction of the reviser portfolio. Panel B shows the impact of excluding recent revisions near the vintage date. For example, when $k > 6$ only funds with revisions that occur six months prior to the date of the vintage are included, and when $k > 12$, only funds which revise returns over a year old are included in the construction of the reviser portfolio. Newey-West heteroskedasticity and autocorrelation robust standard errors (with three lags) are employed to assess statistical significance. Regression betas are shown with t-statistics shown in parentheses beneath coefficients. The significance of the alpha is denoted by stars at the 10% (*), 5% (**) and 1% (***) levels respectively.

Panel A: Significance of Revision (Fung-Hsieh Seven-Factor Model)

	Minimum Significance of Revisions			
Factors	**1 bp**	**10 bp**	**50 bp**	**100 bp**
Constant	0.229***	0.252***	0.241***	0.258***
	(-2.877)	(-3.043)	(-2.768)	(-2.741)
SP500	0.018	0.018	-0.001	-0.004
	(-1.422)	(-1.428)	(-0.086)	(-0.261)
SMB	0.025	0.023	0.011	0.016
	(-1.532)	(-1.26)	(-0.591)	(-0.720)
BOND10YR	-0.288	-0.138	-0.062	0.114
	(-1.049)	(-0.498)	(-0.199)	(-0.297)
CREDSPR	-0.026	0.127	0.367	0.509
	(-0.107)	(-0.482)	(-1.255)	(-1.380)
PTFSBD	-0.288	-0.086	0.175	0.608
	(-0.439)	(-0.137)	(-0.249)	(-0.712)
PTFSFX	0.95	0.95	0.885	1.104
	(-1.763)	(-2.086)	(-1.929)	(-1.546)
PTFSCOM	-1.471	-1.410	-1.266	-1.767
	(-2.147)	(-2.202)	(-1.890)	(-1.627)
N	40	40	40	40
Adjusted R^2	0.074	0.015	0.069	0.152

Panel B: Recency of Revision (Fung-Hsieh Seven-Factor Model)				
	Minimum Recency of Revisions			
Factors	$k = 1$	$k > 3$	$k > 6$	$k > 12$
Constant	0.229***	0.284***	0.301***	0.247***
	(-2.877)	(-3.396)	(-3.459)	(-2.881)
SP500	0.018	0.000	-0.009	-0.009
	(-1.422)	(-0.024)	(-0.673)	(-0.679)
SMB	0.025	0.03	0.041	(0.062)
	(-1.532)	(-2.005)	(-2.276)	(-1.992)
BOND10YR	-0.288	-0.313	-0.268	-0.334
	(-1.049)	(-1.212)	(-1.069)	(-1.257)
CREDSPR	-0.026	-0.023	0.059	(-0.004)
	(-0.107)	(-0.119)	(-0.338)	(-0.024)
PTFSBD	-0.288	-0.232	-0.137	(-0.073)
	(-0.439)	(-0.338)	(-0.189)	(-0.092)
PTFSFX	0.95	0.91	0.856	1.366
	(-1.763)	(-1.637)	(-1.589)	(-1.969)
PTFSCOM	-1.471	-1.017	-1.002	-1.934
	(-2.147)	(-1.509)	(-1.509)	(-2.253)
N	40	40	40	40
Adjusted R^2	0.074	-0.042	0.007	0.084

Table 3.9: *Robustness Checks: Fund Characteristics*

This table conditions the results in Table 3.7 on the cross section of various fund characteristics. We split both revisers and non-revisers by sorting funds on specific characteristics, into groups that are above (Hi) and below (Lo) the cross-sectional median of all funds reporting in each period. These characteristics are ρ_1 (lifetime first return autocorrelation); the lockup period as at the last available vintage; fund size (AUM at the end of the prior period); Return Std. (lifetime return standard deviation); and history length (the number of return observations in the return history of the fund). Returns are equally weighted within portfolios. Newey-West heteroskedasticity and autocorrelation robust standard errors (with three lags) are employed to assess statistical significance. Regression betas are shown with t-statistics shown in parentheses beneath coefficients. The significance of the alpha is denoted by stars at the 10% (*), 5% (**) and 1% (***) levels respectively.

	Characteristic									
	ρ_1		**Lockup**		**Fund Size**		**Return Std.**		**History Length**	
Factors	**Hi**	**Lo**	**Hi**	**Lo**	**Hi**	**Lo**	**Hi**	**Lo**	**Hi**	**Lo**
Constant	0.283***	0.111*	0.331***	0.139*	0.146**	0.337***	0.270**	0.197***	0.108**	0.372***
	(-3.479)	(-1.689)	(-3.646)	(-1.928)	(-2.397)	(-3.335)	(-2.573)	(-3.325)	(-2.073)	(-3.142)
SP500	0.042	0.008	0.07	-0.015	0.027	0.023	-0.046	-0.023	-0.039	-0.01
	(-3.898)	(-0.473)	(-11.591)	(-0.685)	(-2.289)	(-1.230)	(-2.584)	(-2.697)	(-3.205)	(-0.708)
SMB	0.017	0.029	0.025	0.021	0.013	0.021	-0.024	-0.025	-0.017	-0.04
	(-1.043)	(-1.415)	(-1.705)	(-1.126)	(-1.005)	(-0.739)	(-1.283)	(-1.409)	(-1.173)	(-1.716)
BOND10YR	-0.376	-0.463	0.043	-0.569	-0.459	-0.001	-0.387	-0.096	-0.234)	-0.208
	(-1.405)	(-1.778)	-0.151	(-1.727)	(-2.427)	(-0.002)	(-1.133)	(-0.431)	(-1.451)	(-0.490)
CREDSPR	-0.134	-0.438	0.075	-0.436	-0.334	0.276	-0.303	-0.434	-0.191	-0.222
	(-0.563)	(-2.032)	(-0.254)	(-1.921)	(-1.552)	(-0.808)	(-1.066)	(-2.100)	(-1.458)	(-0.544)
PTFSBD	-0.011	-1.291	0.199	-1.398	-0.604	-0.066	-0.937	-0.355	-0.293	-0.249
	(-0.019)	(-1.858)	-0.318	(-1.852)	(-1.382)	(-0.059)	(-1.076)	-0.708	(-0.713)	(-0.263)
PTFSFX	0.954	1.322	1.07	0.769	1.148	1.123	-1.409	-0.26	-0.605	-1.212
	(-1.764)	(-2.487)	(-1.990)	(-1.202)	(-2.762)	(-1.285)	(-1.990)	(-0.613)	(-1.661)	(-1.397)
PTFSCOM	-1.323	-2.449	-0.876	-2.355	-1.38	-2.642	-2.464	-0.154	-1.092	-1.572
	(-1.907)	(-3.373)	(-1.228)	(-2.997)	(-2.717)	(-2.538)	(-2.743)	(-0.242)	(-2.200)	(-1.413)
N	40	40	40	40	40	40	40	40	40	40
Adjusted R^2	0.351	0.284	0.508	0.206	0.352	0.11	0.366	0.468	0.469	-0.015

Table 3.10: *Robustness Checks: Median*

To test for the influence of extreme observations, this table shows the significance of the differences in returns between the Non-reviser and Reviser portfolios using the portfolios median return. The monthly return differences are analysed against different risk models. Panel A uses factors from the Fung-Hsieh model, such as a market model using S&P 500, four of the market related Fung-Hsieh factors, and then the Fung-Hsieh 7 and 8 Factor model. Panel B uses an alternate specification with the Fama-French 3 factor model, and then adds a momentum factor, and finally the Pastor-Stambaugh Liquidity factor. The PS-Liquidity factors are only available to December 2010. Newey-West heteroskedasticity and autocorrelation robust standard errors (with three lags) are used. Regression betas are shown with t-statistics shown in brackets beneath. Alpha significance is denoted by stars at 10% (*), 5% (**) and 1% (***) respectively.

Panel A: Return differences (Fung-Hsieh Model)

Factors	**Constant**	**Market**	**FH 4**	**FH 7**	**FH 8**
Constant	0.207**	0.213***	0.196***	0.200***	0.203***
	(-2.382)	(-3.79)	(-3.318)	(-3.218)	(-3.273)
SP500	-	-0.032	-0.014	-0.012	-0.004
	-	(-2.316)	(-1.349)	(-1.263)	(-0.323)
SMB	-	-	0.019	0.018	0.017
	-	-	(-1.48)	(-1.384)	(-1.161)
BOND10YR	-	-	-0.125	-0.125	-0.145
	-	-	(-0.855)	(-0.542)	(-0.587)
CREDSPR	-	-	0.494	0.431	0.38
	-	-	(-2.605)	(-1.957)	(-1.531)
PTFSBD	-	-	-	0.063	0.062
	-	-	-	(-0.129)	(-0.125)
PTFSFX	-	-	-	0.45	0.466
	-	-	-	(-1.112)	(-1.095)
PTFSCOM	-	-	-	-0.378	-0.414
	-	-	-	(-0.634)	(-0.712)
EMERGING	-	-	-	-	-0.007
	-	-	-	-	(-0.588)
N	40	40	40	40	40
Adjusted R^2		0.21	0.378	0.348	0.334

Panel B: Return differences (Fama-French 3 factors + Momentum + Pastor-Stambaugh Liquidity Model)			
Factors	**FF3**	**FF3 + Mom**	**FF3 + Mom + Liquidity**
Constant	0.224***	0.208***	0.227***
	(-3.557)	(-3.741)	(-4.47)
MKTRF	-4.353	-5.019	-4.216
	(-2.898)	(-3.437)	(-3.544)
SMB	-0.487	-0.163	0.171
	(-0.333)	(-0.108)	-0.129
HML	4.768	3.473	-0.071
	(-2.766)	(-1.906)	(-0.039)
UMD	-	-1.792	-1.628
	-	(-3.479)	(-3.320)
PSLIQ	-	-	-2.809
	-	-	(-2.689)
N	40	40	36
Adjusted R^2	0.107	0.095	0.088

"It's a proprietary strategy. I can't go into it in great detail."
— Bernard Madoff quoted in *Barron's*, 7 May 2001.

4 Hedge Fund Asset Disclosure

In this chapter, I analyse the level and changes in disclosure by hedge funds of their assets under management over the period funds report return information. I find more than a third of funds reporting assets to public hedge fund databases, do so selectively, with asset records falling short of their returns. This shortfall in disclosure is concentrated in funds with lower restrictions and with smoother return profiles. There is evidence of strategic disclosure by funds — with assets reporting drying up after times of fund stress such as poor performance or outflows. Furthermore, investors should take heed of the greater propensity for shortfall funds to trigger fraud performance flags. These results have implications for the way current researchers draw hedge fund samples, and exhort regulators to include assets, and not just returns, in the push for mandatory disclosure by hedge funds.

4.1 Introduction

Assets are the lifeblood of an investment management company. They provide the means for the manager to charge monthly fees to survive, and later thrive with the benefits of scale. Assets enable diversification across investment strategies, improve access for trading and better the terms with service providers. Asset levels signify to investors that the strategy is viable. For those who believe bigger is better, they also provide bragging rights, however, excessive asset levels can undermine prior success.

In this chapter, my focus is voluntary asset disclosure by hedge funds to public databases. The goal is to provide empirical evidence to inform the debate on disclosure around hedge funds, widening this conversation to consider assets under management and not just net returns.

Assessing the value of assets precedes the return calculation for a fund, as the month-end asset valuation determines the fund unit price or net asset value.[58] One would expect that if managers make monthly return data available to hedge fund databases, then asset data would be equally available for publication at this frequency. One might also expect if managers have taken the time to submit returns to a database, then the marginal extra effort of reporting assets under management is relatively costless. However, revelation of information is no guarantee in the case of hedge fund data, where provision of information is voluntary. Hedge fund managers choose where, when and what to present.

To contextualise the problem of asset disclosure, recall the Hedge Fund X case introduced in Chapter 3 on page 56. When this fund's performance weakened, the fund stopped reporting assets, but continued reporting returns for another four years. Such cases of missing assets are not isolated, so what are the incentives driving disclosure?

The incentive for the manager to report assets is the valuable marketing channel provided by public databases.[59] Rising asset levels can signify the manager's strategy is popular and can attract interest from prospective investors. The confirmation also provides comfort to pre-existing investors. Publishing asset values signals that the assets are regularly valued (although this has no promise of accuracy).

[58]In a unitised fund, the monthly fund return is calculated as the ratio of unit prices between the two months less one. In a non-unitised fund, then the growth in month-end asset values, correctly adjusted for flows, is the basis for the return calculation.

[59]As mentioned, see Jorion and Schwarz (2010) for discussion on limits to hedge fund solicitation.

However, asset disclosure has two negative consequences for managers — assets reveal outflows and the extent of smoothing. With return and asset information, investors can estimate flows (Sirri and Tufano, 1998). Negative outflows can distress both prospective and pre-existing investors, whilst large inflows can dismay pre-existing investors who are concerned about the impact of asset growth on future performance. In the absence of flows, then the growth of assets, less expenses such as fees, should tie up with the net return reported. Investors can use this to 'check' the asset values and assess the extent of hedge fund return smoothing. Mismatches can ring due diligence alarm bells. In particular, pre-existing investors are positioned to have more insight on flow mechanics and activity. I will show both outflows and smoother funds do indeed lead to worsening asset disclosure.

These can be strong disincentives to not report assets. Therefore, some managers choose consistently to not provide *any* asset information, although they are willing to share return information to public databases. For example, as many as 15.7% of all funds in the *entire* consolidated database have zero or missing lifetime assets.[60] Since these funds provide no asset information at all they were removed from the investigation.

The funds in question are those willing to report some asset information, and then to test what drives this selective voluntary disclosure. This sample of asset providers contains 18,382 funds across a wide set of consolidated databases. I define the asset disclosure shortfall as the percentage of missing assets over the fund's return track record. For those managers willing to share some asset information, only 59.4% have perfect disclosure (i.e. no shortfall). In the sample, nearly a quarter of funds having a year of asset data missing. Patchy selection of asset values by hedge fund managers, in aggregate, results in around a fifth of asset months missing compared to returns reported.

[60] 4,276 of the overarching 27,127 de-duplicated funds before the sample is constructed.

Since hedge fund managers are leaving assets missing, the focus is on unearthing the drivers of the deterioration of asset information. Are smaller or more liquid funds, where the loss of assets is critical, shy about asset levels? What are the implications of worsening asset disclosure? Should investors vote with their feet when faced with a 'less is more' approach by managers?

To answer these questions, I analyse cross-sectional and time-variation in the coverage of asset data relative to return data, and the reliability of this asset data. As we saw in Chapter 3, data is not only identifiable by fund and reporting month, but also by the month of submission to the databases. The focus of this chapter is *not* on data at the vintage level, however I do take the opportunity to check changes in the asset data reporting between the first and last "vintage" of funds over the period 2007 and 2011.

First, as discussed, the shortfall in asset disclosure (proportion of missing months) is estimated. Fortunately, this shortfall in asset disclosure is improving. There is time variation in asset disclosure, with the annual shortfall falling since 1994, although not at a consistent rate. For example, the LTCM crisis period, 1998 to 1999, had a marked impact on improving average asset disclosure.

Secondly, I look at the dynamic nature of asset reporting under the hypothesis that there is an asymmetric information problem at the core of voluntary disclosure. There are times when managers will not want to reveal their asset levels despite having revealed information in the past. This can happen for the range of reasons posited above — hiding the impact of outflows, or, in contrast, inflows of assets breaching agreed fund capacity targets,[61] or in the case of no flows, monthly asset levels not tying up with reported (over-smoothed?) returns.

[61] Investors are aware hedge funds face capacity constraints (see Fung et al., 2008; Ramadorai, 2012). Since high net inflows precede falling future alphas; investors, in practice, question the manager upfront on the expected fund size the manager would target for soft or hard closing.

The incidence of the shortfall in asset values is not random, and I employ information on fund characteristics and past performance of hedge funds to predict them. For example, Multi-Process (Managed Futures) hedge funds are more likely to withhold (reveal) asset data. Multi-Process funds blend strategies within a single management company. The disincentive to revealing information could arise as these funds are more diversified, and could therefore be a bigger part of pre-existing investors' portfolios, and so face increasing impact of disinvestment risk, for the fund family. Managers seem less wary of disclosing assets and maintain asset reporting when total restrictions are higher, although this is not conclusive. Institutional controls, like high-water mark or hurdles, improve the likelihood of funds maintaining asset reporting. This might be owing to the high-water mark creating a dampening effect on outflows after performance reversals, thus placing managers under less liquidity pressure (Goetzmann et al., 2003).

Disclosure of information might amplify the risk of runs on funds under pressure, which could hasten fund failure. Withdrawing from voluntary disclosure might protect existing investors from asset fire sales (and managers through co-investment and fee revenues). Limiting disclosure could have some social benefit by stabilising markets. However withholding information is not the sole avenue to protect managers, as hedge fund managers have a number of contractual features to cushion them from a run. Hedge fund managers set their terms of their funds in advance, particularly regarding liquidity restrictions, in light of the knowledge of their strategy and its liquidity limitations. Redemption notice periods can slow the pace of withdrawals. In the recent financial crisis, some hedge fund managers went further and turned to, rarely used but in the prospectus, 'gating' provisions to suspend redemptions to protect capital (seen as perhaps disingenious).[62] Weak disclosure might mask fund positions, but it is the contractual features that drive the ability to limit flows.

[62] See "Should hedge funds put a stop to investor withdrawals?", *Financial Times*, 21 January 2009, for an overview of this debate by different industry players.

The key result is the strong evidence I find of strategic timing by managers. Managers resort to decreased asset reporting post periods when funds were under pressure, owing to lower performance and outflows. This is in line with the expectations about the disincentives of disclosure. In aggregate, I observe performance is slightly higher in the months where assets are missing (around 6 basis points per month), than those months where all information is shown. This supports either funds not disclosing are better performers on average, or corroborates that the months funds choose not to report follows a period of weaker performance.

An open question is the potential for reverse casuality to be driving this disclosure. Namely, outflows might lead to sales of assets in illiquid markets making it difficult for managers to price assets and determine valuations. This might be the case with mutual funds, but in a voluntary disclosure framework, a hedge fund manager faced with such a problem can simply suspend reporting. This would allow time to establish the asset valuations and correctly ascertain fund returns (see Aragon and Nanda, 2011, for evidence of such strategic return revelation). However, managers are still reporting returns, although not showing assets. In addition, I am looking at disclosure in the calendar year after poor outflows, not contemporaneously. If the cause of weak disclosure was solely an illiquidity issue, as opposed to strategic behaviour, then we would see shortfalls in disclosure concentrated in illiquid strategies.[63] This is not the case.

Finally, I consider whether weak asset disclosure is a precursor to becoming a potential "problem" fund in the parlance of Brown et al. (2008). Using a host of performance flags, initially developed by Straumann (2009) for database return quality, and, extended by Bollen and Pool (2012) as a tool for fraud prediction, I find that funds with asset disclosure shortfalls trigger these flags with greater consistency. Since these flags have been correlated to the incidence of fraud, investors should take greater heed of funds not providing consistent asset information.

[63]Getmansky et al. (2004, p. 574) suggest illiquid strategies are Relative Value (including Event Driven), Emerging Market, Fixed Income and Funds-of-Funds.

Other implications of this research is the inadvertent bias researchers expose their results to, when focusing *only* on a set of funds with rich and deep asset information. For example, researchers who require regular flow calculations, prize a set of funds with full return and asset histories. As I find that asset reporting dries up in times of poor performance or outflows, and are concentrated in funds with weaker restrictions, excluding them from a set of funds requiring *full* disclosure can lead to significant selection bias. In such a sample, one would only be seeing the most resilient funds, and so this fund set could be expected to have survivorship and return results biased upwards.[64] Some hedge fund studies advocate using funds-of-funds as an investible hedge fund proxy (for example, Fung and Hsieh, 2000). The extent of bias on returns in full samples, and those using strategies with poor asset disclosure, must be appreciated.

Investors and researchers have been reliant on voluntary disclosure by hedge funds for a number of reasons. Initially, hedge funds were restricted with their dealings with the public. Secondly, hedge funds, before the wave of industry consolidation, were generally small focused businesses without much marketing resources as compared to mutual fund retail platforms. Thirdly, rapid growth in the industry in the 2000s would have thwarted any efforts by investors to collect data individually. For example, even if investors had focused on live and not defunct funds, fund numbers soar from just under two thousand in 1994 to a peak of 9,300 funds going into 2007. These factors reinforce the role of databases as the key communication medium, as suggested by Jorion and Schwarz (2010), to reach investors and to circumvent controls over advertising hedge funds to the public. Going forward the relaxation of restrictions on hedge fund marketing, as seen by recent changes in U.S. legislation[65] could require more robust reporting as the investing audience widens. The game is also changing

[64]Examples of studies that drop funds to ensure return and asset matching, or drop funds below a size limit (presumably including zeros and missings), are: Ben-David, Franzoni, and Moussawi (2012a); Fung and Hsieh (1997); Liang (2003); Zhong (2008); although some research, like Kosowski et al. (2007), carefully check this effect and find a fuller sample makes their results stronger.

[65]The JOBS(Jumpstart Our Business Start-ups) Act, signed into law on 5 April 2012, aims at stimulating small businesses by easing securities regulations. For hedge funds, this relaxes rules on general solicitation. See "Green Light for Hedge-Fund Ads Means Caution on Main Street", *Wall Street Journal*, 17 April 2012.

with SEC requiring larger funds to disclose both returns and positions to them in early 2013. Although this information is not for wider consumption, it removes past freedoms for hedge fund managers to stop and start reporting. The policy implications of this research is that some standardisation of reporting is necessary for assets and not just returns.

The remainder of this chapter is structured as follows. In Section 4.2, I review the related literature. In Section 4.3, I describe the data. Section 4.4 outlines the methodology to determine shortfall, changes in asset reporting and performance flags. I present my main empirical results in Section 4.5, and some robustness checks in Section 4.6. I conclude in Section 4.7. The core tables can be found in Section 4.8 and Appendix F contains additional analyses.

4.2 Related literature

The tracks regarding the problems with self-reported hedge fund data are well worn, but predominantly travel through the field of returns. Assets under management reporting has received less attention. From a return quality perspective, Ackermann et al. (1999); Fung and Hsieh (2000, 2009); Liang (2000) lay out the biases in hedge fund returns such as survivorship bias, self-selection and backfill bias. Issues with assets only surface in the motivation of equal-weighted indices for returns, such as in Fung and Hsieh (2000, p. 295), because asset records are noted as "frequently incomplete". This research explores the fund characteristics and dynamics behind this missing asset phenomenon.

Statistical evidence of data manipulation of the hedge fund returns is pervasive. For example, smoothing of hedge fund returns (Getmansky et al., 2004), conditional smoothing with negative returns smoothed more than positive returns (Bollen and Pool, 2008), or discontinuities in the distribution of returns with no small negative returns creating a kink at zero (Bollen and Pool, 2009). Some of these results are useful in assessing the managerial quality of hedge funds which provide limited disclosure of assets.

Direct, as opposed to statistical, evidence on the implications for investors of voluntary disclosure is emerging as well. Although not the focus of this chapter, I draw on the approach pioneered by Patton, Ramadorai, and Streatfield (2012) regarding return revisions for hedge funds. Another approach, asking 'when?' as opposed to 'what?', Aragon and Nanda (2011) explore late reporting of hedge funds using daily snapshots and find hedge funds strategically delay bad return news. This research explores a similar problem in which asset 'news' is delayed, at times indefinitely, although return reporting continues. Questioning the reliability of voluntary submissions of data is not a uniquely hedge fund related phenomenon. Consider, for example, the banks under scrutiny in the current LIBOR scandal.[66]

The removal of asset information makes it harder for investors to verify returns. However, hedge fund managers have flexibility regarding valuation, so these managers have other alternatives to simply not reporting asset data. For example, the valuation of assets under management in hedge funds has often been criticised. Hedge funds gain exemptions, through holding appropriate numbers of accredited investors, enabling them to circumvent the valuation sections required of other investment funds in the Investment Company Act of 1940. Cici et al. (2011) find this freedom leads to the valuation of listed equities reported by hedge funds on 13F forms to differ from conventional valuation policies using prices in CRSP. They find these differences are statistically significant in a quarter of their managers. They claim these deviations in policy are strategic as they are concentrated in managers who smooth returns, who are offshore, have greater return discontinuities around zero, and have higher proportions of fraudulent funds. I demonstrate that managers who choose the path of non-disclosure share similar characteristics. Shortfall managers smooth, are offshore and also have greater propensity to trigger fraud flags.

[66]See initial concerns about banking inputs into the London Interbank Offered Rate (LIBOR) "Study Casts Doubt on Key Rate", *Wall Street Journal*, 29 May 2008.

Assets under management in public hedge fund databases are observed on a monthly basis. Recall hedge fund valuation can be subject to month-end manipulation in which prices, particular of illiquid securities, can be 'bumped' up by last minute trading at key valuation dates (Ben-David et al., 2012b). This is an inevitable downside of non-continuous monitoring, that is exacerbated as funds weaken their asset disclosure.

On the demand side, institutional investors can use a minimum asset size as an investment filter or screen. For investors, assets is a measure easily comparable between managers, ensures organisation has sufficient resources and operational strength, and increases the chance others are co-investing (Allen, 2007). The consequence of this when assets are falling, whether by outflows or performance, is that managers suspending reporting, and leave potential investors with stale, yet inflated, figures in order to remain on investment watchlists.

On the other hand when assets are rising, all investment managers face constraints in alpha generation or rising limits to arbitrage. Allen finds that smaller managers, particularly in less liquid strategies, perform better. Such limitations on manager skill is noted earlier in the investment consulting literature (see the "skill cycle" in Hodgson, Breban, Ford, Streatfield, and Urwin, 2000). Hedge funds in particular face capacity constraints (see Fung et al., 2008; Ramadorai, 2012). Hedge funds price for this accordingly through incentive fee structures to boost revenues, although reported management fees have been on the increase, benefitting managers over investors, as evidenced in Chapter 2. Size of assets can thus play a role in disclosure, with managers temporarily overreaching agreed capacity constraints, to gain the benefit of higher management fee income. This provides an rationale to suspend asset reporting whilst still communicating returns.

Although this work is evaluating voluntary disclosure, there are results from the mandatory side of hedge fund disclosure regarding hedge fund assets. Brown et al. (2008) find over the brief period when SEC required hedge funds to be registered, funds with weaker operational controls had lower incentive fees and lacked high-water marks. They did not find size to be a factor for problem funds. However, protection of confidential trading strategies can be related to size of assets. In Form 13F disclosures by hedge funds, Agarwal et al. (2011b) find large funds with concentrated trades seek confidentiality provisions. There can be benefits in revealing information by stimulating trade volumes from copy-cat traders (Brown and Schwarz, 2011). Of course, mandatory disclosure is no panacea as it does not promise that the data submitted is accurate or complete. Schwarz and Potter (2012) demonstrate this with their discovery of the large, although not economically significant, disparities between mutual fund portfolio filings submitted to the SEC, and the Thomson Mutual Fund Holdings database.

The majority of hedge fund indices tend to be constructed on an equal as opposed to value weighted basis.[67] This goes against the convention in indices in other asset markets. This could point to a lack of confidence in regular collection of asset information in the industry. Teo (2009a) finds smaller funds outperform bigger hedge funds, so an equal-weighted index leads to a harder performance target hurdle. One would expect leading managers to contest these benchmarks, particularly as the market consolidates into larger hedge fund management groups. The impetus could also come through an increasingly institutionalised investor base, with preferences for peer benchmarks.

4.3 Data

4.3.1 Consolidated hedge fund and fund-of-fund data

The data used in this Chapter 4 is broadly similar to that used in Chapter 3, where the process is outlined in Section 3.3. Namely, a large cross-section of hedge funds and funds-of-funds over the period from January 1994 to May 2011 is consolidated from data in the TASS, HFR, CISDM, Morningstar, and BarclayHedge databases. Appendix A contains details of the process followed to consolidate these data. The funds in the combined database are consolidated into 10 main strategy groups: Security Selection, Macro, Relative Value, Directional Traders, Funds-of-Funds, Multi-

[67] Géhin and Vaissié (2005) summarise the features of major hedge fund indices.

Process, Emerging Markets, Fixed Income, Managed Futures, and Other (a catch-all category for the remaining funds).[68] The set contains both live and dead funds. Returns and assets under management (AUM) are reported monthly, and returns are net of management and incentive fees.

4.3.2 Hedge fund database vintages

Hedge fund data providers update their databases from time to time with the entire history of return and asset data. As discussed in Section 3.3 on page 62, this allows one to compare reported histories across vintages of these databases at various points in time. Recall I compare a total of 40 vintages, v, of the different databases between July 2007 and May 2011.

I apply the same standard filters to the data before analysis as explained in Section 3.3 on page 63. These are removing funds with extreme monthly returns outliers, funds reporting data quarterly not monthly, funds with insufficient return history or missing fund level data, and removing the few remaining Morningstar funds that passed the quality filters.

This sample differs slightly from Chapter 3 owing to two additional cleaning measures on assets under management. Firstly, when checking asset changes between vintages, I find some benign revisions mainly attributable to common data entry errors. Two such errors under consideration are decimal place errors (for example, 15 entered as opposed to 15,000,000, i.e. 15 million, so I check whether the revised asset value differs from the original asset value in the preceding earlier vintage by exactly a multiple of 10 such as 10, 100, and going up to 1,000,000) and transposition errors (where adjacent digits in the original asset value are transposed in the revised asset value). The proportion of funds with these errors are small. 140 funds with such errors were removed to improve asset data quality and prevent fund lifetime asset values being skewed.[69]

[68]The mapping between these broad strategies and the detailed strategies provided in the databases is reported in Appendix B.

[69]Panel C of Table F.23 details the decimal place and transposition error counts.

Secondly, some asset values required additional cleaning. The few entries with captured negative AUM values were marked as missing. Older asset data in some databases was marked as zero instead of the correct missing indicator for that database. This affects 6.6% of all asset months. I set these zero asset values to the missing indicator. This prevents spurious deletes if at a later vintage this value is correctly marked as missing.

The final cleaned dataset contains 18,234 unique hedge funds. Table 4.1 shows some characteristics of this sample. Although slightly smaller than the sample used in Chapter 3, the characteristics are broadly the same as before in Table 3.1 on page 91. On average, funds report for five and a half years, have U.S.$103 MM in assets, and generate returns of approximately 0.64% per month. Slightly over a quarter of them are Funds-of-Funds, with Security Selection and Managed Futures being the predominant hedge fund strategies represented in the data. Approximately one-third of the funds are from the TASS database, with the CISDM database accounting for the smallest share of the four databases represented in the final sample, at just under 10% of funds.

In general, the asset and return information at the last available vintage is used, i.e. for most funds this will be the May 2011 vintage. Where data is compared across vintages, this process has been outlined in detail in Chapter 3 on page 63. To recap, we had defined $Ret_{i,t,v}$, as the return for a fund i, at a time t, reported in vintage v of the data. Likewise, one can also consider $AUM_{i,t,v}$, the total assets under management value for a fund i, at a time t, reported in vintage v of the data. Asset revisions are not the focus of this research. However, the asset and return values indexed by vintage v allow the measurement of the extent of missing assets information between vintages.

I study the determinants of asset disclosure behaviour in more detail using a probit model described in the next section.

4.4 Methodology

I begin by documenting the extent to which a managers' asset disclosure lags that of returns. I then explore the persistence and reliability of asset reporting by considering asset history changes over time. Secondly, I identify the characteristics of funds that are prone to a shortfall in asset disclosure, focusing the analysis on worsening disclosure (as opposed to funds that begin asset reporting). Thirdly, I investigate how fund characteristics impact on deterioration of disclosure conditioning on annual information. Finally, I explore for investors in the group of funds with shortfalls whether this heralds other risks, by using fraud performance flags with simulated fund specific percentiles.

A hedge fund manager's decision to stop reporting asset information is not their only option in managing disclosure. As seen in Chapter 3 with returns, managers can, and do, change their historic records across time. This is not the focus of this chapter, but I also investigate how asset histories and disclosure vary by vintage.

4.4.1 How does asset disclosure relate to returns?

I measure asset disclosure relative to returns reported by funds. At a vintage, v, consider the set of months where returns are available for fund i, $T_{i,v}$.[70] The count of non-missing and non-zero data — returns, $\#Ret_{i,v}$, and asset values, $\#AUM_{i,v}$ — are determined over this period respectively. The shortfall in asset disclosure is defined as:

$$Shortfall_{i,v} = \frac{\#Ret_{i,v} - \#AUM_{i,v}}{\#Ret_{i,v}}. \tag{4.1}$$

For example, in the first vintage, July 2007, consider a fund showing eighteen months of returns since inception, but only the last year of asset values. The asset disclosure shortfall for that fund would therefore be a third, (18-12)/18. Likewise, this shortfall concept can be used to explore changes for fund i across the time dimension, t, by fixing the vintage v, and then considering subsets of time periods in $T_{i,v}$, like calendar years, where returns are available.

[70]The cases where assets are reported but returns are not, is minimal (0.039% of asset months).

4.4.2 Which funds have missing assets?

The first step is to assign a '1' to any fund which has a non-zero shortfall in asset disclosure. Assigning a '0' to all other funds, I then estimate a cross-sectional fund level probit regression, conditioning this indicator variable (label $IShortfall_i$ for fund i) on various fund characteristics, which are described below, and denoted by the vector X_i:

$$IShortfall_i = \alpha + X_i'\beta + u_i. \tag{4.2}$$

In the above regression, the right-hand side comprises pure cross-sectional variables, including lifetime summaries of variables such as fund performance and fund size. While it is a useful summary of the fund characteristics associated with weaker disclosure, it offers no causal intuition. I thus estimate a version of the above specification looking at each year of the fund (denoted T) and look at the fund right-hand-side variables computed as at the prior year $T-1$.[71] This examines if relative performance and market conditions in the preceding year affects the disclosure of assets in the following period.[72]

$$IShortfall_{i,T} = \alpha + \gamma X_{i,T-1}'\beta + u_{i,T} \tag{4.3}$$

In these specifications I employ a range of conditioning variables on the right-hand-side. I employ assets under management (AUM) to study whether changes in disclosure are more likely to occur for larger or smaller hedge funds, ranking funds by their lifetime average AUM. I also use the average of lifetime returns for each fund. This is to capture the possibility that weaker performing funds might resort to stopping reporting assets to hide their misfortune. Third, I use the standard

[71] Table F.24 in Appendix F shows the variation in shortfall by calendar month. I chose to analyse by calendar year as disclosure improves slightly over the course of the year.

[72] Standard errors are clustered by database in equation (4.2) and by year in equation (4.3). The former is to control for the possibility that errors across funds are correlated according to the database as there might be unobserved determinants controlling revisions in specific databases. The latter is to control for the possibility that there are certain periods in which unexplained shortfalls are more likely to be prevalent.

deviation of lifetime returns, as more volatile returns can mask any underlying flow information in asset levels reducing any need to suppress asset figures. Finally, I use a measure of return smoothing suggested by Getmansky et al. (2004), namely the first-order autocorrelation coefficient of lifetime returns. In all cases in which ranks are employed, they are standardized between 0 and 1.

In addition to these variables computed from return and AUM histories, I also consider a variety of fund characteristics as explanatory variables. I include strategy fixed effects in the specifications to control for the possibility that differences in volatility and liquidity occasioned by the use of these different strategies, as well as differential access to information about these strategies (for example, Emerging Markets) might lead to differences in the propensity to publish data. I also include database fixed effects as the controls (such as verification of values pre-loading) implemented by each database vendor might differ, and influence the propensity for reporting. I employ an indicator for whether the fund is offshore or onshore, as funds in offshore jurisdictions could be subject to less scrutiny. I condition on the lockup restrictions imposed by the fund on its investors — these restrictions provide liquidity safeguards for the fund manager, and could encourage weak disclosure as lockups prevent outflows. Alternatively, managers with weak restrictions, and unable to hide within the period of the lockup, might more readily omit asset data than face the consequences of disclosure. I also include an indicator for whether a fund has a high-water mark or a hurdle rate, as well as an indicator for whether the fund has any audit information available in the database.[73]

Finally, I include a variable which computes the number of asset values in fund i's lifetime history up to time t. This is to control for the purely mechanical possibility that if there is a small fixed chance of data capture error, then a longer data history provides more exposure to changes and omissions. Despite shortfalls in asset

[73]Underlying databases differ in the types and level of information they provide, with some providing the date of last audit, other providing annual audit flags, and yet others providing auditor names. This indicator takes the value '1' if *any* audit information is available for the fund, and zero otherwise.

reporting, this is also related to the age of a fund, so this variable has multiple interpretations. Table F.22 in Appendix F contains descriptive statistics for eight of these variables.

The specification (4.3) does not capture changes between vintages, so additions — backfilling asset data at a later vintage — will improve shortfall.[74] Such manager behaviour works against these results. The extent of change between two vintages is explored in the results section.

4.4.3 Are changes in asset reporting informative about potential fraud risks?

Hedge fund managers not supplying requisite asset information can simply point to weak oversight, or, it could be covering something more sinister. Some fraudulent structures, like a Ponzi scheme, rely on secrecy, and a continued stream of inflows to fund any outflows.[75] Obscuring asset information might hide outflows to both current and prospective investors to prevent scrutiny, as a run on the fund would unravel the scheme. Some of the current investors might even be aware the scheme is fraudulent, but choose to ride the wave of good returns. Faltering flows would signal to this 'smart money' that it is a good time to exit.[76]

Dubbing funds with asset shortfalls as potential "problem" funds, I explore the incidence of triggering performance flags between shortfall funds and those with consistently good asset reporting. These flags highlight higher risk of frauds through observing unlikely patterns in returns. Investors are not compensated for fraud risk through lower fees or performance (Dimmock and Gerken, 2012). If asset shortfalls point to an increased risk, then circumventing this risk holds substantial value for hedge fund investors. The proviso, of course, is that fraud flags are simply a prediction and do not imply causality, or that fraud will occur. Nevertheless it is a tool for

[74] As noted, Appendix F presents results of the prevalence of AUM changes across vintages: additions, deletions, revisions and 'any change,' a catch-all category encompassing all three types of changes.

[75] See Gregoriou and Lhabitant (2009) for insights into the 17-year long Ponzi scheme operated by Bernard Madoff.

[76] I thank Richard Hills for this suggestion.

signalling additional due diligence to counter the substantial information asymmetries hedge fund investors face.

Drawing on tests motivated by previous hedge fund research, (see Bollen and Pool, 2012; Straumann, 2009), I assess the return histories of funds using a number of data quality tests: (1) 'Perc. Negative', the percentage of returns that are negative, with lower proportions of negatives pointing to manufactured values; (2) 'Count Zeros', the count of exactly zero values; (3) 'String', the count of the longest sequence of repeated data; (4) 'Num. Pairs', the number of repeated blocks of length two, without counting overlaps;[77] (5) 'Perc. Repeats', the proportion of returns that are repeated;[78] and (6) 'Uniform', establishing whether the second digit of the value is uniformly distributed.

This exercise has two challenges. Firstly, the observed returns are rounded, and, secondly, funds have different history lengths. This materially affects the probability of observing some of the flags. For example, consider a hypothetical hedge fund, Fund A, with coarse rounding to 1 decimal place (i.e. 0.1%) with median drift and volatility versus a volatile Fund B with rounding to four decimal places and a longer return history.[79] One might expect a higher relative number of zero returns or repeated returns for Fund A than B, owing to both its lower return volatility and tighter rounding. Table 4.2 confirms this intuition and highlights how the simulated percentiles vary based on these fund inputs. The longer return history in Table 4.2 Panel B would make the chance of longer strings more likely, however, the high volatility and median rounding level mitigate this bearing out in the simulation. Comparison of Panels A and C in Table 4.2 demonstrates the impact of rounding whilst keeping fund return and risk parameters constant. I find increasing the rounding levels limits the chance of finding strings and repeated blocks ('Num. Pairs'), and lowers the zero count owing to the increased precision. Rounding has a negligible

---

[77]So in a sequence of $8.0, 8.0, 8.0$, only one pair of eights would be counted as the middle value leads to an overlap of blocks. The 'String' test tackles repeats of this nature.

[78]Or, as Straumann (2009) describes, the inverse of unique values.

[79]Observed fund statistics are used to calibrate this illustration (see Table F.21 in Appendix F). In Table 4.2 Panel A, Fund A has a median return, risk and history length. In Panel B, Fund B has an upper quartile volatility and history length.

effect on the percentage of negative returns, as might be expected.

Therefore avoiding potentially misleading fixed thresholds, I follow Straumann (2009) and construct fund specific bootstrap percentiles. These percentiles make use of a scaled t-distribution with excess kurtosis of 3.[80] The number of simulations of the return series is 10,000.

A significance level of 10% was used. For most of the tests, apart from the percentage of negative returns, the upper tail marks the critical value as a higher test statistic value indicates a potential problem. For the first five tests, the observed count or percentage is used as the test statistic. For the 'Uniform' test, a goodness of fit statistic is constructed. This could be done with either a Uniform prior, each digit equally likely, or dependent on the simulated probabilities of what the probability of finding each digit is. For this sixth test I simply assume the digits are uniformly distributed, but I follow Straumann's approach of sourcing the critical values of the test from the simulated distribution. An alternative is simply using the $\chi^2_{df=9}$ statistic (see Bollen and Pool, 2012), but this does not take fund history length into account.

Let q_d be the probability the digit of interest is equal to d where $d = 0, 1, \ldots, 9$. $n_{d,i}$ is the count of digit d in the sequence of N_i observed returns for the fund i. The test statistic for testing the distribution of digits (Straumann, 2009, p. 31) for fund i:

$$Z_i = \sum_{d=0}^{9} \frac{(n_{d,i} - N_i q_d)^2}{N_i q_d} \tag{4.4}$$

I add a test of my own as an alternative to a uniform assumption of digits by reflecting on digital analysis in the auditing profession literature. The auditing profession uses reasonability tests to target audit samples. Nigrini and Mittermaier (1997) who outline such tests, make use of Benford's Law, as opposed to a Uniform assumption, noting the empirical frequency of the first digit in practice declines logarithmically.[81] Benford's Law was posited by Benford (1938), although initially discovered in 1881 by Newcomb, after noticing the front pages of logarithmic tables were well-thumbed.

[80]This is implemented by setting the t-distribution degrees of freedom to 6, given the distribution's excess kurtosis is $\frac{v}{v-4}$. The variance, $\frac{v}{v-2}$ for $v > 2$, is also corrected when the scaled t is simulated with the fund's observed volatility. I thank Andrew Patton for this correction.

[81]Nigrini and Mittermaier commend the economist Varian who initially suggests Benford's Law as a "test of reasonableness" of data (Varian, 1972, see p. 65).

This law suggests the frequency of the first non-zero digit d is:

$$\Pr[d] = \log_{10}\left(1 + \frac{1}{d}\right); \text{ where } d = 0, 1, \ldots, 9 \tag{4.5}$$

Benford's Law adjusts the frequency of the first digit, finding with natural growth rates that 1 as the first digit occurs 30% more likely, whereas the expectation of 9 as a first digit is only 4.6%. A rationalisation for this phenomenon is it is a more substantial effort growing from 1 to 2, a 100% growth rate, whereas growth rapid falls as digits increase. For example, moving from 4 to 5 the growth rate has fallen to 25%. This is an empirical law although there have been numerous theoretical explanations (for example, see Hill, 1998).

I use the expected digital frequencies for a second digit in a number calculated by Nigrini and Mittermaier (1997). Although the differentials are not as marked as the first digit (expectations for the second digit range from 11.9% for digit 0 to 8.5% for digit 9), there is some variation beyond the Uniform distribution. De Ceuster, Dhaene, and Schatteman (1998) support the adoption of Benford's Law over the Uniform distribution in assessing psychological barriers in stock indices. Durtschi, Hillison, and Pacini (2004) caution on the conditions for optimal use of Benford's Law tests, such as avoiding assigned numbers like product prices or ATM withdrawals amounts. They table large data sets, mathematical combinations of numbers, and positive skew support a Benford distribution. Hedge fund returns fit these criteria, although purists might question the sufficiency of the span of returns to support the distribution. The seventh test, 'Benford,' assesses the fit of the second digit as per equation (4.4) using the Benford's Law second digit probabilities for q_d. This acts as a complementary robustness check on the Uniform test.

The final set of tests are on the smoothing properties of returns. 'AR(1)' tests the unconditional serial correlation in returns (Getmansky et al., 2004). Fund returns are regressed on their first lag, as in equation (4.6). The test flag is triggered if the β coefficient from this regression is positive and significantly different from zero at the 10% significance level. The essence of this 'AR(1)' test is that manufactured returns may rely on excessive smoothing of returns to lower volatility in order to appear 'safer' and to boost Sharpe ratios. Thus, a highly significant and positive first order autocorrelation coefficient will indicate the possibility of gratuitous smoothing, and trigger the flag.

$$R_{i,t} = \alpha + \beta R_{i,t-1} + \epsilon_t; \text{ where } R_{i,t} \text{ is observed return for fund } i \text{ at time } t \tag{4.6}$$

The 'CAR(1)' test extends this return smoothing investigation by anticipating smoothing may be uni-directional with downside volatility more likely to be smoothed by managers. Whereas, gains will be fully reported due to competition for capital (Bollen and Pool, 2008, 2012). 'CAR(1)' thus tests the conditional serial correlation in returns to separate illiquidity from intentional misreporting following losses. Fund returns are regressed on their lag and an interaction term, β^-, that is dependent on the level of return being below the mean of its fitted value, as in equation (4.7).

$$R_{i,t} = \alpha + \beta^+ R_{i,t-1} + \beta^-(1 - I_{i,t-1})R_{i,t-1} + \epsilon_t \tag{4.7}$$

This technique makes use of an assumption on the underlying actual, unobserved, return. This return is proxied by the fitted return from a Fung-Hsieh seven-factor model. The indicator $I_{i,t-1}$ is set to 1, if the fitted return for fund i is above its mean, or zero otherwise. The test flag is triggered if the conditional β^- is positive and significantly different from zero at the 10% significance level. The presence of significant incremental serial correlation following negative performance is consistent with the intentional avoidance of reporting losses.

In summary, fraud flags are determined by assessing each fund's return track record and analysing the characteristics of the returns, such as the proportion of positive returns, or digit frequencies. The hypothesis is that manufactured or manipulated return series would have characteristics that look unlikely to happen by chance. To assess this, some of these tests rely on establishing a threshold statistic at a set significance level. This is determined through bootstrapping the returns series using the risk, return and rounding level of each fund. So, for example, 10,000 return series are simulated for a specific fund, and the characteristics like the distribution of the proportion of positive returns is determined for that fund. The estimate of the proportion of positive returns, say 32%, at the 10% significance level can then be derived and compared to that fund's actual proportion given its observed returns, say 15%. In this case, the number of observed negative months is below what is expected, and so is flagged as potentially suspicious. For the smoothing tests, 'AR(1)' and 'CAR(1)', the direction and strength of significance of the autocorrelation coefficients, will determine if the fraud flag is triggered.

4.5 Results

4.5.1 How does asset disclosure relate to returns?

For many funds, assets are not reported as regularly as returns. This shortfall in coverage is explored, first, on an aggregate basis over time, secondly, at the fund-level, and then, dynamically at the fund-level between vintages. In Table 4.3 Panel A, the aggregate shortfall in assets to return disclosure is calculated on an annual basis for funds reporting assets at the start of each calendar year. So, for example, in this table, in 2000, 4,406 funds reported returns in January. These funds reported 50,304 months of returns for that year in aggregate, but only 41,156 asset values. The proportion of missing assets, or shortfall, is 18.2% for that calendar year.

The voluntary disclosure of assets on an annual basis varies over time, and has been generally improving each year. The shortfall has fallen from a high of 29.9% in 1994 to a low of 7.9% in 2010. Figure 4.1 shows this trend in disclosure has not been uniform. The average monthly return figure gives an impression of market conditions for the year. The shortfall ticks up again from a low in 1999, perhaps owing to the weaker market conditions post the dotcom and LTCM crises. Although in the next difficult market period, it seems 2008 and 2009 were years of marked improvement in disclosure. This crisis period was severe for hedge funds as they suffered substantial deleveraging relative to other investment funds (Ben-David et al., 2012a). The average months of returns reported, that is, return count divided by the number of starting funds, confirm this by falling over this period to below 11

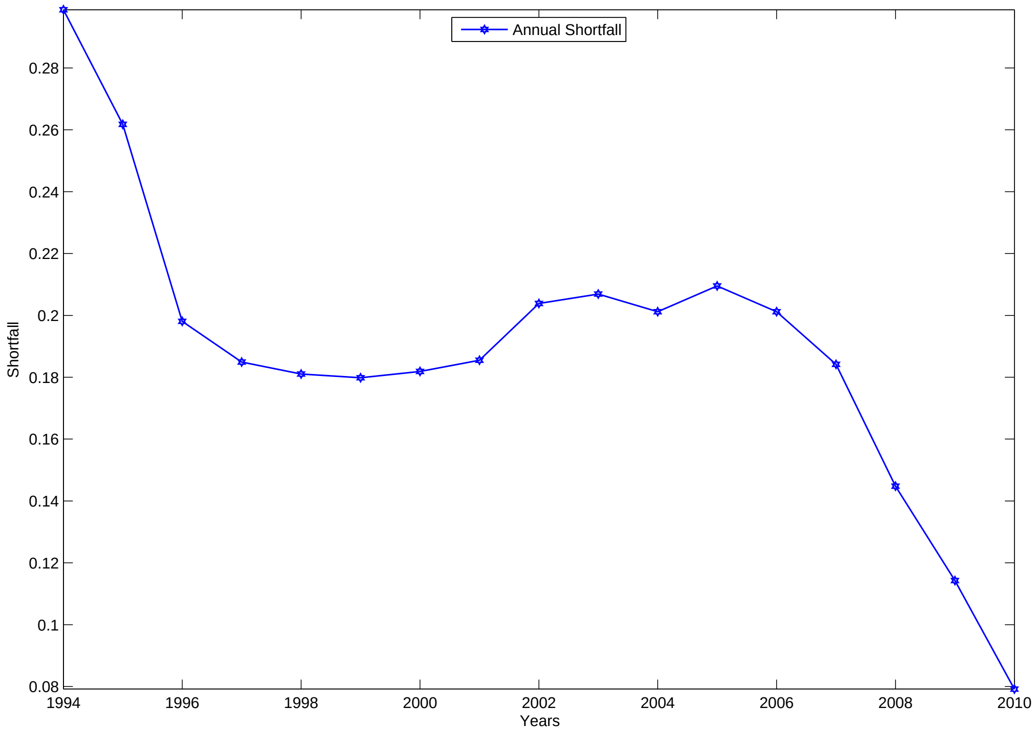


Figure 4.1: *Aggregate Asset Shortfall*
The figure reflects the asset to return disclosure as a proportion of missing asset months over the period returns were reported, on an annual basis. Annual shortfall measures the aggregate annual shortfall for funds reporting returns in January at the start of the year.

months (10.96 and 10.35 months) for these two years. This indicates the tough conditions for hedge funds led to fund closures (or cessation of all reporting). One interpretation is that the funds that did not survive these periods seem to have poorer asset coverage, since their exit has improved relative disclosure. Naturally, 'dead' funds leaving databases can continue, and not close. Aiken et al. (2012) study SEC funds-of-funds data and confirm this, however funds de-selecting, and exiting from the database, underperform those reporting to the databases. In contrast, Edelman et al. (2012) document amongst the larger hedge fund management companies, the extent to which their funds do not report to any database, although the unobserved performance of these larger funds can be inferred.

In Table 4.3 Panel B, the average and median returns conditional on whether assets are reported for the month are calculated. If it is believed not showing asset returns is simply due to chance, then we would expect no difference in these conditional averages. I find that the returns in months where assets are missing are better than those months where all information is shown. The annual average different in returns conditional on asset availability is around 6 basis points per month higher for the months where assets are not disclosed.[82] This could mean that the funds not disclosing are better performers on average, or that the months they choose not to report follows a period of weaker performance. (The respective average returns are higher than the median return, which suggests a positive skew for these 'recovery' returns.) The macroeconomic cycle does have an influence on hedge fund return predictability (Avramov, Kosowski, Naik, and Teo, 2011). There is weak evidence of reversals around times of crisis like 1998 and 2007. I explore this later with probit regressions.

Table 4.4 turns the attention to the fund unit to determine shortfall at the fund level. In Panel A, a rather substantial 40.6% of funds are missing asset information over the period the funds report returns. In Panel B, the depth of this shortfall is high with on average more than a third, 39.4%, of assets are missing for these funds. A substantial part of the asset history is unavailable to investors.

In Table 4.4 Panel C, I show the sensitivity to increasing the number of months missing before a fund is deemed in shortfall.[83] For some funds this shortfall in asset disclosure is slight, but a core of funds with substantial months of data missing exists. Nearly a quarter of all funds in our sample have more than a year of assets missing. These results indicate this is more than a minor delay in reporting issue. In Panel D the shortfall is conditioned by fund size measured by lifetime assets under management. It seems larger funds have more funds in shortfall than smaller funds. This is tested later in Section 4.5.3 to see if this persists in a multivariate setting after controlling for effects like strategy and other fund characteristics.

[82]This improvement is 8.9% of the average monthly return, shown to be 0.64% in Table 4.1.
[83]Alternatively, a histogram of these missing months counts, Figure F.2, is in Appendix.

4.5.2 How does asset disclosure vary by vintage?

Although not the crux of this chapter, I do find substantial changes to *asset* histories using the same methodology used in the chapter on returns in Section 3.3.3 on page 63 over a very similar sample.[84] This digression reinforces the central message, namely the need for asset disclosure to be given as much attention as returns.

Table 4.5 captures improvements in disclosure by funds, by indicating how shortfall at the fund level changes between vintages. In Panel A, the fund's shortfall is calculated over the fund's lifetime at the first and last available vintage for each fund. In a small number of cases, 115 funds, there is only one vintage available so these funds are excluded from the analysis. Shortfall for the majority of funds, 72.0%, does not change between vintages. These 'No Change' funds have a low average shortfall of 7.2%. The first vintage is July 2007, so this group does include many defunct funds no longer making additions to their reporting.[85] Also in this group, more than three quarter of all funds have perfect matching of assets with returns in both first and last vintages (zero shortfall). These results show, even over the volatile period of the last financial crisis, there were hedge funds able to maintain reporting of their assets.

For those funds where asset disclosure changes, twice as many funds are improving their shortfall as opposed to worsening it. Where shortfall worsens, the average deterioration is greater at 21.5% versus the improvement of 15.9%. Just as we saw before in the yearly analysis, there is an improvement in overall shortfall for funds at the vintage level. The average shortfall has improved across the vintages, the 2007 to 2011 period, by about 0.9%. In Table 4.5 Panel B, the percentiles show shortfall is quite positively skewed. The mean number of missing months rose to about 33 months whether the shortfall improved or worsened.

[84]I estimate the prevalence of three different types of changes to funds' asset histories as per equation (3.1). Panel A of Table F.23 shows over 48% of funds have altered, deleted or added to their asset histories. Of these, revisions of pre-existing asset data are the most frequent, at 36.5%, followed by deletions at 19.2%, and additions at 10.6%. We observed similar levels of revisions when considering returns in Table 3.2. More notable for assets, is the higher level of deletions and additions of fund asset values (compared to 5.9% deletions, 2.0% additions in returns over the same period). These results on asset changes can be found in Appendix F.

[85]I estimate, given a six-month late reporting window, that 28.7% of total funds can be considered defunct at this first vintage.

Changes in disclosure across vintages puts the other analysis relying on a single vintage of data into perspective. Future shortfall is not only affected by new data points, but also by both asset data *and* fund returns being subject to revisions, deletions, and additions. For example, one might find a fund's shortfall is improved at a later time through the backfilling of assets by the manager.[86] A future research direction is to explore the interaction between return and asset changes across vintages.

For now focusing on a single vintage of information leads to the next stage of fund level regression analysis, to establish the determinants of changes in fund disclosure of asset values across time and by fund.

4.5.3 Which funds have missing assets?

Table 4.6 shows the results of estimating the probit equation (4.2) for fund shortfall. The regression presents the marginal effects of each continuous right hand side variable, namely, the change in probability of the dependent variable that results from an infinitesimal change in each of these variables at the variable's mean. The discrete variables, like the offshore indicator, show the effect of a change in the variable from 0 to 1.

Table 4.6 reveals funds with lower volatility and higher levels of return smoothing (as measured by return autocorrelation) have significance at the 10% level as determinants of a fund's propensity to not show asset information. Managers might be resorting to withhold asset values to prevent flows, or their extent of smoothing, being so plainly revealed to investors through the thin veil of more predictable returns. (The strategic incentive is, as Agarwal et al. (2011a) suggest, for managers to hold back value and smooth returns within the calendar year to optimise incentive fees.)

[86] Appendix F has probits on changes of asset data, similar to those in Chapter 3 on page 95. For example, probits on Table F.28, and F.27, show the results of estimating the probit for deletions and additions of asset values, respectively, according to equation (3.1). For example, funds that are larger, less liquid as measured by return autocorrelation, with strategy of Funds-of-Funds or Multi-Process, and are onshore, are more likely to add or delete past asset information.

As anticipated, the incentive to withhold information is stronger in funds with weak restrictions. The total restrictions period is negatively and significantly related to shortfall. (Table 4.4 Panel F reports the lowest tercile of fund restrictions has both a higher incidence and higher average levels of shortfalls.) Furthermore, in Table 4.6, having a high-water mark dramatically reduces the propensity by 14.2% of having a shortfall. High-water marks can create a lock-in when funds underperform and are ‘underwater’ (Goetzmann et al., 2003), and curtail risk-shifting (Aragon and Nanda, 2012; Panageas and Westerfield, 2009). This suggests that hedge fund managers without high-water marks, or facing lower lock-ups, are more concerned with the risk of investors withdrawing and so suppress information. In Chapter 3, the opposite is found for return disclosure. Revisions were more likely in funds with higher restrictions. This suggests asset shortfalls are capturing a different response by hedge fund managers to manage their disclosure.

The return history length is significant but as discussed this can be a mechanical owing to more months of data being exposed to change. There is significant variation between databases, with BarclayHedge and HFR having lower relative impact on shortfall. This is consistent with findings on relative database quality (Joenväärä et al., 2012). Turning to strategy indicators, Multi-Process funds have the highest incidence of shortfall, and Managed Futures funds the lowest (perhaps owing to this strategy’s better liquidity and simpler valuation). Since Multi-Process funds are diversified across strategies but are tied to one manager (as opposed to a fund-of-funds), this suggest once again a managerial incentive to mask asset growth when returns have lower volatility.

The evidence on fund performance over the fund’s lifetime is not clear, so I turn to the specification in equation (4.3) that looks at the most recent past fund performance and its impact on asset disclosure in the following year. In Table 4.7, the prior return and flow coefficients are negative and highly significant. This suggests funds that have experienced outflows (lower flow ranks) and are weaker performers are prone to weakening their disclosure. Relative size of the fund does not seem to be a factor affecting shortfall.

In this specification, again the removal of fund restrictions weakens disclosure. The sign on total lockup restrictions remains negative but is no longer significant. The lack of a high-water mark is highly significant as before. Being offshore now affects the propensity for assets to be withheld. Offshore funds are considered to be under less scrutiny, whilst onshore funds have more organisational constraints (Aragon, Liang, and Park, 2008). Funds with higher incentive fees are more likely to stop reporting, however this impact is marginal as the coefficient is small.[87]

The audit flag, has a large, significant negative coefficient suggesting unaudited funds would be more able to change asset reporting. The presence of an audit might be because of past concerns by investors requesting more scrutiny by auditors and thus more pressure on a fund's reporting. However, Cassar and Gerakos (2011) in looking at manager behaviour to smooth returns, suggest not reading too much into the impact of auditing as it is felt on an longer term basis. The crisis period fixed effects were only significant over the 1998 to 1999 period. A particular feature of this LTCM crisis period was leverage, and financial leverage can exacerbate operational risks (Brown, Goetzmann, Liang, and Schwarz, 2009). However, since the other crisis period fixed effects are not significant this suggests shortfall is driven more by individual manager behaviour than broad market stresses.

Table 4.8 uses the size of the annual shortfall in an OLS regression. The results are similar to those reflected in the probit but tests which factors would lead to more assets being reported missing. It is the poor flows over poor past relative performance that stands out as affecting more months of the disclosure. A lack of a high-water mark, being offshore and higher incentive fees remain significant influences on the extent of shortfalls. Interestingly, larger funds are more likely to show less months of information when in a shortfall position. One explanation is that larger funds have a wider client base and so are better able to weather withdrawals allowing them to take more risk with non-disclosure. Another is that larger funds might take longer to change portfolio strategy given their size, and coupled with being less conservative in

[87] The role of fees was estimated in an earlier specification of Table 4.6 but had no effect on the fund's shortfall over the fund's lifetime.

risk-shifting (Aragon and Nanda, 2012), require more time 'under the radar.'

In summary, poor relative performance, prior fund outflows, lower scrutiny (offshore and no audit) and a decrease in fund restrictions (total lockup restrictions or no watermark) increases the chance of a fund witholding asset values. This could be because the fund is being more exposed to withdrawal risk, resulting in the pre-emptive need to obfuscate the fund's asset position.

4.5.4 Are changes in asset reporting informative about potential fraud risks?

Table 4.9 presents results of the performance flags and shows the percentages of non-problem and problem funds triggering the flags. The p-value of the test of differences between the groups rejection rates is also provided. The significance level chosen for the tests is 10% so this forms a base line rate of rejection even if returns are accurate. Attention should be focused on trigger rates above this level. Ideally, the tests would, as Bollen and Pool (2012) suggest, trigger at a higher level in problem funds to minimise Type I error and to trigger at a lower rate for non-problem funds to minimise Type II error.

Majority of the tests show the problem funds, those with weaker asset disclosure, are triggering the flags at a greater frequency than non-problem funds in Table 4.9. An exception is the 'String', 'Num. Pairs', and 'CAR(1)' tests which trigger at a slightly lower rate for the problem funds but these results are not much greater than the significance level of 10%. The tests of the second digit with the Uniform and Benford distribution were not markedly difference between the groups. The results that are further away from the base significance level, namely 'Perc. Negative', 'Count Zeros', and 'Perc. Repeats', are more informative. In these cases the problem funds are being triggered at a much higher rate and this difference is highly significant.

These results suggest some funds with shortfall in their asset disclosure face a higher risk of fraud. I must qualify that these tests are overlapping to some extent, and yet might not be exhaustive.[88]

As discussed in Section 4.4, critical levels for these tests are determined on a fund specific basis through simulation. This mitigates concerns this effect is owing to group characteristics like different rounding levels or fund history lengths. Table F.21 in Appendix F reports the summary statistics of the simulation inputs for the two groups and finds they are broadly similar. The higher fund age of problem funds is consistent with Brown et al. (2009)'s findings.

For non-problem funds, the percentage of negative returns being lower than expected is triggering at a much higher rate, 22.2%, than the expected 10%. The 'AR(1)' trigger rate is also high. This is not unexpected for hedge funds given findings of return smoothing in hedge funds (Bollen and Pool, 2008; Getmansky et al., 2004), and tweaking of negative returns reported by Bollen and Pool (2009). The simulation did not build in such serial correlation.

Furthermore there might be other concerns regarding the distributional assumptions for the simulation. Although alternative heavy tailed distributions could be postulated, it is the process to discretize the returns to deal with the nature of the tests that is the essential element.[89]

Naturally, it can be argued that a fund committing fraud might do all in its power not to raise any suspicions. These funds would maintain asset reporting and remain in the problem funds group. An alternative is for a manager to choose to simply not report anything at all to the databases — neither returns, nor assets. These are certainly possibilities, however such behaviour would bias against what is observed, narrowing the differences between the groups.

[88]An assessment of the correlations between flags, and a probit analysis of the shortfall conditional on fraud flags can be found in Appendix F, Table F.34.

[89]Alternative specifications using other distributions, such as a symmetric scaled t-distribution, did not change the results.

It remains to ask, whether shortfall in asset disclosure is a different phenomenon from the unreliable disclosure of returns discovered in Chapter 3? I have already mentioned funds in shortfall have lower, not higher, total restrictions compared to the funds that revise returns. Of the 6,654 funds in this sample that revise returns, only 32.7% are in shortfall. This suggests that funds weakening asset disclosure are not necessarily the same funds as those revising. Applying the fraud frequency flags to the revisers records in Table 4.10 that reviser funds do not trigger the same flags with the same intensity. This is summarised by the signs in the last column. 'Perc. Negative' and 'AR(1)' is triggered at a higher rate, 'Perc. Repeats' lower, but the 'Count Zeros' flag is not triggered. The trigger rates for distribution of digits is significantly smaller than non-revisers. 'String' and 'Num. Pairs' are also much lower and is consistent with the finding about their higher volatility of returns of revisers. It seems that a shortfall in asset disclosure is distinct, and should be considered in addition to return revisions.

The next section explores the robustness of the findings.

4.6 Robustness checks

In this section the main findings are put through a number of robustness checks to validate them. I vary parameters in the analysis, consider removing funds-of-funds, and examine the individual results for databases separately. Appendix F presents these additional robustness checks and analyses.

4.6.1 Varying the minimum months missing threshold

I consider allowing for a minimum number of months 'grace' of missing assets before a fund is labelled as being in shortfall. In other words, a shortfall is counted only if m months or more of assets are missing. So $m - 1$ months of grace are given, and this amount of missing assets months are added back, if needed, to the asset count in calculating the shortfall. For example, if m is 3, and asset count is 10 and return count is 12, then two months are missing and the fund is not classified as being in shortfall. Since disclosure is dependent on discrete months, this route was chosen as opposed to varying by a percentage of shortfall. Although a consequence of this is the percentage shortfall forgiven will vary on a fund-by-fund basis depending on the length of the fund's return count. Although it seems unusual that a manager can report returns but not assets, this tests the sensitivity of the shortfall to occasional lapses in asset reporting.

In Table 4.11 Panel A, the missing months threshold, m, varies from one, three, six, and then twelve months. The results on the drivers of weak asset disclosure remain consistent even out to a year of missing assets being ignored. This suggests that those engaging in weakening asset disclosure are either hiding substantial asset information, or alternatively, funds with weak access restrictions and smooth return profiles are likely to hide information even over relatively short horizons.

The fraud flag frequencies are also consistent if m varies, triggering for the flags of interest, those greater than the 10% significance base level, at similar or slightly higher rates, as can be seen in the Appendix F, Table F.31 Panel A.

4.6.2 Varying the delay asset reporting threshold

I also consider widening the return to asset reporting gap to 'forgive' some late reporting on assets. It is not clear how a manager can be confident enough to report returns, a quantity which will determine remuneration through incentive fees, but not the asset value, the key input into determining the return. Nevertheless, further analysis is considered in which assets can lag reporting by k months, where k varies from one, three, six and then twelve months, without being considered as a shortfall in asset reporting. This was implemented by shortening the period in which shortfall was estimated to remain at least k months away from vintage date.

In Table 4.11 Panel B the effect of changing the recency of reporting shows the results remain consistent. The baseline results are shown for $k = 1$. The disclosure of assets is not purely affected by short-term delays in asset reporting.

The fraud flag frequencies are also consistent if recency, k, varies, triggering at similar rates to the base case, $k = 1$, as can be seen in the Appendix F, Table F.31 Panel B.

4.6.3 Excluding funds-of-funds

As mentioned in the robustness Section 3.6 in Chapter 3, the information reported by funds of hedge funds (FOFs) are of course a function of the data supplied to them from individual hedge funds in which the FOF is invested. If an individual fund suppresses asset information, then the FOF might be unable to update its own asset information.[90] To test if the shortfall effect is owing to such a knock-on effect, I test whether these results are robust to the removal of FOFs from the analysis.

Table F.32 in Appendix F replicate the results presented in Table 4.6 by excluding funds-of-funds from the estimation of the probit regressions on shortfall at the fund-level. The results are robust to this alternative specification. The smoothing result is stronger now via the standard deviation on returns than autocorrelation. The result on the negative effect on total restrictions is more significant for hedge funds. Thus the lack of institutional controls remains a significant predictor of poor asset disclosure for both individual hedge funds and funds-of-funds.

4.6.4 Empirical results for single databases

As discussed before in Chapter 3, the analysis not only tracks vintages of databases but also consolidates funds across the four larger hedge fund databases (TASS, HFR, CISDM, BarclayHedge). This process, described in detail in Appendix A, involves identifying funds that report to more than one database. If this occurs, then a single fund is chosen, and the fund's data, both assets and returns, are selected from a single database (choosing the database with the longest history for that fund, if more than one database is available). This rules out the possibility of vintage deletion results being driven by switching between database sources. I avoid such an approach as it

[90] In practice, a FOF has other alternatives such as contacting the manager they are invested in directly, estimating the figure given this hedge fund manager is a single holding and the FOF might have information on the fund's peers in similar strategies, or delaying their reporting cycle.

would be exposed to a manager reporting different levels of details to more than one vendor. However, as a robustness check to address any concerns with this process, I show the results from using just a single database.

Although incidence of shortfall does differ markedly between databases, the results are not driven by behaviour solely in one database. The probit in equation (4.2) for fund shortfall is estimated for each of the four databases. These give weaker but quantitatively similar results. A noticeable feature is the significance on the asset rank emerging at this level, although it differs by database and is not consistent on sign. These results are in the Appendix F in Table F.33.

4.7 Conclusions

In this chapter I examine the extent and reliability of voluntary disclosure of assets under management by hedge funds. The focus is the impact of deterioration of asset information to investors. I first compare the level of asset to return disclosure by funds. Furthermore, I track changes in disclosure through annual comparisons, and to historic information across time by comparing versions of the published fund data collected over time.

I find there is a shortfall of asset data relative to return data, and that in aggregate around a fifth of assets are missing. The shortfall in asset disclosure at the fund level has been improving over time, but it remains high and concentrated in funds with lower restrictions and with smoother return profiles. Interestingly, there is strategic timing by managers as weakening asset disclosure is related to times of poor relative performance and outflows.

To assess the benefit of disclosure and to assess manager 'quality,' I test whether the propensity to trigger fraud performance flags varies by disclosure. Managers with shortfall in asset disclosure trigger these flags with greater intensity. This suggests some benefit in consistent asset disclosure for investors., particularly as most of the funds in asset shortfall are not the same as returns revisers.

A further extension for this research is to test how assets under management disclosure behaviour differs within management companies. In the spirit of Nanda et al. (2004), management companies might be steering investors' attention to a subset of their funds to optimise inflows. In this 'flagship' hypothesis, a management company with more than one offering has more flexibility in the selection of funds to market. For example, by halting production of information on a fund that is not in favour, or, has reached capacity. This could be tested by checking for variation in shortfall within a management company, and for evidence of flows and performance influencing inconsistencies in asset information disclosure within the same company.

There is recent interest by regulators to require hedge funds to disclose more information. Given that larger funds, when withholding asset information, do so with more intensity, regulators can expect some lobbying resistance. This analysis suggests that mandatory, audited disclosures by hedge funds, such as those recently proposed by the SEC, and due to be implemented in 2012/2013, would be beneficial to both regulators and hedge fund investors. I recommend that equal regard be given to asset reporting and not just that of returns.

4.8 Appendix: Tables

Table 4.1: *Summary Statistics, Overall Dataset*

This table shows summary statistics on funds employed in the analysis, with time-series statistics in Panel A computed only using the May 2011 (final) vintage of the 40 vintages of data captured. AUM refers to assets under management. Panel A shows broad statistics on returns and AUM, Panel B shows the strategies into which the funds are classified, and Panel C shows the databases from which the funds are sourced.

Panel A: Fund Summary Statistics

	Num. Funds	Average Fund AUM U.S.$ MM	Average Fund Return	Average Fund History Length (years)
	18,243	103.19 m	0.639	5.508

Panel B: Fund Strategies

	Fund Count	Count%
Security Selection	2,981	16.34%
Macro	1,194	6.54%
Relative Value	249	1.36%
Directional Traders	2,346	12.86%
Funds-of-Funds	4,805	26.34%
Multi-Process	1,852	10.15%
Emerging	812	4.45%
Fixed Income	952	5.22%
Other	174	0.95%
Managed Futures	2,878	15.78%
	18,243	100.00%

Panel C: Funds by Database

	Fund Count	Count%
TASS	6,521	35.75%
HFR	4,709	25.81%
CISDM	1,694	9.29%
BarclayHedge	5,319	29.16%
	18,243	100.00%

Table 4.2: *Illustration of Simulated Percentiles*

This table illustrates the fund-specific bootstrapped percentiles for seven of the data-quality tests outlined in Section 4.4. Each Panel shows the result of different fund characteristic inputs used in the Monte Carlo simulation of 10,000 iterations of return series. A scaled t-distribution with degrees of freedom of six was used to simulate returns. In Panel A, Fund A is a hypothetical hedge fund with median return length, but tightly rounded. Panel B reflects a more volatile Fund B with a upper quartile volatility and fund length. Panel C re-considers Fund A differing only by rounding of the percentage to five, and not one, decimal places.

Panel A: Rounded Fund A ($\mu = 8\%$ p.a., $\sigma = 12\%$, rounded to 0.1%, $n = 60$)

	Percentiles								
	1	**5**	**10**	**25**	**50**	**75**	**90**	**95**	**99**
Perc. Negative	0.267	0.300	0.317	0.367	0.400	0.450	0.483	0.517	0.550
Count Zeros	0	0	0	0	1	1	2	2	3
String	1	1	1	1	1	2	2	2	2
Num. Pairs	0	0	0	0	0	0	1	1	1
Perc. Repeats	0.217	0.267	0.300	0.350	0.400	0.450	0.500	0.517	0.567
Uniform	4.000	6.333	7.667	10.333	13.667	17.667	22.333	25.333	31.667
Benford	3.824	5.905	7.291	9.937	13.308	17.591	22.214	25.657	33.274

Panel B: Volatile Fund B ($\mu = 12\%$ p.a., $\sigma = 18\%$, rounded to 0.001%, $n = 96$)

	Percentiles								
	1	**5**	**10**	**25**	**50**	**75**	**90**	**95**	**99**
Perc. Negative	0.292	0.333	0.344	0.375	0.406	0.448	0.479	0.490	0.531
Count Zeros	0	0	0	0	0	0	0	0	0
String	1	1	1	1	1	1	1	1	1
Num. Pairs	0	0	0	0	0	0	0	0	0
Perc. Repeats	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.021	0.042
Uniform	1.917	3.375	4.208	5.875	8.375	11.500	14.833	16.917	21.708
Benford	2.286	3.726	4.811	6.701	9.514	12.934	16.896	19.448	24.603

Panel C: Fund A Revisited ($\mu = 8\%$ p.a., $\sigma = 12\%$, rounded to 0.00001%, $n = 60$)

	Percentiles								
	1	**5**	**10**	**25**	**50**	**75**	**90**	**95**	**99**
Perc. Negative	0.275	0.319	0.333	0.377	0.406	0.449	0.493	0.507	0.551
Count Zeros	0	0	0	0	0	0	0	0	0
String	1	1	1	1	1	1	1	1	1
Num. Pairs	0	0	0	0	0	0	0	0	0
Perc. Repeats	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Uniform	2.159	3.319	4.188	5.928	8.246	11.435	14.623	16.942	21.870
Benford	2.179	3.560	4.532	6.489	9.143	12.513	16.161	18.807	24.527

Table 4.3: *Asset Disclosure Shortfall Through Time*

This table assesses the aggregate shortfall of missing asset data, relative to return data, published by funds in each calendar year. Shortfall is calculated over the period returns are reported, and, is the proportion of asset months missing over the number of available return data months reported by the fund. Data is taken from the last available vintage of a fund. Panel A shows the contribution on an annual basis. The Funds at Start column is the total of fund's reporting returns as at the start of the calendar year. Return and Asset Counts are the total months of data reported in the year by these funds, over the period where funds reported returns. Averages are calculated on an aggregate basis across all months. Note 2011 is a partial year and so is not shown. Panel B shows the average returns and assets conditional on asset value availability, and the difference (Diff.) between the Assets Missing and Assets Not Missing values.

Panel A: Annual Proportions of Missing Assets

Year	Funds at Start	Return Count	Asset Count	Aggregate Annual Shortfall	Avg Monthly Return	Avg Monthly AUM U.S.$ MM
1994	1,967	23,595	16,544	0.299	0.247	58.207
1995	2,605	30,619	22,604	0.262	1.315	50.779
1996	2,998	35,105	28,152	0.198	1.334	58.825
1997	3,369	39,358	32,081	0.185	1.298	74.063
1998	3,745	43,171	35,355	0.181	0.409	86.622
1999	4,040	46,592	38,213	0.180	1.860	80.961
2000	4,406	50,304	41,156	0.182	0.780	98.613
2001	4,668	53,911	43,912	0.185	0.486	104.403
2002	5,306	61,471	48,939	0.204	0.253	121.947
2003	6,069	70,490	55,907	0.207	1.347	130.659
2004	6,950	80,709	64,471	0.201	0.694	177.658
2005	7,995	92,523	73,140	0.209	0.731	229.759
2006	8,796	101,205	80,844	0.201	0.912	259.294
2007	9,336	105,829	86,338	0.184	0.859	309.557
2008	9,332	102,338	87,522	0.145	-1.322	301.901
2009	8,101	83,809	74,234	0.114	1.291	227.069
2010	6,844	75,606	69,619	0.079	0.645	215.422
Aggregate		1,096,635	899,031	0.180	0.773	152.102

Panel B: Returns Conditional on Asset Availability						
	Average Return			**Median Return**		
Year	**Assets Missing**	**Assets Not Missing**	**Diff.**	**Assets Missing**	**Assets Not Missing**	**Diff.**
1994	0.016	0.345	-0.329	0.092	0.130	-0.038
1995	1.498	1.250	0.248	1.200	0.859	0.341
1996	1.570	1.276	0.294	1.260	1.018	0.242
1997	1.420	1.270	0.150	1.160	1.000	0.160
1998	0.343	0.423	-0.081	0.701	0.610	0.091
1999	1.933	1.844	0.088	1.175	1.010	0.165
2000	0.904	0.752	0.152	0.800	0.670	0.130
2001	0.575	0.466	0.109	0.620	0.522	0.098
2002	0.332	0.233	0.099	0.350	0.290	0.060
2003	1.386	1.337	0.049	0.900	0.850	0.050
2004	0.716	0.688	0.027	0.559	0.520	0.039
2005	0.782	0.718	0.064	0.759	0.640	0.119
2006	0.902	0.914	-0.012	0.870	0.821	0.049
2007	0.838	0.863	-0.025	0.850	0.800	0.050
2008	-0.971	-1.381	0.410	-0.010	-0.490	0.480
2009	1.203	1.303	-0.100	0.849	0.760	0.089
2010	0.479	0.659	-0.180	0.450	0.570	-0.120
Average	0.819	0.762	0.057	0.740	0.622	0.118

Table 4.4: *Asset Disclosure Shortfall by Fund*

This table shows the extent of asset disclosure at the fund level. Shortfall is calculated over the period returns are reported, and, is the proportion of asset months missing over the number of non-missing return data months reported by the fund. Data is at the last available vintage of data captured for the fund. The final vintage is collected in May 2011. Panel A shows the number of funds with a shortfall in asset disclosure. Panel B shows broad statistics on the range of asset disclosure shortfall for those funds missing at least 1 month of asset data. Panel C shows the proportion of funds if the minimum threshold in missing months of asset data relative to return months is varied. For example, 6,026 funds have a disclosure shortfall of 3 or more months of missing asset data. Panel D conditions the proportion of funds in shortfall by fund size, measured by lifetime assets under management and grouped into terciles. For example, in the largest third of funds by size 2,627 funds are in shortfall (i.e. 43.2% of funds in this tercile is missing asset data). Panel E conditions the proportion of funds in shortfall by the total number of funds within the fund's management company (mgtco), grouped into terciles. So the high tercile indicates the fund has lots of sibling funds within the same management company. Panel F conditions the proportion of funds in shortfall by total restrictions, measured as the sum of the reported lockup and redemption notice periods, grouped into terciles.

Panel A: Proportion of Funds with Shortfall

		Shortfall	
	Fund Count	**No**	**Yes**
Funds	18,243	10,841	7,402
% of Total Funds		59.4%	40.6%

Panel B: Size of Shortfall

Funds with Shortfall	**All**
Count	7,402
Mean	0.394
Median	0.355
99th perc	0.970
95th perc	0.898
75th perc	0.667
25th perc	0.094
5th perc	0.016
1st perc	0.007

Panel C: Minimum Missing AUM Months Count

At Least	**Fund Count**	**% of Total Funds**	**Avg. Shortfall**
1 month	7,402	40.6%	39.4%
3 months	6,027	33.0%	47.5%
6 months	5,320	29.2%	52.3%
1 year	4,443	24.4%	57.8%
2 years	3,228	17.7%	64.3%
3 years	2,239	12.3%	70.3%
5 years	1,155	6.3%	76.6%

Panel D: Funds with Shortfall by Size			
Size Tercile	**Shortfall Fund Count**	**% of Funds in Shortfall**	**Avg. Shortfall**
High	2,627	43.2%	40.1%
Medium	2,505	41.2%	40.1%
Low	2,270	37.3%	37.6%
Panel E: Funds with Shortfall by Funds in Mgtco Count			
Management Company Fundcount Tercile	**Shortfall Fund Count**	**% of Funds in Shortfall**	**Avg. Shortfall**
High	2,870	50.1%	44.2%
Medium	1,875	39.0%	36.9%
Low	2,657	34.5%	35.9%
Panel F: Funds with Shortfall by Total Restrictions			
Total Restrictions Tercile	**Shortfall Fund Count**	**% of Funds in Shortfall**	**Avg. Shortfall**
High	1,963	33.4%	36.6%
Medium	2,588	41.8%	37.6%
Low	2,851	46.2%	42.9%

Table 4.5: *Asset Disclosure Shortfall Across Vintages*

This table assesses the change across vintages of the extent of shortfall in fund-level asset disclosure published by funds. Shortfall is calculated over the period returns are reported, and, is the proportion of asset months missing over the number of non-missing return data months reported by the fund. An increase in shortfall indicates worsening asset disclosure. First indicates the first vintage at which there is non-zero and non-missing asset values. Last shows the coverage at the fund's last vintage where assets were reported. Single Vintage means the funds first and last vintage where assets were reported is the same. Since there is no change by definition, these funds were excluded from the broader 'No Change' category. The 'No Change' category has the first and last vintage being distinct and having the same percentage of missing values (10,295 funds in this group had perfect matching of assets with returns, i.e. no assets missing at distinct vintages, a shortfall of 0%. This is 56.4% of all funds).

Panel A: Comparison of Missing Asset Proportions at First and Last Vintages

			Avg. Shortfall		
	Fund Count	**Count%**	**First**	**Last**	**Avg. Change**
Shortfall Worse	1,664	9.12%	0.180	0.395	0.215
Shortfall Better	3,326	18.23%	0.545	0.386	-0.159
No Change	13,138	72.02%	0.072	0.072	-
Single Vintage	115	0.63%	0.217	0.217	-
All Funds	18,243	100.00%	0.169	0.160	-0.009

Panel B: Asset Shortfall Summary Statistics Across Vintages

	No Change		**Shortfall Worse**		**Shortfall Better**	
	First	**Last**	**First**	**Last**	**First**	**Last**
Shortfall						
Mean	0.072	0.072	0.180	0.395	0.545	0.386
Median	0.000	0.000	0.025	0.342	0.629	0.386
Std Deviation	0.196	0.196	0.262	0.310	0.320	0.308
75th Percentile	0.000	0.000	0.291	0.697	0.826	0.667
25th Percentile	0.000	0.000	0.000	0.092	0.234	0.054
Missing Months						
Mean	4.1	4.1	10.9	32.6	34.7	33.1
Median	0	0	1	17	24	23

Table 4.6: *Probit Regression for Shortfall*

The table shows the marginal effects from a probit regression. The dependent variable, the Aggregate Shortfall Indicator, takes the value of 1 if a fund's reported asset months are below that of the period of the reported returns over the fund's lifetime (i.e. the funds aggregate shortfall is greater than zero). Data is taken from the funds last available vintage. The independent variables are lifetime average returns and lifetime average AUM, standard deviation of returns, and the autocorrelation of returns, all measured as ranks relative to the other funds in the data; and the number of return observations in the return history of the fund. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located Offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods), and a flag which takes the value of 1 if there is any information pertaining to audits available in any of the databases. I also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Z-stat**
Lifetime Avg. AUM rank	-0.076	(-0.791)
Lifetime Avg. Return rank	0.028	(0.770)
Lifetime Ret. Std. rank	-0.068**	(-2.327)
Return Autocorrelation rank	0.025**	(1.994)
Return History Length	0.015***	(2.986)
Offshore	0.021	(0.459)
Total Restrictions	-0.020***	(-3.222)
Audit	0.026	(0.557)
High-Water Mark or Hurdle	-0.142***	(-3.448)
Database Fixed Effects		
HFR	-0.441***	(-39.461)
CISDM	-0.148***	(-7.499)
BarclayHedge	-0.311***	(-45.282)
Strategy Fixed Effects		
Macro	0.026	(0.823)
Relative Value	-0.036**	(-2.162)
Directional Traders	0.095***	(3.333)
Funds-of-Funds	0.061**	(2.111)
Multi-Process	0.141***	(5.290)
Emerging	-0.019	(-0.851)
Fixed Income	0.002	(0.042)
Other	-0.017	(-0.931)
Managed Futures	-0.161***	(-4.459)
Observations	18,243	
Pseudo R^2	0.192	

Table 4.7: *Probit Regression for Annual Shortfall*

This table runs essentially the same specification as in Table 4.6, the difference is that I employ the panel structure of the data, and the fund-year is now the unit of analysis. The dependent variable, the Annual Shortfall Indicator, takes the value of 1 if the fund's reported asset months are below that of the period of the reported returns over that year (i.e. the shortfall in year T is greater than zero). Data is taken from the fund's last available vintage. The ranks of the fund variables are therefore now measured using data in prior year $T-1$ on assets under management, returns and flows. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located offshore, a total restrictions variable, a flag which takes the value of 1 for the fund if there is any information pertaining to audits available, and the fund's management and incentive fee. Database, strategy and crisis period fixed-effects are included in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by year. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx $\times 100$	**Z-stat**
Asset prior year t-1 rank	0.100	(0.321)
Return prior year t-1 rank	-1.211***	(-4.175)
Flow prior year t-1 rank	-0.976***	(-4.673)
Offshore	0.488**	(3.156)
Total Restrictions	-0.048	(-1.387)
Audit	-1.007***	(-3.333)
High-Water Mark or Hurdle	-1.201***	(-3.528)
Management Fee	0.033	(0.316)
Incentive Fee	0.041***	(3.461)
Database Fixed Effects		
HFR	-6.722***	(-12.091)
CISDM	-1.675***	(-3.492)
BarclayHedge	-4.647***	(-10.486)
Strategy Fixed Effects		
Macro	-0.359	(-1.205)
Relative Value	-0.210	(-0.309)
Directional Traders	-0.552	(-1.785)
Funds-of-Funds	0.038	(0.081)
Multi-Process	0.118	(0.609)
Emerging	-0.642	(-1.609)
Fixed Income	-0.708**	(-2.688)
Other	-1.358***	(-3.842)
Managed Futures	-1.912***	(-6.843)
Period Fixed Effects		
Crisis period (1998–1999)	2.002***	(5.316)
Crisis period (2000–2001)	0.029	(0.068)
Crisis period (2008–2009)	0.371	(1.055)
Observations	51,775	
Pseudo R^2	0.180	

Table 4.8: *Explaining Size of Annual Shortfall*

This table conditions the annual shortfall on various fund characteristics and period fixed effects. The dependent variable is the annual shortfall, for all calendar years in which a fund reported returns in January, based on the last available vintage of data for the fund. I only include periods in which the fund had at least six months of return observations, to reduce the noise in the dependent variable. Period dummies include crisis dummies for the 1998–1999 period, the 2000–2001 period, and the 2008–2009 period. The independent variables are the fund's AUM at the end of the prior year, prior year annual return, and the prior year standardised flows. These are all measured as ranks relative to the other funds in the data. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods), a flag which takes the value of 1 if there is any information pertaining to audits available in any of the databases, and the management fee and the incentive fee of the fund. I also include strategy fixed-effects in the regressions. Shortfall was expressed as a percentage, i.e. 75 for 75%. t-statistics, shown in parentheses, are robust to heteroskedasticity and clustered at the fund-level. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	(1)			(2)			(3)		
	Coeff.	***t*-stat**		**Coeff.**	***t*-stat**		**Coeff.**	***t*-stat**	
Constant	3.970	(14.725)	***	1.733	(5.156)	***	4.240	(9.760)	***
Crisis dummy1: 1998–99	2.089	(7.622)	***	1.765	(6.600)	***	1.869	(6.984)	***
Crisis dummy2: 2000–01	0.177	(1.017)		0.025	(0.142)		0.015	(0.087)	
Crisis dummy3: 2008–09	-0.202	(-1.598)		-0.163	(-1.274)		-0.083	(-0.660)	
Asset t-1 rank				1.129	(3.852)	***	0.653	(2.383)	**
Return prior year t-1 rank				-0.243	(-1.421)		-0.271	(-1.562)	
Flow prior year t-1 rank				-1.000	(-5.771)	***	-0.932	(-5.437)	***
Offshore				0.811	(4.129)	***	0.448	(2.249)	**
Total Restrictions				0.125	(2.639)	***	0.049	(0.994)	
High-Water Mark or Hurdle				-1.202	(-5.463)	***	-1.304	(-5.660)	***
Audit				0.341	(1.365)		-0.297	(-1.074)	
Management Fee				-0.374	(-3.912)	***	-0.093	(-0.939)	
Incentive Fee				0.044	(4.245)	***	0.055	(4.292)	***
Strategy Fixed Effects	Y			N			Y		
N	51,775			51,775			51,775		
Adjusted R^2	0.016			0.008			0.020		

Table 4.9: *Flag Frequencies*

This table shows the proportion of funds triggering the performance flags. The problem funds group are the funds that experienced a shortfall in asset values at the date of the last available vintage. The non-problem group is the remaining funds showing full asset histories. Returns are taken over the funds full history. Funds require a minimum of 24 months of returns. The stars indicate the results of a test of differences in trigger rejection rates of problem and non-problem funds. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively. The sign in the last column indicates the direction the flag is triggered when there is a significant difference. For example, a positive sign indicates the fraud flag is triggered at a higher rate in the problem group compared to the non-problem group.

	Non-problem Funds	**Problem Funds**		
Flag	$(N = 8,971)$	$(N = 6,624)$	***p*-value**	
Perc. Negative	0.221	0.283***	0.000	+
Count Zeros	0.137	0.231***	0.000	+
String	0.113	0.104**	0.050	−
Num. Pairs	0.055	0.049	0.113	
Perc. Repeats	0.174	0.200***	0.000	+
Uniform	0.140	0.141	0.880	
Benford	0.119	0.119	0.877	
AR(1)	0.368	0.436***	0.000	+
CAR(1)	0.121	0.111*	0.069	−

Table 4.10: *Flag Frequencies for Revisers and Non-revisers*

This table shows the proportion of funds triggering the performance flags. The problem funds group are now the group of funds that have revised historic returns at least once over the 40 vintages, i.e. revisers. The non-problem group is the remaining funds showing no changes in return histories. Returns are taken over the funds full history. Funds require a minimum of 24 months of returns. The stars indicate the results of a test of differences in trigger rejection rates of problem and non-problem funds. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively. The sign in the last column indicates the direction the flag is triggered when there is a significant difference. For example, a positive sign indicates the fraud flag is triggered at a higher rate in the problem group compared to the non-problem group.

	Non-reviser Funds	**Reviser Funds**		
Flag	$(N = 9,469)$	$(N = 6,126)$	***p*-value**	
Perc. Negative	0.189	0.338***	0.000	+
Count Zeros	0.180	0.173	0.217	
String	0.124	0.086***	0.000	−
Num. Pairs	0.065	0.032***	0.000	−
Perc. Repeats	0.180	0.193**	0.033	+
Uniform	0.150	0.127***	0.000	−
Benford	0.128	0.105***	0.000	−
AR(1)	0.324	0.509***	0.000	+
CAR(1)	0.116	0.117	0.875	

Table 4.11: *Robustness Checks: Missing Threshold and Recency*

This table conditions the results in Table 4.6 after adjusting for an allowance of some months of missing asset data, and the recency, k, of disclosure. Panel A shows the impact of 'forgiving' $m-1$ months of non-disclosure of assets before the shortfall indicator is flagged. For example, the first column $m \geq 1$ of Panel A reproduces the results from Panel A of Table 4.6, and $m \geq 3$ only treats funds with three months missing as being in shortfall. Panel B shows the impact of excluding recent shortfalls in disclosure near the vintage date. For example, when $k > 3$ only funds with shortfalls that occur more than three months prior to the date of the vintage are included, and when $k > 12$, only funds which stop reporting assets over a year old are included in the shortfall indicator. The columns show the regression marginal effects, dF/dx. This is the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

Panel A: Minimum Months Missing Shortfall Threshold

	Minimum Missing Months			
	$m \geq 1$	$m \geq 3$	$m \geq 6$	$m \geq 12$
Lifetime Avg. AUM rank	-0.076	-0.047	-0.039	-0.037
Lifetime Avg. Return rank	0.028	0.050	0.049	0.058
Lifetime Ret. Std. rank	-0.068**	-0.082***	-0.084***	-0.075***
Return Autocorrelation rank	0.025**	0.030***	0.028	0.050***
Return History Length	0.015***	0.016***	0.018***	0.021***
Offshore	0.021	0.017	0.007	0.003
Total Restrictions	-0.020***	-0.018***	-0.018***	-0.018***
Audit	0.026	0.075**	0.065*	0.052
High-Water Mark or Hurdle	-0.142***	-0.096**	-0.078**	-0.065**
Database Fixed Effects				
HFR	-0.441***	-0.344***	-0.284***	-0.220***
CISDM	-0.148***	-0.071***	-0.055**	-0.056**
BarclayHedge	-0.311***	-0.255***	-0.202***	-0.154***
Strategy Fixed Effects				
Macro	0.026	0.022	0.003	0.016
Relative Value	-0.036**	-0.031	-0.029	-0.017
Directional Traders	0.095***	0.100***	0.099***	0.101***
Funds-of-Funds	0.061**	0.071**	0.076**	0.083***
Multi-Process	0.141***	0.126***	0.119***	0.112***
Emerging	-0.019	-0.040**	-0.039**	-0.040***
Fixed Income	0.002	-0.011	-0.015	-0.017
Other	-0.017	-0.085***	-0.064***	-0.056***
Managed Futures	-0.161***	-0.196***	-0.183***	-0.157***
N	18,243	18,243	18,243	18,243
Pseudo R^2	0.192	0.191	0.175	0.178

Panel B: Recency of Shortfall				
	Minimum Recency of Shortfall			
	$k = 1$	$k > 3$	$k > 6$	$k > 12$
Lifetime Avg. AUM rank	-0.076	-0.078	-0.070	-0.064
Lifetime Avg. Return rank	0.028	0.021	0.014	0.009
Lifetime Ret. Std. rank	-0.068**	-0.060**	-0.057**	-0.046**
Return Autocorrelation rank	0.025**	0.025*	0.028**	0.031**
Return History Length	0.015***	0.015***	0.016***	0.016***
Offshore	0.021	0.017	0.019	0.020
Total Restrictions	-0.020***	-0.019***	-0.019***	-0.018***
Audit	0.026	0.022	0.018	0.007
High-Water Mark or Hurdle	-0.142***	-0.141***	-0.140***	-0.141***
Database Fixed Effects				
HFR	-0.441***	-0.429***	-0.420***	-0.405***
CISDM	-0.148***	-0.139***	-0.134***	-0.130***
BarclayHedge	-0.311***	-0.304***	-0.300***	-0.296***
Strategy Fixed Effects				
Macro	0.026	0.024	0.020	0.007
Relative Value	-0.036**	-0.033**	-0.030*	-0.032
Directional Traders	0.095***	0.088***	0.084***	0.076**
Funds-of-Funds	0.061**	0.061**	0.059**	0.060**
Multi-Process	0.141***	0.139***	0.129***	0.129***
Emerging	-0.019	-0.030	-0.040***	-0.065***
Fixed Income	0.002	0.003	0.000	-0.012
Other	-0.017	-0.019	-0.019	-0.030
Managed Futures	-0.161***	-0.159***	-0.156***	-0.152***
N	18,243	18,243	18,243	18,243
Pseudo R^2	0.192	0.186	0.183	0.181

"Sunlight is said to be the best of disinfectants; electric light the most efficient policeman."
— Louis D. Brandeis *Other People's Money*, 1914.

5 Conclusion

This thesis has ranged across reported investment and management fees, changes in returns, and finally, the extent of asset disclosure. Hedge fund data, including funds-of-funds and CTAs, from a large consolidated dataset, has been used as the backdrop. The central thread between these three empirical works is the forensic nature of digging, beneath the voluntarily published data, to unearth the elements of interest.

In the chapter on fees, the time variation of reported management fees and incentive fees is detected through making use of fund launches over the period 1995 to 2009. This allows us to condition on management company characteristics to gain more insight into fee pricing levels. Management fees in the hedge fund industry are lower on average than the oft-quoted 2%, but this work shows management fees in new fund launches have been rising in the 2000s. It is the better performing and larger management companies that are exercising this pricing power. More importantly, the value capture of higher fees rewards skilled managers over their investors in the case of management fees, although there is still some value shared for incentive fees for raw returns. Also techniques from mutual fund and labour economics literature are employed, using management companies and interval regression respectively, to deepen our understanding of hedge funds and funds-of-funds.

The core of this thesis is the chapter on return revisions, which through the use of snapshots of hedge fund data, over the period from 2007 to 2011, reveals voluntarily disclosure by hedge funds is unreliable and prone to change. This is an important result that extends our thinking about hedge fund performance data, as it shows that returns are affected by the 'when' of data capture, the vintage dimension, and not just the traditional 'which month', the time dimension of returns. The time horizon (going back as periods as far back as fifteen years) and breadth (40% of 18,382 hedge funds and 45% of assets) of these return changes underlines this is not a short-term revision story, typically seen with economic data like GDP revisions. Revisions are also not isolated to a single database.

From an asset pricing perspective, on average these returns are revised downwards, showing that the effect of revisions is to dampen initial hedge fund manager optimism. Interestingly, the performance of funds that revise historic returns *subsequently* underperforms those that never revise their returns. This differential is robust to risk adjustments, extreme values, fund type, database, and, recency and size of revisions. Although fund illiquidity plays a role in revisions, this is not a sole driver given the age of revisions, the positive bias in initial estimates, and variation in performance still existing in more liquid funds. Detractors might point out that the performance result is over a modest interval, at a time where funds experienced extreme volatility. Nevertheless, the return differential is secondary to the contribution of this thesis, which is to highlight direct evidence of the unreliability of voluntary disclosure.

A future extension of the return revisions work would be to marry return with asset revisions, to determine if, and when, changes in returns are echoed in asset valuations. Further research investigating changes in administrative characteristics, such as management company or auditor switches, could also prove fruitful in outlining a subset of potential revision triggers. Likewise, as the hedge fund industry consolidates into large management companies, the role of revisions and consistency of disclosure at the management company, and not mere fund level, is a substantial future line of enquiry.

In the last chapter, a connection between the publication of assets in step with the returns these assets generated, is explored. Although aware that returns observed in the most recent vintage are not concrete, this work turns instead to 'what is missing?' with asset disclosure. A significant shortfall in the level of asset to return disclosure is found. Unlike return revisions, these shortfalls are more likely to be concentrated in funds with less restrictions, i.e. no high-water marks, or lower total restrictions. Periods of stress for funds precedes strategic timing by their managers to weaken asset disclosure. The consequences of voluntary disclosure of assets are two-fold. For researchers, pragmatic reliance on dual return and asset disclosure can lead to reaching for more resilient sets of funds. For investors, the spirit of *caveat emptor* prevails with inconsistent asset disclosure, as funds in shortfall trigger fraud flags with a greater propensity.

Throughout this thesis, I have emphasized the unreliability of voluntary disclosure. What you see is not always what you get, and, what is not shown has financial consequences for investors. This does not detract from the success and growth of hedge funds as a valuable investment vehicle, but cautions investors to take heed of verifying the information provided, and, to query omissions and changes of return and asset information.

The policy implications of this work are clear. Recent moves by regulators, like the SEC, for mandatory disclosures by hedge funds, would be beneficial to investors, regulators, and, the hedge fund industry. Disclosure should cover returns and assets. The disclosure framework should have a mechanism to manage changes and restatements. However, it cannot be forgotten that disclosure has a cost, often borne by investors, and there is no guarantee of accuracy. Regulators should strive for consistency, and not frequency, to ensure high-quality information for investor protection is the end game.

Appendices

A
The Consolidated Hedge Fund Database

As hedge funds can report to one or more databases, the use of any single source will fail to capture the complete universe of hedge fund data. We therefore aggregate data from TASS, HFR, CISDM, BarclayHedge, and Morningstar, which together have 74,742 records of fund entries that comprise administrative information as well as returns and AUM data for hedge funds, funds-of-funds, and CTAs. However this number hides the fact that there is significant duplication of information, as multiple providers often cover the same fund. To identify all unique entities, we must therefore consolidate the aggregated data. To do so, we adopt the following steps:

1. Group the Data: Records are grouped based on reported management company names. To do so, we first create a 'Fund name key' and a 'Management company key' for each data record, by parsing the original fund name and management company name for punctuations, filler words (e.g., 'Fund', 'Class'), and spelling errors. We then combine the fund and management name keys into 8,390 management company groups.

2. De-Duplication: Within a management company group, records are compared based on returns data (converted into U.S. dollars), and 27,395 match sets are created out of matching records, allowing for a small error tolerance limit (10% deviation) to allow for data reporting errors.

3. Selection: Once all matches within all management company groups are identified, a single record representing the unique underlying fund is created for each match set. We select the record with the longest returns data history available from the match set, and fill in any missing administrative information using the remaining records in the match set. The process thus yields 27,395 representative funds.

We filter the fund data in a few ways to ensure data integrity. For example, removing return outliers and quarterly reporting funds, and ensuring funds have sufficient return or asset information. We also remove the Morningstar funds (as less than a third passed these filters), to ensure sufficient depth by database. The result is around 18,382 funds.

B
Strategy Mappings

Table B.1 shows the broad strategies to which the underlying source strategies of the database vendors, HFR, TASS, CISDM, and BarclayHedge, are mapped. Examples of strategies are shown in the second column; the full set of more than 600 mappings is not shown. We also make use of fund type in the source database to aid in allocating an appropriate mapping. For example, a CTA with a source strategy dubbed 'Other' will be allocated to the Managed Futures strategy with the other CTAs, and not into the Other hedge fund category.

Table B.1: *Strategy Mapping Examples*

Mapped Strategy	**Examples of Source Strategies**
Security Selection	Equity Long/Short, Equity Arbitrage, Equity Long/Short - Growth Bias, Equity Market Neutral, Equity Market Neutral - U.S. Value Long/Short
Macro	Global Macro, Global Macro - FX only, Global Macro - Quantitative, Macro - Active Trading
Relative Value	Merger Arbitrage, Equity Market Neutral - Relative Value, Single Strategy - Event Driven Risk Arbitrage, Statistical Arbitrage
Directional Traders	Dedicated Short Bias, Equity Long Only, Equity Long/Short - Long biased, Market Timing, Single Strategy - Tactical trading
Funds-of-Funds	(By fund type) , Fund of Funds, Fund of Funds - Strategic, Conservative - Absolute Return Fund of Funds, Fund of Funds - Nondirectional, Fund of Funds - Derivatives
Multi-Process	Multi-process, Multi Strategy - Arbitrage, Equity Hedge - Multi-Strategy, Event Driven Multi Strategy
Emerging	Emerging Markets, Emerging Markets - Central Asia focus, Equity Long/Short - Emerging Markets, Emerging Markets - Directional, Emerging Markets - Global
Fixed Income	Convertible Arbitrage, Fixed Income - Arbitrage, Fixed Income - ABS/Sec. Loans, Fixed Income - Structured Credit, Global Debt, Distressed Securities - Stressed High Yield Bonds
Other	Other, Undefined
Managed Futures	(By CTA fund type), Managed Futures, Global trend, Systematic - Systematic arbitrage & counter-trend, Discretionary - CTA Managed Futures

C
Region Mappings

In Chaper 2, the country headquarters of the funds are mapped to specific regions. The countries allocated to the regional mapping are outlined below.

North America:

U.S., Canada and Mexico

Offshore:

Bahamas, Bermuda, British Virgin Islands, Cayman Islands, Channel Islands, plus other offshore islands.

Europe:

United Kingdom, Switzerland, Italy, France, Ireland, Luxembourg, Sweden, plus other European countries.

Other:

Australia, China, Hong Kong plus other countries — such as from Middle East and Africa.

D
Interval Regression

In outlining the interval regression approach in Chapter 2, we draw on work by Breen (1996)(p. 50-53) on multiple thresholds, and Maddala (1983)(p. 160-161) who considers a double threshold of minimum and maximum daily price limits in commodity trading.

Let y_{it}^o be the observed fee for fund i launching at time t (this is the reported discrete fee).

We assume that y^* is the unobserved underlying, continuous measure of the fee. This can be modelled as:

$$y_{it}^* = \beta' x_{it-1} + u_{it}$$

where x_{it-1} are appropriate regressors (lagged, as in our specifications) and u_{it} are mean zero error terms which are initially considered to be independent and normally distributed with variance σ^2. We do this for both management and then incentive fee for fund launches through time.

We assume that the reported fee lies within an interval I_i, with the interval breakpoints modeled as time-invariant. This is formed by selecting thresholds c_j from $j = 1...M$ and assigning y_{it} to its respective interval I_i:

$$\begin{aligned} I_i &= c_1 \text{ if } y^o_{it} \le c_1 \\ &= c_2 \text{ if } c_1 < y^o_{it} \le c_2 \\ &\quad \dots \\ &= c_M \text{ if } y^o_{it} > c_{M-1} \end{aligned}$$

In constructing the intervals, we consider two cases. First, we take the typical breakpoints as midpoints for intervals; and, second, we consider a 'rounding up' scenario in which a fee maximising manager would set fees by rounding upwards — this is equivalent to taking the right-hand-side of the intervals to be the typical breakpoints. We were parsimonious with the bands to both minimize the number of parameters and to ensure sufficient data in each interval. We kept the highest band to isolate the impact of the extreme fee values, setting c_M to the maximum fee observed in the data. Table F.3 shows the fees for the all funds aggregated into broad fee bands under two different bases for the purpose of the interval regression.

Let ϕ and Φ be the standardised density and cumulative distribution of the error distribution. The probability that the latent fee exceeds threshold c_j is:

$$\begin{aligned} \Pr[y^*_{it} &> c_j] = \Pr[\beta' x_{it-1} + u_{it} > c_j] \\ &= \Pr[u_{it} > c_j - \beta' x_{it-1}] \\ &= 1 - \Phi\left(\frac{c_j - \beta' x_{it-1}}{\sigma}\right) \end{aligned}$$

Then for the interval ending with c_j, using the notation $\Phi_{it-1}(c_j)$ for $\Phi\left(\frac{c_j - \beta' x_{it-1}}{\sigma}\right)$:[91]

$$\Pr[c_{j-1} < y^*_{it} \le c_j] = \Phi_{it-1}(c_j) - \Phi_{it-1}(c_{j-1})$$

[91] In the case $j = 1$, set $\Phi_{it-1}(c_0) = 0$.

Extending the simple Tobit case, Breen (1996) shows the log likelihood for these multiple thresholds is:

$$
\begin{aligned}
& L(\beta, \sigma \mid y_{it}, x_{it-1,}\ c_1 \ldots c_M) \\
= \ & \sum_{y \in I_1} \ln(\Phi_{it-1}(c_1)) + \sum_{y \in I_2} \ln(\Phi_{it-1}(c_2) - \Phi_{it-1}(c_1)) + \ldots + \sum_{y \in I_M} \ln(1 - \Phi_{it-1}(c_{M-1}))
\end{aligned}
$$

We then employ maximum likelihood estimation to back out the parameters β, σ.

E

Cross-Sectional Correlation and Autocorrelation Consistent Bootstrap

A non-parametric bootstrap is employed to compute standard errors in Chapter 2. Using the parametric bootstrap (drawing residuals) leaves the standard errors virtually unchanged (if anything they are slightly smaller on average). The bootstrap draws are conducted as follows: First, a random number is drawn from the sample $t = 1; ...; T$ (this is a cross-sectional correlation consistent bootstrap, so all cross-sections are automatically drawn for each time period). For the second resample observation, following Politis and Romano (1994), a uniform random variable is drawn from (0; 1). If it is less than a pre-set value Q, then the consecutive observation to the first bootstrap draw is also included (i.e., if $t = 4$ was first drawn, $t = 5$ is also included). If the draw is at the end of the sample, the procedure starts from the beginning again. If the uniform draw is greater than Q, a new observation is drawn (with replacement). Values of $Q = 0; 0.1; 0.5$, are tried, and this work reports the results for $Q = 0.5$. The other two choices for Q do not materially affect the results. 5,000 draws are used to compute the final bootstrap standard errors. The presence of a number of dummy variables in the data occasionally results in a null vector in the bootstrap trial X matrix. In such (rare) cases the point estimate of the bootstrap coefficient on the offending variable is set to zero (the null) and the remaining coefficients are estimated using the bootstrap sample.

F
Supplementary Tables and Figures

This appendix holds additional tables, for example robustness checks, supporting the work in the respective chapters.

F.1 Chapter 1

F.1.1 Tables

Table F.1: *Data for Illustration of Database Fragmentation*

This table contains the data underlying Figure 1.1, showing the percentage of funds covered by the consolidation of five underlying databases (as outlined in Appendix A). Namely, the merger of BarclayHedge, the Center for International Securities and Derivative Markets (CISDM), Hedge Fund Research (HFR), Morningstar, and Tremont Advisory Shareholder Services (TASS) (now Lipper) hedge fund databases. Panel A summarises the percentage of funds covered by a single database, and shows the numeric database identification (ID) code. Panel B shows all potential combinations of fund entry into these databases (labelled by these numeric ID codes as opposed to database names).

Panel A: Fund Coverage by a Single Database			
Database ID	**Database**	**Coverage%**	**Missing%**
1	TASS	43.12%	56.88%
2	HFR	46.41%	53.59%
4	CISDM	36.56%	63.44%
6	BarclayHedge	47.32%	52.68%
7	Morningstar	28.06%	71.94%

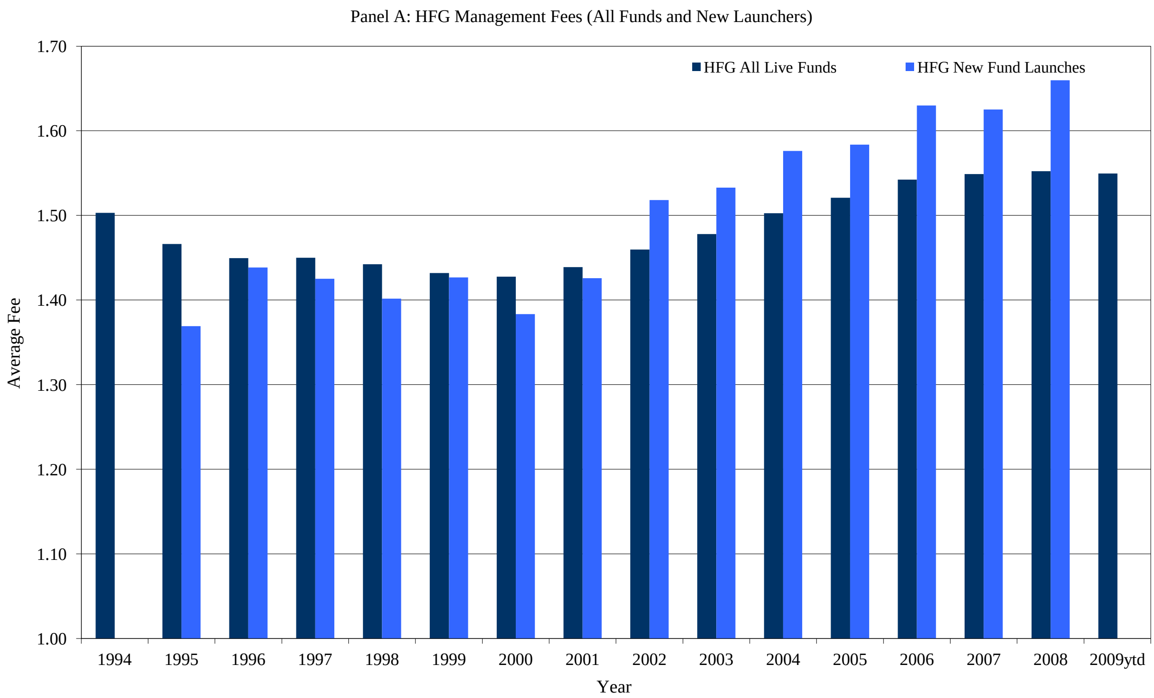
Panel B: Fund Coverage by Database		
Source Combination	**Fund Count**	**Fund%**
1, 2, 4, 6, 7	1,760	6.42%
1, 2, 4, 6	1,595	5.82%
1, 2, 4, 7	179	0.65%
1, 2, 4	509	1.86%
1, 2, 6, 7	635	2.32%
1, 2, 6	805	2.94%
1, 2, 7	240	0.88%
1, 2	618	2.26%
1, 4, 6, 7	152	0.55%
1, 4, 6	254	0.93%
1, 4, 7	61	0.22%
1, 4	382	1.39%
1, 6, 7	194	0.71%
1, 6	380	1.39%
1, 7	600	2.19%
1	3,449	12.59%
2, 4, 6, 7	424	1.55%
2, 4, 6	750	2.74%
2, 4, 7	88	0.32%
2, 4	478	1.74%
2, 6, 7	363	1.33%
2, 6	1,146	4.18%
2, 7	217	0.79%
2	2,906	10.61%
4, 6, 7	133	0.49%
4, 6	823	3.00%
4, 7	115	0.42%
4	2,313	8.44%
6, 7	249	0.91%
6	3,301	12.05%
7	2,276	8.31%
Total of Merged Funds	27,395	100.00%

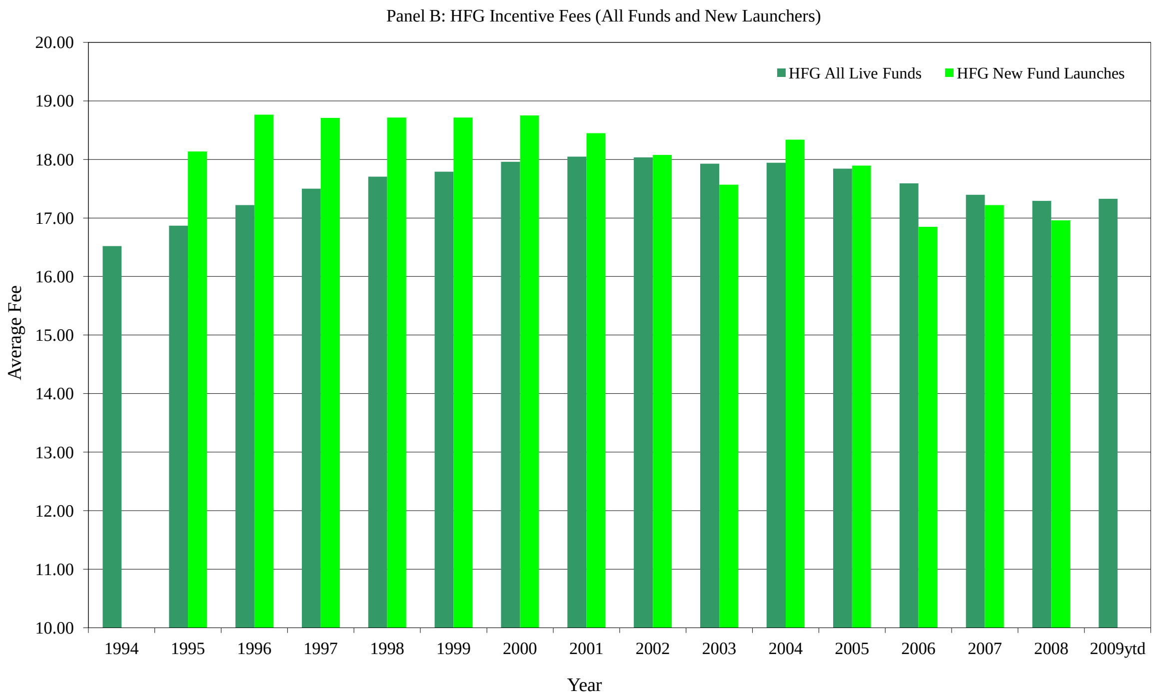
F.2 Chapter 2

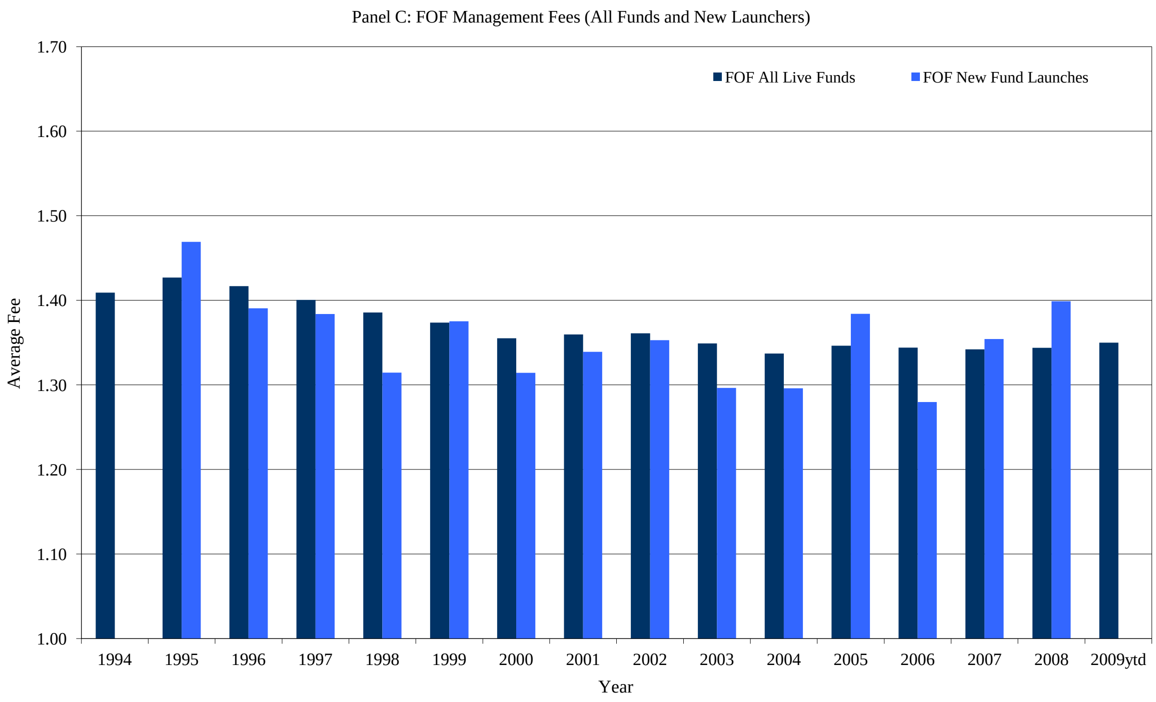
F.2.1 Figures

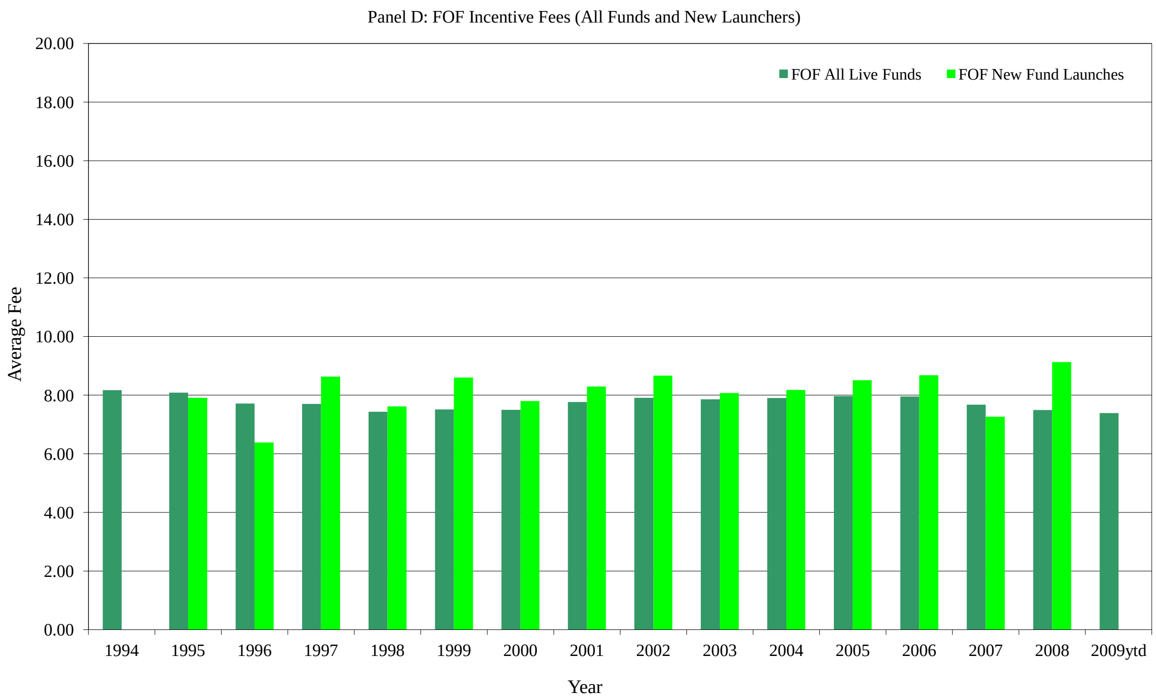
Figure F.1: *Fees Across Time by Fund Group*

These figures show the average fee values (first bar) for different types of funds across time. Namely, the Hedge Fund Group (HFG), and then the Funds-of-Funds (FOF). All live funds are all funds currently reporting returns at the end of each year. New fund launches (second lighter bar) refers to funds launching in that calendar year and gives an indication of the current fee setting trend. Panel A and C plots management fees, and Panel B and D plots incentive fees for the HFG and FOF groups respectively.






Panel C: FOF Management Fees (All Funds and New Launchers)
FOF All Live Funds
FOF New Fund Launches
Average Fee
1.70
1.60
1.50
1.40
1.30
1.20
1.10
1.00
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009ytd
Year


Panel D: FOF Incentive Fees (All Funds and New Launchers)
FOF All Live Funds
FOF New Fund Launches
Average Fee
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009ytd
Year

F.2.2 Tables

Table F.2: *Dataset Composition by Country*

This table shows the composition of the database by fund type and the underlying country headquarters for all funds with fee information up to June 2009, including defunct funds. Countries are grouped by their region.

	Hedge Fund Group (HFG)				
	Hedge Funds	**CTAs / CPOs**	**Total HFG**	**Funds-of-Funds**	**Total Funds**
Funds	7,008	963	7,971	2,676	10,647
% of Total Funds	65.82%	9.04%	74.87%	25.13%	100.00%
Country% by Fund Type					
North America					
Canada	1.30%	1.87%	1.37%	0.41%	1.13%
USA	42.84%	39.36%	42.42%	27.80%	38.74%
Offshore					
Bahamas	1.58%	3.43%	1.81%	1.94%	1.84%
Bermuda	4.81%	6.96%	5.07%	6.05%	5.32%
British Virgin Islands	6.29%	4.78%	6.11%	7.59%	6.48%
Cayman Islands	23.66%	12.36%	22.29%	17.75%	21.15%
Channel Islands	2.08%	1.45%	2.01%	6.80%	3.21%
Island Offshore	1.07%	1.04%	1.07%	2.24%	1.36%
Europe					
France	1.61%	3.84%	1.88%	5.12%	2.70%
Ireland	2.54%	3.01%	2.60%	2.99%	2.70%
Italy	0.50%	0.21%	0.46%	2.77%	1.04%
Luxembourg	2.10%	2.91%	2.20%	6.80%	3.35%
Sweden	0.61%	0.31%	0.58%	0.60%	0.58%
Switzerland	1.08%	1.45%	1.13%	3.85%	1.81%
UK	2.38%	3.12%	2.47%	2.91%	2.58%
Other Europe	1.43%	1.97%	1.49%	2.28%	1.69%
Other					
Australia	1.00%	1.45%	1.05%	0.60%	0.94%
China	0.60%	0.21%	0.55%	0.15%	0.45%
Other Countries	2.51%	10.28%	3.45%	1.35%	2.92%
	100.00%	100.00%	100.00%	100.00%	100.00%

Table F.3: *Interval Regression: Grouping Fees into Fee Bands*

This table shows the aggregated fund counts within the broad fee bands under two different bases for the purpose of the interval regression. The fund count is over all funds and the panels are split between management and incentive fees.

Panel A: Management Fee Bands			
Band I		**Band II**	
Feeband	**% of Funds**	**Fee Band**	**% of Funds**
0 to 0.75	6.20%	0 to 0.5	4.90%
0.75 to 1.25	29.60%	0.5 to 1	25.80%
1.25 to 1.75	32.30%	1 to 1.5	33.90%
1.75 to 2.25	28.00%	1.5 to 2	30.80%
2.25+	4.00%	2+	4.60%
	100.00%		100.00%
Fund Count	4,657		

Panel B: Incentive Fee Bands			
Band I		**Band II**	
Fee Band	**% in Band**	**Fee Band**	**% in Band**
0 to 7.5	16.90%	0 to 5	16.47%
7.5 to 12.5	17.56%	5 to 10	17.63%
12.5 to 17.5	4.64%	10 to 15	4.90%
17.5 to 22.5	57.12%	15 to 20	57.05%
22.5 +	3.78%	20 +	3.95%
	100.00%		100.00%
Fund Count	4,657		

Table F.4: *Live and Defunct Fund Proportions by Database*

This table shows the breakdown of fund reporting status by underlying database. Live funds indicate the proportion of funds still reporting at the end of the period. (CISDM is affected by its six monthly update cycle). Defunct means the funds are no longer reporting but had been reporting during the period since January 1994.

	All Funds		
	Total Funds	**Defunct%**	**Live%**
TASS	5,395	45.0%	55.0%
HFR	3,295	45.9%	54.1%
CISDM	1,081	74.7%	25.3%
Other incl BarclayHedge	876	15.6%	84.4%
All Funds	10,647	45.9%	54.1%

F.3 Chapter 3

This section holds supplementary figures and tables for Chapter 3.

F.3.1 Tables

Table F.5: *Listing of Vintage Dates*

This table shows the vintage dates of the 40 snapshots.

Number	Vintage date
1	Jul-07
2	Jan-08
3	Feb-08
4	Mar-08
5	Apr-08
6	May-08
7	Jun-08
8	Jul-08
9	Aug-08
10	Sep-08
11	Oct-08
12	Nov-08
13	Dec-08
14	Jan-09
15	Mar-09
16	Apr-09
17	May-09
18	Jun-09
19	Jul-09
20	Aug-09
21	Sep-09
22	Oct-09
23	Dec-09
24	Jan-10
25	Feb-10
26	Mar-10
27	Apr-10
28	May-10
29	Jun-10
30	Jul-10
31	Aug-10
32	Sep-10
33	Oct-10
34	Nov-10
35	Dec-10
36	Jan-11
37	Feb-11
38	Mar-11
39	Apr-11
40	May-11

Table F.6: *Summary Statistics for Lifetime Variables*

This table shows summary statistics of lifetime AUM and return averages, medians and standard deviations; the number of observations in the asset history of the fund; and the first sample autocorrelation of returns, ρ_1. (Data used to construct these variables is taken from the final vintage of the data.)

	Lifetime AUM			**Lifetime Return**			**Return Auto-**	**Fund History**
	Average	**Std. Dev.**	**Median**	**Average**	**Std. Dev.**	**Median**	**correlation ρ_1**	**Length (years)**
Observations	18,382	18,382	18,382	18,382	18,382	18,382	18,382	18,382
Mean	149,289,134	79,707,015	135,439,957	0.644	4.102	0.677	0.139	5.535
Std dev	1,491,667,969	744,463,251	1,413,991,918	1.18	3.638	1.008	0.222	3.778
99th perc	1,723,491,752	972,471,117	1,595,549,937	4.652	18.278	3.77	0.655	17.25
75th perc	73,538,781	35,962,608	64,020,000	1.008	5.152	1.02	0.284	7.333
Median	22,754,853	9,070,742	19,444,848	0.552	3.032	0.61	0.139	4.5
25th perc	5,891,644	2,018,060	4,574,000	0.181	1.813	0.24	0.005	2.667
1st perc	101,520	-	-	2.283	0.437	2.047	0.415	1.083

Table F.7: *Probit Regression for Any Changes*

The table shows the marginal effects from a probit regression. The dependent variable takes the value of 1 if a fund had any change (Deletion, Revision, or Addition) in returns over any of the 40 vintages that we capture, and 0 otherwise. The independent variables are lifetime average returns, lifetime average AUM, standard deviation of returns, and the autocorrelation of returns, all measured as ranks relative to the other funds in the data; and the number of return observations in the return history of the fund. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located Offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods) and a flag which takes the value of 1 for the fund if there is any information pertaining to audits available in any of the databases. We also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Mean**	**Z-stat**	
Lifetime Avg. AUM (Rank)	0.227	0.500	4.100	***
Lifetime Avg. Return (Rank)	-0.093	0.500	-1.720	*
Lifetime Ret. Std. (Rank)	0.071	0.500	1.790	*
Return Autocorrelation (Rank)	0.114	0.500	8.610	***
Return History Length	0.025	5.535	6.140	***
Offshore	-0.030	0.501	-8.740	***
Total Restrictions	0.007	1.829	1.460	
Audit	0.138	0.712	1.660	*
High-Water Mark or Hurdle	0.193	0.604	5.87	***
Database Fixed Effects				
HFR	-0.067	0.258	-3.550	***
CISDM	-0.059	0.092	-0.870	
BarclayHedge	0.076	0.290	3.900	***
Strategy Fixed Effects				
Macro	0.092	0.065	13.330	***
Relative Value	0.186	0.014	3.140	***
Directional Traders	0.008	0.128	0.510	
Funds-of-Funds	0.262	0.264	14.030	***
Multi-Process	0.090	0.102	4.390	***
Emerging	0.125	0.045	11.750	***
Fixed Income	0.052	0.052	1.520	
Other	0.156	0.009	1.500	
Managed Futures	0.175	0.157	5.090	***
N	18,382			
Pseudo R^2	0.135			

Table F.8: *Probit Regression for Additions*

The table shows the marginal effects from a probit regression. The dependent variable takes the value of 1 if a fund had added past return data over any of the 40 vintages that we capture, and 0 otherwise. (Additions exclude fund launches; the first time a return appears for a fund; and additions within 12 months of the vintage $v-1$ date so as to avoid picking up late reporting.) The independent variables are lifetime average returns, lifetime average AUM, standard deviation of returns, and the autocorrelation of returns, all measured as ranks relative to the other funds in the data; and the number of return observations in the return history of the fund. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located Offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods) and a flag which takes the value of 1 if there is any information pertaining to audits available in any of the databases. We also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Mean**	**Z-stat**	
Lifetime Avg. AUM (Rank)	-0.002	0.500	-1.850	*
Lifetime Avg. Return (Rank)	-0.005	0.500	-0.740	
Lifetime Ret. Std. (Rank)	0.006	0.500	1.440	
Return Autocorrelation (Rank)	0.003	0.500	0.720	
Return History Length	0.002	5.535	6.230	***
Offshore	0.000	0.501	-0.100	
Total Restrictions	0.000	1.829	-0.290	
Audit	0.008	0.712	1.780	*
High-Water Mark or Hurdle	0.007	0.604	3.720	***
Database Fixed Effects				
HFR	-0.007	0.258	-5.010	***
CISDM	-0.013	0.092	-5.220	***
BarclayHedge	-0.004	0.290	-3.410	***
Strategy Fixed Effects				
Macro	-0.004	0.065	-1.080	
Relative Value	0.003	0.014	0.360	
Directional Traders	-0.004	0.128	-0.870	
Funds-of-Funds	0.009	0.264	4.180	***
Multi-Process	-0.003	0.102	-2.540	**
Emerging	0.003	0.045	1.580	
Fixed Income	0.006	0.052	0.780	
Other	0.044	0.009	14.400	***
Managed Futures	0.005	0.157	1.430	
N	18,382			
Pseudo R^2	0.095			

Table F.9: *Probit Regression for Deletions*

The table shows the marginal effects from a probit regression. The dependent variable takes the value of 1 if a fund had deleted return data over any of the 40 vintages that we capture, and 0 otherwise. The independent variables are lifetime average returns, lifetime average AUM, standard deviation of returns, and the autocorrelation of returns, all measured as ranks relative to the other funds in the data; and the number of return observations in the return history of the fund. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located Offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods) and a flag which takes the value of 1 if there is any information pertaining to audits available in any of the databases. We also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Mean**	**Z-stat**	
Lifetime Avg. AUM (Rank)	0.011	0.500	2.270	**
Lifetime Avg. Return (Rank)	-0.030	0.500	-1.200	
Lifetime Ret. Std. (Rank)	0.010	0.500	1.680	*
Return Autocorrelation (Rank)	-0.006	0.500	-0.480	
Return History Length	0.003	5.535	28.990	***
Offshore	0.016	0.501	2.590	***
Total Restrictions	-0.002	1.829	-0.880	
Audit	0.015	0.712	2.560	**
High-Water Mark or Hurdle	0.021	0.604	9.250	***
Database Fixed Effects				
HFR	-0.012	0.258	-6.210	***
CISDM	-0.029	0.092	-7.910	***
BarclayHedge	-0.023	0.290	-10.100	***
Strategy Fixed Effects				
Macro	0.005	0.065	0.850	
Relative Value	0.049	0.014	4.120	***
Directional Traders	0.007	0.128	1.640	
Funds-of-Funds	0.027	0.264	7.620	***
Multi-Process	-0.009	0.102	-1.510	
Emerging	0.020	0.045	2.670	***
Fixed Income	0.017	0.052	1.200	
Other	0.020	0.009	1.080	
Managed Futures	0.014	0.157	2.380	**
N	18,382			
Pseudo R^2	0.046			

Table F.10: *Explaining Revision Return Differences — Interactions Detail*

This table conditions the return differences occasioned by revisions on various fund characteristics and period fixed effects. (This table, similar to Table 3.6, holds the details of the interactions between strategy and crisis periods). The dependent variable is the average difference, for all years in which a fund experienced return revisions, between the final set of annual returns provided by a fund and the first set of annual returns provided by the same fund for the same year. For example, if fund Y initially reported 4% average annual return for year t, and at the final vintage, this average stood at 6%, then the return difference variable would be 2%. We only include periods in which the fund had at least 6 months of return observations, to reduce the noise in the dependent variable. Panel A takes the absolute value of all such differences as the dependent variable, and Panel B conditions the signed revisions on the independent variables. Period dummies include crisis dummies for the 1998–1999 period, the 2000–2001 period, and the 2008–2009 period. The remaining regressors have been described earlier in these tables, with three new additions, namely the rank of flows experienced by the fund relative to all other funds in the same year; the Management fee and the Incentive fee of the fund. t-statistics, shown in parentheses, are robust to heteroskedasticity and clustered at the fund-level. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

Panel A: Absolute Value of Differences

	Coeff.	***t*-stat**		**Coeff.**	***t*-stat**	
Constant	1.109	(-24.984)	***	1.285	(-6.830)	***
Crisis1 * Security Selection	1.869	(-1.525)		1.878	(-1.572)	
Crisis1 * Macro	2.332	(-1.800)	*	2.355	(-1.848)	*
Crisis1 * Relative Value	-			-		
Crisis1 * Directional Traders	4.662	(-2.260)	**	4.554	(-2.254)	**
Crisis1 * Funds-of-Funds	1.100	(-1.073)		1.302	(-1.276)	
Crisis1 * Multi-Process	-0.565	(-1.296)		-0.773	(-1.505)	
Crisis1 * Emerging	-			-		
Crisis1 * Fixed Income	-			-		
Crisis1 * Managed Futures	-0.451	(-1.606)		-0.507	(-1.298)	
Crisis2 * Security Selection	0.348	(-0.553)		0.454	(-0.724)	
Crisis2 * Macro	1.628	(-1.481)		1.527	(-1.397)	
Crisis2 * Relative Value	-			-		
Crisis2 * Directional Traders	1.553	(-1.707)	*	1.474	(-1.692)	*
Crisis2 * Funds-of-Funds	0.712	(-0.732)		0.824	(-0.862)	
Crisis2 * Multi-Process	-0.395	(-1.147)		-0.406	(-1.153)	
Crisis2 * Emerging	1.566	(-1.643)		1.216	(-1.297)	
Crisis2 * Fixed Income	-			-		
Crisis2 * Managed Futures	0.444	(-0.718)		0.317	(-0.528)	
Crisis3 * Security Selection	0.717	(-3.763)	***	0.610	(-3.128)	***
Crisis3 * Macro	0.427	(-1.937)	*	0.315	(-1.411)	
Crisis3 * Relative Value	0.285	(-0.999)		0.210	(-0.758)	
Crisis3 * Directional Traders	0.563	(-3.226)	***	0.515	(-2.965)	***
Crisis3 * Funds-of-Funds	0.479	(-5.592)	***	0.641	(-7.191)	***
Crisis3 * Multi-Process	0.678	(-3.643)	***	0.684	(-3.667)	***
Crisis3 * Emerging	2.084	(-4.885)	***	2.005	(-4.703)	***
Crisis3 * Fixed Income	0.639	(-2.388)	**	0.644	(-2.425)	**
Crisis3 * Managed Futures	0.158	(-1.272)		-0.090	(-0.651)	

Panel A: Absolute Value of Differences (Continued)					
	Coeff.	***t*-stat**	**Coeff.**	***t*-stat**	
Offshore			0.175	(-2.076)	**
Total Restrictions			-0.019	(-1.401)	
High-Water Mark or Hurdle			-0.211	(-2.126)	**
Audit			0.161	(-1.314)	
Management Fee			0.013	(-0.181)	
Incentive Fee			0.022	(-3.605)	***
Asset t-1 rank			-0.938	(-6.132)	***
Return prior year t-1 rank			-0.075	(-0.597)	
Flow prior year t-1 rank			-0.082	(-0.789)	
N	10,004		10,004		
Adjusted R^2	0.021		0.032		

Panel B: Return Differences						
	Coeff.	***t*-stat**		**Coeff.**	***t*-stat**	
Constant	-0.012	(-0.291)		-0.080	(-0.467)	
Crisis1 * Security Selection	-0.209	(-0.148)		-0.179	(-0.127)	
Crisis1 * Macro	-3.429	(-2.639)	***	-3.518	(-2.809)	***
Crisis1 * Relative Value	-			-		
Crisis1 * Directional Traders	0.954	(-0.343)		1.028	(-0.370)	
Crisis1 * Funds-of-Funds	-0.369	(-0.326)		-0.431	(-0.373)	
Crisis1 * Multi-Process	-0.358	(-0.817)		-0.281	(-0.551)	
Crisis1 * Emerging	-			-		
Crisis1 * Fixed Income	-			-		
Crisis1 * Managed Futures	0.162	(-0.401)		0.084	(-0.205)	
Crisis2 * Security Selection	-0.406	(-0.636)		-0.393	(-0.620)	
Crisis2 * Macro	-1.656	(-1.441)		-1.612	(-1.416)	
Crisis2 * Relative Value	-			-		
Crisis2 * Directional Traders	-1.732	(-1.683)	*	-1.669	(-1.635)	
Crisis2 * Funds-of-Funds	-0.269	(-0.396)		-0.318	(-0.485)	
Crisis2 * Multi-Process	-0.441	(-1.246)		-0.391	(-1.088)	
Crisis2 * Emerging	-0.596	(-0.481)		-0.531	(-0.431)	
Crisis2 * Fixed Income	-			-		
Crisis2 * Managed Futures	-0.321	(-0.346)		-0.400	(-0.413)	
Crisis3 * Security Selection	0.041	(-0.236)		0.101	(-0.573)	
Crisis3 * Macro	-0.330	(-1.331)		-0.323	(-1.274)	
Crisis3 * Relative Value	-0.571	(-1.965)	**	-0.527	(-1.794)	*
Crisis3 * Directional Traders	-0.090	(-0.522)		-0.039	(-0.228)	
Crisis3 * Funds-of-Funds	-0.719	(-7.937)	***	-0.774	(-8.295)	***
Crisis3 * Multi-Process	-0.265	(-1.420)		-0.260	(-1.377)	
Crisis3 * Emerging	-0.713	(-1.616)		-0.684	(-1.541)	
Crisis3 * Fixed Income	0.378	(-1.538)		0.417	(-1.685)	*
Crisis3 * Managed Futures	0.008	(-0.056)		0.036	(-0.253)	
Offshore				-0.049	(-0.709)	
Total Restrictions				0.018	(-1.272)	
High-Water Mark or Hurdle				-0.033	(-0.352)	
Audit				-0.076	(-0.679)	
Management Fee				0.138	(-2.203)	**
Incentive Fee				-0.011	(-1.878)	*
Asset t-1 rank				0.242	(-2.108)	**
Return prior year t-1 rank				0.127	(-0.959)	
Flow prior year t-1 rank				-0.162	(-1.341)	
N	10,004			10,004		
Adjusted R^2	0.008			0.009		

Table F.11: *Multinomial Logistic Regression on Revision Direction*

These are coefficients from a multinomial logit regression on revision direction relative to no change at all. Revision direction is the net number of positive or negative revisions experienced by a fund. The base case of zeros refers to funds having no revisions at all. Funds with exactly equal positive and negative revisions were dropped (4.6% of funds). Regressors are as in Table 3.4. Standard errors are estimated by clustering by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

Panel A: More Negative Revisions			
-1 to 0	**Coeff.**	**Z-stat**	
Lifetime Avg. AUM (Rank)	1.079	5.55	***
Lifetime Avg. Return (Rank)	-0.788	-2.64	***
Lifetime Ret. Std. (Rank)	0.51	4.07	***
Return Autocorrelation (Rank)	0.555	8.59	***
Return History Length	0.009	4.16	***
Offshore	-0.095	-2.03	**
Total Restrictions	0.001	4.19	***
Audit	0.934	1.73	*
Database Fixed Effects			
HFR	0.1	3.27	***
CISDM	-0.027	-0.06	
BarclayHedge	0.768	24.34	***
Strategy Fixed Effects			
Macro	0.326	5.39	***
Relative Value	0.668	4.24	***
Directional Traders	-0.161	-2.04	**
Funds-of-Funds	0.884	9.47	***
Multi-Process	0.136	1.46	
Emerging	0.429	6.74	***
Fixed Income	-0.084	-0.45	
Other	0.295	0.95	
Managed Futures	0.548	2.12	**
Constant	-4.073	-9.17	***

Panel B: More Positive Revisions			
+1 to 0	**Coeff.**	**Z-stat**	
Lifetime Avg. AUM (Rank)	1.100	3.38	***
Lifetime Avg. Return (Rank)	0.071	0.57	
Lifetime Ret. Std. (Rank)	0.065	0.27	
Return Autocorrelation (Rank)	0.587	6.600	***
Return History Length	0.008	4.89	***
Offshore	-0.167	-4.34	***
Total Restrictions	0.001	5.04	***
Audit	0.690	1.43	
Database Fixed Effects			
HFR	-0.201	-7.59	***
CISDM	-0.467	-1.200	
BarclayHedge	0.262	4.43	***
Strategy Fixed Effects			
Macro	0.415	15.03	***
Relative Value	0.882	2.24	**
Directional Traders	0.088	2.34	**
Funds-of-Funds	0.946	15.15	***
Multi-Process	0.359	2.85	***
Emerging	0.651	9.22	***
Fixed Income	0.160	0.870	
Other	0.663	1.32	
Managed Futures	0.519	2.93	***
Constant	-3.832	-12.43	***
Panel C: Regression Statistics			
N	17,587		
Pseudo R^2	0.092		

Table F.12: *Change in Predictions for Revision Direction*

The panels below show changes in predicted probabilities in the revision direction multinomial logit regression, where -1 indicates more negative revisions, 1 for more positive revisions in the fund and 0 for no revisions at all. Panel A shows the impact of the Audit flag dummy, and Panel B shows a change from 1st to 3rd quartile in lifetime ranks. Confidence intervals are estimated by the delta method.

Panel A: Audit				
	Audit flag			
	Audit	**No Audit**	**Diff**	**95% CI for Diff**
$\Pr(y=-1\|x)$	0.189	0.093	0.095	[0.0810, 0.1098]
$\Pr(y=1\|x)$	0.182	0.115	0.067	[0.0518, 0.0824]
$\Pr(y=0\|x)$	0.630	0.792	-0.163	[-0.1821, -0.1428]
Panel B: Change in Quartiles				
	Lifetime Average AUM			
	AUM 0.75	**AUM 0.25**	**Diff**	**95% CI for Diff**
$\Pr(y=-1\|x)$	0.186	0.129	0.057	[0.0462, 0.0679]
$\Pr(y=1\|x)$	0.194	0.133	0.061	[0.0496, 0.0719]
$\Pr(y=0\|x)$	0.620	0.738	-0.118	[-0.1323, -0.1032]
	Lifetime Return Average			
	Ret 0.75	**Ret 0.25**	**Diff**	**95% CI for Diff**
$\Pr(y=-1\|x)$	0.131	0.184	-0.053	[-0.0636, -0.0421]
$\Pr(y=1\|x)$	0.168	0.154	0.015	[0.0036, 0.0258]
$\Pr(y=0\|x)$	0.700	0.662	0.038	[0.0238, 0.0524]
	Lifetime Return Standard Deviation			
	Std 0.75	**Std 0.25**	**Diff**	**95% CI for Diff**
$\Pr(y=-1\|x)$	0.173	0.140	0.033	[0.0217, 0.0438]
$\Pr(y=1\|x)$	0.160	0.162	-0.002	[-0.0133, 0.0092]
$\Pr(y=0\|x)$	0.667	0.698	-0.031	[-0.0455, -0.0159]
	Lifetime Return First Autocorrelation			
	Rho 0.75	**Rho 0.25**	**Diff**	**95% CI for Diff**
$\Pr(y=-1\|x)$	0.171	0.142	0.029	[0.0184, 0.0397]
$\Pr(y=1\|x)$	0.178	0.146	0.033	[0.0219, 0.0435]
$\Pr(y=0\|x)$	0.651	0.713	-0.062	[-0.0759, -0.0477]

Table F.13: *Characteristics of the Reviser and Non-reviser Funds*

This table shows the differences in characteristics between the reviser and non-reviser groups of funds using the status of the funds at the last vintage. The non-reviser funds at this stage have never revised between vintages. Once a fund revises a return it joins the reviser portfolio and it stays out of the non-reviser group. Lifetime AUM and return measures are used for the funds, not the period in which they belonged to the group. There are 11,476 non-reviser funds out of the 18,382 funds. t-statistics of the differences between groups assume a common variance.

	Revisers		**Non-Revisers**			
Variable	**Mean**	**Std Dev**	**Mean**	**Std Dev**	***t*-stat diff**	***p*-value**
Lifetime AUM Average $m	180.908	1,479.514	130.262	1,498.677	2.230	0.026
Lifetime Return Average	0.636	0.987	0.649	1.282	-0.680	0.497
Lifetime Return Std.	3.726	3.098	4.327	3.909	-10.884	0.000
Return Autocorrelation	0.186	0.218	0.111	0.219	22.508	0.000
Return History Length (years)	6.635	4.201	4.873	3.329	31.420	0.000
Total Restrictions (quarters)	2.205	2.908	1.603	2.377	15.251	0.000

Table F.14: *Robustness Check (excl. FOFs): Probit Regression for Revisions*

The table shows the marginal effects from a probit regression on the sample but excluding Funds-of-Funds (FOFs). (We remove funds marked with this strategy). The dependent variable takes the value of 1 if a fund had revised data over any of the 40 vintages that we capture, and 0 otherwise. The independent variables are lifetime average returns, lifetime average AUM, standard deviation of returns, and the autocorrelation of returns, all measured as ranks relative to the other funds in the data; and the number of return observations in the return history of the fund. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods) and a flag which takes the value of 1 for the fund if there is any information pertaining to audits available in any of the databases. We also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Mean**	**Z-stat**	
Lifetime Avg. AUM (Rank)	0.187	0.482	4.000	***
Lifetime Avg. Return (Rank)	-0.023	0.544	-0.710	
Lifetime Ret. Std. (Rank)	0.066	0.557	3.250	***
Return Autocorrelation (Rank)	0.074	0.457	5.060	***
Return History Length	0.021	5.339	5.410	***
Offshore	-0.040	0.459	-6.450	***
Total Restrictions	0.002	1.792	0.660	
Audit	0.139	0.675	1.810	*
High-Water Mark or Hurdle	0.186	0.620	4.170	***
Database Fixed Effects				
HFR	-0.044	0.272	-1.840	*
CISDM	-0.023	0.088	-0.380	
BarclayHedge	0.089	0.313	3.530	***
Strategy Fixed Effects				
Macro	0.060	0.089	6.530	***
Relative Value	0.154	0.018	3.070	***
Directional Traders	-0.004	0.174	-0.190	
Multi-Process	0.097	0.139	4.870	***
Emerging	0.110	0.061	20.160	***
Fixed Income	0.042	0.071	1.070	
Other	0.141	0.013	1.450	
Managed Futures	0.157	0.213	5.120	***
N	13,536			
Pseudo R^2	0.109			

Table F.15: *Robustness Check (excl. FOFs): Probit Regression for Revisions at Vintage Level*

This table runs essentially the same specification as in Table F.14, excluding Funds-of-Funds (FOFs), the difference is that we employ the panel structure of the data, and the fund-vintage is now our unit of analysis. The dependent variable takes the value of 1 if a fund revised data between the last available vintage $v-1$ and the current vintage v. The ranks of the lifetime variables are therefore now measured using data in vintage $v-1$ on assets under management, and returns. We also add an independent variable that takes the value of 1 if the fund experienced a data revision in the prior vintage, and 0 otherwise. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods) and a flag which takes the value of 1 for the fund if there is any information pertaining to audits available in any of the databases. We also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by vintage. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Mean**	**Z-stat**	
Lifetime Avg. AUM (Rank) (v-1)	0.022	0.480	13.920	***
Lifetime Avg. Return (Rank) (v-1)	0.014	0.541	8.590	***
Lifetime Ret. Std. (Rank) (v-1)	0.002	0.560	1.390	
Return Autocorrelation (Rank) (v-1)	0.006	0.458	3.590	***
Return History Length (v-1)	0.001	5.185	7.240	***
Prior Vintage Revision Indicator	0.202	0.047	18.500	***
Offshore	-0.004	0.459	-5.470	***
Total Restrictions	0.000	1.855	0.690	
Audit	0.014	0.655	8.860	***
High-Water Mark or Hurdle	0.014	0.616	9.790	***
Database Fixed Effects				
HFR	0.000	0.269	0.260	
CISDM	-0.017	0.092	-1.710	*
BarclayHedge	0.010	0.301	2.250	**
Strategy Fixed Effects				
Macro	0.010	0.088	6.800	***
Relative Value	0.010	0.017	5.290	***
Directional Traders	-0.002	0.174	-1.460	
Multi-Process	0.008	0.126	5.050	***
Emerging	0.008	0.059	6.110	***
Fixed Income	0.005	0.069	4.830	***
Other	0.014	0.012	7.190	***
Managed Futures	0.021	0.217	12.660	***
N	413,343			
Pseudo R^2	0.179			

Table F.16: *Robustness Check (excl. FOFs): Regressions on Return Differences between Portfolios*

This table shows the significance of the differences in returns between the non-reviser and reviser portfolios (on the sample excluding Funds-of-Funds). The monthly return differences are analysed against different risk models. Panel A uses factors from the Fung-Hsieh model, such as a market model using S&P 500, four of the market related Fung-Hsieh factors, and then the Fung-Hsieh 7 and 8 Factor model. Panel B uses an alternate specification with the Fama-French 3 factor model, and then adds a momentum factor, and finally the Pastor-Stambaugh Liquidity factor. The PS-Liquidity factors are only available to December 2010. Newey-West heteroskedasticity and autocorrelation robust standard errors (with three lags) are used. Regression betas are shown with t-statistics shown in brackets beneath. Alpha significance is denoted by stars at 10% (*), 5% (**) and 1% (***) respectively.

Panel A: Return differences (Fung-Hsieh Model)					
Factors	**Constant**	**Market**	**FH 4**	**FH 7**	**FH 8**
Constant	0.204***	0.198***	0.183***	0.165***	0.166***
	(2.278)	(2.805)	(2.247)	(2.417)	(2.470)
SP500	-	0.034	0.026	0.024	0.026
	-	(2.841)	(1.338)	(1.538)	(1.247)
SMB	-	-	0.026	0.022	0.022
	-	-	(1.716)	(1.246)	(1.145)
BOND10YR	-	-	-0.135	-0.395	-0.399
	-	-	(-0.861)	(-1.479)	(-1.423)
CREDSPR	-	-	-0.115	-0.179	-0.190
	-	-	(-0.581)	(-0.727)	(-0.638)
PTFSBD	-	-	-	-0.697	-0.697
	-	-	-	(-1.113)	(-1.112)
PTFSFX	-	-	-	1.289	1.292
	-	-	-	(2.612)	(2.576)
PTFSCOM	-	-	-	-2.355	-2.363
	-	-	-	(-3.251)	(-3.238)
EMERGING	-	-	-	-	-0.002
	-	-	-	-	(-0.092)
N	40	40	40	40	40
Adjusted R^2		15.47%	10.98%	26.58%	24.23%

Panel B: Return differences (Fama-French 3 factors + Momentum + Pastor-Stambaugh Liquidity Model)			
Factors	**FF3**	**FF3 + Mom**	**FF3 + Mom + Liquidity**
Constant	0.182***	0.145***	0.175***
	(2.214)	(2.637)	(3.224)
MKTRF	2.504	1.017	1.852
	(1.819)	(1.167)	(1.963)
SMB	2.149	2.873	3.301
	(1.086)	(1.325)	(1.502)
HML	1.307	-1.586	-4.670
	(0.973)	(-1.462)	(-2.950)
UMD	-	-4.005	-3.779
	-	(-6.025)	(-5.134)
PSLIQ	-	-	-2.251
	-	-	(-2.354)
N	40	40	36
Adjusted R^2	21.94%	15.18%	15.38%

Table F.17: *Robustness Check (Single Database Check): Probit Regression for Revisions*

The table shows the marginal effects from a probit regression on the sample focusing on each database in turn. (We drop other funds not from the database in each case.) The dependent variable takes the value of 1 if a fund had revised data over any of the 40 vintages that we capture, and 0 otherwise. The remaining regressors have been described earlier in these tables such as Table 3.4. We also include strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	TASS		**HFR**		**CISDM**		**BarclayHedge**	
	dF/dx		**dF/dx**		**dF/dx**		**dF/dx**	
Lifetime Avg. AUM (Rank)	0.183	***	0.382	***	0.063	*	0.214	***
Lifetime Avg. Return (Rank)	-0.140	***	-0.090	***	-0.157	***	0.006	
Lifetime Ret. Std. (Rank)	0.001		0.101	***	0.073	*	0.154	***
Return Autocorrelation (Rank)	0.125	***	0.058	**	0.142	***	0.113	***
Return History Length	0.029	***	0.023	***	0.023	***	0.013	***
Offshore	-0.024	*	-0.038	**	-0.016		-0.042	**
Total Restrictions	0.003		0.015	***	0.021	***	0.009	***
Audit	0.243	***	-0.105	***	0.138	***	0.210	***
High-Water Mark or Hurdle	0.180	***	0.117	***	0.118	***	0.265	***
Strategy Fixed Effects								
Macro	0.071	**	0.050		0.024		0.088	
Relative Value	0.390	***	0.202	***	0.058		0.076	
Directional Traders	0.046		-0.030		-0.039		0.015	
Funds-of-Funds	0.229	***	0.197	***	0.200	***	0.277	***
Multi-Process	0.112	***	0.053		0.105	**	-0.018	
Emerging	0.125	***			0.061		0.097	**
Fixed Income	-0.056	*	0.076	*	0.021		0.020	
Other	0.214	***			0.157		-0.179	*
Managed Futures	0.107	***	0.470	***	0.276	***	0.231	***
N	6,604		4,742		1,698		5,338	
Pseudo R^2	0.147		0.122		0.188		0.157	

Table F.18: *Robustness Check (Single Database Check): Probit Regression for Revisions at Vintage Level*

This table runs essentially the same specification as in Table F.17, the difference is that we employ the panel structure of the data, and the fund-vintage is now our unit of analysis. We also focus on each database in turn. The dependent variable takes the value of 1 if a fund revised data between the last available vintage $v-1$ and the current vintage v. The remaining regressors have been described earlier in tables such as Table 3.5. We also include strategy fixed-effects. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by vintage. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	TASS		**HFR**		**CISDM**		**BarclayHedge**	
	dF/dx		**dF/dx**		**dF/dx**		**dF/dx**	
Lifetime Avg. AUM (Rank) (v-1)	0.023	***	0.049	***	0.003	*	0.032	***
Lifetime Avg. Return (Rank) (v-1)	0.012	*	0.008		-0.001		0.027	***
Lifetime Ret. Std. (Rank) (v-1)	-0.005		0.007		-0.001		0.004	*
Return Autocorrelation (Rank) (v-1)	0.011	***	0.011	***	0.002		0.008	**
Return History Length (v-1)	0.001	***	0.001	***	0.000	***	-0.001	***
Prior Vintage Revision Indicator	0.199	***	0.224	***	-0.005		0.341	***
Offshore	-0.004	*	-0.006	***	0.000		-0.003	**
Total Restrictions	0.000		0.001	***	0.001	***	0.002	***
Audit	0.031	***	0.002		0.003		0.027	***
High-Water Mark or Hurdle	0.011	***	0.014	***	0.004	***	0.025	***
Strategy Fixed Effects								
Macro	0.005	*	0.01	***	0.001		0.023	***
Relative Value	0.035	***	0.027	***	-0.001		-0.008	*
Directional Traders	0.007		-0.007	***	-0.002		0.001	
Funds-of-Funds	0.037	***	0.029	***	0.010	***	0.057	***
Multi-Process	0.011	***	0.004		0.005	***	0.004	
Emerging	0.014	***			0.004	**	0.007	***
Fixed Income	0.003		0.003		0.004	***	0.003	
Other	0.031	***			0.005		-0.034	***
Managed Futures	0.027	***	0.08	***	0.013	***	0.037	***
N	202,051		143,632		55,026		159,719	
Pseudo R^2	0.179		0.209		0.052		0.277	

Table F.19: *Robustness Check (Single Database Check): Regressions on Return Differences between Portfolios*

This table shows the significance of the differences in returns between the non-reviser and reviser portfolios (focusing on each database in turn). CISDM is not shown due to the slower updating of the database. The monthly return differences are analysed against different risk models. Panel A analyses return differences against the Fung-Hsieh seven-factor model. Panel B uses an alternate specification with the Fama-French 3 factor model, with a momentum factor, and the Pastor-Stambaugh Liquidity factor. The PS-Liquidity factors are only available to December 2010. Newey-West heteroskedasticity and autocorrelation robust standard errors (with three lags) are used. Regression betas are shown with t-statistics shown in brackets beneath. Alpha significance is denoted by stars at 10% (*), 5% (**) and 1% (***) respectively.

Panel A: Return Differences (Fung-Hsieh 7 Factor Model)			
	Database Selection		
Factors	**TASS**	**HFR**	**BarclayHedge**
Constant	0.161*	0.214***	0.528***
	(-1.858)	(-2.784)	(-4.578)
SP500	-0.033	0.04	0.044
	(-2.114)	(-3.350)	(-2.637)
SMB	0.026	0.038	0.048
	(-1.376)	(-2.490)	(-2.030)
BOND10YR	-0.45	-0.111	0.074
	(-1.614)	(-0.415)	(-0.209)
CREDSPR	-0.188	-0.043	0.343
	(-0.979)	(-0.183)	(-1.007)
PTFSBD	-0.431	-0.017	0.002
	(-0.615)	(-0.027)	(-0.002)
PTFSFX	0.574	0.666	1.116
	-0.915	(-1.063)	(-1.626)
PTFSCOM	-1.062	-0.938	-0.867
	(-1.233)	(-1.358)	(-0.988)
N	32	32	40
Adjusted R^2	-0.001	0.412	0.133

Panel B: Return differences (Fama-French 3 factors + Momentum + Pastor-Stambaugh Liquidity Model)			
	Database Selection		
Factors	**TASS**	**HFR**	**BarclayHedge**
Constant	0.138***	0.239***	0.588***
	(-2.297)	(-5.077)	(-7.723)
MKTRF	-3.368	2.909	0.621
	(-3.712)	(-3.566)	(-0.381)
SMB	2.520	3.188	3.070
	(-1.453)	(-2.595)	(-1.278)
HML	-3.584	-0.900	-0.133
	(-1.677)	(-0.541)	(-0.050)
UMD	-3.554	-1.778	-3.226
	(-6.474)	(-7.679)	(-5.961)
PSLIQ	-1.533	-2.199	-4.126
	(-1.469)	(-3.454)	(-3.016)
N	28	28	36
Adjusted R^2	0.138	0.049	0.156

Table F.20: *Robustness Check (Other Quantiles): Probit Regression for Any Changes*

The table shows the marginal effects from a probit regression similar to Table F.7, but marginal effects are calculated at different quantile ranks, rather than the mean, for the continuous ranked variables. The dependent variable takes the value of 1 if a fund had any change (Deletion, Revision, or Addition) in returns over any of the 40 vintages that we capture, and 0 otherwise. regressors are as described in the prior table. Panel A shows results at the 0.75 rank, and Panel B at the 0.25 rank. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

Panel A: Marginal Effects of Ranks at 0.75		
	dF/dx	**Z-stat**
Lifetime Avg. AUM (Rank)	0.236***	(4.055)
Lifetime Avg. Return (Rank)	-0.096*	(-1.727)
Lifetime Ret. Std. (Rank)	0.074*	(1.775)
Return Autocorrelation (Rank)	0.118***	(8.541)
Return History Length	0.026***	(6.308)
Offshore	-0.031***	(-8.697)
Total Restrictions	0.007	(1.448)
Audit	0.146*	(1.716)
High-Water Mark or Hurdle	0.203***	(5.796)
Database Fixed Effects		
HFR	-0.070***	(-3.480)
CISDM	-0.062	(-0.880)
BarclayHedge	0.078***	(3.943)
Strategy Fixed Effects		
Macro	0.093***	(13.437)
Relative Value	0.184***	(3.278)
Directional Traders	0.009	(0.510)
Fund-of-Funds	0.263***	(15.900)
Multi-Process	0.091***	(4.389)
Emerging	0.125***	(11.554)
Fixed Income	0.054	(1.522)
Other	0.156	(1.539)
Managed Futures	0.176***	(5.261)
N	18,382	
Pseudo R^2	0.135	

Panel B: Marginal Effects of Ranks at 0.25		
	dF/dx	**Z-stat**
Lifetime Avg. AUM (Rank)	0.209***	(4.297)
Lifetime Avg. Return (Rank)	-0.086	(-1.639)
Lifetime Ret. Std. (Rank)	0.066*	(1.905)
Return Autocorrelation (Rank)	0.105***	(8.858)
Return History Length	0.023***	(5.163)
Offshore	-0.028***	(-8.443)
Total Restrictions	0.006	(1.501)
Audit	0.126*	(1.708)
High-Water Mark or Hurdle	0.177***	(7.528)
Database Fixed Effects		
HFR	-0.061***	(-4.144)
CISDM	-0.054	(-0.918)
BarclayHedge	0.071***	(3.447)
Strategy Fixed Effects		
Macro	0.086***	(12.952)
Relative Value	0.181***	(2.949)
Directional Traders	0.008	(0.511)
Fund-of-Funds	0.251***	(18.437)
Multi-Process	0.084***	(3.835)
Emerging	0.119***	(9.476)
Fixed Income	0.049	(1.528)
Other	0.150	(1.398)
Managed Futures	0.167***	(5.011)
N	18,382	
Pseudo R^2	0.135	

F.4 Chapter 4

This section holds supplementary figures and tables for Chapter 4.

F.4.1 Figures

Figure F.2: *Frequency of Missing Months for Shortfall Funds*

This figure shows, for funds in shortfall, the frequency count of the number of missing months of assets under management (AUM) data relative to the returns reported by the fund. The bins with ranges include the lower bound, but exclude the upper bound. For example, the 1-2 years column counts the funds with 12 to 23 months of missing AUM data.

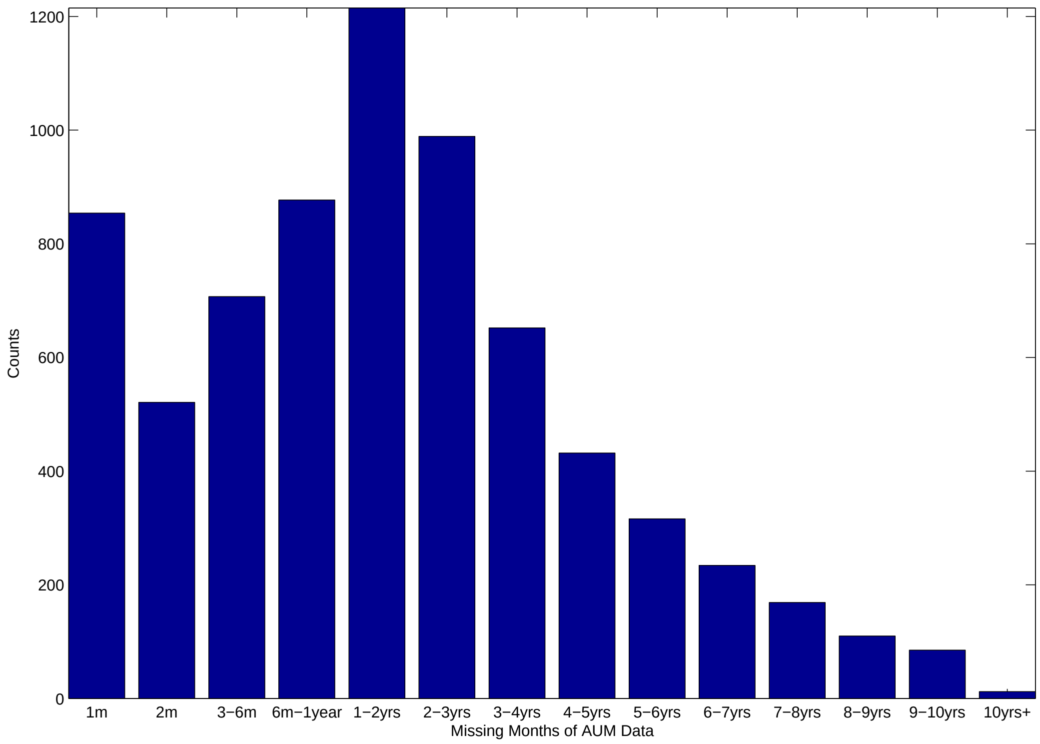


Figure F.3: *Portfolio Performance — Stoppers and Continuers*

This figure shows the cumulative performance of stopper and continuer portfolios. The continuer portfolio holds performance of funds that maintain or improve their asset to return reporting levels between vintages plus the early records of funds before they become stoppers (show deterioration in asset reporting). For example, if a fund worsens its asset reporting at vintage v; its earlier performance will be included in the non-reviser portfolio as it had not yet been classified as a stopper. But once it joins the stopper portfolio it stays out of the continuer portfolio. The index is based to 100 at 31 December 2007, just before the second vintage starts. Returns are equally weighted in portfolios.

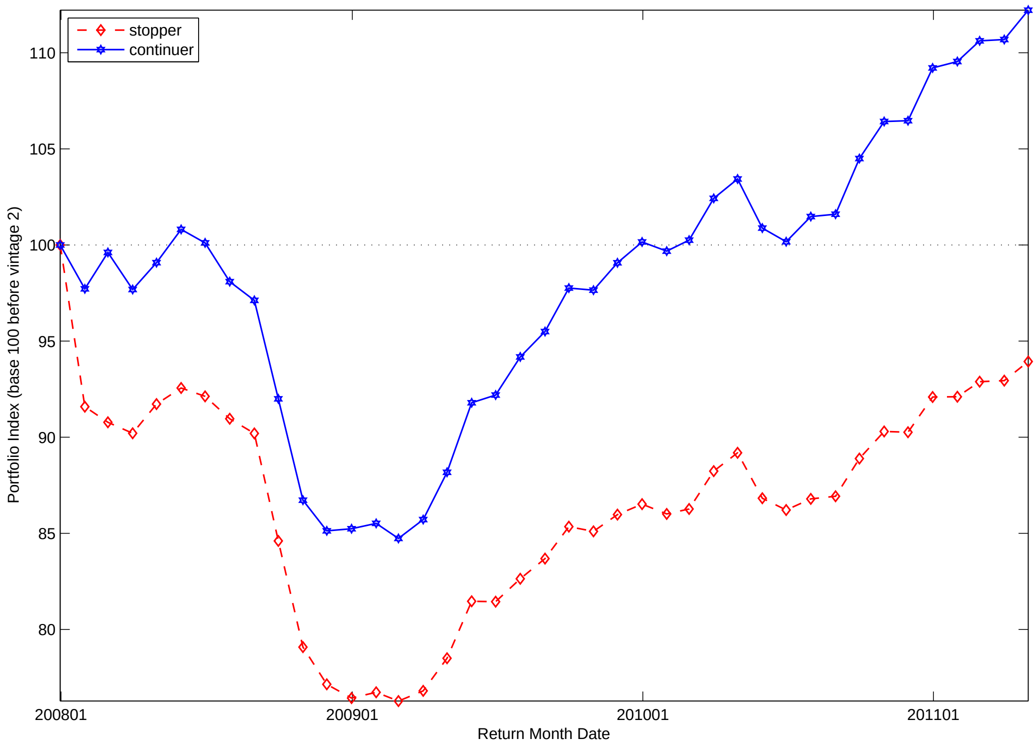
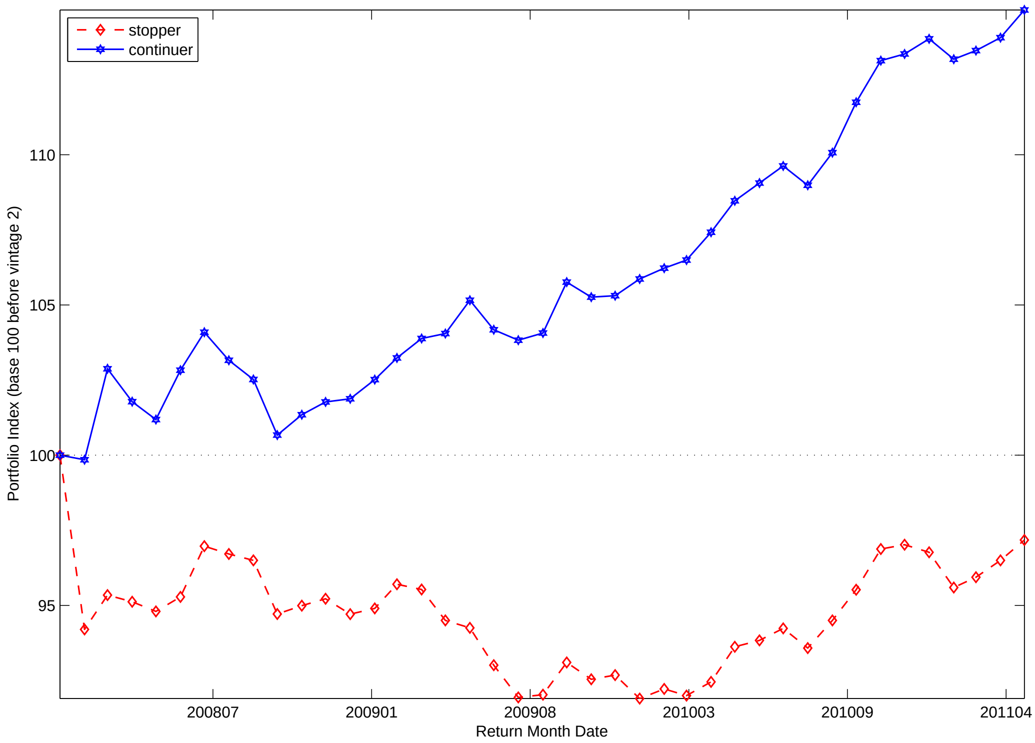


Figure F.4: *Cumulative Alpha*

The figure plots cumulative alpha (i.e. $\alpha + \epsilon_t$ for the portfolio regression of the return difference) estimated using the Fung-Hsieh seven-factor model for the stopper portfolio and continuer asset disclosure portfolios. The index is based to 100 at 31 December 2007, just before the second vintage starts.



F.4.2 Tables

Table F.21: *Summary Statistics of Fund Simulation Inputs*

This table shows the fund summary statistics of the inputs into the fund specific bootstrap percentiles simulation process. Namely, the fund average returns over its lifetime, standard deviation of these returns, the mode of the rounding level of the fund's returns and the return history length for the fund in months. The minimum return length was 24 months hence the total fund count is below the full sample size. The statistics are also shown for sub-groups of funds. Problem funds indicates the funds had missing assets over their return histories.

	Mean	**Std. Dev.**	**Rounding level**	**Return Count**
All Funds				
Fund Count	15,595	15,595	15,595	15,595
Mean	0.688	4.058	2.645	73.705
Std. Deviation	0.972	3.449	1.294	43.673
Median	0.574	3.046	2	61
75th Percentile	1.013	5.168	3	95
25th Percentile	0.232	1.838	2	40
Non-Problem Funds				
Fund Count	8,972	8,972	8,972	8,972
Mean	0.733	4.288	2.285	69.762
Std. Deviation	1.020	3.569	0.958	42.457
Median	0.590	3.234	2	57
75th Percentile	1.043	5.437	2	88
25th Percentile	0.253	1.986	2	37
Problem Funds				
Fund Count	6,623	6,623	6,623	6,623
Mean	0.627	3.746	3.133	79.047
Std. Deviation	0.899	3.255	1.511	44.720
Median	0.553	2.774	2	67
75th Percentile	0.978	4.785	5	102
25th Percentile	0.205	1.677	2	44

Table F.22: *Summary Statistics for Lifetime Variables*

This table shows summary statistics of lifetime AUM and return averages, medians and standard deviations; the number of observations in the asset history of the fund; and the first sample autocorrelation of returns, ρ_1. (Data used to construct these variables is taken from the final vintage of the data.)

	Lifetime AUM			**Lifetime Return**			**Return Auto-**	**Fund History**
	Average	**Std. Dev.**	**Median**	**Average**	**Std. Dev.**	**Median**	**correlation ρ_1**	**Length (years)**
Observations	18,243	18,243	18,243	18,243	18,243	18,243	18,243	18,243
Mean	148,488,799	79,008,827	134,830,249	0.643	4.100	0.675	0.139	5.508
Std. Dev.	1,496,057,989	746,064,257	1,418,089,604	1.182	3.636	1.009	0.222	3.757
95^{th} perc	404,236,410	236,665,963	372,510,500	2.292	10.885	2.197	0.505	13.667
75^{th} perc	73,195,462	35,767,121	63,804,250	1.007	5.153	1.015	0.283	7.250
Median	22,686,042	9,016,363	19,350,000	0.551	3.029	0.610	0.139	4.500
25^{th} perc	5,875,698	2,006,219	4,542,750	0.180	1.810	0.240	0.005	2.667
5^{th} perc	629,388	137,302	200,000	0.745	0.874	0.520	0.225	1.417

Table F.23: *Summary Statistics of Asset Changes across Vintages*

This table shows the breakdown of the changes in asset values between successive vintages where data is available for that database. Let $AUM_{i,t,v}$ be the asset value for a fund i, for month t, at vintage v. Deletion (Del) means an asset value goes missing between vintages: $AUM_{i,t,v-1}$ was available but $AUM_{i,t,v}$ is not available. (Del excludes last vintage to mitigate late reporting.) Addition (Add) means a value appears in a later vintage: $AUM_{i,t,v-1}$ was missing but $AUM_{i,t,v}$ is available. (Add excludes fund launches, first time a return appears for that fund, and funds entering within 12 months from vintage $v-1$ date to avoid picking up late reporting.) Revision (Rev) means both values are available but $AUM_{i,t,v-1}$ and $AUM_{i,t,v}$ are not equal to each other. (Rev excludes absolute revisions ≤ 1 to avoid spurious changes in significant digits in reporting e.g. from 2 to 0 decimal places.) Any Change means the fund experienced at least one of the change types (Del, Add, Rev) in the period of analysis. Panel A shows counts of these types of changes, Panel B looks at proportions of missing asset values. Panel C explores potential reasons for innocuous revisions: decimal errors capture mistakes in decimal places where the asset values between vintages differ by various multiples of 10. For example, 1,750,000 and 175,000, and digit transpositions identify revisions which are re-orderings of adjacent digits (e.g., 2,500,000 and 5,200,000).

Panel A: Changes Breakdown at Fund Level					
	Fund Count	**Any Change Count**	**Deletions Count**	**Additions Count**	**Revisions Count**
Funds	18,243	8,772	3,509	1,934	6,654
% of Total Funds		48.1%	19.2%	10.6%	36.5%
Panel B: Missing Data					
		Missing Fund Months	**Permanently Missing**	**Found Again and Match**	**Different Value Found**
Fund Months		153,594	63,999	85,913	3,682
% of Missing Months		100.0%	41.7%	55.9%	2.4%
Panel C: Statistics for Potential Types of Revisions					
	Reviser Count	**Decimal Place**	**Digit Transposition**	**Decimal *or* Transpose error**	
Funds	4,846	83	66	140	
% of Total Funds	26.58%	0.47%	0.37%	0.79%	
Revisions	145,458	651	88	739	
% of Total Revisions	100%	0.45%	0.06%	0.51%	
Decimal Place errors	**10**	**100**	**1,000**	**10,000**	**1 m**
Fund count	42	3	28	2	9
% of Total Funds	0.24%	0.02%	0.16%	0.01%	0.05%

Table F.24: *Composition by Calendar Month of Missing Months*

The table shows the seasonality of shortfall by calculating, in aggregate, the proportion of missing months of asset data by calendar month. Data is for complete years from 1994 to 2010, using the last available vintage for a fund.

Month	**Missing Months**	**Return Count**	**Shortfall%**
January	18,251	96,527	0.189
February	18,649	96,902	0.192
March	18,066	97,273	0.186
April	18,469	97,796	0.189
May	18,365	97,690	0.188
June	17,690	98,104	0.180
July	18,055	98,695	0.183
August	17,928	98,839	0.181
September	17,916	98,971	0.181
October	18,392	99,236	0.185
November	18,458	99,441	0.186
December	17,389	98,996	0.176
Average	18,136	98,206	0.185

Table F.25: *Probit Regression for AUM Any Changes*

The table shows the marginal effects from a probit regression. The dependent variable takes the value of 1 if a fund had any change in assets (Deletion, Revision, or Addition) over any of the 40 vintages that we capture, and 0 otherwise. The independent variables are lifetime average returns, lifetime average AUM, standard deviation of returns, and the autocorrelation of returns, all measured as ranks relative to the other funds in the data; and the number of return observations in the return history of the fund. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located Offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods) and a flag which takes the value of 1 for the fund if there is any information pertaining to audits available in any of the databases. We also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Mean**	**Z-stat**	
Lifetime Avg. AUM (Rank)	0.227	0.500	4.100	***
Lifetime Avg. Return (Rank)	-0.093	0.500	-1.720	*
Lifetime Ret. Std. (Rank)	0.071	0.500	1.790	*
Return Autocorrelation (Rank)	0.114	0.500	8.610	***
Return History Length	0.025	5.535	6.140	***
Offshore	-0.030	0.501	-8.740	***
Total Restrictions	0.007	1.829	1.460	
Audit	0.138	0.712	1.660	*
High-Water Mark or Hurdle	0.193	0.604	5.870	***
Database Fixed Effects				
HFR	-0.067	0.258	-3.550	***
CISDM	-0.059	0.092	-0.870	
BarclayHedge	0.076	0.290	3.900	***
Strategy Fixed Effects				
Macro	0.092	0.065	13.330	***
Relative Value	0.186	0.014	3.140	***
Directional Traders	0.008	0.128	0.510	
Funds-of-Funds	0.262	0.264	14.030	***
Multi-Process	0.090	0.102	4.390	***
Emerging	0.125	0.045	11.750	***
Fixed Income	0.052	0.052	1.520	
Other	0.156	0.009	1.500	
Managed Futures	0.175	0.157	5.090	***
N	18,234			
Pseudo R^2	0.121			

Table F.26: *Probit Regression for AUM Revisions*

The table shows the marginal effects from a probit regression. The dependent variable takes the value of 1 if a fund had revised asset data over any of the 40 vintages that we capture, and 0 otherwise. The independent variables are lifetime average returns, lifetime average AUM, standard deviation of returns, and the autocorrelation of returns, all measured as ranks relative to the other funds in the data; and the number of return observations in the return history of the fund. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located Offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods) and a flag which takes the value of 1 if there is any information pertaining to audits available in any of the databases. We also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Mean**	**Z-stat**	
Lifetime Avg. AUM (Rank)	0.245	0.500	19.480	***
Lifetime Avg. Return (Rank)	-0.180	0.500	-2.800	***
Lifetime Ret. Std. (Rank)	0.120	0.500	4.900	***
Return Autocorrelation (Rank)	0.081	0.500	12.920	***
Return History Length	0.026	5.508	4.060	***
Offshore	-0.004	0.500	-0.200	
Total Restrictions	-0.001	1.823	-0.260	
Audit	0.136	0.711	1.130	
High-Water Mark or Hurdle	0.159	0.603	5.520	***
Database Fixed Effects				
HFR	0.194	0.258	10.280	***
CISDM	-0.136	0.093	-1.510	
BarclayHedge	0.266	0.292	36.030	***
Strategy Fixed Effects				
Macro	0.096	0.065	7.740	***
Relative Value	0.145	0.014	1.590	
Directional Traders	0.027	0.129	1.760	*
Funds-of-Funds	0.202	0.263	7.520	***
Multi-Process	0.051	0.102	3.760	***
Emerging	0.093	0.045	6.440	***
Fixed Income	0.031	0.052	1.000	
Other	0.100	0.010	0.820	
Managed Futures	0.075	0.158	1.200	
N	18,234			
Pseudo R^2	0.168			

Table F.27: *Probit Regression for AUM Additions*

The table shows the marginal effects from a probit regression. The dependent variable takes the value of 1 if a fund had added past asset data over any of the 40 vintages that we capture, and 0 otherwise. (Additions exclude fund launches; the first time a return appears for a fund; and additions within 12 months of the vintage $v-1$ date so as to avoid picking up late reporting.) The independent variables are lifetime average returns, lifetime average AUM, standard deviation of returns, and the autocorrelation of returns, all measured as ranks relative to the other funds in the data; and the number of return observations in the return history of the fund. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located Offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods) and a flag which takes the value of 1 if there is any information pertaining to audits available in any of the databases. We also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Mean**	**Z-stat**	
Lifetime Avg. AUM (Rank)	0.063	0.500	2.440	**
Lifetime Avg. Return (Rank)	-0.007	0.500	-0.570	
Lifetime Ret. Std. (Rank)	-0.029	0.500	-4.210	***
Return Autocorrelation (Rank)	0.034	0.500	4.830	***
Return History Length	0.005	5.508	3.850	***
Offshore	-0.056	0.500	-3.060	***
Total Restrictions	-0.012	1.823	-2.460	**
Audit	0.057	0.711	2.260	**
High-Water Mark or Hurdle	-0.015	0.603	-1.420	
Database Fixed Effects				
HFR	-0.070	0.258	-9.270	***
CISDM	-0.049	0.093	-3.090	***
BarclayHedge	-0.073	0.292	-16.540	***
Strategy Fixed Effects				
Macro	0.011	0.065	1.350	
Relative Value	0.012	0.014	1.100	
Directional Traders	0.013	0.129	1.380	
Funds-of-Funds	0.045	0.263	8.910	***
Multi-Process	0.054	0.102	11.040	***
Emerging	0.036	0.045	2.570	**
Fixed Income	-0.006	0.052	-0.310	
Other	0.031	0.010	3.450	***
Managed Futures	-0.004	0.158	-0.290	
N	18,234			
Pseudo R^2	0.277			

Table F.28: *Probit Regression for AUM Deletions*

The table shows the marginal effects from a probit regression. The dependent variable takes the value of 1 if a fund had deleted asset data over any of the 40 vintages that we capture, and 0 otherwise. The independent variables are lifetime average returns, lifetime average AUM, standard deviation of returns, and the autocorrelation of returns, all measured as ranks relative to the other funds in the data; and the number of return observations in the return history of the fund. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located Offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods) and a flag which takes the value of 1 if there is any information pertaining to audits available in any of the databases. We also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Mean**	**Z-stat**	
Lifetime Avg. AUM (Rank)	0.095	0.500	2.830	***
Lifetime Avg. Return (Rank)	-0.068	0.500	-1.780	*
Lifetime Ret. Std. (Rank)	-0.050	0.500	-1.820	*
Return Autocorrelation (Rank)	0.032	0.500	4.500	***
Return History Length	0.007	5.508	3.170	***
Offshore	-0.063	0.500	-1.200	
Total Restrictions	-0.023	1.823	-1.620	
Audit	0.107	0.711	2.850	***
High-Water Mark or Hurdle	0.003	0.603	0.060	
Database Fixed Effects				
HFR	-0.118	0.258	-6.300	***
CISDM	0.417	0.093	7.470	***
BarclayHedge	-0.143	0.292	-6.510	***
Strategy Fixed Effects				
Macro	0.016	0.065	0.810	
Relative Value	0.048	0.014	1.670	*
Directional Traders	0.046	0.129	3.160	***
Funds-of-Funds	0.118	0.263	6.730	***
Multi-Process	0.101	0.102	3.040	***
Emerging	0.038	0.045	4.110	***
Fixed Income	0.017	0.052	0.990	
Other	0.069	0.010	1.360	
Managed Futures	0.068	0.158	0.810	
N	18,234			
Pseudo R^2	0.213			

Table F.29: *Probit Regression for AUM Deletions at Vintage Level*

This table runs essentially the same specification as in Table F.28, the difference is that we employ the panel structure of the data, and the fund-vintage is now our unit of analysis. The dependent variable takes the value of 1 if a fund deleted asset data between the last available vintage $v-1$ and the current vintage v. The ranks of the lifetime variables are therefore now measured using data in vintage $v-1$ on assets under management, and returns. We also add an independent variable that takes the value of 1 if the fund experienced a data deletion in the prior vintage, and 0 otherwise. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods) and a flag which takes the value of 1 for the fund if there is any information pertaining to audits available in any of the databases. We also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Coefficients are multiplied by 100 for readability. Robust standard errors control for heteroskedasticity, and cluster by vintage. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Mean**	**Z-stat**	
Lifetime Avg. AUM (Rank) (v-1)	0.544	0.500	5.390	***
Lifetime Avg. Return (Rank) (v-1)	-0.186	0.500	-1.030	
Lifetime Ret. Std. (Rank) (v-1)	-0.200	0.500	-2.920	***
Return Autocorrelation (Rank) (v-1)	0.108	0.500	2.660	***
Return History Length (v-1)	0.002	5.280	0.320	
Prior Vintage Deletion Indicator	7.267	0.017	4.390	***
Offshore	-0.542	0.502	-5.610	***
Total Restrictions	-0.188	1.905	-6.400	***
Audit	0.623	0.705	7.220	***
High-Water Mark or Hurdle	-0.134	0.611	-2.030	**
Database Fixed Effects				
HFR	-0.707	0.263	-5.160	***
CISDM	0.787	0.077	1.000	
BarclayHedge	-0.952	0.293	-7.830	***
Strategy Fixed Effects				
Macro	0.018	0.066	0.320	
Relative Value	0.109	0.013	1.020	
Directional Traders	0.192	0.130	2.440	**
Funds-of-Funds	0.476	0.261	8.500	***
Multi-Process	0.440	0.093	2.820	***
Emerging	0.159	0.042	3.770	***
Fixed Income	-0.082	0.051	-1.820	*
Other	0.427	0.009	7.400	***
Managed Futures	0.048	0.161	0.630	
N	542,807			
Pseudo R^2	0.276			

Table F.30: *Robustness Check(Flag Frequencies): Alternative distribution*

This table is a robustness check for flag frequencies in Table 4.9 with an alternative simulated return distribution. In this table the distribution of simulated returns is the t-distribution with degrees of freedom dependent on fund history length, and not set to six. The table shows the proportion of funds triggering some of the performance flags. Problem funds are the funds that experienced a shortfall in asset values at the date of the last available vintage. The non-problem fund group are the remaining funds showing full asset histories. Returns are taken over the funds full history. Funds require a minimum of 24 months of returns. The stars indicate the results of a test of differences in trigger rejection rates of problem and non-problem funds. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	Non-Problem Funds	**Problem Funds**	
Flag	$(N = 8,972)$	$(N = 6,623)$	***p*-value**
Perc. Negative	0.274	0.341***	0.0000
Count Zeros	0.144	0.237***	0.0000
String	0.120	0.108**	0.0239
Num. Pairs	0.056	0.050*	0.0984
Perc. Repeats	0.201	0.231***	0.0000

Table F.31: *Robustness Check(Flag Frequencies): Missing Threshold and Recency*

This table is a robustness check for flag frequencies in Table 4.9 after adjusting funds in shortfall with an allowance of some months of missing asset data, and the recency, k, of disclosure. Panel A shows the impact of 'forgiving' $m-1$ months of non-disclosure of assets before the Shortfall indicator is flagged. For example, the first column $m \geq 1$ of Panel A reproduces the results from Panel A of Table 4.9, and $m \geq 3$ only treats funds with three months missing as being in shortfall. Panel B shows the impact of excluding recent shortfalls in disclosure near the vintage date. For example, when $k > 3$ only funds with shortfalls that occur more than three months prior to the date of the vintage are included, and when $k > 12$, only funds which stop reporting assets over a year old are included in the shortfall indicator. Problem funds are the funds that experienced a shortfall in asset values at the date of the last available vintage. The non-problem fund group are the remaining funds showing full asset histories. Returns are taken over the funds full history. Funds require a minimum of 24 months of returns. The stars indicate the results of a test of differences in trigger rejection rates of problem and the initial set of non-problem funds. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

Panel A: Flag Frequencies by Minimum Missing Months

	Non-Problem	**Problem Funds**			
	Funds	$m \geq 1$	$m \geq 3$	$m \geq 6$	$m \geq 12$
Flag	**(N=8971)**	**(N=6624)**	**(N=5524)**	**(N=4982)**	**(N=4321)**
Perc. Negative	0.221	0.283***	0.297***	0.304***	0.317***
Count Zeros	0.137	0.231***	0.232***	0.236***	0.244***
String	0.113	0.104**	0.091***	0.091***	0.093***
Num. Pairs	0.055	0.049	0.043***	0.043***	0.046**
Perc. Repeats	0.174	0.200***	0.199***	0.204***	0.209***
Uniform	0.140	0.141	0.137	0.139	0.140
Benford	0.119	0.119	0.116	0.117	0.119
AR(1)	0.368	0.436***	0.459***	0.468***	0.485***

Panel B: Flag Frequencies by Minimum Recency of Shortfall

	Non-Problem	**Problem Funds**			
	Funds	$k = 1$	$k > 3$	$k > 6$	$k > 12$
Flag	**(N=8971)**	**(N=6624)**	**(N=6548)**	**(N=6491)**	**(N=6374)**
Perc. Negative	0.221	0.283***	0.282***	0.283***	0.282***
Count Zeros	0.137	0.231***	0.232***	0.233***	0.234***
String	0.113	0.104**	0.104*	0.104*	0.105*
Num. Pairs	0.055	0.049	0.049	0.049	0.050
Perc. Repeats	0.174	0.200***	0.200***	0.201***	0.202***
Uniform	0.140	0.141	0.141	0.141	0.140
Benford	0.119	0.119	0.119	0.120	0.119
AR(1)	0.368	0.436***	0.435***	0.434***	0.435***

Table F.32: *Robustness Check (excl. FOFs): Probit Regression for Shortfall*

The table shows the marginal effects from a probit regression on the sample but excluding Funds-of-Funds (FOFs). (I remove funds marked with this strategy). The dependent variable takes the value of 1 if a fund's reported asset months are below that of the period of the reported returns over the fund's lifetime (i.e. the funds aggregate shortfall is greater than zero). Data is taken from the funds last available vintage. The independent variables are lifetime average returns and lifetime average AUM, standard deviation of returns, and the autocorrelation of returns, all measured as ranks relative to the other funds in the data; and the number of return observations in the return history of the fund. Other relevant fund variables are a dummy variable which takes the value of 1 if the fund is located Offshore, a total restrictions variable (measured as the sum of the reported lockup and redemption notice periods), and a flag which takes the value of 1 if there is any information pertaining to audits available in any of the databases. I also include database and strategy fixed-effects in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity, and cluster by database. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	dF/dx	**Z-stat**
Lifetime Avg. AUM rank	-0.054	(-0.604)
Lifetime Avg. Return rank	0.027	(0.729)
Lifetime Ret. Std. rank	-0.055***	(-3.082)
Return Autocorrelation rank	0.014	(1.634)
Return History Length	0.014***	(2.858)
Offshore	0.046	(1.674)
Total Restrictions	-0.015***	(-3.985)
Audit	-0.001	(-0.028)
High-Water Mark or Hurdle	-0.143***	(-2.784)
Database Fixed Effects		
HFR	-0.388***	(-43.949)
CISDM	-0.149***	(-9.830)
BarclayHedge	-0.305***	(-50.527)
Strategy Fixed Effects		
Macro	-0.001	(-0.052)
Relative Value	-0.042*	(-1.698)
Directional Traders	0.065**	(2.393)
Multi-Process	0.126***	(6.080)
Emerging	-0.019	(-1.190)
Fixed Income	-0.008	(-0.196)
Other	-0.013	(-0.539)
Managed Futures	-0.149***	(-4.659)
N	13,438	
Pseudo R^2	0.166	

Table F.33: *Robustness Check (Single Database Check): Probit Regression for Shortfall*

The table shows the marginal effects from a probit regression on the sample focusing on each database in turn. (I drop other funds not from the database in each case.) The dependent variable takes the value of 1 if a fund's reported asset months are below that of the period of the reported returns over the fund's lifetime, and 0 otherwise. The remaining regressors have been described earlier in these tables such as Table 4.6. Strategy fixed-effects are included in the regressions. dF/dx shows the change in the independent variable for a discrete change in any independent dummy variable from 0 to 1, and the slope at the mean for continuous independent variables. Robust standard errors control for heteroskedasticity. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

	TASS	**HFR**	**CISDM**	**BarclayHedge**
	dF/dx	**dF/dx**	**dF/dx**	**dF/dx**
Lifetime Avg. AUM rank	0.068***	-0.133***	0.063	-0.222***
Lifetime Avg. Return rank	-0.009	0.118***	-0.050	-0.001
Lifetime Ret. Std. rank	-0.082***	0.029	-0.147***	-0.077***
Return Autocorrelation rank	0.047**	0.015	-0.036	0.007
Return History Length	0.009***	0.012***	0.006	0.023***
Offshore	-0.053***	0.044***	0.139***	0.073***
Total Restrictions	-0.025***	-0.007***	-0.014**	-0.006**
Audit	0.056***	-0.044***	-0.005	-0.072**
High-Water Mark or Hurdle	-0.179***	-0.093***	0.042	-0.083***
Strategy Fixed Effects				
Macro	0.002	-0.034	-0.180***	0.018
Relative Value	-0.083	-0.062**	-0.073	-0.017
Directional Traders	-0.026	-0.013	-0.065	0.094***
Funds-of-Funds	0.062***	-0.081***	-0.043	0.122***
Multi-Process	0.100***	0.052*	-0.008	0.108**
Emerging	0.007		-0.115**	0.010
Fixed Income	-0.018	-0.027	-0.208***	0.072*
Other	-0.031		-0.146	0.076
Managed Futures	-0.090***	-0.142***	-0.169***	-0.165***
N	6,521	4,709	1,694	5,319
Pseudo R^2	0.083	0.066	0.042	0.074

Table F.34: *Fraud Flag Correlations and Shortfall Probit Regression*

These tables explore the linkages between fraud flags. Panel A shows the correlations between the fraud flags. Num. Pairs is correlated to other data quality flags. Benford and Uniform distribution are also highly correlated. Panel B shows the coefficients from a probit regression. The dependent variable takes the value of 1 if a fund's reported asset months are below that of the period of the reported returns over the fund's lifetime, and 0 otherwise. The regressors are the fraud flags indicators, as in Table 4.9. The first specification uses all the fraud flag tests, and the second drops the two correlated flags, Num. Pairs and Benford. Note both Pseudo-R^2 figures are unadjusted for the number of variables. Robust standard errors control for heteroskedasticity. *, **, *** denote significance at the 10%, 5%, and 1% levels respectively.

Panel A: Correlation of Fraud Flags

Flag	**Perc. Negative**	**Count Zeros**	**String**	**Num. Pairs**	**Perc. Repeats**	**Uniform**	**Benford**	**AR(1)**	**CAR(1)**
Perc. Negative	1.000								
Count Zeros	0.128	1.000							
String	0.095	0.327	1.000						
Num. Pairs	0.106	0.296	0.442	1.000					
Perc. Repeats	0.227	0.308	0.371	0.319	1.000				
Uniform	0.080	0.132	0.198	0.247	0.195	1.000			
Benford	0.031	0.078	0.150	0.210	0.147	0.696	1.000		
AR(1)	0.290	-0.023	-0.047	-0.028	0.060	-0.008	-0.009	1.000	
CAR(1)	0.007	0.026	0.002	0.000	0.007	-0.001	0.005	0.012	1.000

Panel B: Shortfall Probit on Fraud Flags				
	All Flags		**Selected Flags**	
Flag	**Coeff.**	**Z-stat**	**Coeff.**	**Z-stat**
Constant	-0.325***	(-20.785)	-0.321***	(-20.651)
Perc. Negative	0.120***	(4.730)	0.117***	(4.601)
Count Zeros	0.516***	(17.191)	0.502***	(16.920)
String	-0.179***	(-4.480)	-0.217***	(-5.725)
Num. Pairs	-0.170***	(-3.047)		
Perc. Repeats	0.012	(0.388)	0.001	(0.019)
Uniform	-0.042	(-0.984)	-0.020	(-0.648)
Benford	0.056	(1.261)		
AR(1)	0.151***	(6.856)	0.153***	(6.939)
CAR(1)	-0.058*	(-1.839)	-0.057*	(-1.809)
N	15,242		15,242	
Estrella R^2	0.028		0.027	
McFadden R^2	0.020		0.020	

References

Ackermann, Carl, Richard McEnally, and David Ravenscraft, 1999, The performance of hedge funds: Risk, return, and incentives, *Journal of Finance* 54, 833–874.

Agarwal, Vikas, Naveen D. Daniel, and Narayan Y. Naik, 2005, Flows, performance, and managerial incentives in hedge funds, Technical report, Unpublished working paper, London Business School.

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

Agarwal, Vikas, Naveen D. Daniel, and Narayan Y. Naik, 2011a, Do hedge funds manage their reported returns?, *Review of Financial Studies* 24, 3281–3320.

Agarwal, Vikas, Vyacheslav Fos, and Wei Jiang, 2012, Inferring reporting-related biases in hedge fund databases from hedge fund equity holdings, *Management Science* forthcoming.

Agarwal, Vikas, Wei Jiang, Yuehua Tang, and Baozhong Yang, 2011b, Uncovering hedge fund skill from the portfolio holdings they hide, *Journal of Finance* forthcoming.

Agarwal, Vikas, and Narayan Y. Naik, 2000, Multi-period performance persistence analysis of hedge funds, *Journal of Financial and Quantitative Analysis* 35, 327–342, Copyright 2000 University of Washington School of Business Administration.

Agarwal, Vikas, and Sugata Ray, 2011, Determinants and implications of fee changes in the hedge fund industry Unpublished working paper.

Aiken, Adam L., Christopher P. Clifford, and Jesse A. Ellis, 2012, Out of the dark: Hedge fund reporting biases and commercial databases Unpublished working paper.

Allen, Gregory C., 2007, Does size matter?, *Journal of Portfolio Management* 33, 57–62.

Aragon, George O., 2007, Share restrictions and asset pricing: Evidence from the hedge fund industry, *Journal of Financial Economics* 83, 33–58.

Aragon, George O., Bing Liang, and Hyuna Park, 2008, Onshore and offshore hedge funds: Are they twins? Unpublished working paper.

Aragon, George O., and Vikram Nanda, 2011, Strategic delays and clustering in hedge fund reported returns Unpublished working paper.

Aragon, George O., and Vikram Nanda, 2012, Tournament behavior in hedge funds: High-water marks, fund liquidation, and managerial stake, *Review of Financial Studies* 25, 937–974.

Asness, Cliff S., Robert Krail, and John M. Liew, 2001, Do hedge funds hedge?, *Journal of Portfolio Management* 28, 6–19.

Avramov, Doron, Robert Kosowski, Narayan Y. Naik, and Melvyn Teo, 2011, Hedge funds, managerial skill, and macroeconomic variables, *Journal of Financial Economics* 99, 672–692.

Baquero, Guillermo, Jenke Ter Horst, and Marno Verbeek, 2005, Survival, look-ahead bias, and persistence in hedge fund performance, *Journal of Financial and Quantitative Analysis* 40, 493–517.

Baquero, Guillermo, and Marno Verbeek, 2009, A portrait of hedge fund investors: Flows, performance and smart money, *SSRN eLibrary* .

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2012a, Hedge fund stock trading in the financial crisis of 2007–2009, *Review of Financial Studies* 25, 1–54.

Ben-David, Itzhak, Francesco A. Franzoni, Augustin Landier, and Rabih Moussawi, 2012b, Do hedge funds manipulate stock prices?, *Journal of Finance* forthcoming.

Benford, Frank, 1938, The law of anomalous numbers, *Proceedings of the American Philosophical Society* 78, 551–572.

Berk, Jonathan B., and Richard C. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Berk, Jonathan B., and Richard Stanton, 2007, Managerial ability, compensation, and the closed-end fund discount, *Journal of Finance* 62, 529–556.

Bhattacharya, Utpal, Jung Hoon Lee, and Veronika K. Pool, 2012, Conflicting family values in mutual fund families, *Journal of Finance* forthcoming.

Bollen, Nicolas P. B., and Veronika K. Pool, 2008, Conditional return smoothing in the hedge fund industry, *Journal of Financial and Quantitative Analysis* 43, 267–298.

Bollen, Nicolas P. B., and Veronika K. Pool, 2009, Do hedge fund managers misreport returns? evidence from the pooled distribution, *Journal of Finance* 64, 2257–2288.

Bollen, Nicolas P. B., and Veronika K. Pool, 2012, Suspicious patterns in hedge fund returns and the risk of fraud, *Review of Financial Studies* 25, 2673–2702.

Boyson, Nicole M., 2009, The impact of hedge fund family membership on performance and market share, *Journal of Financial Transformation* 25, 123–129.

Breen, Richard, 1996, *Regression Models : Censored, Sample Selected, or Truncated Data — Sage University Paper Series on Quantitative applications in the Social Sciences, 07-111*, Sage University Paper series on Quantitative applications in the social sciences, 07-111 (Sage, Thousand Oaks, CA).

Brown, Stephen J., William Goetzmann, Bing Liang, and Christopher Schwarz, 2012, Trust and delegation, *Journal of Financial Economics* 103, 221–234.

Brown, Stephen J., William N. Goetzmann, and Roger G. Ibbotson, 1999, Offshore hedge funds: Survival and performance, 1989-95, *Journal of Business* 72, 91–117.

Brown, Stephen J., William N. Goetzmann, Bing Liang, and Christopher Schwarz, 2008, Mandatory disclosure and operational risk: Evidence from hedge fund registration, *Journal of Finance* 63, 2785–2815.

Brown, Stephen J., William N. Goetzmann, Bing Liang, and Christopher Schwarz, 2009, Estimating operational risk for hedge funds: The ω-score, *Financial Analysts Journal* 65, 43–53.

Brown, Stephen J., William N. Goetzmann, and James Park, 2001, Careers and survival: Competition and risk in the hedge fund and cta industry, *Journal of Finance* 56, 1869–1886.

Brown, Stephen J., and Christopher Schwarz, 2011, The impact of mandatory hedge fund portfolio disclosure Unpublished working paper.

Buraschi, Andrea, Robert Kosowski, and Worrawat Sritrakul, 2012, Incentives and endogenous risk taking: A structural view of hedge funds alphas Unpublished working paper.

Carhart, Mark M., Jennifer N. Carpenter, Anthony W. Lynch, and David K. Musto, 2002, Mutual fund survivorship, *Review of Financial Studies* 15, 1439–1463.

Cassar, Gavin, and Joseph Gerakos, 2011, Hedge funds: Pricing controls and the smoothing of self-reported returns, *Review of Financial Studies* 24, 1698–1734.

Chen, Qi, Itay Goldstein, and Wei Jiang, 2010, Payoff complementarities and financial fragility: Evidence from mutual fund outflows, *Journal of Financial Economics* 97, 239–262.

Chevalier, Judith, and Glenn Ellison, 1999, Career concerns of mutual fund managers, *Quarterly Journal of Economics* 114, 389–432.

Christoffersen, Susan E. K., 2001, Why do money fund managers voluntarily waive their fees?, *Journal of Finance* 56, 1117–1140.

Cici, Gjergji, Alexander Kempf, and Alexander Puetz, 2011, The valuation of hedge funds' equity positions Unpublished working paper.

Coval, Joshua, and Erik Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 479–512.

Croushore, D., 2011, Frontiers of real-time data analysis, *Journal of Economic Literature* 49, 72–100.

Danielsson, Jon, Paul Embrechts, Charles Goodhart, Con Keating, Felix Muennich, Olivier Renault, Hyun Song Shin, et al., 2001, An academic response to basel ii, *LSE Financial Markets Group Working Paper No. 130* .

De Ceuster, Marc J.K., Geert Dhaene, and Tom Schatteman, 1998, On the hypothesis of psychological barriers in stock markets and benford's law, *Journal of Empirical Finance* 5, 263–279.

de Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, Positive feedback investment strategies and destabilizing rational speculation, *Journal of Finance* 45, 379–395.

Dechow, Patricia, Weili Ge, and Catherine Schrand, 2010, Understanding earnings quality: A review of the proxies, their determinants and their consequences, *Journal of Accounting and Economics* 50, 344–401.

Dimmock, Stephen G., and William C. Gerken, 2012, Predicting fraud by investment managers, *Journal of Financial Economics* 105, 153 – 173.

Durtschi, C., W. Hillison, and C. Pacini, 2004, The effective use of benford's law to assist in detecting fraud in accounting data, *Journal of Forensic Accounting* 5, 17–43.

Edelman, D., William Fung, and David A. Hsieh, 2012, Exploring uncharted territories of the hedge fund industry: Empirical characteristics of mega hedge fund firms Unpublished working paper.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 2003, Incentive fees and mutual funds, *Journal of Finance* 58, 779–804.

Elton, Edwin J., Martin J. Gruber, and T. Clifton Green, 2007, The impact of mutual fund family membership on investor risk, *Journal of Financial and Quantitative Analysis* 42, 257–277.

Evans, Richard B., 2010, Mutual fund incubation, *Journal of Finance* 65, 1581 – 1611.

Foster, Dean P., and H. Peyton Young, 2010, Gaming performance fees by portfolio managers, *The Quarterly Journal of Economics* 125, 1435–1458.

French, K. R., 2008, Presidential address: The cost of active investing, *Journal of Finance* 63, 1537–1573.

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

Fung, William, and David A. Hsieh, 2000, Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases, *Journal of Financial and Quantitative Analysis* 35, 291–307.

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

Fung, William, and David A. Hsieh, 2006, Hedge funds: An industry in its adolescence, *Federal Reserve Bank of Atlanta Economic Review* 1–34.

Fung, William, and David A. Hsieh, 2009, Measurement biases in hedge fund performance data: An update, *Financial Analysts Journal* 65, 36–38.

Fung, William, David A. Hsieh, Narayan Y. Naik, and Tarun Ramadorai, 2008, Hedge funds: Performance, risk, and capital formation, *Journal of Finance* 63, 1777–1803.

Gallaher, Steven, Ron Kaniel, and Laura Starks, 2006, Madison avenue meets wall street: Mutual fund families, competition and advertising Unpublished working paper.

Gaspar, J. M., M. Massa, and P. Matos, 2006, Favoritism in mutual fund families? evidence on strategic cross-fund subsidization, *Journal of Finance* 61, 73–104.

Géhin, Walter, and Mathieu Vaissié, 2005, Lighthouses or tricks of light? an in-depth look at creating a quality hedge fund benchmark, *Journal of Indexes* .

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

Goetzmann, William N., Jr. Ingersoll, Jonathan E., and Stephen A. Ross, 2003, High-water marks and hedge fund management contracts, *Journal of Finance* 58, 1685–1717.

Gregoriou, Greg N., and François-Serge Lhabitant, 2009, Madoff: A flock of red flags, *Journal of Wealth Management* 12, 89–97.

Guasoni, Paolo, and Jan K. Obloj, 2011, The incentives of hedge fund fees and high-water marks Unpublished working paper.

Guidotti, Ivan, and Istvan G. Nagy, 2011, Rational behavior of hedge fund managers and investors Unpublished working paper.

Hausman, Jerry, 2001, Mismeasured variables in econometric analysis: Problems from the right and problems from the left, *Journal of Economic Perspectives* 15, pp. 57–67.

Hill, T. P., 1998, The first digit phenomenon: A century-old observation about an unexpected pattern in many numerical tables applies to the stock market, census statistics and accounting data, *American Scientist* 86, pp. 358–363.

Hodder, James E., and Jens Carsten Jackwerth, 2007, Incentive contracts and hedge fund management, *Journal of Financial and Quantitative Analysis* 42, 811–826.

Hodgson, Tim M., Stephan J. Breban, Chris L. Ford, Michael P. Streatfield, and Roger C. Urwin, 2000, The concept of investment efficiency and its application to investment management structures, *British Actuarial Journal* 6, 451–545.

Hsieh, David A., 2006, The search for alpha - sources of future hedge fund returns, *CFA Institute Conference Proceedings Quarterly* September 2006, 79–89.

Jin, Ginger Zhe, and Phillip Leslie, 2003, The effect of information on product quality: Evidence from restaurant hygiene grade cards, *Quarterly Journal of Economics* 118, 409–451.

Jin, Ginger Zhe, and Phillip Leslie, 2009, Reputational incentives for restaurant hygiene, *American Economic Journal: Microeconomics* 1, 237–267.

Joenväärä, Juha, Robert Kosowski, and Pekka Tolonen, 2012, Revisiting 'stylized facts' about hedge funds - insights from a novel aggregation of the main hedge fund databases Unpublished working paper.

Jorion, Philippe, and Christopher Schwarz, 2010, Strategic motives for hedge fund advertising Unpublished working paper.

Jotikasthira, Pab, Christian T. Lundblad, and Tarun Ramadorai, 2012, Asset fire sales and purchases and the international transmission of funding shocks, *Journal of Finance* forthcoming.

Jylha, Petri, 2011, Hedge fund return misreporting: Incentives and effects Unpublished working paper.

Khorana, Ajay, Henri Servaes, and Peter Tufano, 2009, Mutual fund fees around the world, *Review of Financial Studies* 22, 1279–1310.

Kolokolova, Olga, 2010, Strategic behavior within families of hedge funds, *Journal of Banking and Finance* 35, 1645—1662.

Kosowski, Robert, Narayan Y. Naik, and Melvyn Teo, 2007, Do hedge funds deliver alpha? a bayesian and bootstrap analysis, *Journal of Financial Economics* 84, 229–264.

Lan, Yingcong, Neng Wang, and Jinqiang Yang, 2012, The economics of hedge funds: Alpha, fees, leverage, and valuation Unpublished working paper.

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

Liang, Bing, 2003, The accuracy of hedge fund returns, *Journal of Portfolio Management* 29, 111–122.

Liang, Bing, and Hyuna Park, 2008, Share restrictions, liquidity premium, and offshore hedge funds Unpublished working paper.

Ljungqvist, Alexander, Christopher Malloy, and Felicia Marston, 2009, Rewriting history, *Journal of Finance* 64, 1935–1960.

Maddala, G. S., 1983, *Limited-dependent and Qualitative Variables in Econometrics*, Econometric Society monographs in quantitative economics (Cambridge University Press, Cambridge).

Massa, Massimo, 2003, How do family strategies affect fund performance? when performance-maximization is not the only game in town, *Journal of Financial Economics* 67, 249–304.

Naik, Narayan Y., Tarun Ramadorai, and Maria Stromqvist, 2007, Capacity constraints and hedge fund strategy returns, *European Financial Management* 13, 239–256.

Nanda, Vikram, Z. Jay Wang, and Lu Zheng, 2004, Family values and the star phenomenon: Strategies of mutual fund families, *Review of Financial Studies* 17, 667–698.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Nigrini, Mark J., and Linda J. Mittermaier, 1997, The use of benford's law as an aid in analytical procedures, *Auditing* 16, 52–67.

Panageas, S., and M. M. Westerfield, 2009, High-water marks: High risk appetites? convex compensation, long horizons, and portfolio choice, *Journal of Finance* 64, 1–36.

Pástor, Ľuboš, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Patton, Andrew J, and Tarun Ramadorai, 2012, On the high-frequency dynamics of hedge fund risk exposures, *Journal of Finance* forthcoming.

Patton, Andrew J., Tarun Ramadorai, and Michael P. Streatfield, 2012, Change you can believe in? hedge fund data revisions Unpublished working paper.

Politis, Dimitris N., and Joseph P. Romano, 1994, The stationary bootstrap, *Journal of the American Statistical Association* 89, 1303–1313.

Pollet, J. M., and M. Wilson, 2008, How does size affect mutual fund behavior?, *Journal of Finance* 63, 2941–2969.

Ramadorai, Tarun, 2011, The secondary market for hedge funds and the closed-hedge fund premium, *Journal of Finance* 67, 479–512.

Ramadorai, Tarun, 2012, Capacity constraints, investor information, and hedge fund returns, *Journal of Financial Economics* forthcoming.

Ramadorai, Tarun, and Michael P. Streatfield, 2011, Money for nothing? understanding variation in reported hedge fund fees Unpublished working paper.

Rogers, William H., 1983, Analyzing complex survey data, *Unpublished paper, Rand Corporation, Santa Monica, CA* .

Rogers, William H., 1993, Regression standard errors in clustered samples, *Stata Technical Bulletin Reprints* STB-13: STB-18, 88–94.

Schwarz, Christopher, 2007, Hedge fund fees Unpublished working paper.

Schwarz, Christopher, and Mark E. Potter, 2012, Revisiting mutual fund portfolio disclosure Unpublished working paper.

Sirri, Erik R., and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.

Stewart, Mark B., 1983, On least squares estimation when the dependent variable is grouped, *Review of Economic Studies* 50, 737–753.

Stigler, George J., 1961, The economics of information, *Journal of Political Economy* 69, 213–225.

Straumann, Daniel, 2009, Measuring the quality of hedge fund data, *Journal of Alternative Investments* 12, 26–40.

Stulz, René M., 2007, Hedge funds: Past, present, and future, *Journal of Economic Perspectives* 21, 175–194.

Teo, Melvyn, 2009a, Does size matter in the hedge fund industry? Unpublished working paper.

Teo, Melvyn, 2009b, The geography of hedge funds, *Review of Financial Studies* 22, 3531–3561.

Varian, Hal R., 1972, Benford's law, *American Statistician* 26, 65–6.

Verhoeff, J., 1969, Error detecting decimal codes, *Mathematical Centre Tracts* 29, 1–118.

White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817–838.

Zhong, Zhaodong (Ken), 2008, Why does hedge fund alpha decrease over time? evidence from individual hedge funds Unpublished working paper.