

Exploring Uncharted Territories of the Hedge Fund Industry: Empirical Characteristics of Mega Hedge Fund Firms*

By

Daniel Edelman**

William Fung***

David A. Hsieh****

Date of this version:

October 29, 2012

Abstract: This paper investigates mega hedge fund management companies that manage over 50% of the industry's assets, incorporating previously unavailable data from those that do not report to commercial databases. We document similarities among mega firms that report performance to commercial databases compared to those that do not. We show that the largest divergences between the performance of reporting and non-reporting can be traced to differential exposure to credit markets. Thus the performance of hard-to-observe mega firms can be inferred from observable data. This conclusion is robust to delisting bias and the presence of serially correlated returns.

Keywords: hedge funds, asset management

JEL Classifications: G11, G12, G23

* We are grateful for comments from Stephen Brown, Will Goetzmann, Philippe Jorion, Juha Joenväärä, Robert Kosowski, Narayan Naik, Melvyn Teo, Robert Whaley, participants of the Oxford-Man Hedge Fund Conference in November 2010, Q-Group Fall 2011 meeting in October 2011, Financial Markets Research Center Conference on Hedge Funds at Owen, Vanderbilt in October 2011, and Kellogg-JOIM Fall 2012 Conference in October 2012.

** Head of Quantitative Research and Development, Alternative Investment Solutions.

*** Visiting Research Professor, London Business School. Correspondence author: William Fung, London Business School, Regent's Park, London NW1 4SA, United Kingdom, email: bfung@london.edu.

****Professor, Fuqua School of Business, Duke University.

1. Introduction

For over half a century wealthy individuals have invested their capital alongside *talented* traders in opaque, lightly regulated private investment vehicles. These exclusive investment vehicles managed by eclectic investment professionals have come to be referred to as hedge funds.¹ While hedge funds may have benefited from limited regulatory purview, privacy does come with a cost. Decades of voluntary reporting to publicly available commercial databases have left the industry with no standardized, comprehensive source of performance data comparable to mutual funds. Even for those managers who choose to report performance to a commercial database, they often do so selectively, reporting some but not all of the product offerings in their firms.² And for those funds whose monthly return data are available in commercial databases, there are often gaps in their reported asset under management (“AUM”) histories. Today hedge funds researchers face two critical data gaps—missing data on “*non-reporting hedge fund firms*” (hedge fund management firms that do not participate in commercial databases), incomplete data records from “reporting firms” (those that report data to commercial databases)—both of which entail the collection of thousands of missing funds and the cleaning of hundreds of thousands of data records spanning multiple years. This begs the question “is there a manageable way of bridging these data gaps so that we can better assess the capital formation process and performance characteristics of the hedge fund industry?”

This paper documents extensive empirical evidence to address these two questions on this data gap that overshadows the extant literature on hedge fund research. First we establish a tractable way to measure the capital formation process of the hedge fund industry which substantially expands the AUM coverage of commercial databases. Specifically we identify previously unexplored data sources on non-reporting firms and show that their inclusion adds 42% to the total AUM of firms reporting to a consolidation of commonly used data sources such as BarclayHedge, HFR, and Lipper TASS (collectively referred to as the “commercial data set”) at the end of 2001 rising to 65% by the end of 2010. Second we explore the question “can performance characteristics of non-reporting firms be inferred from reporting firms?” We

¹ It is generally accepted that Alfred W. Jones started the first hedge fund well over fifty years ago in 1949. Exclusivity is typically achieved by imposing a much higher minimum investment and carrying less than generous liquidity terms to potential investors.

² As an example, the ADV filings of Paulson & Company and Paulson Management, LLC reveal that this firm operates some six different hedge funds, none of which is in the Lipper TASS database as of the end of 2010; three are not included in the BarclayHedge database, and reporting of data to HFR started well after the inception of a number of the funds.

manually collected performance data on non-reporting firms as well as missing returns from reporting firms. There is no complete registry of hedge funds. Our coverage of the large hedge fund firms is limited by the coverage of two surveys published by *Institutional Investor* and *Absolute Return+Alpha*. Although these two sources alone significantly increase the AUM coverage of commonly used commercial hedge fund databases, a complete census of large hedge fund firms cannot be accurately determined. At the other end of the spectrum, the lack of a complete registry of hedge funds also implies that we cannot verify the number of smaller hedge fund firms that do not report to the commercial databases used in this study. Effort to improve AUM coverage of smaller non-reporting hedge fund firms can be challenging as the number of firms needed to create the AUM coverage by a similar amount is likely to be substantially larger.³

To facilitate easy comparison to conventional hedge fund indices we construct equally weighted as well as AUM-weighted indices of non-reporting (“NRIs”) and reporting (large) firms controlling for size (“RLIs”). Our paper is the first to directly compare the performance characteristics of NRIs to commonly used hedge fund indices such as the Hedge Fund Research Fund Weighted Composite Index (“HFRI”) and the Dow Jones Credit Suisse Hedge Fund Index (“DJCSI”).⁴ There is, however, an important point of departure in our data construction method from the extant literature in the way we identify large hedge funds. The unit of analysis in our paper is the hedge fund management company (“hedge fund firm”). Therefore, size and performance data pertains to the level of the hedge fund management company as a whole as distinct from previous studies which typically utilize individual hedge funds as their basic unit of analysis.

Measuring industry performance using firm level data grants new insights into the decision-making process of investors as well as fund managers. From the demand side, large institutional investors, due to scale economy, can exert influence on the structuring of a hedge fund product. These investors’ decision criteria are motivated not only by the features of individual funds, but also by the characteristics of the hedge fund management company—which encompasses the strategy dimension as well as a host of other operational characteristics. Consequently, for institutional investors, proper due diligence should embrace both individual hedge funds characteristics and the track record of the management company itself as an asset management business. Table 1 and Figure 1 illustrate the growing presence of institutional

³ We are able to reach the 60% increase of AUM coverage by collecting data on some 350 large hedge fund firms.

⁴ See www.hfr.com and www.hedgeindex.com for details.

investors in the hedge fund industry. From the supply side, hedge fund managers, like other active managers, often face capacity constraints and increasingly so when strategies are leveraged. Consistent with the theoretical predictions of Berk and Green (2004), when the AUM of a hedge fund firm grows, the array of products it offers typically broadens to absorb rising capital inflow and to diversify its business. Although capacity considerations can limit the size of an individual fund, they may only have a muted impact on the AUM growth of hedge fund firms. Therefore, the estimated growth of large hedge funds will understate the growth of large hedge fund firms leading to a downward bias in estimates of AUM concentration in the industry. These observations on the operating characteristics of the hedge fund industry, when taken together, suggest that the hedge fund firm itself rather than the individual hedge fund it manages may be a more natural unit of analysis of the industry. However, using firms rather than funds as the basic unit of analysis entails a significant data management exercise which calls for the conversion of commercially available databases (typically organized by individual funds) to a consolidated database of unique hedge fund firms whose data is publicly available.⁵ To the best of our knowledge, our paper is the first to do this utilizing the three major commercial databases.

Non-reporting firms' data, by definition, are not available through commercial databases, in this study we manually collected data from a number of data sources which include industry surveys, regulatory filings, fund managers themselves and the funds' investors.⁶ We use the term "privately collected data" to distinguish them from the data in the commercial data set. Section 2 provides a detail description of helpful sources of industry "AUM" data with long histories. We describe how monthly return data is collected and the extent to which this additional source of data contributes to fill the data gap on the performance of non-reporting hedge fund firms. The section ends with how firm-wide data is constructed for the hedge fund firms in our expanded

⁵ It is a non-trivial task to overcome the myriad of naming conversions of funds and their management companies from one commercial database to another so as to avoid double counting and duplicated data entries in our consolidated database.

⁶ Regulatory data is one source of performance information that can augment the offerings of commercial vendors. For example, Och-Ziff Capital Management Group, which is ranked #7 in the 2009 survey of the largest hedge fund firms (see Section 2 for more details on survey data), does not report to the three commercial databases in our paper. Each month, however, Och-Ziff files a Form 8-K with the SEC, with monthly returns of four master funds, as well as the firm's AUM. The same is true of Fortress Investment Group. Another example is the US Senate's Permanent Subcommittee on Investigations, which published a report entitled "Excess Speculation in the Natural Gas Market"; from this one can obtain the monthly returns of Amaranth. In Aiken et al (2012) used publicly traded funds-of-hedge funds which must disclose the performance of their underlying sub managers to obtain returns for large hedge funds that do not report to commercial databases. Beyond commercial databases and regulatory information, hedge fund investors themselves are by far the best source of performance data—investors may be privy to private information coming directly from prime brokers, administrators, third-party marketers, and fund managers themselves. However, obtaining data from this source entails strict confidentiality and non-disclosure agreements.

data set which combines both public and private sources of data. Section 3 of the paper provides insight on the unconditional return distributions of the non-reporting firms to their observable counterparts. Controlling for firm size, we document evidence which shows the unconditional return distribution of NRIs are similar to their reporting counterparts as well as publicly available hedge fund indices such as the HFRI and the DJCSI. Continuing with our analysis of data gaps in hedge fund returns, we examine the effect of missing returns from reporting firms in commercial databases. A benefit of having access to privately collected data is the ability to locate returns of hedge funds that stopped reporting to commercial databases but continued to report performance to investors. This provides us with a rare opportunity to glean insight on the behavior of “missing returns.” Successful hedge fund firms often voluntarily exit from commercial databases as they are no longer actively seeking new investors. It is reasonable to suspect that their missing returns may be better than those of the average hedge fund in the commercial data set. This “missing return bias” was first noted in Ackerman, McEnally and Ravenscraft (1999). However, an opposite conjecture exists for a delisted, liquidated fund whose missing returns may well be the final months of its existence. Including these missing returns may lead to lower average returns and a rise in return volatility compared to the observable data set. This is a major contention raised in Posthuma and van der Slius (2003) and Malkiel and Saha (2005). Thus we may have two opposing biases from missing returns which together constitute a “delisting bias” in observable hedge fund returns. In practice, there are many other operational reasons why a firm elects not to participate in commercial databases and equally firms are delisted by database vendors for reasons other than liquidation. For these delisted firms, there is no obvious reason to suspect any systemic bias in their missing returns. To date, the academic literature has not been able to determine the net effect of delisting biases because missing returns have a tendency to, as the description suggests, remain unobservable. We contribute to the extant literature by directly estimating the net impact of missing returns on the hedge fund industry’s performance. Our estimates suggest that the different motivations for delisting might have neutralized each other over the sample period 2002-2010.

Given the laborious effort required to gather performance data from non-reporting firms the results in Section 3 naturally motivate the question: do existing and future research results on reporting large firms be extend to hard-to-observe non-reporting firms? A stumbling block to inferring the return behavior of NRIs based on the observable data such as RLIs and publicly available indices is the dynamic nature of hedge fund strategies. Studies such as Fung, Hsieh, Naik, and Ramadorai (2008) and Bollen and Whaley (2009) document discrete time-varying risk

taking behavior of hedge fund managers. Section 4 is devoted to analyze the relationship between NRIs and LRIs allowing for differences in time-varying risk exposures. Specifically, we address the key question: *can NRIs' returns be predicted based on observable data?*

Section 4 begins with addressing a well-known empirical regularity in hedge fund returns—the existence of return serial correlation which raises the question: do NRI returns differ from observable returns in this regard? To answer this question, we disentangle the effect of hedge fund managers smoothing their returns from return serial correlation arising from exposure to illiquid risk factors. Section 4 of the paper provides empirical evidence pointing to observed serial correlation from illiquid risk factors being the main driver of observed return serial correlations in the hedge fund portfolios (indices) we constructed for this study. Accordingly we apply a less commonly used version of the Getmansky, Lo and Makarov (2004) model to obtain unbiased beta estimates of the Fung, Hsieh, Naik, and Ramadorai (2008) model and its extension in Edelman, Fung, Hsieh, and Naik (2012). This allows us to document the time series properties of the spread between NRIs and observable hedge fund indices—RLIs and commonly used indices such as HFRI and DJCSI—in the presence of return serial correlation. Applying a similar modeling approach as in Fung, Hsieh, Naik, and Ramadorai (2008) we find no sample-break points in the return spread between the NRIs and their observable counterparts. In addition, Section 4 documents empirical evidence that time-varying risk taking behavior leading to performance differences between non-reporting and reporting firms do occur during large moves in illiquid securities in the credit markets.

Our paper makes five contributions to the literature. First, we identify previously unexplored sources of reliable, survey data on the size of the hedge fund industry which help us to extend significantly the AUM coverage of commercially available hedge fund databases using only a manageable number of mega hedge fund firms. Second, using firm-level data rather than individual hedge funds as the basic unit for analyzing the hedge fund industry, we are able to create historical AUM series from our consolidated commercial database which, when combined with the survey data we collected, better depict the capital formation process of the hedge fund industry. Third, we manually collect missing returns from non-reporting hedge fund firms from public as well as private sources. This unexplored performance data set provides important insight on how non-reporting firms' return compare to reporting firms of comparable size as well as to commonly used hedge fund indices. Our expanded data set also allows us to estimate directly the effect of delisting bias which supports the conclusion that missing returns from

liquidating funds tend to be offset by missing returns from successful funds over our sample period. Fourth, we explore the drivers of commonly reported return serial correlation in hedge funds and show empirically that return serial correlations for diversified portfolios of hedge funds such as hedge fund indices are primarily driven by their exposures to the return serial correlation of illiquid risk factors to which they are exposed. This allows us to estimate the risk exposures of reporting and non-reporting firms free of biases that may arise from the effect of manager smoothing. Fifth, we document evidence that non-reporting mega firms' returns can be deduced from their reporting counterparts during normal market conditions. We identify market conditions in which significant divergences between reporting versus non-reporting large hedge fund firms can occur. Taken together our results provide researchers a way to extend research conclusions based on commercially available data to previously unobserved large hedge fund firms.

The paper is organized as follows. Section 2 provides a description on our expanded data set combining both commercial databases and missing data set we have collected and reports on the capital formation trend in the hedge fund industry over the past decade. Section 3 analyzes the performance characteristics of reporting and non-reporting hedge fund firms. Section 4 examines the impact of time-varying risk taking behavior on reporting and non-reporting hedge fund firms. Concluding remarks are presented in Section 5.

2. Data description

2.1. Firm-level survey data on assets under management

Our data collection process begins with the printed history of two industry surveys: *Institutional Investor's* annual "Hedge Fund 100," which is a list of the 100 largest hedge fund firms ("*III100*"), and *Absolute Return+Alpha* magazine's semiannual "Billion Dollar Club" which is a compendium of all American firms managing one billion USD or more in assets ("*ARBDC*"). These two surveys have the advantage of a long operating history in collecting data and are therefore free of backfilled bias.⁷ Taken together these two sources of data provides a ten year

⁷ To get a better idea of the data construction process of our two sources, we contacted both organizations. Using the recent survey as an example, the authors of the ARBDC article tell us that 75% of their AUMs are provided directly by fund managers, and the rest are obtained from investment documents, investment sources, and SEC filings. Only in two cases did they have to estimate the assets from the previous surveys where they could not obtain updates. To ensure that the surveys did not miss any firms above the one billion dollar cutoff, the authors sent questionnaires to dozens of firms below the threshold, firms that

history of the top 100 hedge fund firms and a record of all hedge fund firms that managed at least one billion dollars at some point in their operating history, as well as each firm's respective AUM at the end of each year. The choice of a one billion dollar cut-off to identify mega hedge fund firms is somewhat artificial and hindsight suggests this may be on the low side for identifying mega hedge fund firms. Nonetheless a similar rule-of-thumb was used in Eichengreen et al (1998), Fung and Hsieh (2000) and Fung, Hsieh and Tsataronis (2000) who analyze the market impact of hedge funds during different market events. All of these studies rely on manually collected data and we know of no updates to the data sets used in these studies. To the best of our knowledge the II100 and ARBDC surveys are among the most reliable sources of data on mega hedge fund firms and ones that both have continuous and unbiased histories.

Institutional Investor obtains its data from questionnaires filled out by hedge fund managers directly. These questionnaires are supplemented with extensive follow-up calls from reporters, an examination of public filings, and other staff research. According to the description of their survey methodology, II100 includes only comingled assets managed by hedge fund managers excluding separate, bespoke accounts and assets managed in derivative structures, such as CBOs or CDOs.⁸ Also excluded are intermediaries, such as funds-of-hedge funds as well as firms dedicated to the marketing of single hedge funds and have no direct asset management role.⁹ The II100 is widely acknowledged by investment professionals as a complete record of this segment of the hedge fund industry. The first II100 survey, published in June 30, 2002 (based on December 31, 2001 data) captured over USD264 billion in hedge fund assets. At that time, the

recently dropped off the list, and any firms in the one billion dollar neighborhood that they come across after considerable research and their extensive network of advisers. To check that the AUM is correctly reported, ARBDC routinely reviews with the firms the exact source of the assets, ensuring none are funds-of-hedge funds to avoid double counting of AUM, nor notional capital amounts so as to avoid exaggerated AUMs from leveraged investment vehicles. We also contacted the editor at *Institutional Investor* in charge of the II100 rankings who confirmed that this publication uses a similarly rigorous process to determine the largest 100 hedge fund firms. Despite having the same corporate owner—Euromoney—the ARBDC article is prepared completely separate from the II100 article, under different departments with different staff each applying independent methodology for obtaining and verifying assets. When we compare the firms that are common to both lists, we do not find material differences. According to *Institutional Investor*, its circulation is around 115,000; see <http://www.institutionalinvestor.com/Institutional-Investor-Magazine.html?StubID=1033400>.

⁸ Collateralized Bond (Debt) Obligations—CB(D)Os—in which the underlying collateral are hedge funds. Although the II100 survey methodology excludes hedge fund vehicles that are dedicated feeders to another hedge fund so as to avoid double counting, some of the feeder vehicles are included in the earlier surveys. With hindsight, this may seem inappropriate but we adhere to the way the survey methodology is applied so as to avoid ex-post adjustments that may arguable create its own sample selection biases.

⁹ It should be noted that some of the II100 surveys included Madoff feeders such as Fairfield Sentry and Fairfield Greenwich which remain in our dataset. Our data selection criteria preclude ex-post adjustment to data collection that arguably can only be justified with the benefit of hindsight.

largest firm on the list managed eight billion dollars of assets and the 100th (smallest) reported USD645 million AUM; the median II100 firm had USD1.95 billion AUM. The tenth survey, published May, 2011 (based on January 1, 2011 valuations) reports a combined USD1.231 trillion in AUM among these top firms, ranging from the largest at USD58.9 billion to the smallest at USD4.7 billion while the median II100 firm has USD8.05 billion AUM. Table 2 panel B tabulates these statistics. As a firm-level oriented survey the II100 survey does not provide a reconciliation of the reported firm-wide AUM to the individual underlying funds' AUM managed by each firm for all surveyed firms.¹⁰ Consistent with the II survey methodology, the AUM statistics in Table 2 panel B are based on these firm-wide AUM figures from the II100 annual surveys.

Commercially available hedge fund databases capture only a fraction of the entire universe of hedge funds which is consistent with the fact that participation is entirely voluntary. Of the largest 100 hedge fund firms contained in the II100 surveys, over 25% have *never* reported performance to the three commercial hedge fund databases used in this paper. Year-on-year, the number of non-reporting II100 firms ranges from a low of 37 at the end of 2001 to a high of 54 by the end of 2010; see Table 3 panel 3.A. These non-reporting II100 firms managed USD103 billion of AUM in 2001, increasing to USD592 billion of AUM in 2010, as shown in Table 3 panel B. As the II100 surveys provide the names and the AUM of the largest 100 hedge fund firms, it is an important source of data to locate a significant amount of the AUM missing from the commercial databases based on only a limited number of hedge fund firms. Accordingly we began our search for missing data at the end of 2001 to align with the inaugural II100 ranking.

Another source for identifying large hedge fund firms that are missing from the mainstream commercial databases is the ARBDC. The first "Billion Dollar Club" article was written in May 2003 based on December 31, 2002 AUM data.¹¹ The May 2011 survey, reflecting January 1, 2011 assets, reveals 330 firms managing a combined USD1.7 trillion in assets. Like the II100 survey, the ARBDC is also a firm-level oriented survey in which only firm-wide AUM figures are reported. The underlying funds' AUM data in each surveyed firm are not part of the published statistics. Of these 330 firms, 110 firms are outside the mainstream commercial

¹⁰ It is entirely possible that survey staffs did, as part of their due diligence, reconciled firm-wide AUM to the AUM of the funds managed by each surveyed firm annually. We are not able to locate publicly available records of such reconciliations.

¹¹ The May 2003 ARBDC survey contained information about firms in 2001, which we used to estimate the "Billion Dollar Club" of 2001.

databases and the II100 list—see Table 2 panel 2C for this and additional statistics of the ARBDC firms. Table 3 shows that these 110 non-reporting firms collectively manage USD270 billion. Consistent with the survey methodology, both Table 2 panel C and Table 3 refer to firm-level AUM of the annual ARBDC surveys.

Together, II100 and ARBDC provide the names and AUMs of 164 mega hedge fund firms controlling some USD863 billion of assets in 2010 outside the mainstream commercial databases. Table 2 shows that by the end of 2010 the median AUM of firms in the II100 survey reached USD 8,750 billion which is nearly 151 times larger than the median AUM firm in the commercial data set. The comparable figure for the ARBDC survey is just under 45 times. If one were to expand the AUM coverage of the commercial data set using only reporting firms, the number of missing firms to identify will quickly rise to several thousands. Extrapolating from the data in Table 1 and its graphic representation in Figure 1, institutional investors are likely to remain a major source of capital for the hedge fund industry. The size of the capital institutional investors have to deploy is likely to motivate a preference for larger hedge fund firms.¹² This provides a consistent explanation for the more than four-fold AUM increase for the median II100 firm over our sample period. Taking these observations together leads to the conclusion that effort to gain insight on the performance of assets invested in the hedge fund industry may be better placed seeking out data on mega hedge fund firms. Our data analysis shows this to be a tractable and efficient approach compare to the alternative of expanding database coverage by adding a large number of small funds.

2.2. Commercially available data

The majority of empirical studies on hedge funds rely on return data and descriptive information in publicly available commercial databases. In this paper, we merge the entries of three major providers of hedge fund data—BarclayHedge, HFR, and Lipper TASS, to arrive at a single *commercial data set*. Each provider offers both active and graveyard databases of funds past and present. As of this writing, these vendors collectively account for over 41,000 raw data entries. However, the actual number of unique funds is considerably smaller. Consistent with the II100 approach, we remove all intermediaries including funds-of-hedge funds. We then group all

¹² If only on a margin return to due diligence basis, the effort to review a small hedge fund firm may not be materially less than that of a large one which adds to the search cost of seeking talented small managers relative to the capital that the investor can deploy.

funds across the three databases by the respective management firms carefully avoiding duplications of firms that may arise from widely different naming conventions used by different database vendors. Following II100's methodology, we stay with the concept that a *hedge fund firm* refers to the fund's investment advisor, as opposed to the marketing company, sponsor, or other manager not responsible for actual trading decisions.

For each firm, we create a set of unique funds. Here, a *fund* refers to a specific strategy or investment objective, rather than *share classes*. To appeal to different clientele, a firm often offers multiple share classes of a fund. Offshore share classes cater to non-US investors, while onshore share classes typically target US investors there may also be share classes denominated in different currencies, such as USD, EUR, JPY, etc. to broaden the appeal of the fund to global investors. While data providers often refer to share classes as distinct funds, we aggregate multiple share classes pursuing the same strategy into a single fund. Duplicate share classes from different databases with nearly identical names and essentially the same returns are removed. A single AUM is then created for the fund by aggregating the AUM of all the unique share classes. We create a single return series for the fund using the returns of a representative share class. Typically, the offshore share class is preferred over the onshore vehicle and similarly for USD share class over non-USD share classes. If we must use a non-USD share classes, we convert its return into USD adjusting for the interest differential between the reporting currency and the US Dollar. We refer to the firms in the commercial data set as "reporting firms" and those firms that do not participate in the commercial data set as "non-reporting firms." Table 2 panel A tabulates the reporting firms' statistics. At the end of 2010, our commercial data set has 2,443 firms with AUM of USD1,322 billion, as shown in Table 2. By adding just the 164 non-reporting mega firms in II100 (54 firms) and ARBDC (110 firms) which is a mere 2.62% increase in the number of reporting firms we increase the AUM coverage of the commercial data set by 65% in 2010. This is an important insight to the capital formation process of the hedge fund industry—that substantial amount of the hedge fund industry's AUM growth can be traced to the AUM increases of a small fraction of the number of management firms in the industry. It also underscores the important role non-reporting mega firms' returns play in measuring the performance of assets invested in the industry.¹³

2.3. *Privately collected data on fund performance*

¹³ As distinct from measuring the performance of funds a large number of which are simply too small to attract the influx of institutional investors' capital over the past decade.

We create an expanded data set by first loading the reporting firms from the commercial data set. Next we add the data from the II100 and ARBDC surveys. If a firm in the surveys is a reporting firm, we substitute its year-end AUM in the expanded data set with the year-end AUM from the surveys. We then add the non-reporting firms' names and AUM figures to the expanded data set and we procure monthly returns of its funds using a variety of sources including hedge fund investors, fund managers, fund administrators, prime brokers, third-party marketers, government documents, and regulatory filings. For simplicity, we refer to the additional data of the non-reporting firms as our *privately collected data*. To the best of our knowledge, this is the first paper that reports the empirical characteristics of hedge funds drawn from such a comprehensive data set that integrates both commercially available and privately collected sources.

To provide an idea of the coverage of the privately collected data, some terminology is needed. Define a survey firm as a component firm in on of the annual II100 or ARBDC surveys and a survey fund as a fund managed by a survey firm. First, we use the commercial dataset as a reference point to assess the marginal contribution of collecting performance data from private sources. For *reporting firms* in all annual surveys, we are able to collect more monthly return data (both in terms of missing returns from funds in the commercial data set as well as omitted funds).¹⁴ Second, the II100 annual survey does provide an incomplete list of survey funds targeting the biggest AUM funds in each surveyed firm depending on availability. The number of reported II-survey funds in the entire sample is 2,131 of which we have performance data on 2,032 survey funds or a coverage rate of 95.35%. The lowest percentage survey funds coverage is 94.22% in 2002 and the highest is 97.57% in 2005. The ARBDC survey does not contain any survey fund information so there is no equivalent coverage calculation. Third, there are firms in each annual survey for which we are unable to procure performance data. For the II100 survey, the lowest coverage of surveyed firms is 93% and the highest is 98%. For the ARBDC survey, the lowest coverage of surveyed firms is 91% and the highest is 96%.¹⁵

¹⁴ For example, at the end of 2010, there were 12,110 share classes in our commercial data set. We aggregated them into 6,182 unique funds. The commercial data set captured 946 funds managed by II100 firms that are in the surveys over our sample period. Our privately collected dataset increased this to 1,554 funds. For all the reporting mega firms in our sample our private sources uncover fund data beyond what is reported in the commercial data set.

¹⁵ In terms of AUM coverage, the lowest and highest for the II100 surveyed firms are also 93% and 98% but they don't occur in the same year as when coverage is measured in number of firms. The corresponding AUM coverage range for the ARBDC survey is a low of 92% and high of 96%.

2.4. Missing Returns from reporting funds

Another benefit of having access to data outside of commercial databases is the ability to deal with a data gap among reporting firms—those missing returns from funds delisted by commercial database vendors. There are two types of hedge funds that stop reporting returns to a database. At one extreme, successful hedge fund businesses operating at close to capacity limit and are no longer actively marketing their services may stop reporting to commercial databases. As a result, the data providers must delist them—but as a consequence of the hedge fund manager’s choice. Let us call this voluntary delisting. *Prima fascia*, their missing returns are likely to be better than those of a typical hedge fund. Adding back their returns post delisting would raise the average return in the commercial data set. This observation was noted in one of the earliest research in hedge funds by Ackerman et al (1999). At the other extreme, consider funds that undergo liquidation, which typically occurs when a fund incurs substantial losses leading to capital withdrawal from investors. These funds often only provide patchy reporting to commercial databases during the final liquidation phase. In these cases, often database providers may actively delist these funds prior to the final liquidation point thereby missing the worst returns from the databases. Refer to these cases as involuntary delisting. Adding back missing returns from involuntary delisting could lower the average returns in the commercial data set. This was a major contention raised in Posthuma and van der Slius (2003) and Malkiel and Saha (2005). Thus we may have two opposite biases from delisted funds that exited databases. To date, the academic literature has not been able to resolve this debate on delisting bias, since the answer depends on being able to observe the missing, post delisting returns. Our expanded data set can shed light on this issue. Section 3.3 provides empirical insight to this unresolved issue in the literature on the effect of delisting bias on hedge fund returns.

2.5. Measuring the size of the data gap: an index of non-reporting hedge fund firm—NRI

The choice of using annual AUM data is motivated by three issues. First, monthly AUM figures from the commercial databases are incomplete in that not every reported monthly return is accompanied by the corresponding AUM figure. Over our sample period, BarclayHedge has 197,067 monthly returns but only 164,346 monthly AUM figures. For HFR, there are 345,370 monthly returns but only 257,881 AUM figures. For Lipper TASS, there are 250,705 monthly returns but only 155,262 AUM figures. Second, year-end AUM figures tend to coincide with accounting audits, fee calculations as well as redemption periods which make them more reliable

than monthly AUM figures. Third, AUM figures for the non-reporting II100 (ARBDC) firms are only available once (twice) a year.

Since the complete universe of hedge funds is not directly observable we appeal to another commonly used industry survey as a reference check on the scope of our expanded data set's industry coverage. Post Madoff, most hedge fund vehicles are independently audited and engage institutional service providers who act as custodian of the fund's assets and administrative manager of the hedge fund vehicle. HedgeFund.Net and Advent Software conduct an annual survey of hedge fund assets under administration ("AUA"). This is a comprehensive survey covering nearly all hedge fund administrators (*AUA survey*). Panel 3.A in Table 3 shows the 2010 AUA survey reporting a total hedge fund industry AUM of USD2.83 trillion; within which USD1.32 trillion is managed by reporting firms and USD0.86 trillion is managed by non-reporting firms we have identified.¹⁶

Figure 2 illustrates the magnitude of the missing assets as well as the concentration of assets in large firms. The green bars represent the total AUM of reporting firms each sample year. The blue bars represent the AUM of the non-reporting firms while the purple bars show the additional assets in the AUA survey. Consistently, the AUM of non-reporting mega firms (in blue bars) together with the reporting firms (the green bars) in the commercial data set represents the bulk of the assets in the industry.¹⁷ Figure 2 confirms that the non-reporting hedge fund firms in our expanded data set represent a reasonable proxy of the hedge fund data gap which we defined as the set of all non-reporting hedge fund firms. Accordingly, we use this non-reporting firm set to construct an index of non-reporting hedge fund firms—the NRI.¹⁸

¹⁶ The AUA survey may be biased upwards in terms of assets under administration as they do not eliminate those assets that utilize more than one administrator. Total hedge fund industry assets would consist of AUA plus the assets of self-administered hedge funds, minus any double counting such as cross-investments of hedge funds into other funds. After the recent hedge fund frauds and the market meltdown of 2008, hedge fund investors have put pressure on the few funds that self-administer to seek better reporting standards. We believe the amount of self-administered AUM has fallen substantially over the past years. Hence AUA is a likely a very good proxy of the true aggregate industry AUM.

¹⁷ Notwithstanding the fact that the AUA survey data may lead to an upward bias of total industry AUM.

¹⁸ It should be noted that while we have significantly narrowed the data gap between the commercial data set and the AUA survey in terms of AUM, we have yet to completely eliminate the gap. How much of the unaccounted for AUM in Figure 2 is simply due to an upward bias from measurement issues (AUA survey methodology versus that of II100 and ARBDC) or a result of missing large hedge fund firms beyond our data set remains an open question. It is also possible that there are numerous missing small hedge fund firms, which are naturally not in the surveys of large hedge fund firms, but have also elected not to report to data bases.

3. The effect of missing returns on observed return distributions of hedge fund returns

Since non-reporting firms' data are hard to observe requiring a substantial amount of data collection effort, it will be helpful to know if sufficient similarity exists between non-reporting and reporting firms' returns such that research conclusions based on the observable universe of hedge fund data extends to non-reporting firms. We begin by comparing the unconditional distribution of returns from reporting and non-reporting firms.

3.1. Indexes of reporting and non-reporting firms

In order to contrast the return properties of non-reporting hedge fund firms to the observable reporting universe, we create a series of comparisons to publicly available hedge fund indices like the DJCSI and HFRI as well as statistical averages from our commercial data sets. These comparisons help to link return statistics of non-reporting firms to the observable hedge fund universe without having to undertake the data construction tasks we described in section 2. We begin by comparing non-reporting firms to public hedge fund indices—the HFRI and DJCSI. We create indices of non-reporting firms in a manner consistent with the index construction method of these hedge fund indices.

First, for each non-reporting hedge fund firm, a firm-wide composite return is computed as follows. At the start of each year, an equal amount of capital is invested into each of the fund managed by the firm. We then track the return of this portfolio for the next twelve months without reallocation across funds. Our choice of annual rebalancing is dictated by a lack of confidence in individual fund's intra-year AUM data. Averaging across product offerings of a firm also avoids the problem of overweighting matured strategies that may have gathered more assets over time compared to new strategies that are at the beginning of their life cycle. If a fund stops reporting return during the year, the fund is assumed to be in liquidation and the proceeds are reinvested in Treasury bills until the end of the year.¹⁹ If the firm launches a new fund in the middle of the year, it is included in the portfolio of the following year. In this calculation, only funds that report returns net of all management fees and incentive fees are used. Applying the

¹⁹ The other commonly used alternative is to reinvest the liquidated funds into the remaining operating funds in the following month. However this requires the assumption of capital being released from liquidating funds in a time frame that may be unrealistic. While there is no realistic way to estimate the return of a fund during the final days of its operating history, we make the reasonable assumption that as risky assets are sold, proceeds are kept liquid and returned to investors as the fund closes down.

same methodology we create returns series for reporting firms using data from the commercial data set.

Second, using the individual non-reporting firm's return series we construct two indices of non-reporting firms using standard index construction methodology—NRI-EW which is an equally weighted index of the constituent non-reporting firms (for comparison to the equally weighted HFRI) and NRI-AW which is an index weighted according to the respective firm-wide AUM of the constituent non-reporting firms (for comparison to the AUM weighted DJCSI index). Both indices are rebalanced at the end of each sample year using end-of-year statistics for the respective firms.

Third, to augment the two broad-based hedge fund indices we divide the reporting firms into two subsets according to AUM so as to control for the impact of firm size in the comparisons with NRI. At the beginning of each formation year, we divide the commercial dataset into a reporting large subset—RL—and its complement the RS subset. The RL subset contains reporting large firms which is comprised of those firms in the commercial dataset with AUM exceeding the smallest AUM of firms in the II100 and ARBDC lists at the end of the previous year. Similarly, the RS subset of reporting small firms contains all other firms in the commercial data set not in the RL subset. Given the definition of the RL and RS subsets, we create two additional indices of reporting firms using the same methodology as in the construction of NRI-EW and NRI-AW. These indices refer to large firms in the RL subset, RLI-EW which is equally weighted and RLI-AW which is weighted by the firm-wide AUM of constituent firms.²⁰

Before proceeding to analyze the statistical properties of these indices, we take note of some relevant return measurement biases well-known in the hedge fund research literature. For a reporting firm to be included in our expanded data set, it has to be in the commercial data set at the end of the formation year. For a non-reporting firm to be included, it has to be in the II100 or ARBDC surveys for that year. This avoids any backfill bias cause by a fund entering a database with its historical returns prior to the entry date being incorporated *pari passu* to subsequent data collected in real time; see Fung and Hsieh (2009), and Aggarwal and Jorion (2010) for further discussions on this bias. Our method does not suffer from survivorship bias, since we use all known firms and funds at the formation year, without regard to their fate at any future point in

²⁰ Monthly return series of these five portfolios of non-reporting firms can be found in http://faculty.fuqua.duke.edu/~dah7/MegaHF_Data.htm

time. By using only the AUM in the formation year, we avoid any look-ahead bias that involves AUM in future years.

3.2. Comparing the return characteristics of NRI to observable data: 2002 to 2010

Table 4 contains summary statistics for the excess returns (net of the risk-free rate) of the hard to observe non-reporting hedge fund firms, conventional hedge fund indices and reporting hedge fund firms sorted into comparable AUM size-adjusted indices. Table 4, panel A contains descriptive statistics for each index—the annualized mean, standard deviation, skewness, kurtosis, and Sharpe ratio of each index. The Jarque-Bera (1980) test of normality reveal p-values rejecting normality for all return distributions all of the indices. Consequently the tests for differences between non-reporting firms’ returns and the returns from the observable universe of hedge fund, we employ methods that are robust with respect to returns that are not normally distributed.

First, we investigate if the NRI has a different mean than the other indices as follows. Take for example the p-value of 0.9447 corresponding to the entry in the row “pval(equal means)” under the column labeled HFRI in Table 4, panel A. This is obtained by taking the difference of the returns between NRI-EW and HFRI, regressing it on a constant term, calculating the standard error using the Newey-West (1987) method with 6 lags, forming the t-statistic, and using the standard normal distribution to find the p-value. In this case, the equality of means between NRI-EW and HFRI is not rejected at any standard conventional level. A similar conclusion extends to the comparison of means between NRI-AW and the DJCSI which has a p-value of 0.9782—see column “DJCSI” Table 4, panel A. The only marginal exception to this conclusion is the mean return comparison between large non-reporting firms in the NRI-EW and large reporting firms in the RLI-EW. Here, the p-value is 0.0596, which rejects equality at the 10% significance level, but not the 5% level. Taken together, these tests do not reveal a material difference between the mean excess returns of non-reporting firms to reporting firms after controlling of comparable AUM.

Next, we employ the Brown and Forsythe (1974) test for the equality of variance which is robust against departure from normality. In the row labeled “pval(equal variance)”, we test if NRI-EW has the same variance as the other indices. In the column under HFRI, the p-value is 0.1547, which does not reject the equality of variance at standard conventional levels. This is also true when we compare the variance of NRI-EW to RLI-EW, where the p-values is 0.8682. This

conclusion extends to the comparisons of variance between NRI-AW and that of DJCSI, RLI-AW. Taken together the empirical evidence fails to reject the equality of variances between non-reporting and reporting firms adjusting for size.²¹ .

Lastly, Table 4, panel B contains the correlation matrix of the monthly excess returns of these indices. Both these panels indicate that all the indices are highly correlated with each other. Thus far the empirical evidence does not reveal significant return differences between those firms that elected not to participate in commercial databases—non-reporting firms—and those that do—reporting firms. Next we investigate the effect of missing returns from reporting firms.

3.3. Does delisting bias matter?

Section 2.4 identifies the two sources of missing returns from reporting firms—those that are delisted by the databases and those that elect to stop reporting their performance. The open empirical question is whether these two sources of missing returns impact observable data differently? In order to provide insight to this empirical question we digress from our central methodology of analyzing non-reporting firms to focus on missing returns at a more micro level—missing returns of specific share classes of individual hedge funds. From our expanded data set we can identify 9,839 delisted share classes in reporting firms with at least USD50 million AUM. Of these delisted share classes, we identify 1,903 that had 13,561 additional monthly returns (beyond their last return in our commercial data set). We refer to this set of observations as the missing returns from reporting firms. The highest missing monthly return is 84.15%, while the lowest missing monthly return is -100%. The average missing monthly return is 0.30%. Interestingly, 322 of these 1,903 share classes that were delisted from our commercial data are still alive (alive in the sense of reporting NAV to investors) at the end of 2010.

3.3.1 Delisted Funds: Live versus Dead Funds

The Posthuma and van der Slius (2003) and Malkiel and Saha (2005) conjecture is motivated by the potential return difference from funds exiting databases which are in distress

²¹ We also perform tests for the equality of distribution using the procedure in Kendall and Stuart (1967), by comparing the histograms of two random variables. In the row labeled “pval(equal distribution)”, the p-value of 0.6569 under the column labeled HFRI shows that the distribution of NRI-EW is not statistically different from that of HFRI. Similarly, it is not different from that of RLI-EW. In addition, the distribution of NRI-AW is not different from that of RLI-AW and DJCSI.

and on their way to liquidation. The issue being whether the missing, post-delisting returns leading up to complete liquidation are systemically worse than the returns reported to the databases. This conjecture is in stark contrast to Ackerman et al (1999) who noted earlier the possibility of missing returns from successful funds which stop reporting to commercial databases that may well have unobserved returns superior to reporting firms. Here, we can provide additional insight to these two contrasting empirical conjectures. Our expanded data set allows us to identify missing returns from funds which are delisted by database vendors prior to the end of 2010—refer to these as “missing dead funds.” The complementary set of missing returns are from those funds which we have reported returns in our expanded dataset post delisting but not listed as liquidated by database vendors—refer to these as “missing live funds.”

Table 5, panel A reports the summary statistics of these two sets of missing returns—from missing live and dead funds. The total numbers of 13,561 missing returns is comprised of 6,106 from missing live funds and 7,455 from missing dead. Table 5 panel A tabulates estimates of delisting bias for all reporting firms year by year as well as for the entire sample period. These estimates are obtained by following conventional procedure for measuring survivorship bias in the empirical literature. We first calculate the average returns of the firms using only reported returns in our commercial data set, in the column labeled “Database” in Table 5 panel A. Then we recalculate the average returns of these firms after we appended the missing returns of the delisted funds from our expanded data set incrementally. First we append the missing returns from missing dead funds to align the data with the Posthuma and van der Slijs (2003), and Malkiel and Saha (2005) conjecture, in the column labeled “Database + Missing Dead”. Next we append the missing returns from the missing live funds separately to the returns in the “Database” column to align with the Ackerman et al (1999) conjecture which is under the column “Database + Missing Live”. Finally we append all the missing returns (dead or alive) to the returns in the Database column which we label “Database + All Missing”. The column labeled “Delisting bias” is obtained by subtracting the returns in the “Database” column from the “Database + All Missing” column. The overall delisting bias averages at -0.022% per annum and there is an increase in the annualized return standard deviation of 0.051%. Consistent with the economic intuition of the two conjectures, adding back missing dead fund returns does lower the annual mean return by 5.2 basis point on average and adding back missing returns from live funds does increase annual mean return by 1.3bp per on average. But are these changes statistically significant?

To see if the delisting bias is statistically important, we follow the test procedure used in Section 3.2. First, we apply the Jarque-Bera (1980) test of normality. For the returns of the reporting firms that use all available data, the p-value of the JB test statistic is 0.0000, which rejects normality of the return series in all cases. This confirms that we need to employ statistical procedures that are robust to non-normality to compare these return series. Next, we check to see if adding back the missing returns has a significant impact to mean returns from the commercial data set. Accordingly, we test if the mean of the “Data + Missing” series (Dead, Alive and All) is different from the mean of the data in the “Database” series using the same procedure as in Table 4 by regressing their differences on a constant term and estimating the standard error of the differences using the Newey-West (1987) estimator with 6 lags. The p-values of this test of equity of mean are respectively 0.2254 (for the returns from the missing dead) 0.0095 (for the returns from the missing Live) and 0.4330 (for the returns from the combined), which collectively affords the conclusion that the mean return of the funds from the commercial databases is not affected by missing returns from delisted funds at conventional levels of statistical significance. Next we employ the Brown and Forsythe (1974) test for the equality of variance between the “Data in database” series and the respective series with the additional missing returns. The p-values of the test statistic are 0.9985, 0.9982 and 0.9972 respectively which reject the hypothesis that the standard deviation of returns from the Database is different when missing returns are appended. Finally, we employ a test in Kendall and Stuart (1967) to see if the distribution of the “Database” return series is different from those with the missing returns appended. The p-values of this test are uniformly 1.000, which is consistent with the return distribution of the funds from the commercial databases is not affected by delisting.

3.3.2. Delisting bias: the overall picture

Before concluding we extend the Ackerman et al (1999) conjecture on delisting bias to the context of our paper. Here we ask the question “do missing returns from mega hedge fund firms that stopped reporting to databases differ systematically from those mega firms which continued to report performance?” Table 5 panel B tabulates the empirical results relevant to this question. The “Data in Database” column now only refers to those reporting firms in the RL subset of the commercial data set that continued to report data to the end of 2010. The next column under “Database + Missing Large” includes the missing returns for those in the RL subset that stopped reporting to databases during our sample period and became a non-reporting firm. The “Delisting Bias” column is similarly defined as in section 3.3.A. Similar test procedures

show that the inclusion of missing returns from large firms does not yield return characteristics that are statistically different than those in the RL subset. These results taken together tell us that delisting bias does not affect the measurement of the unconditional return distribution of the hedge fund universe.

The empirical results in this section do not reveal significant differences between non-reporting large firms' returns and that of their reporting counter parts over our sample period as individual distributions of monthly returns. Similar conclusions hold for missing returns for funds in the commercial dataset that are managed by firms with greater than USD50 million. However, previous studies such as Fung, Hsieh, Naik and Ramadorai (2008) and Bollen and Whaley (2009) reported empirical evidence on discrete, time-varying risk taking behavior from hedge fund managers. Accordingly we investigate whether the conclusions in section is robust to dynamic risk taking behavior of hedge fund managers.

4. Can non-reporting firms' performance be predicted using observable data?

Given the hurdles involved in gathering non-reporting firms' data and the similarity of the unconditional return distribution of NRIs and their observable counterpart, we turn now to the time-series properties of the NRIs. In particular, we tackle the key challenge of estimating the hard-to-observe NRIs returns from the observable hedge fund universe and market factors. We begin by examining the behavior of the residuals from the regression model in equation (1) below:

$$y_t = a_0 + b_0 x_t + e_t \tag{1}$$

In equation (1), the dependent variable y is the time series of returns from the non-reporting firm indices, NRI-EW and NRI-AW. The regressors are respectively the time series of returns of the comparable reporting firms, RLI-EW and RLI-AW as well as the equally weighted hedge fund index HFRI and AUM-weighted index DJCSI. In total, we have two pairs of regressions, one pair for the equally weighted index of non-reporting firms NRI-EW (versus RLI-EW and HFRI respectively) and one pair for the AUM weighted index of non-reporting firms NRI-AW (versus RLI-AW and DJCSI respectively). Table 6 shows that over the sample period, the empirical results fail to reject a zero intercept term and a slope coefficient which is different from one at the 1%-confidence level. However, in the comparison between NRI-EW and RLI-EW, the intercept is different from zero at the 5%-confidence level. Also, in the comparison of NRI-EW and HFRI,

the slope coefficient is different from one at the 5%-confidence level. These regression results are consistent with the results in Section 3 which compares the unconditional return distribution of non-reporting firms to the observable data. We now turn our attention to the stability of these conclusions. Given the insignificant intercepts and slope estimates from equation (1) which are statistically not different from one, we shall focus on the time-series properties of the four monthly return spreads between non-reporting and observable data. These are respectively, NRI-EW minus RLI-EW, NRI-EW minus HFRI for the equally weighted indices and NRI-AW minus RLI-AW. NRI-AW minus DJCSI for the AUM weighted indices.

Consistent with Asness et al (2001) and Getmansky et al (2004), Table 7 panel A1 reports statistically significant first and second order serial correlation in the monthly returns of all indices used in this study. For equally weighted indices, the first order serial correlations over our sample period (2002-2010) are 0.374 for HFRI, 0.493 for non-reporting mega firms (NRI-EW), and 0.388 for reporting mega firms (RLI-EW). For AUM weighted indices, the first order serial correlations are 0.430 for DJCSI, 0.502 for non-reporting mega firms (NRI-AW), and 0.359 for reporting mega firms (RLI-AW). All of these are statistically different from zero at conventional levels. The more troublesome results are the return serial correlations for the four return spread series of non-reporting firms versus observable data in Table 7 panel A2. Three of the four return spread series have statistically significant first order serial correlation. If the return serial correlation of these spread series is driven by differences in manager return smoothing then inferring non-reporting firm returns from observable reporting firms will be a much more difficult task.

However, the Getmansky et al (2004) model admits two different scenarios which can result in observed return serial correlations. In the first scenario, a fund is exposed to a risk factor that is not serially correlated but the fund returns are serially correlated due to smoothing by the manager. We refer to this type of serial correlation in fund returns as “manager-driven.” In the second scenario, a fund is exposed to a risk factor whose returns are serially correlated but there is no return smoothing by the manager. We refer to this type of serial correlation in hedge fund returns as “factor-driven.” If the observed serial correlation in the spread series is factor-driven then we will be able to examine the conditional return distribution of these spread series with respect to observable risk factor exposures of different manager groups (indices) to adjust for serially correlated returns. Section 4.1 investigates this empirical issue with respect to our indices of non-reporting firms.

4.1. Return serial correlation in NRIs—manager smoothing or risk factor illiquidity?

Asness et al (2001) and Getmansky et al (2004) showed that serial correlation in hedge fund returns can affect the estimates of hedge fund risk exposure in regression models. Getmansky et al (2004) considered key findings for two different scenarios—“manager-driven” as distinct from “factor-driven” return serial correlation. In the first scenario, the regression of fund returns on the risk factor does not give the true exposure and the residuals are serially correlated. Under this scenario, there are two established correction procedures: one, filter out the manager-driven serial correlation, and then regress the “unsmoothed” returns on the risk factor; or two, regress the fund returns on the risk factor and lags of the risk factor, and then sum up the slope coefficients. In the second, factor-driven scenario the regression of the fund returns on the risk factor will give the true risk exposure and the residuals of the regression should be serially uncorrelated. Under this scenario, there is no need to “unsmooth” the fund returns prior to the factor regression.²² In this subsection, we document empirical evidence that for the indices used in this study, the observed serial correlation in monthly returns is mainly factor-driven and therefore it is the second scenario in Getmansky et al (2004) that applies to our analysis of non-reporting firms’ indices.

We begin with the observation that hedge fund strategy indices exhibit different degrees of serial correlation. Table 7 panel B shows that over our sample period (2002-2010), the first order serial correlation of the DJCS Managed Futures index is 0.118, and that for the DJCS Equity Market Neutral index is 0.044. Both are statistically insignificant at conventional levels. In contrast, the serial correlation for the DJCS Convertible Arbitrage, Distressed Securities, and Fixed Income Arbitrage indices are 0.547, 0.55, and 0.564 respectively. All of these coefficients are statistically significant at conventional levels. Managed Futures and Equity Market Neutral are strategies that deploy mainly liquid instruments (e.g. futures and equities) which have low serial correlation in their returns. Convertible Arbitrage, Distressed Securities, and Fixed Income Arbitrage are strategies that employ less liquid instruments and exhibit higher return serial correlations (e.g. corporate bonds and collateralized high yield debt). While the observed return serial correlation can still be both manager-driven as well as factor-driven, the empirical evidence

²² In fact, regressing the unsmoothed fund returns on the (serially correlated) risk factor will bias the true risk exposure of the fund. We provide a simple simulation in Appendix A to highlight the effect of choosing the wrong model.

does suggest that return smoothing, for whatever reason, is more prominent among strategies involving illiquid securities.²³ In order to estimate the relative importance of manager-driven versus factor-driven return serial correlation we appeal to control comparisons of hedge funds and mutual funds managers.

Table 7 panel C provides the median serial correlation of mutual funds in the CRSP Mutual Fund database drawn from various Standard and Poor's strategy category classifications. The first order serial correlation ranges from a low of 0.121 for Equity Market Neutral mutual funds to a high of 0.393 for High Yield Bond mutual funds. It is reasonable to assume that in a highly regulated industry such as mutual funds, manager-driven serial correlation is much less of a concern. Next we compare return serial correlation of the median hedge fund in similar hedge fund strategies per the hedge fund databases' classification scheme. In parentheses are the correlation coefficient between the median hedge fund and their counterpart from the respective mutual fund categories—for example, the correlation between the median convertible arbitrage fund and convertible bond funds is 0.865. The range of return correlation between hedge funds and mutual funds which operate similar, but not identical, strategies has a low of 0.702 for Equity Market Neutral funds and a high of 0.941 for Emerging Market funds. For Long/Short Equity, the median first order serial correlation for hedge funds is 0.176, while it is 0.141 for mutual funds. For Distressed Securities hedge funds, the median serial correlation is 0.384, while it is 0.393 for High Yield bond funds. These comparable levels of return serial correlation apply to other strategies such as Emerging Market and Merger Arbitrage. While this does not rule out return-smoothing by individual hedge fund managers, at an aggregated index level, factor-driven return serial correlation is the more likely cause for the observed return serial correlation in the indices used here. Given these results we can proceed to estimate risk exposures of the NRIs and LRIs in the presence of return serial correlation.

To estimate risk factor exposures of the various indices we apply the model in Edelman, Fung, Hsieh and Naik (2012) which is an extension of the models in Fung, Hsieh, Naik and Ramadorai (2008).²⁴

²³ Similar results obtain for comparable HFRI sub strategy indices.

²⁴ Specifically, the model has two equity factors—the excess return of large cap stocks (proxy by the excess return of the S&P 500 index: SP-Rf) and the return of small cap stocks (proxy by the total return of the Russell 2000 index) minus large cap stocks (RL-SP); they represent key exposures of equity hedge funds (see Fung and Hsieh (2010)) and correspond to the first two factors in the Fama-French three factor model for public equities. There are two bond factors: the excess return of ten-year Treasuries (TY-Rf) and the return of Moody's BAA corporate bonds minus ten-year Treasuries (BAA-TY); they represent key

$$y = a + b_1(SP - R_f) + b_2(RL - SP) + b_3(TY - R_f) + b_4(BAA - TY) + b_5(PTFSBD - R_f) + b_6(PTFSFX - R_f) + b_7(PTFSCOM - R_f) + b_8(IFC - R_f) + u \quad (2)$$

As a consistency check that at the index level observed return serial correlation from hedge funds are primarily factor-driven, Table 7 panel D provides the return serial correlation of the eight risk factors in our regression in our sample period (2002-2010). In particular, the credit spread factor, BAA-TY, has first order serial correlation of 0.423. This is comparable to the serial correlation in Distressed Securities, Convertible Arbitrage, and Fixed Income Arbitrage strategies. Certainly at the portfolio (or index) level, serially correlated returns appear to be primarily factor-driven. Applying equation (2), the factor-model regressions results in Table 8 indicate that all of the indices we constructed for non-reporting firms as well as the publicly available indices such as the HFRI and DJCSI have statistically significant exposures to both emerging market equities (IFC-Rf) and the credit spreads (BAA-TY) factors; the latter factor has the highest return serially correlation of the eight risk factors. The Durbin-Watson (1950, 1951) DW statistics in the last row of Table 8 can now be used as a test for the presence of serial correlation in the regression residuals.²⁵ Overall the preponderance of empirical evidence points to return serial correlation of the hedge fund indices in our sample is primarily factor-driven rather than manager-driven.

4.2. How stable are the return spreads between non-reporting and observable data?

We now return to the risk exposure estimates in Table 8 for the various indices. As hedge funds are dynamic asset allocators, consistent with prevailing empirical evidence on time-varying

exposures of fixed income hedge funds (see Fung and Hsieh (2002)). Three factors represent trend-following styles of managed futures, respectively the excess returns of look-back straddles on bonds (PTFSBD-Rf), currencies (PTFSFX-Rf), and commodities (PTFSCOM-Rf). See Fung and Hsieh (2001) for details on the construction of these straddle returns. Lastly, excess return on emerging market stocks (proxy by the total return of the International Finance Corporation index, IFC-Rf) represents key exposures of emerging market hedge funds (see Fung and Hsieh (2006)).

²⁵ As a robustness check, we perform the factor regression at the firm level. In appendix B, Table B panel A contains the medians of the regression coefficients of the excess returns of non-reporting mega firms, by formation year. Take, for instance, 2001. There were 35 non-reporting mega firms that had 24 monthly returns—12 from the formation year (2001) and 12 from the subsequent year (2002). Their excess returns are then regressed on eight risk factors. The medians of the regression coefficients, adjusted R², and Durbin-Watson (1950, 1951) statistic are reported. None of these regressions resulted in serially correlated residuals—all the D.W. statistics are above the lower bound, d_L. In all the firm regressions for the 10 formation years, only 1 regression resulted in serially correlated residuals. Table B panel B contains the regression results for the reporting mega firms, and Panel C has the regression results for the reporting small firms. There is very little evidence of serial correlation in the residuals of all the factor regressions.

factor betas, we interpret these regressions as estimating the average of changing exposures over the sample period (2002-2010). There are several interesting features in the regression coefficients reported in Table 8. Beginning with the equity-related factors, SP-Rf represents directional exposure to equities. The indices of reporting large firms RLI-EW and NRI-EW have beta estimates 0.0206 and 0.0306 respectively which are not statistically different from zero. RL-SP represents the spread of small firms over large firms. None of the AUM weighted indices, RLI-AW, NRI-AW and DJCSI, has equity betas which are statistically different from zero. Finally, IFC-Rf represents exposure to emerging market equities. Here all indices have similar exposures which are all statistically different from zero at the one percent level. In terms of fixed income risk factor exposures none of the indices have statistically significance betas with respect to the directional factor TY-Rf. In contrast, all indices have statistically significant factor beta with respect to the credit factor, BAA-TY. In terms of exposure to three volatility factors PTFSBD-Rf, PTFSFX-Rf and PTFSKOM-Rf none of the indices have statistically significant beta.

The adjusted R-squares of these regressions have a low of 0.6381 for RLI-AW and a high of 0.8282 for the HFRI indicating reasonable levels of explanatory power of equation (1) over this sample period. Here we interpret a statistically significant factor beta being indicative of persistent but time-varying exposure to that risk factor whose average over this sample period is the beta estimate in equation (2). While the lack of statistically significance beta with respect to a risk factor does not reject dynamic exposures to that risk factor, it is consistent with highly variable exposures to that risk factor over time.²⁶ With respect to the central question of this study—do non-reporting mega firms differ from their reporting counterparts—the last four columns of Table 8 reports the regression results of the four return spreads on the eight risk factors—the spread regressions. First, the DW statistics are well behaved consistent with our conjecture that return series correlation of these spread series are primarily factor-driven. Second, the adjusted R-squares of the spread regressions are substantially lower than the individual index-level regressions. Third, the intercepts of the spread regressions are not different from zero at the 1%-level. For the equally weighted spread series, NRI-EW minus RLI-EW, and NRI-EW minus HFRI, both regressions show statistically significant beta to the credit factor (BAA-TY) and to the emerging market equity risk factor. In addition, the NRI-EW minus HFRI regression show

²⁶ Appendix B reports individual firm-level regressions and report empirical evidence supporting the assertion that risk-factor betas do vary over time. In terms of addressing the overarching research question in this paper of whether non-reporting mega firms are different reporting mega firms, equation (2) remains the simplest way of summarizing the comparability of overall risk characteristics.

statistically significant betas to the equity factors (SP-Rf and RL-SP). For the AUM weighted spread series, NRI-AW minus RLI-AW and NRI-AW minus DJCSI, Table 8 finds that the NRI-AW minus RLI-AW has a significant credit factor beta whereas NRI-AW minus DJCSI has a significant beta with respect to the trend-following factor on commodities. Taken together, these results suggest that there may well be time-varying differences between non-reporting firms and the observable data series driven by different risk factor variations. Next we investigate the impact of significant betas on these return spreads.

4.3. Comparing non-reporting mega firms to reporting large firms in different states of the world

To investigate whether different discrete changes in the managers' risk exposure between non-reporting and reporting firms, we follow the procedure in Fung, Hsieh, Naik and Ramadorai (2008) and apply a CUSUM test to each of the four spread regressions in Table 8 panel .²⁷ Table 9 reports the results of one-step-ahead forecasts of the 8-factor model on four spread regressions. Entries to Table 9 are expressed in standardized residuals. The results in Table 9 tell us about the reliability of the regression relationships reported in Table 6. For example, the column in Table 9 corresponding to the spread series NRI-EW minus RLI-EW tells us that there are only four instances forecast errors exceeding one percent confidence limit.²⁸ The two largest forecast errors occurred in October 07 and October 08 respectively—in the start and the midst of the 2008 financial crisis. From Table 8, under the column NRI-EW minus RLI-EW we can conclude that the likely cause of large forecast errors is a large change in the BAA-TY factor, or the credit market. Similar analysis allows us to assess the reliability of predicting the performance of non-reporting indices based on the respective reporting series (index). Figures 3 A to D depict these tests. The blue line in each of the four figures depicts the CUSUM plots. For the two NRI-EW regressions, Figure 3A, and 3B show violations of the spread regressions tend to be transitory which is consistent with time-varying risk exposures taken by hedge fund managers which may differ between non-reporting and reporting managers from time to time.

Overall, our results show that non-reporting and reporting large hedge fund firms do not behave differently in general albeit transitory differences can occur which can be traced backed to significant changes in specific risk factors. For example, divergences between the NRIs' returns

²⁷ See Ploberger and Kramer (1992) and Brown, Durbin and Evans (1975),

²⁸ There are eleven instances of forecast errors exceeding 5%-confidence limit. Similar results are tabulated in Table 9 for the other spread series.

and that of the comparable reporting large firms are likely to be significant during large moves in the credit market. If the same inference the NRIs returns is to be drawn base on the returns of the commonly used indices HFRI and DJCSI, then significant divergence can also occur during large moves in the equity market. The good news is that in majority of the months in our sample, an index of reporting large firms can be used as an indicator of non-reporting mega firms.

5. Summary, conclusion, and suggestions for future research

The unchecked growth of the hedge fund industry with its rich tapestry of leveraged strategies has captured the imagination of many academic researchers. The arrival of electronically available hedge fund databases provided the critical fuel spurring the equally impressive growth in hedge fund research. Since the early papers on the subject, such as Fung and Hsieh (1997) and Ackerman et al (1999) almost 15 years ago, a nagging issue looms over conclusions of many hours of hard work and pages of scholarly research. Given voluntary reporting of hedge fund managers to these commercial databases, an outstanding issue for researchers is this: “Just how much of the assets invested in the hedge fund industry am I missing in my analysis and what is the difference if they are included?”

While it is virtually impossible to obtain information on the entire universe of hedge funds, in this paper we identify previously unexplored data sources whereby collecting data on fewer than 200 large hedge fund management firms which do not participate in major commercial databases adds to the observable industry AUM by as much as 34% in 2001 rising to 65% by the end of 2010. This represents some USD877 billion of AUM which is missing from the reported USD 1,322 billion of AUM manage by firms in the three major commercial databases combined. To extend the coverage of commercial hedge fund databases, we obtain the names and AUMs of large hedge fund firms that do not participate in commercial databases from surveys published by *Institutional Investor* and *Absolute Return+Alpha* magazines which are good sources of information with almost a decade of continuous history. These previously untapped sources of data provide valuable insight into the capital formation process of the industry over the past decade. While commercial databases have successfully depicted data on the growing trend of hedge fund industry’s AUM, from USD 278 billion in 2001 to USD 1,322 billion in 2010, there is a more important trend in the capital formation process of the industry that has not been documented in the research literature. Over this past decade, the AUM of non-reporting mega hedge fund firms has grown from USD 118 billion (2001) to USD 863 billion (2010). This points

to a much faster growing rate of mega hedge fund companies “opting for privacy” dropping out of the voluntary system of reporting to commercial databases. The results in Section 2 point to a limited group of mega hedge fund firms that manage the bulk of the assets in the industry. Taken together this implies that the assets of the hedge fund industry is concentrated in the hands of a small number of mega management firms with rising opacity as their AUM increases. Given the survey data sources we have identified in this paper going forward, the phenomenon of continuing AUM concentration in the hedge fund industry can be monitored by tracking a small number of mega hedge fund firms.

Collecting performance data of non-reporting firms and integrating them with commercially available data is a challenging task. Accordingly the balance of the paper is focused on the question: can hard-to-observe return characteristics of non-reporting firms be inferred from observable data? We begin the task of answering this question by constructing indices of non-reporting firms using both equally weighted as well as an AUM-weighted methodology so as to facilitate comparison to commonly used indices of hedge funds—HFRI and DJCSI. Section 3 reports empirical evidence on the unconditional distributions of non-reporting firms and their observable counterparts. Conventional tests fail to reject equality of the first two moments of the unconditional distributions between non-reporting and reporting firms after controlling for size. The Kendall and Stuart (1967) tests also failed to detect significant differences comparing between the respective histograms. Before moving on to examine the time-series properties of non-reporting and reporting firms’ returns, we report evidence on a long-standing empirical issue on the missing returns from reporting firms in commercial databases—the delisting bias. By adding back the missing returns of reporting funds with greater than USD 50M AUM, we show that delisting bias is numerically small and statistically insignificant. Specifically, while the Ackerman et al (1999) conjecture that missing returns from successful firms that voluntarily stopped reporting data do exhibit marginally higher mean returns, Table 5 panel B shows that the difference is statistically insignificant. The same observation applies to the conjectures of Posthuma and van der Sliuis (2003) and Malkiel and Saha (2005) which imply that delisted firms heading for liquidation are likely to have inferior returns. While Table 5 panel A reports data consistent with this implication, the differences are not statistically significant. More importantly when these two opposite reasons for firms to stop reporting are combined, the effect of missing returns is neutralized for all practical purposes.

Section 4 returns to the central question of the paper. Here we focus on the return spread series between non-reporting and reporting firms. In order to draw inference on hard-to-observe non-reporting firms' returns based solely on observable data we need to show that return serial correlation on these spread series is not driven by manager-smoothing (which will be manager specific and hard-to-observe) but is market factor-driven. The results in Table 7 and 8 confirm that this is indeed the case and accordingly we are able to estimate the risk factor exposures of these spread series using an extended version of the Fung, Hsieh, Naik and Ramadorai (2008) model applying the correct version of the Getmansky et al (2004) procedure which admits factor-driven serially correlated returns. Section 4.3 tabulates the results of applying the CUSUM test on the return spread series between non-reporting and reporting firms following the same procedure as in Fung, Hsieh, Naik and Ramadorai (2008). Table 9 tabulates the one-step-ahead forecast results on using observable market factors to forecast the respective return spread series. Consistent with the plots in Figure 3, there are only a few instances in which indices of non-reporting firms returns cannot be inferred from reporting firms after controlling for size as well as using commonly used hedge fund indices. Table 9 provides insight on the type of risk factor moves that cause non-reporting firms' return to diverge from their observable counterpart. Overall, there is a high degree of similarity between the return characteristics of non-reporting mega firms and their, observable, reporting counterparts with the largest divergence between them occurred during the 2008 financial crisis feeding through differences in their respective exposure to the credit related factor.

References

Absolute Return+Alpha. Billion dollar club. Various issues.

Ackermann, C., McEnally R., Ravenscraft, D., 1999. The performance of hedge funds: Risk, return, and incentives. *Journal of Finance* 54, 833–74.

Aggarwal, R. K., Jorion, P., 2010. The performance of emerging hedge funds and managers, *Journal of Financial Economics* 96, 238-256.

Aiken, A., Clifford, C.P., Ellis, J., 2012. Out of the dark: hedge fund reporting biases and commercial databases. *Review of Financial Studies*, forthcoming.

Asness, C., Krail, R., Liew, J. 2010. Do hedge funds hedge? *Journal of Portfolio Management* 28, 6-19.

Berk, J. B., Green R., 2004. Mutual fund flows and performance in rational markets. *Journal of Political Economy* 112, 1269–95.

Bollen N., R. E. Whaley, 2009, Hedge Fund Risk Dynamics: Implications for Performance Appraisal, *Journal of Finance*, 64, 987-1037.

Brown, R. L., J. Durbin, J. M. Evans, 1975, Techniques for Testing the Constancy of Regression Relationships over Time, 1975, *Journal of the Royal Statistical Society, Series B* 37, 2, 149-192.

Brown, M., Forsythe, A., 1974. Robust tests for the equality of variances. *Journal of the American Statistical Association* 69, 364-367.

Durbin, J., G. Watson, 1950. Testing for serial correlation in least squares regression, I. *Biometrika* 37, 409-428.

Durbin, J., G. Watson, 1951. Testing for serial correlation in least squares regression, II. *Biometrika* 38, 159-179.

Edelman, D., Fung, W., Hsieh, D.A., Naik, N., 2012. Funds of hedge funds: performance, risk and capital formation: 2005 to 2010. *Financial Markets and Portfolio Managements* 26, 87-108.

Eichengreen, B., Mathieson, D., Chadha, B., Jensen, A., Kodres, L., Sharma, S., 1998. Hedge fund and financial market dynamics. Occasional Paper 166. Washington: International Monetary Fund.

- Fung, W., Hsieh, D. A., 1997. Empirical characteristics of dynamic trading strategies: The case of hedge funds. *Review of Financial Studies* 10, 275–302.
- Fung, W., Hsieh, D.A., 2000. Measuring the market impact of hedge funds. *Journal of Empirical Finance* 7, 1-36.
- Fung, W., Hsieh, D. A., 2001. The risk in hedge fund strategies: theory and evidence from trend followers. *Review of Financial Studies* 14, 313-341.
- Fung, W., Hsieh, D. A., 2002. The risk in fixed-income hedge fund styles. *Journal of Fixed Income* 12, 6-27.
- Fung, W., Hsieh, D. A., 2006. Hedge funds: an industry in its adolescence. *Federal Reserve Bank of Atlanta Economic Review* 91, 1-33.
- Fung, W., Hsieh, D. A., 2009. Measurement biases in hedge fund performance data: an update. *Financial Analyst Journal* 65, 36-38.
- Fung, W., Hsieh, D.A., 2010. The risk in hedge fund strategies: theory and evidence from long/short equity hedge funds. *Journal of Empirical Finance* 18, 547-569.
- Fung, W., Hsieh, D.A., Tsatsaronis, K., 2000. Do hedge funds disrupt emerging markets? *Brookings-Wharton Papers on Financial Services*, 377-421.
- Fung, W., Hsieh D. A., Naik N., Ramadorai, T., 2008. Hedge funds: performance, risk and capital formation. *Journal of Finance* 63, 1777-1803.
- Getmansky, M., Lo, A., and Makarov, I., 2004. An econometric mode of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics* 74, 529-609.
- HFN / HedgeFund.net. HFN administrator survey results. Channel Capital Group, Inc. Various issues.
- Hedge Fund Research. HFR Global Hedge Fund Industry Report. Various issues.
- Institutional Investor. Hedge Fund 100. Various issues.
- Jarque, B., A. Bera, 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economic Letters* 6, 255-259.

Kendall, M., Stuart, A. 1967. *The Advance Theory of Statistics: Volume 2*. New York: Hafner Press.

Malkiel, B., Saha, A., 2005. Hedge funds: risk and return. *Financial Analyst Journal* 61, 80-88.

National Association of Colleges and University Business Officers. Endowment study. Various years.

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

Pension and Investments. P&I 1000. Various years.

Ploberger, W., W. Kramer, 1992. The CUSUM Test With OLS Residuals, *Econometrica* 60, 2, 271-285.

Posthuma, N., Van der Sluis, P. J., 2003. A reality check on hedge fund returns. Unpublished Working Paper. Available at <<http://ssrn.com/abstract=438840>> .

Table 1 AUM by US Institutional Investors and Their Allocation to Hedge Funds

Year	US University Endowments		200 Largest US Defined Benefit Pension Plans	
	Assets* (\$b)	% invested in hedge funds	Assets** (\$b)	% invested in hedge funds
1998	178.3	4.7%		
1999	195.4	5.1%		
2000	241.1	4.7%		
2001	236.2	6.1%	3,561.0	0.1%
2002	222.5	11.3%	3,172.5	0.3%
2003	230.5	13.5%	3,653.7	0.4%
2004	267.0	14.7%	4,048.1	0.5%
2005	298.9	16.6%	4,522.4	0.7%
2006	341.2	18.0%	4,911.0	1.0%
2007	411.2	18.2%	5,597.5	1.4%
2008	412.8	21.0%	3,688.6	2.2%
2009	306.0	21.9%	3,475.4	2.0%
2010	346.0	21.3%	3,693.4	3.0%

* June of each year.

** September of each year.

Sources: NACUBO, *Pension & Investments*.

Table 2. Statistics of the Firm in Commercial Databases, II100, and ARBDC

At the end of each year, three commercial databases (BarclayHedge, HFR, Lipper TASS) are merged. The firms in the merged database are in Panel A. Firms in the *Institutional Investor* “Hedge Fund 100” list are in Panel B. Non-reporting firms are outside the three commercial databases. Firms in the *Absolute Return+Alpha* “Billion Dollar Club” list are in Panel C. Non-reporting firms are those outside the three commercial databases and the *Institutional Investor* “Hedge Fund 100” list.

A. Firms in Three Commercial Databases (BarclayHedge, HFR, Lipper TASS)

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
No. of Firms	1,026	1,132	1,269	1,421	1,650	1,761	2,049	2,140	2,185	2,443
Total AUM (\$b)	278	341	502	679	874	1,129	1,522	1,027	1,061	1,322
Largest (\$m)	7,794	10,000	11,500	11,500	21,023	32,531	36,000	38,600	34,600	58,900
Median (\$m)	46	46	53	65	67	83	95	58	54	58
Smallest (\$m)	0.001	0.014	0.007	0.015	0.001	0.025	0.016	0.016	0.019	0.001

B. Firms in II100

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
No. of firms	100	100	100	100	100	100	100	100	100	100
Non-reporting	37	41	39	42	41	48	52	52	51	54
Total AUM (\$b)	264	344	439	556	720	1,004	1,353	1,034	1,082	1,231
Non-reporting	103	147	169	244	317	453	688	530	547	592
Largest (\$m)	8,000	10,000	11,500	12,500	21,023	32,531	36,000	38,600	45,100	58,900
Median (\$m)	1,950	2,619	3,600	4,685	5,829	8,141	11,500	8,000	8,050	8,750
Smallest (\$m)	645	1,400	2,009	2,900	3,550	4,874	6,252	4,000	4,000	4,691

C. Firms in ARBDC

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
No. of firms	86	100	150	176	204	243	262	220	214	330
Non-reporting	7	16	39	50	66	75	89	73	66	110
Total AUM (\$b)	268	310	465	645	851	1,233	1,605	1,167	1,182	1,701
Non-reporting	16	27	64	92	146	205	291	180	159	270
Largest (\$m)	8,000	10,500	12,100	12,500	22,000	34,000	36,000	38,600	43,600	58,900
Median (\$m)	2,600	2,300	2,300	2,650	2,850	3,000	3,905	3,200	3,070	2,600
Smallest (\$m)	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Note: Non-reporting II100 firms are outside of the three commercial databases. Non-reporting ARBDC firms are those outside of the three commercial databases and II100. ARBDC firms in 2001 are estimated from the ARBDC 2002 survey.

Table 3 AUM in Hedge Fund Firms (billion USD) by Year

Reporting firms are the firms in three commercial databases (BarclayHedge, HFR, Lipper TASS). Non-reporting firms are those firms in the *Institutional Investor* “Hedge Fund 100” list and the *Absolute Return+Alpha* “Billion Dollar Club” list that do not participate in the three commercial databases. At the end of each formation year, the assets under management (AUM) of these two groups of firms are aggregated. The *HFN Hedge Fund Administrator Survey* of assets under administration (AUA) is in the last row of Panel A.

A. Firm AUM

	Formation Year									
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Reporting Firms	278	341	502	679	874	1,129	1,522	1,027	1,061	1,322
Non-reporting Firms	118	174	233	336	468	657	988	709	706	863
In II100	103	147	169	244	317	453	688	530	547	592
In ARBDC but not in II100	16	27	64	92	146	205	291	180	159	270
All Firms	396	515	734	1,015	1,342	1,787	2,509	1,737	1,767	2,185
HFN Administrator Survey AUA	NA	NA	845	1,293	1,531	2,153	2,861	1,932	2,172	2,826

B. Number of Firms

	Formation Year									
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Reporting firms	1,026	1,132	1,269	1,421	1,650	1,761	2,049	2,140	2,185	2,443
Non-reporting firms	44	57	78	92	107	123	141	125	117	164
In II100	37	41	39	42	41	48	52	52	51	54
In ARBDC but not in II100	7	16	39	50	66	75	89	73	66	110
All Firms	1,070	1,189	1,347	1,513	1,757	1,884	2,190	2,265	2,302	2,607

Sources: BarclayHedge, HFR, Lipper TASS, *Institutional Investor*, *Absolute Return+Alpha*, Hedgefund.net
Notes: N.A. for years in which survey data is not available. ARBDC firms in 2001 are estimated from the ARBDC 2002 survey.

Table 4 Characteristics of Excess Returns for Portfolios of Non-reporting Firms and Hedge Fund Indices

Reporting firms are the firms in three commercial databases (BarclayHedge, HFR, Lipper TASS). RLI and RSI are reporting mega (small) firms that have assets under management more than (less than) the smallest firm in the II100 list and the ARBDC list. NRI are non-reporting mega firms which are the firms in the II100 and ARBDC lists that are outside the three commercial databases. For each firm, an equally-weighted portfolio is formed for the funds in operation at the end of the formation year and held for the subsequent year. If a fund's return stops, it is assumed to be liquidated and the proceeds are invested in treasury bills during the rest of the year. If a firm launches a new fund, it is included in the portfolio in the following year. EW represents equally weighted portfolios, and AW represents AUM weighted portfolios. DJCSI is the Dow-Jones Credit Suisse Broad Hedge Fund Index, and HFRI is the HFR Fund Weighted Index. Panel A provides annualized means, standard deviation, skewness, kurtosis, and Sharpe ratios of the excess returns of these portfolios. Panel B contains the correlation matrix for these portfolio excess returns. Panel C contains the annual excess returns for each year.

A. Descriptive Statistics	RLI-EW	NRI-EW	HFRI	RLI-AW	NRI-AW	DJCSI
Mean	3.76%	4.95%	5.01%	3.96%	5.28%	5.27%
Standard Deviation	5.88%	5.99%	6.55%	5.34%	6.27%	5.90%
Skewness	-1.789	-2.099	-1.027	-1.696	-2.226	-1.515
Kurtosis	8.794	10.580	5.608	8.505	11.404	7.635
Sharpe Ratio	0.640	0.826	0.766	0.742	0.843	0.892
JB test of normality, pval	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Tests of equality	NRI-AW versus Others					
pval(equal mean)	0.0596		0.9447	0.1876		0.9782
pval(equal variance)	0.8682		0.1547	0.4854		0.9461
pval(equal distribution)	0.9871		0.6173	0.9941		0.8106

B. Correlation Matrix	RLI-EW	NRI-EW	HFRI	RLI-AW	NRI-AW	DJCSI
RLI-EW	1					
NRI-EW	0.964	1				
HFRI	0.967	0.928	1			
RLI-AW	0.992	0.948	0.954	1		
NRI-AW	0.966	0.988	0.920	0.950	1	
DJCSI	0.972	0.951	0.945	0.971	0.956	1

C. Annual Excess Returns/Year	RLI-EW	NRI-EW	HFRI	RLI-AW	NRI-AW	DJCSI
2002	-0.45%	0.42%	-3.04%	0.20%	2.64%	1.39%
2003	13.42%	10.73%	18.37%	12.25%	10.74%	14.32%
2004	6.03%	8.80%	7.57%	6.09%	8.72%	8.17%
2005	4.72%	7.13%	5.96%	4.05%	8.50%	4.31%
2006	8.00%	7.91%	7.74%	6.99%	8.55%	8.67%
2007	5.04%	9.69%	4.68%	4.63%	8.43%	7.18%
2008	-22.19%	-23.80%	-20.84%	-18.28%	-25.15%	-20.87%
2009	14.53%	20.15%	19.71%	13.19%	19.72%	18.31%
2010	8.92%	9.24%	10.08%	9.65%	11.65%	10.79%

Table 5 Estimates of Delisting Bias in Reporting Firms

Reporting firms are the firms in three commercial databases (BarclayHedge, HFR, Lipper TASS). Reporting large firms have assets under management more than (less than) the smallest firm in the *Institutional Investor* “Hedge Fund 100” list and the *Absolute Return+Alpha* “Billion Dollar Club” list. In a given year, two sets of returns are calculated. First, we calculate the annual return of the reporting firms using only data in three commercial databases (BarclayHedge, HFR, and Lipper TASS). Second, we calculate the annual return of the reporting firms after augmenting the returns of delisted funds with additional data beyond the delisting date from private sources, whenever possible. We then subtract the first annual return from the second annual return to obtain the delisting bias for that year. Pval(t-stat) is the p-value of the t-test for the null hypothesis that the delisting bias has a zero mean.

A. Missing Return from the Dead and the LivingB. Reporting Large Firms (RLI)

Year>Returns from	Database	Database + Missing Dead	Database + Missing Live	Database + All Missing	Delisting Bias	Database	Database + Missing Large	Delisting Bias
2002	-1.495%	-1.510%	-1.470%	-1.486%	0.009%	1.048%	1.178%	0.130%
2003	18.886%	18.862%	18.973%	18.948%	0.062%	13.750%	14.582%	0.833%
2004	7.981%	7.960%	7.940%	7.920%	-0.061%	7.522%	7.477%	-0.046%
2005	8.407%	8.370%	8.407%	8.370%	-0.037%	8.021%	8.019%	-0.002%
2006	11.395%	11.360%	11.420%	11.384%	-0.011%	13.053%	13.173%	0.121%
2007	9.472%	9.459%	9.491%	9.477%	0.004%	10.313%	10.344%	0.031%
2008	-18.985%	-19.097%	-19.034%	-19.155%	-0.169%	-20.182%	-20.427%	-0.244%
2009	18.341%	18.304%	18.380%	18.343%	0.002%	14.636%	14.819%	0.183%
2010	9.668%	9.669%	9.671%	9.672%	0.004%	8.924%	9.038%	0.114%
Mean	7.074%	7.042%	7.087%	7.053%	-0.022%	6.343%	6.467%	0.124%
Std Dev	10.815%	10.837%	10.842%	10.866%	0.051%	10.148%	10.311%	0.278%
Skewness	-0.9636	-0.9675	-0.9637	-0.9661		-0.9637	-0.9675	
Kurtosis	5.1227	5.1432	5.1286	5.1317		5.1286	5.1432	
p-value(normality)	0.0000	0.0000	0.0000	0.0000		0.0000	0.0000	
p-value(equal mean)		0.2254	0.0095	0.4330			0.1033	
p-value(equal var)		0.9985	0.9982	0.9972			0.9494	
p-value(equal dist)		1.0000	1.0000	1.0000			1.0000	

Table 6 Regression Analysis of Non-reporting Firms Returns

This tables reports the results of applying equation (1), $y_t = a_0 + b_0x_t + e_t$, on two dependent variable on two pairs of independent variables. Panel A reports the results in the regressions NRI-EW on RLI-EW and NRI-EW on HFRI. Panel B reports the results in the regressions NRI-AW on RLI-AW and NRI-AW on DJCSI. Standard errors are in italics and t-statistics are in bold face. Pval(t-stat) is the p-value of the t-test for the null hypothesis of a zero intercept and separately for a slope of one.

	Panel A		Panel B	
$y_t =$	NRI-EW	NRI-EW	NRI-AW	NRI-AW
$x_t =$	RLI-EW	HFRI	RLI-AW	DJCSI
Constant	0.0010 <i>0.0006</i> 1.7745	0.0006 <i>0.0010</i> 0.5676	0.0007 <i>0.0009</i> 0.8206	-0.0001 <i>0.0007</i> -0.0793
Slope	0.9821 <i>0.0506</i> 19.4028	0.8517 <i>0.0847</i> 10.0588	1.1166 <i>0.0838</i> 13.3184	1.0161 <i>0.0656</i> 15.4808
Adj R ²	0.9292	0.8641	0.9041	0.9150
pval(constant=0)	0.0380	0.2851	0.2059	0.5316
pval(slope=1)	0.3621	0.0399	0.9179	0.5966

Table 7 Serial Correlation Analysis of Hedge Fund Returns: 2002-2010

First, second, and third order serial correlations are computed for hedge funds and mutual funds. HFRI is the HFR Fund Weighted Index. DJCSI is the Dow Jones/Credit Suisse Broad Index. Reporting firms are the firms in three commercial databases (BarclayHedge, HFR, Lipper TASS). RLI (RSI) are reporting mega (small) firms that have assets under management more than (less than) the smallest firm in the II100 list and the ARBDC list. NRI are non-reporting firms that are the firms in the II100 and ARBDC lists that do not report to the three commercial databases. Strategy indices are hedge fund indices based on peer groupings of hedge funds that employ similar strategies. The eight risk factors are: equity factors—the excess total return of the S&P 500 index (SP-Rf) and the return of the Russell 2000 index minus the return of the S&P 500 index (RL-SP); bond factors—the excess return of ten-year Treasuries (TY-Rf) and the return of Moody’s BAA corporate bonds minus ten-year Treasuries (BAA-TY); three factors representing trend-following styles of managed futures—the excess returns of look-back straddles on bonds (PTFSBD-Rf), currencies (PTFSFX-Rf), and commodities (PTFSCOM-Rf); an emerging market factor—the excess return from the International Finance Corporation’s index of emerging stocks (IFC-Rf).

Serial Correlation	$\rho(1)$	$\rho(2)$	$\rho(3)$		$\rho(1)$	$\rho(2)$	$\rho(3)$
A. Broad based Indices	A1 Individual Return Series				A2 Return Differences from:		
HFRI	0.374	0.232	0.159	NRI-EW - RLI_EW			
NRI-EW	0.493	0.313	0.146		0.187	-0.001	0.180
RLI-EW	0.388	0.239	0.153	HFRI	0.151	-0.007	0.067
DJCSI	0.430	0.243	0.141	NRI-AW - RLI_AW			
NRI-AW	0.502	0.314	0.155		0.415	0.239	0.197
RLI-AW	0.359	0.185	0.116	DJCSI	0.067	-0.015	0.059
B. Strategy Indices	HFRI Strategy Indices				DJCS Strategy Indices		
Equity Market Neutral	0.148	0.176	0.102		0.044	0.019	0.157
Managed Futures	0.028	-0.128	-0.113		0.118	-0.206	-0.239
Global Macro	0.003	-0.172	0.006		0.233	-0.021	0.028
Long/Short Equity	0.366	0.230	0.165		0.312	0.153	0.054
Emerging Market	0.420	0.307	0.197		0.366	0.194	0.169
Convertible Arbitrage	0.595	0.293	0.188		0.547	0.249	0.149
Fixed Income Arbitrage	0.609	0.349	0.212		0.564	0.225	0.129
Distressed Securities	0.610	0.411	0.337		0.550	0.402	0.341
C. Median Fund from Strategy Sub Sets :	Hedge Fund Database Classification				Lipper Mutual Fund Classification		
Equity Market Neutral (0.702)	0.258	0.168	0.127		0.121	-0.088	0.074
Long/Short Equity (0.925)	0.176	0.052	0.093		0.141	-0.021	0.091
Emerging Market (0.941)	0.258	0.168	0.127		0.244	0.118	0.119
Convertible Arbitrage/ Convertible Bonds (0.865)	0.482	0.146	0.091		0.321	-0.002	0.104
Distressed Securities/High Yield* (0.777)	0.384	0.260	0.216		0.393	-0.018	0.141
Merger Arbitrage/Merger Fund** (0.725)	0.116	0.030	0.185		0.137	0.049	0.102
D. Eight Risk Factors							
SP-Rf	0.205	-0.001	0.152				
RL-SP	-0.154	0.041	-0.003				
TY-Rf	0.000	-0.232	0.110				
BAA-TY	0.423	0.046	-0.006				
PTFSBD-Rf	0.048	0.004	0.158				
PTFSFX-Rf	0.080	0.074	-0.190				
PTFSCOM-Rf	0.028	-0.018	-0.021				
IFC-Rf	0.167	0.193	0.135				

Table 8 Regression of Excess Returns of Hedge Fund Portfolios on Eight Risk Factors: 2002-2010

Excess returns of hedge fund portfolios above Tbills are regressed on eight risk factors. The NRI-EW (NRI-AW) portfolio is an equally weighted (AUM weighted) portfolio of non-reporting firms in the II100 and the ARBDC. The RLI-EW and RLI-AW are equally weighted and AUM weighted portfolios of firms in the commercial databases with AUM more than the smallest firm in the II100 and ARBDC lists. The DJCSI is the Dow-Jones Credit Suisse Broad Hedge Fund Index, and HFRI is the HFR Fund Weighted Index. The eight risk factors are: equity factors—the excess total return of the S&P 500 index (SP-Rf) and the return of the Russell 2000 index minus the return of the S&P 500 index (RL-SP); bond factors—the excess return of ten-year Treasuries (TY-Rf) and the return of Moody’s BAA corporate bonds minus ten-year Treasuries (BAA-TY); three factors representing trend-following styles of managed futures—the excess returns of look-back straddles on bonds (PTFSBD-Rf), currencies (PTFSFX-Rf), and commodities (PTFSCOM-Rf); an emerging market factor—the excess return from the International Finance Corporation’s index of emerging stocks (IFC-Rf). Standard errors are in italics, based on Newey-West with 6 lags, t-statistics and chi-square statistics are in bold.

	RLI-EW	NRI-EW	HFRI	RLI-AW	NRI-AW	DJCSI	NRI-EW- RLI-EW	NRI-EW- HFRI	NRI-AW- RLI-AW	NRI-AW- DJCSI
Constant	-0.0002 <i>0.0010</i> -0.2046	0.0009 <i>0.0012</i> 0.7350	0.0007 <i>0.0008</i> 0.8972	0.0003 <i>0.0012</i> 0.2644	0.0007 <i>0.0014</i> 0.4797	0.0009 <i>0.0013</i> 0.7263	0.0011 <i>0.0005</i> 2.1629	0.0002 <i>0.0008</i> 0.2284	0.0004 <i>0.0008</i> 0.5172	-0.0002 <i>0.0005</i> -0.5276
SP-Rf	0.0206 <i>0.0348</i> 0.5920	0.0306 <i>0.0331</i> 0.9241	0.0732 <i>0.0269</i> 2.7263	0.0036 <i>0.0376</i> 0.0946	0.0060 <i>0.0380</i> 0.1575	0.0246 <i>0.0375</i> 0.6574	0.0100 <i>0.0144</i> 0.6925	-0.0427 <i>0.0201</i> -2.1182	-0.0134 <i>0.0214</i> -0.6255	-0.0187 <i>0.0207</i> -0.9020
RL-SP	0.0129 <i>0.0465</i> 0.2779	-0.0121 <i>0.0509</i> -0.2375	0.0567 <i>0.0433</i> 1.3103	0.0049 <i>0.0473</i> 0.1032	-0.0036 <i>0.0500</i> -0.0715	0.0100 <i>0.0443</i> 0.2258	-0.0250 <i>0.0112</i> -2.2307	-0.0688 <i>0.0150</i> -4.5771	-0.0147 <i>0.0105</i> -1.3997	-0.0136 <i>0.0141</i> -0.9619
TY-Rf	-0.0188 <i>0.0442</i> -0.4247	-0.0268 <i>0.0593</i> -0.4517	-0.0481 <i>0.0411</i> -1.1694	-0.0073 <i>0.0505</i> -0.1451	-0.0094 <i>0.0621</i> -0.1515	-0.0128 <i>0.0572</i> -0.2243	-0.0080 <i>0.0231</i> -0.3470	0.0213 <i>0.0239</i> 0.8919	-0.0086 <i>0.0302</i> -0.2855	0.0034 <i>0.0149</i> 0.2292
BAA-TY	0.1830 <i>0.0548</i> 3.3431	0.2330 <i>0.0564</i> 4.1279	0.1428 <i>0.0319</i> 4.4702	0.1904 <i>0.0592</i> 3.2157	0.2273 <i>0.0626</i> 3.6317	0.1967 <i>0.0545</i> 3.6085	0.0499 <i>0.0214</i> 2.3331	0.0902 <i>0.0331</i> 2.7277	0.0585 <i>0.0219</i> 2.6689	0.0306 <i>0.0192</i> 1.5889
PTFSBD - Rf	-0.0065 <i>0.0067</i> -0.9691	-0.0129 <i>0.0103</i> -1.2557	-0.0023 <i>0.0054</i> -0.4232	-0.0093 <i>0.0081</i> -1.1472	-0.0143 <i>0.0112</i> -1.2747	-0.0130 <i>0.0097</i> -1.3420	-0.0064 <i>0.0049</i> -1.3161	-0.0106 <i>0.0058</i> -1.8271	-0.0082 <i>0.0073</i> -1.1191	-0.0013 <i>0.0030</i> -0.4412
PTFSFX - Rf	0.0038 <i>0.0034</i> 1.1271	0.0045 <i>0.0044</i> 1.0254	0.0048 <i>0.0031</i> 1.5560	0.0041 <i>0.0036</i> 1.1648	0.0056 <i>0.0035</i> 1.5832	0.0074 <i>0.0043</i> 1.7085	0.0007 <i>0.0035</i> 0.1887	-0.0004 <i>0.0031</i> -0.1208	0.0015 <i>0.0026</i> 0.5664	-0.0018 <i>0.0022</i> -0.8405
PTFSCOM - Rf	-0.0014 <i>0.0069</i> -0.2067	-0.0042 <i>0.0071</i> -0.5907	0.0007 <i>0.0057</i> 0.1238	-0.0037 <i>0.0074</i> -0.5030	-0.0082 <i>0.0079</i> -1.0330	0.0018 <i>0.0073</i> 0.2420	-0.0027 <i>0.0030</i> -0.9017	-0.0049 <i>0.0029</i> -1.7108	-0.0075 <i>0.0034</i> -2.1920	-0.0099 <i>0.0036</i> -2.7378
IFC-Rf	0.1550 <i>0.0214</i> 7.2519	0.1274 <i>0.0220</i> 5.7849	0.1714 <i>0.0151</i> 11.3771	0.1434 <i>0.0233</i> 6.1596	0.1515 <i>0.0216</i> 7.0025	0.1390 <i>0.0193</i> 7.2090	-0.0276 <i>0.0082</i> -3.3653	-0.0440 <i>0.0134</i> -3.2859	0.0156 <i>0.0096</i> 1.6140	0.0125 <i>0.0105</i> 1.1915
Adj R ²	0.7137	0.6739	0.8282	0.6381	0.6536	0.6482	0.0912	0.4157	0.1440	0.0816
DW	1.7995	1.9366	1.9387	1.7694	1.8245	1.6904	1.6357	1.6076	1.7345	2.0548

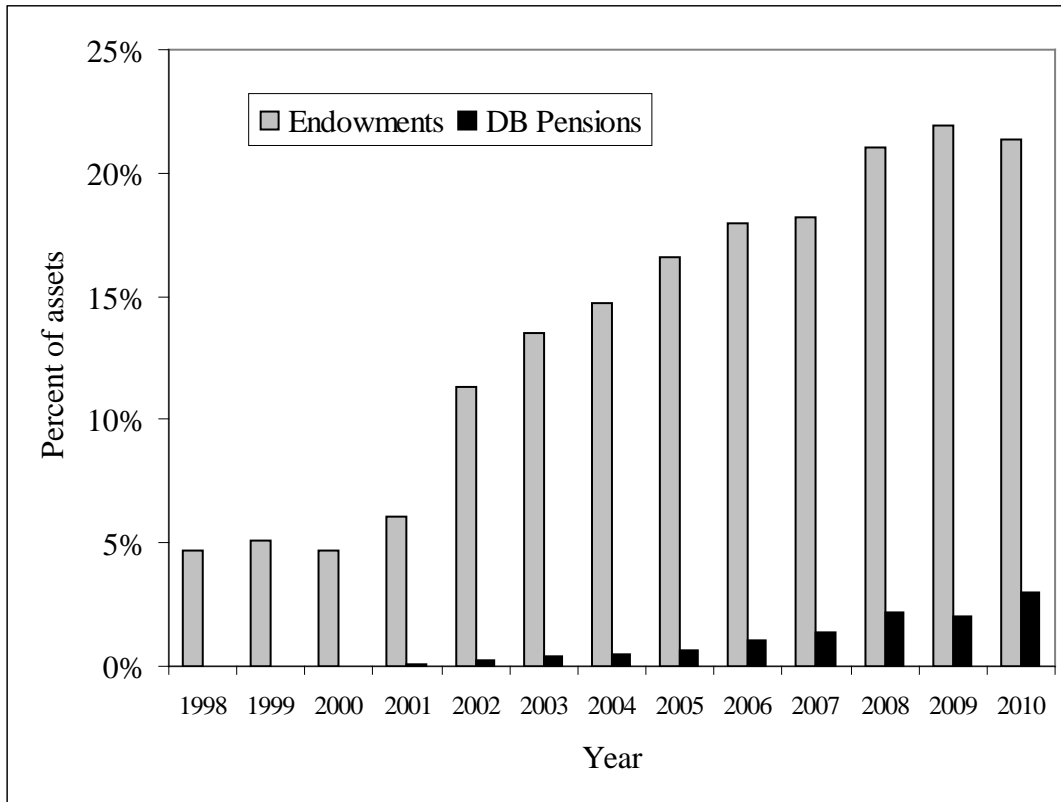
Table 9 One-Step-Ahead Forecast Errors using Standardized Residuals

The column NRI-EW minus RLI-EW contains the standardized one-step-ahead forecast errors using the 8 factor model. For example a value of 1.41 (-1.14) represents a forecast error 1.41 standard errors above zero (below zero). The CUSUM graphs are generated by cumulating these standardized forecast errors over time. Data prior to Jan-04 are used to generate the first prediction. The rows “# Obs > 5% (1%)” represent respective the number of months CUSUM exceeded the 5% (1%) confidence level that the regression relationship is stable.

	NRI-EW- RLI-EW	NRI-EW- RLI-EW	NRI- AW- RLI-AW	NRI- AW- DJCSI		NRI-EW- RLI-EW	NRI-EW- RLI-EW	NRI- AW- RLI-AW	NRI- AW- DJCSI
Jan-04	1.01	-0.59	-0.11	-0.45	Aug-07	-0.58	0.43	-1.13	-1.05
Feb-04	1.00	1.18	0.57	-0.13	Sep-07	0.04	-0.50	-0.40	-1.30
Mar-04	0.57	1.03	0.08	1.20	Oct-07	2.54	0.65	0.06	-0.93
Apr-04	0.50	0.73	0.34	-0.07	Nov-07	-0.31	1.91	1.86	1.52
May-04	-0.12	0.01	1.06	0.42	Dec-07	-2.42	-0.12	0.89	0.92
Jun-04	-0.86	-0.38	0.44	0.34	Jan-08	-1.04	-4.06	-2.72	-4.27
Jul-04	-0.98	-0.93	-1.21	-1.64	Feb-08	0.22	-1.52	-1.80	-1.09
Aug-04	0.03	-0.50	-0.41	-0.33	Mar-08	1.56	-0.38	-1.05	-0.09
Sep-04	1.70	-0.56	0.08	-0.48	Apr-08	-0.25	2.13	2.27	3.87
Oct-04	1.40	1.76	1.09	0.42	May-08	-0.24	0.10	0.45	0.02
Nov-04	0.04	2.56	1.13	1.17	Jun-08	0.59	-1.17	-2.36	-3.16
Dec-04	1.22	0.17	0.11	0.54	Jul-08	1.22	0.30	0.55	0.82
Jan-05	-0.39	0.06	1.30	2.26	Aug-08	-2.05	1.68	0.56	1.60
Feb-05	0.97	-0.06	-0.52	-0.38	Sep-08	0.22	-3.68	-4.06	-2.05
Mar-05	-0.03	0.29	1.80	0.23	Oct-08	-3.90	-1.37	0.14	-0.68
Apr-05	0.70	-0.76	-0.53	-0.65	Nov-08	-1.39	-3.51	-5.25	0.65
May-05	0.33	1.08	1.31	3.96	Dec-08	1.35	-2.31	-3.30	-3.54
Jun-05	0.73	0.74	1.48	1.16	Jan-09	-0.54	1.05	-0.25	0.74
Jul-05	0.41	0.69	0.78	-0.04	Feb-09	0.02	-0.79	-1.13	0.16
Aug-05	-0.55	-0.02	1.16	-0.01	Mar-09	0.39	-0.17	0.73	-0.07
Sep-05	-0.04	-0.14	-1.50	-0.70	Apr-09	1.75	-1.73	0.82	0.30
Oct-05	0.32	-0.25	-0.27	1.52	May-09	1.48	-1.37	0.79	-0.21
Nov-05	0.21	1.42	0.33	0.98	Jun-09	0.23	0.75	0.68	1.12
Dec-05	-0.99	-0.50	0.53	0.08	Jul-09	0.39	0.14	0.00	-0.95
Jan-06	-0.07	0.05	0.39	-1.64	Aug-09	0.30	0.78	0.91	0.30
Feb-06	-0.38	-0.05	-0.28	0.72	Sep-09	0.19	-0.19	-0.25	-1.66
Mar-06	0.04	-0.22	-0.03	-0.64	Oct-09	0.34	-0.72	0.05	-0.35
Apr-06	0.62	0.05	1.09	-0.46	Nov-09	-0.53	0.35	-0.44	-1.64
May-06	0.42	-0.74	-1.13	-0.55	Dec-09	1.19	-0.37	-1.27	-1.37
Jun-06	0.01	0.79	-0.37	0.23	Jan-10	0.83	0.90	1.04	0.27
Jul-06	-0.08	0.13	-0.70	-0.55	Feb-10	-2.15	0.67	0.19	0.42
Aug-06	-1.69	0.52	1.22	1.29	Mar-10	-0.52	-0.35	-1.73	-0.22
Sep-06	-1.57	-0.42	-2.00	-1.04	Apr-10	2.85	-0.23	-0.47	-0.46
Oct-06	0.69	-0.01	-0.53	-0.24	May-10	-0.19	1.73	4.10	3.73
Nov-06	-0.80	0.12	1.07	0.30	Jun-10	-1.22	-0.88	0.43	0.34
Dec-06	1.38	0.69	-0.05	0.08	Jul-10	0.77	-0.15	-0.71	-0.18
Jan-07	0.85	1.95	1.96	1.82	Aug-10	-0.94	0.57	0.94	0.35
Feb-07	1.23	0.28	0.25	-0.50	Sep-10	-1.07	-0.40	-0.58	-1.23
Mar-07	-0.08	1.29	0.59	0.07	Oct-10	0.73	-0.66	-0.77	-0.22
Apr-07	1.49	0.45	0.29	0.06	Nov-10	-2.15	0.50	1.06	1.00
May-07	-0.25	2.42	2.14	1.31	Dec-10	0.00	-1.22	-1.56	-2.04
Jun-07	1.11	0.18	-1.66	-0.52	# Obs >5%	10	13	15	11
Jul-07	1.33	-0.64	-1.56	-1.10	# Obs >1%	4	5	6	6

Figure 1 Hedge Fund Allocation in Endowments and Pensions by Year (as a % of Total Assets)

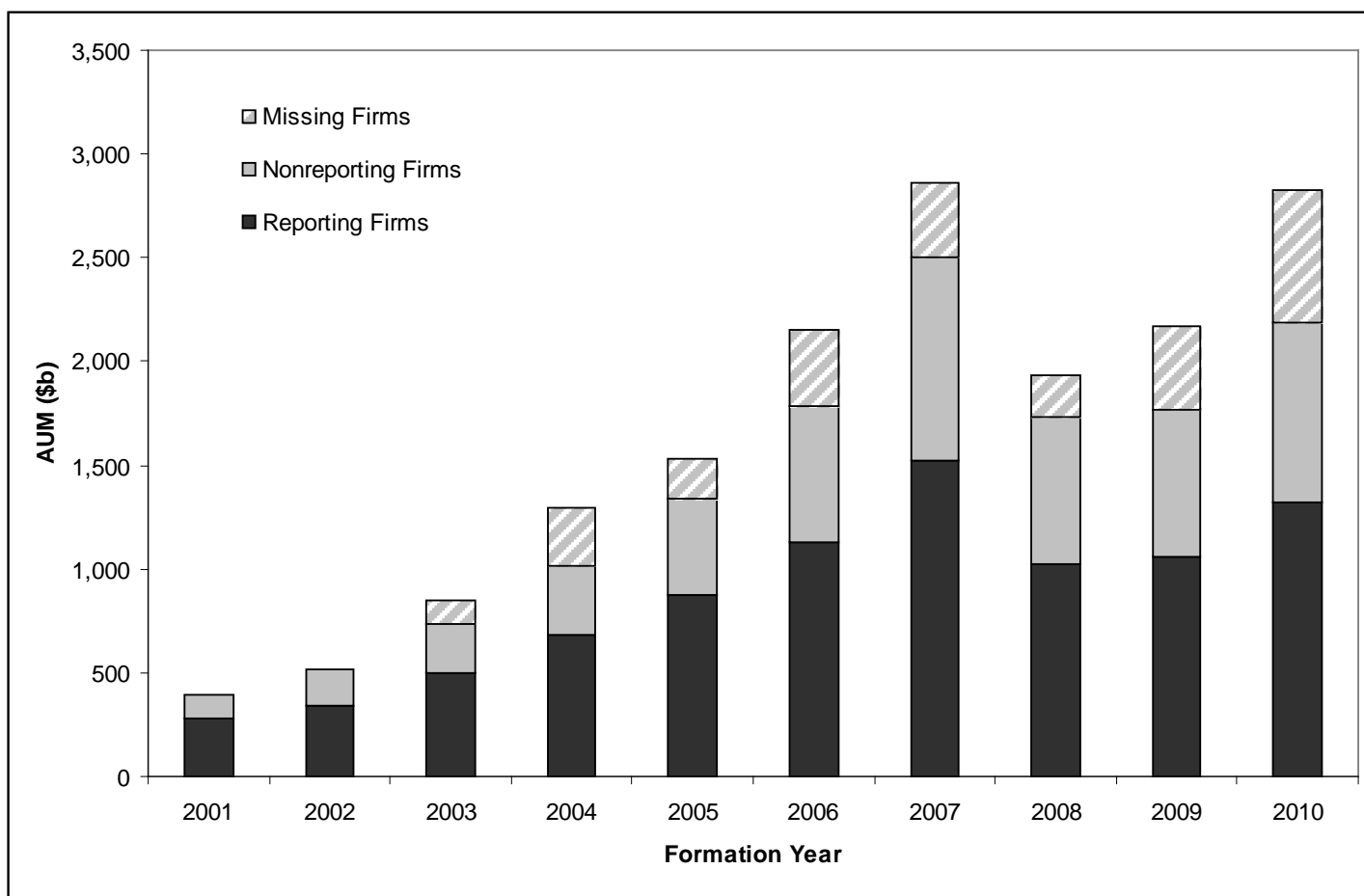
The blue bars represent the percentage of assets allocated to hedge funds in university endowments allocated to hedge funds (based on the Annual Endowment Surveys of the National Association of College and University Business Officers). The purple bars represent the percentage of assets allocated to hedge funds in the top 200 defined benefit pension plans in the US (based on annual surveys from Pension & Investments).



Source: *Pension & Investments*, NACUBO.

Figure 2 AUM in Hedge Fund Firms by Year

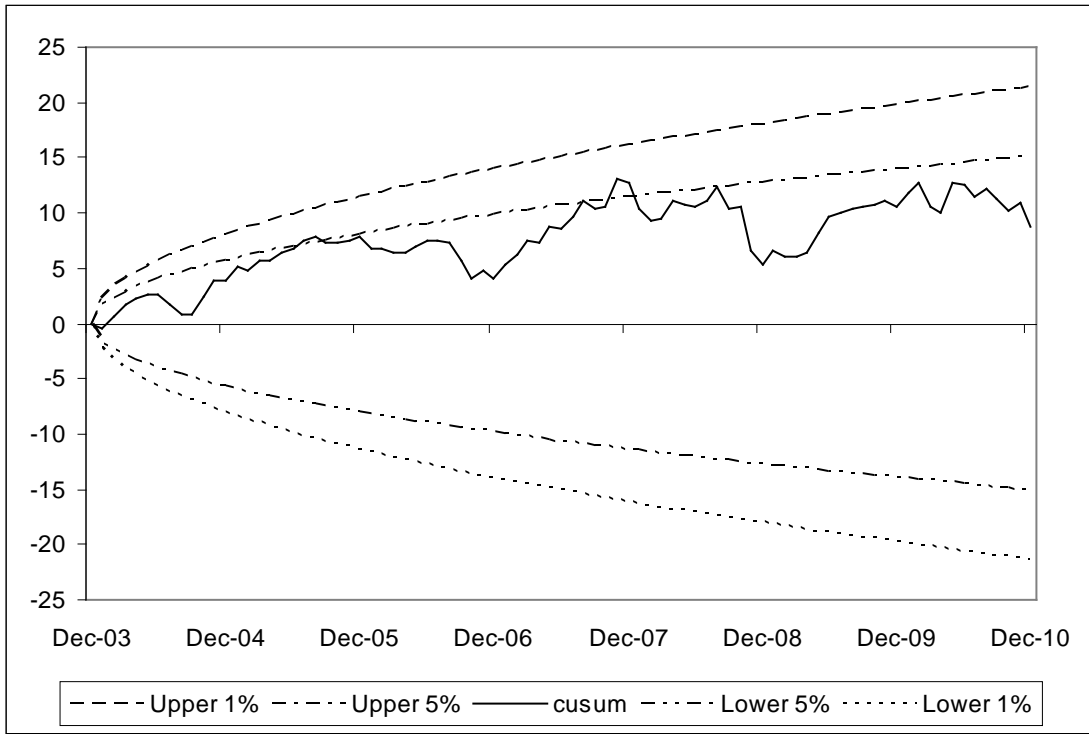
The green bars represent the assets under management (AUM) of reporting firms in the three commercial databases (BarclayHedge, HFR, and Lipper TASS). The blue bars represent the AUM of the nonreporting firms, which are the firms in the *Institutional Investor* “Hedge Fund 100” list and the *Absolute Return+Alpha* “Billion Dollar Club” list outside the three commercial databases. The purple bars represent the missing assets under administration (AUA) in the *HFN Hedge Fund Administrator Survey*.



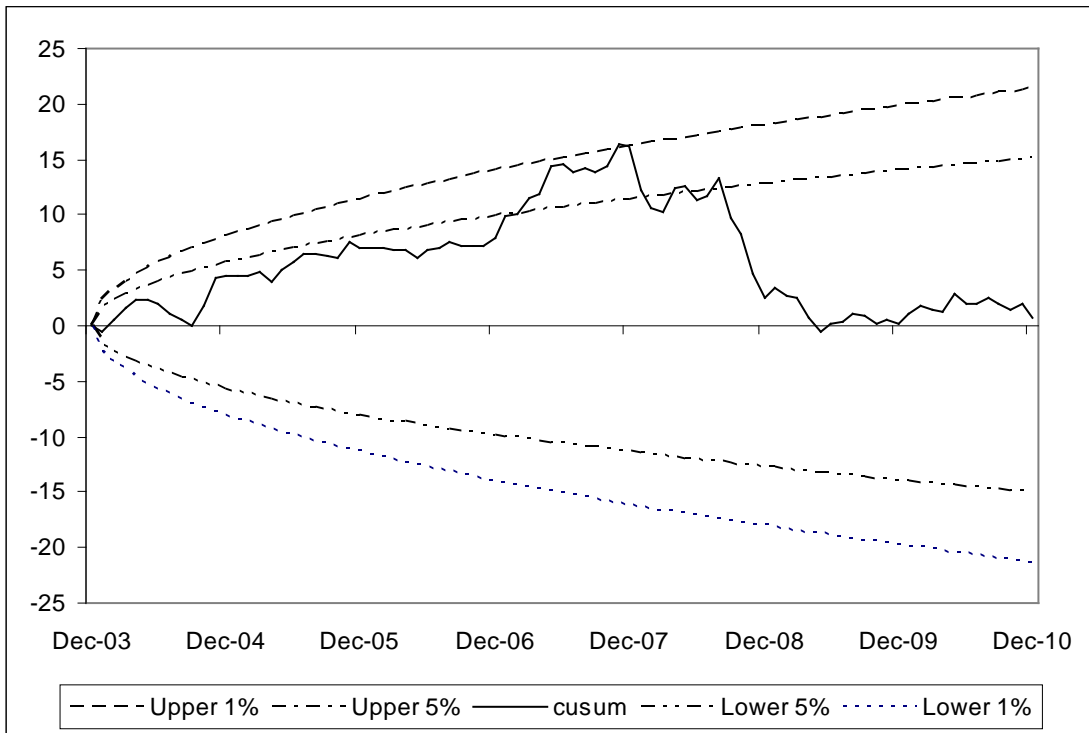
Source: *Institutional Investor*, *Absolute Return+Alpha*, BarclayHedge, HFR, Lipper TASS, HFN Hedgefund.net

Figure 3

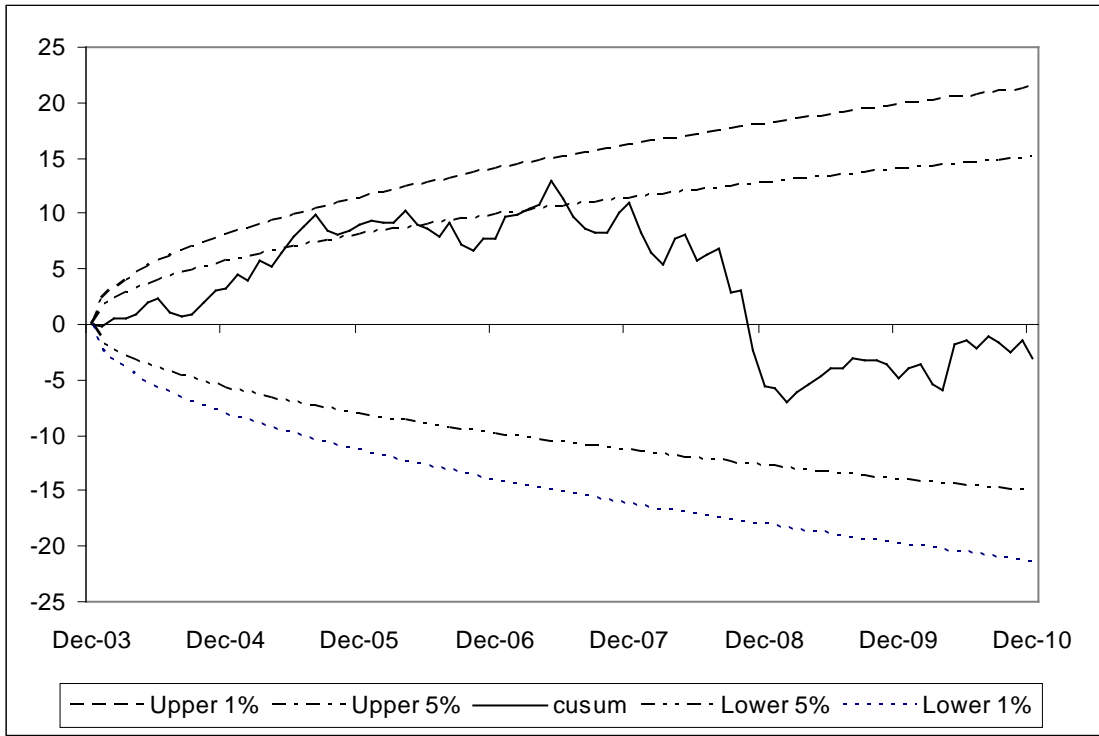
3 A. CUSUM test of NRI_EW-RLI_EW versus 8-factors



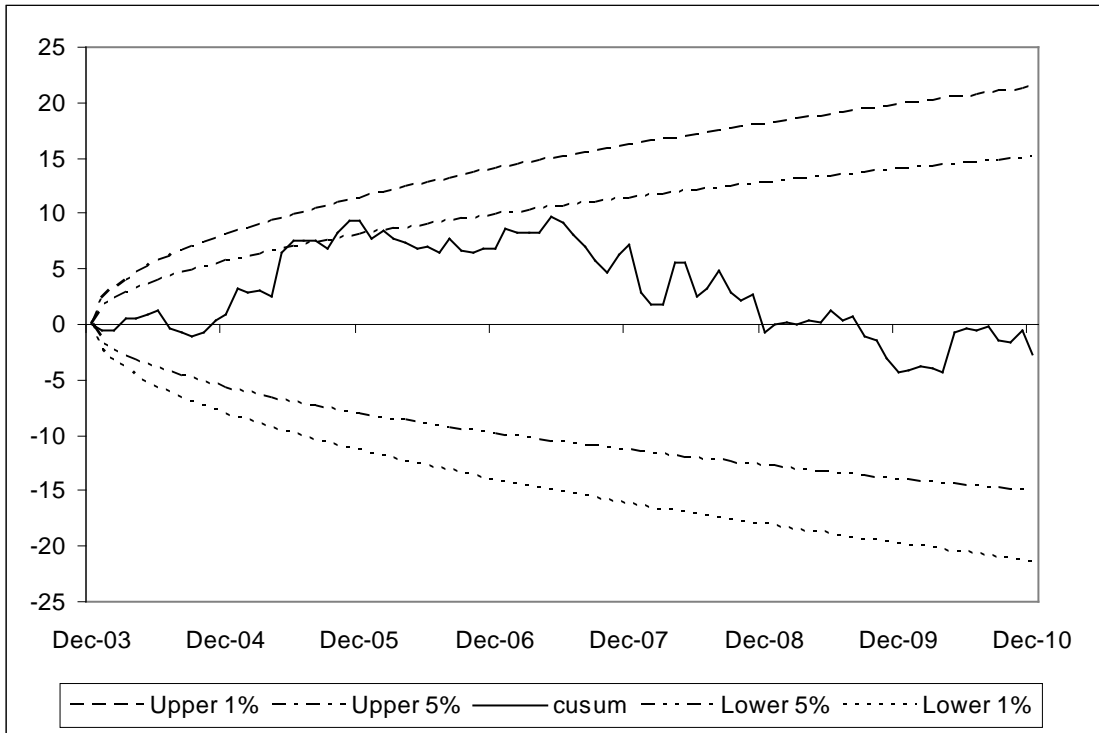
3 B. CUSUM test of NRI_EW- HFRI versus 8-factors



3 C. CUSUM test of NRI_AW-RLI_AW versus 8-factors



3 D. CUSUM test of NRI_AW- DJCSI versus 8-factors



Appendix A

Simulation of Three Models of Serial Correlation in Hedge Fund Returns

In this appendix, we provide a simple simulation experiment to demonstrate the key results for the three models of serial correlation in Getmansky et al (2004).

Model 1: This model corresponds to “scenario one” in the main text of our paper. The true returns are given by:

$$R_t = \beta\Lambda_t + \varepsilon_t, \text{ where } \Lambda_t \text{ and } \varepsilon_t \text{ have zero mean and are IID.}$$

The observed returns are given by an MA(1) process:

$$R_t^o = \theta_0 R_t + \theta_1 R_{t-1}, \text{ where } \theta_0 + \theta_1 = 1$$

As advocated in Getmansky et al (2004), one can obtain correct estimates of β by first applying a moving average filter to the observed returns, and then regress the “unsmoothed” returns on the risk factor. We call this the MA estimator.

Model 2: This model corresponds to “scenario two” in the main text of the paper. The true returns are given by:

$$R_t = \beta\Lambda_t + \varepsilon_t, \text{ where } \varepsilon_t \text{ has zero mean and IID.}$$

The factor follows an AR(1) process

$$\Lambda_t = \gamma\Lambda_{t-1} + z_t, \text{ where } z_t \text{ has zero mean and IID.}$$

The observed returns is the true return:

$$R_t^o = R_t.$$

Here, one can obtain correct estimates of β by regressing the observed returns on the risk factor. We call this the OLS estimator.

Model 3: This is a hybrid model. The true returns are given by:

$$R_t = \beta\Lambda_t + \varepsilon_t, \text{ where } \varepsilon_t \text{ has zero mean and IID.}$$

The factor follows an AR(1) process

$$\Lambda_t = \gamma\Lambda_{t-1} + z_t, \text{ where } z_t \text{ has zero mean and IID.}$$

The observed returns are given by an MA(1) process:

$$R_t^o = \theta_0 R_t + \theta_1 R_{t-1}, \text{ where } \theta_0 + \theta_1 = 1$$

As discussed in Asness et al (2001, Journal of Portfolio Management) and Getmansky et al (2004), one can regress the observed returns on the risk factor and lags of the risk factor. The sum of the regression coefficients gives an

estimate of the risk exposure. In the general case, the lag length of the risk factor should exceed the length of the moving average return smoothing process, which is 1 in our case, so we use 3 lags of the risk factor. We call this the AR estimator.

We pick parameter values for these models so the variance of the observed return roughly matches the variance of the Dow Jones-Credit Swiss Broad Hedge Fund Distressed Securities Index, and the risk factor's variance is close to that of the credit factor (Moody's Baa return - 10Y treasury return, or "BAA-TY") in the 8 factor model. Further, we specify the risk exposure, β , to be 0.4, and the length of the sample to be 100, roughly equal to the number of monthly observations in the size portfolio regression in our paper.

For each of the 3 data generating models, we simulated 10,000 trials of 100 observations each. For each trial, we perform all three estimators: MA, OLS, and AR. The table below provides the median estimate of the exposure, β :

	MA	OLS	AR
Model 1	0.3913	0.2994	0.3998
Model 2	0.4369	0.4004	0.3999
Model 3	0.4281	0.3382	0.4001

The root mean squared errors are given as follows:

	MA	OLS	AR
Model 1	0.0645	0.1068	0.0718
Model 2	0.0672	0.0437	0.0622
Model 3	0.1302	0.0735	0.0590

The MA estimator gives the correct estimate of the risk exposure only for Model 1, when there is return smoothing but the risk factor has no serial correlation. The MA estimator is biased in Models 2 and 3, where the risk factor has serial correlation. The OLS estimator gives the correct estimate only in Model 2, when there is no return smoothing. But it is biased in Models 1 and 3, where there is return smoothing. The AR estimator is unbiased in all 3 cases.

Appendix B

Factor Regressions at the Firm Level

In this appendix, we analyze the estimated risk exposure at the firm level, given in Table 10 below. Since these estimates are done in 2 year intervals, so we can observe the time variation in risk exposures across firm size.

Given the large amount of information, we shall only summarize a few key variables to illustrate the results. From the average exposure estimates in Table 6, we know that the small/large firm spread (RL-SP), the credit spread (BAA-TY), and the emerging market equities (IFC-Rf) are the most important risk factors in our data.

Figure 4 graphs the exposure to the Small/Large equity spread (RL-SP) over time. Figure 5 graphs the exposure to the credit spread factor (BAA-TY). Figure 6 graphs the exposure to the emerging market factor (IFC-RF).

Figure 4. Time Varying Risk Exposure to Small/Large Equity Spread Factor (RL-SP)

NRI are firms in the II100 and ARBDC lists that do not report to commercial databases. RLI (RSI) are the firms that report to commercial databases with assets under management greater (less) than that of the smallest NRI firm. RL-SP is the return of the Russell 2000 index minus the return of the S&P 500 index. The solid line graphs the median exposure of NRI firms to RL-SP; the dashed (dotted) line is the 75th (25th) percentile of the exposure of NRI firms to RL-SP. The solid line with crosses graphs the median exposure of RLI firms to RL-SP; the dashed (dotted) line with crosses is the 75th (25th) percentile of the exposure of RLI firms to RL-SP.

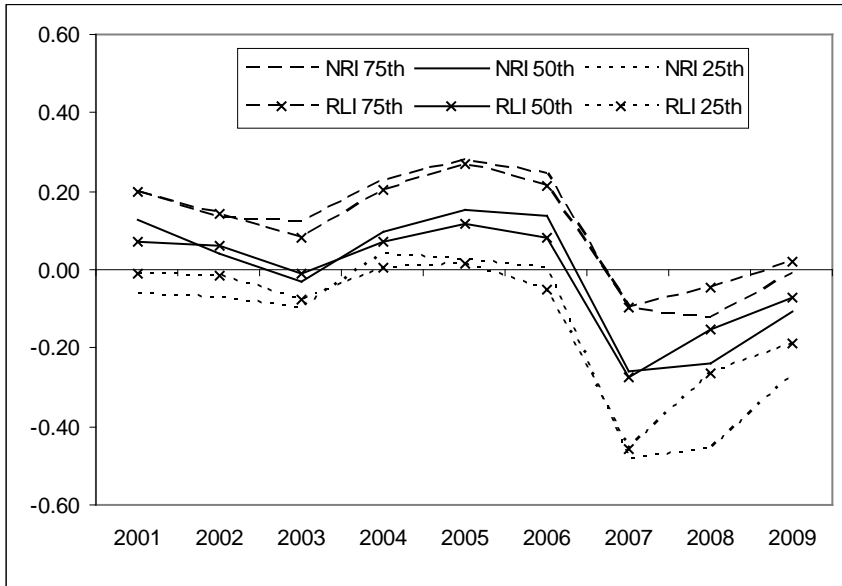


Figure 5. Time Varying Risk Exposure to Credit Spread Factor (BAA-TY)

NRI are firms in the II100 and ARBDC lists that do not report to commercial databases. RLI (RSI) are the firms that report to commercial databases with assets under management greater (less) than that of the smallest NRI firm. BAA-TY is the return of Moody's BAA bonds minus the return of 10 year treasuries. The solid line graphs the median exposure of NRI firms to BAA-TY; the dashed (dotted) line is the 75th (25th) percentile of the exposure of NRI firms to BAA-TY. The solid line with crosses graphs the median exposure of RLI firms to BAA-TY; the dashed (dotted) line with crosses is the 75th (25th) percentile of the exposure of RLI firms to BAA-TY.

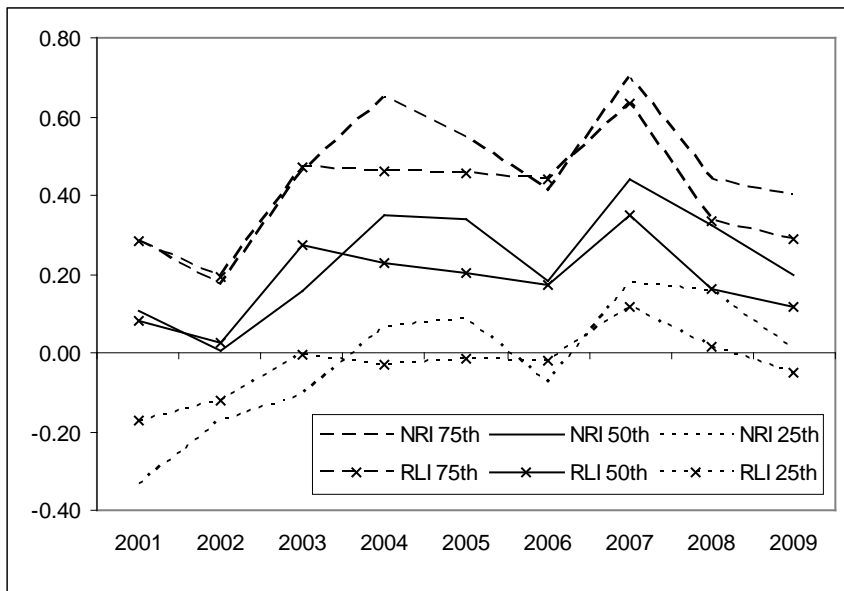


Figure 6. Time Varying Risk Exposure to Emerging Market Equities Factor (IFC-RF)

NRI are firms in the II100 and ARBDC lists that do not report to commercial databases. RLI (RSI) are the firms that report to commercial databases with assets under management greater (less) than that of the smallest NRI firm. IFC-Rf is the excess return of the IFC Emerging Market Index above T-bills. The solid line graphs the median exposure of NRI firms to IFC-RF; the dashed (dotted) line is the 75th (25th) percentile of the exposure of NRI firms to IFC-RF. The solid line with crosses graphs the median exposure of RLI firms to IFC-RF; the dashed (dotted) line with crosses is the 75th (25th) percentile of the exposure of RLI firms to IFC-RF.

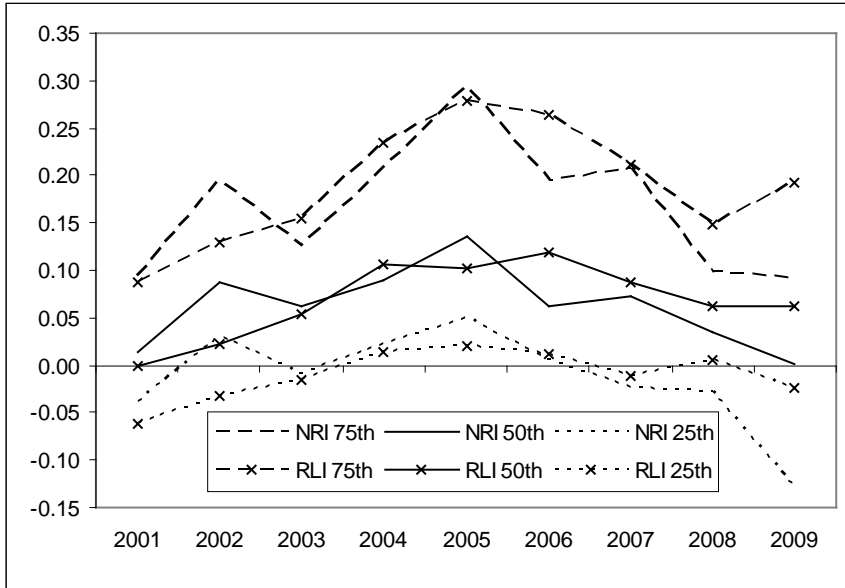


Table 10. This table contains the medians of the regression coefficients, adjusted R², and Durbin-Watson (D.W.) statistics of firm excess returns on eight risk factors. Reporting firms are the firms in three commercial databases (BarclayHedge, HFR, Lipper TASS). Reporting mega (small) firms have assets under management (AUM) more than (less than) the smallest firm in the II100 list and the ARBDC list. Non-reporting mega firms are those in the II100 and ARBDC lists that are outside the three commercial databases. In each formation year, the return of each firm is calculated as an equally weighted portfolio of the unique funds of the firm, for the 12 months of the formation year, and the 12 months of the subsequent year. The firm returns are regressed on eight risk factors: equity factors—the excess total return of the S&P 500 index (SP-Rf) and the return of the Russell 2000 index minus the return of the S&P 500 index (RL-SP); bond factors—the excess return of ten-year Treasuries (TY-Rf) and the return of Moody’s BAA corporate bonds minus ten-year Treasuries (BAA-TY); three factors representing trend-following styles of managed futures—the excess returns of look-back straddles on bonds (PTFSBD-Rf), currencies (PTFSFX-Rf), and commodities (PTFSCOM-Rf); an emerging market factor—the excess return from the International Finance Corporation’s index of emerging stocks (IFC-Rf).

Formation												Number of Firms	
Year	Constant	SP-Rf	RL-SP	TY-Rf	BAA-TY	PTFSBD-Rf	PTFSFX-Rf	PTFSCOM-Rf	IFC-Rf	Adj R ²	DW	Total	DW < d _L
A. Non-reporting Mega Firms:													
2001	0.0054	0.0918	0.1009	0.1152	0.1212	-0.0032	-0.0073	0.0145	-0.0033	0.1796	1.9183	35	0
2002	0.0045	-0.0001	0.0246	0.1854	0.0274	0.0027	0.0055	0.0075	0.0939	0.3022	1.9333	38	0
2003	0.0027	0.0777	0.0068	0.1022	0.1634	-0.0071	0.0176	0.0039	0.1152	0.2639	1.9488	50	0
2004	0.0032	0.0641	0.1277	-0.0578	0.3150	0.0059	0.0023	0.0048	0.1311	0.3626	1.9761	62	0
2005	0.0056	-0.1344	0.1860	-0.1573	0.3342	0.0258	0.0015	-0.0001	0.1989	0.3768	1.7508	57	0
2006	0.0046	0.0433	0.1451	-0.1564	0.1409	0.0033	0.0121	-0.0109	0.1176	0.3280	1.7335	70	0
2007	0.0041	0.0497	-0.3236	-0.2385	0.4896	-0.0270	0.0039	-0.0153	0.1098	0.4060	1.9832	73	1
2008	-0.0031	0.0808	-0.2952	-0.0369	0.3655	-0.0464	0.0080	-0.0141	0.0400	0.3358	1.9529	65	0
2009	0.0091	0.1179	-0.1416	-0.1061	0.2152	-0.0084	0.0212	-0.0118	-0.0026	0.3136	2.0201	57	0
B. Reporting Mega Firms:													
2001	0.0017	0.0428	0.0978	0.0040	0.0903	-0.0054	0.0016	0.0060	0.0234	0.2461	1.8868	92	0
2002	0.0031	0.0623	0.0922	0.1345	0.0511	-0.0053	0.0060	-0.0021	0.0764	0.2650	1.9274	75	0
2003	0.0019	0.0874	0.0045	0.0843	0.2719	0.0072	0.0151	-0.0149	0.1064	0.3423	1.9046	115	0
2004	-0.0007	0.0411	0.1077	-0.0459	0.2117	-0.0078	0.0082	0.0014	0.1610	0.3621	1.8840	151	0
2005	0.0042	-0.0246	0.1555	-0.1559	0.2050	0.0189	0.0031	0.0019	0.1848	0.3930	1.9308	182	1
2006	0.0022	0.0188	0.1066	-0.2593	0.2136	-0.0021	0.0113	-0.0018	0.1734	0.3879	1.8491	210	0

2007	0.0022	-0.0169	-0.3322	-0.2036	0.4211	-0.0111	-0.0065	0.0002	0.1342	0.4211	1.9761	252	0
2008	0.0002	0.0299	-0.1755	-0.0383	0.1791	-0.0051	-0.0096	0.0057	0.0936	0.3844	2.0256	175	0
2009	0.0060	0.0882	-0.0958	-0.1442	0.1311	0.0017	-0.0021	0.0075	0.1097	0.4046	2.0234	182	0

C. Reporting Small Firms:

2001	0.0008	0.2063	0.1402	0.1033	0.1402	0.0034	0.0060	0.0144	0.0846	0.3139	1.9571	862	1
2002	0.0023	0.1627	0.1037	0.1170	0.0924	0.0067	0.0065	0.0084	0.1186	0.3105	1.9530	970	0
2003	0.0004	0.1690	0.1099	0.0843	0.1769	0.0111	0.0131	0.0004	0.1680	0.3052	1.9731	1052	1
2004	-0.0016	0.1172	0.1814	-0.0045	0.2337	-0.0045	0.0049	0.0077	0.1687	0.3221	1.9099	1129	2
2005	0.0024	0.0678	0.2373	-0.1313	0.1779	0.0267	0.0050	0.0009	0.1693	0.3344	1.8725	1317	7
2006	0.0018	0.0121	-0.1859	-0.2042	0.3953	-0.0020	-0.0002	0.0020	0.1534	0.3698	1.9057	1434	0
2007	0.0018	0.0121	-0.1859	-0.2042	0.3953	-0.0020	-0.0002	0.0020	0.1534	0.3698	1.9057	1659	0
2008	-0.0002	0.1005	-0.1386	-0.0757	0.1778	-0.0048	0.0015	-0.0027	0.1345	0.3384	1.9652	1780	3
2009	0.0052	0.1452	-0.0423	-0.1446	0.1012	0.0097	0.0088	-0.0049	0.1191	0.3270	2.0734	1848	0