The Fix is In: Properly Backing out Backfill Bias

Philippe Jorion and Christopher Schwarz*

December 22, 2017

ABSTRACT

Hedge fund researchers have long known about backfill bias, typically correcting for it by truncating a fixed number of returns from the beginning of each fund's return series. However, we document that this practice decreases the percentage of backfilled returns by only 25%. Thus, empirical conclusions using this correction are still biased by backfill, including average performance and performance's relation with size, age, and other fund characteristics. Unfortunately, many databases do not include the listing dates needed to properly control for this bias (now including TASS.) We therefore propose a novel method to infer listing dates when not available.

JEL Classifications: G11 (portfolio choice), G23 (private financial institutions), G32 (financial risk management)

Keywords: hedge funds, public database, performance evaluation, backfill bias, listing dates

* Philippe Jorion (phone: 949-824-5245, email: pjorion@uci.edu) is at The Paul Merage School of Business at the University of California at Irvine and Pacific Alternative Asset Management Company (PAAMCO). Christopher Schwarz (phone: 949-824-0936, email: cschwarz@uci.edu) is at The Paul Merage School of Business at the University of California at Irvine.

The hedge fund industry has grown rapidly, now reaching over $3 trillion in assets under management (AUM) by some estimates.[1] This AUM growth has been accompanied by an expanding body of empirical research on hedge funds, including topics relating to performance, investor flows, fund characteristics, as well as the relation between fund characteristics and both flows and performance.

One significant issue with such empirical work is how hedge fund databases are constructed. Hedge fund managers only report to commercial databases on a voluntary basis. This leads to several types of biases, which were discussed early on by Park (1995), Ackermann et al. (1999), and Fung and Hsieh (2000). The two major biases, backfill and survivorship, are non-trivial.[2] Malkiel and Saha (2005), for example, argue that the difference between backfilled and non-backfilled performance averages 7.3% per year and that the difference between live and dead performance averages 8.4% per year. Database vendors have largely solved survivorship bias by preserving dead fund information, which are now routinely part of the sample examined.[3] In contrast, approaches to dealing with backfill bias have been more diverse, with effects that have not been systematically analyzed so far. The most common approach to correct for backfill bias is to truncate a fixed number of months such as 12 or 24 at the beginning of each fund's return series.

---

[1] This is the estimated provided by Hedge Fund Research (HFR) as of 2016. In 1994, hedge funds accounted for $167 billion in AUM only.

[2] Survivorship bias arises when funds that no longer report are dropped from the database of "live" funds. This generates a bias in performance measurement because funds that liquidate are more likely to have had poor returns that led to their closure. Survivorship bias can be controlled by keeping information on "dead" funds. This has become a common procedure of most databases starting in 1994, which largely fixes this first bias. Backfill bias arises when the fund's performance is not made public during some incubation period but then is added to the database presumably following good performance since the listing decision is voluntary. This generates a bias because the fund manager's decision to include the fund or not is most likely correlated with past performance. This bias is sometimes called "instant-history" bias. Hedge fund databases can also have other biases, such as the missing assets information, which is important for value weighted tests (i.e., Joenvaara et al. (2016).)

[3] Even so, managers whose funds are performing poorly have no incentive to report performance when they are planning to close the fund. Omission of end-of-life returns creates another type of bias, called "delisting bias", which is hard to control for, as discussed by Jorion and Schwarz (2014b).

In this paper, we investigate how well this technique controls for backfill bias, using databases which keep track of the actual date that funds are added to the database (i.e. the "listing date"). In summary, we find that the usual truncation method still leads to significant biases in empirical conclusions. This occurs because this adjustment retains at least approximately 75% of backfilled returns.[4] The reason for this high level of backfill return retention is the considerable dispersion in the distribution of the backfill periods, which contradicts the homogenous backfill assumption behind ad hoc adjustments of fixed length. Indeed, Fung and Hsieh (2009) note that backfill periods can extend to ten years.

To demonstrate the shortcoming of this ad hoc technique, Figure 1 plots the cumulative frequency distribution of backfill periods by fund.[5] It also graphs average returns, in excess of style and annualized, for funds with fixed backfill lengths, over the fund's entire return period.

<Insert Figure 1 about here>

The figure shows that 60% of funds have backfill periods greater than 12 months (40% have more than 24). Average fund returns increase sharply to 36 months but then remain very high after that, reflecting high levels of backfill bias. Hence, while the truncation technique does work for some funds, it still leaves a large fraction of backfill for funds with the largest amount of bias. In fact, 12 (24) months of truncation only reduces the percentage of backfilled returns by 18% (28%).[6] Thus arbitrarily truncating returns leaves most of this backfill bias in place.

---

[4] See Table A-1 for a selected catalog of hedge fund research and their backfill adjustments. A large majority of hedge fund research either uses no backfill adjustment or uses the truncation method. Only 25% of all papers use the listing date.

[5] As will be explained in a later section, this uses data from Hedge Fund Research (HFR) and Lipper Trading Advisor Selection System (TASS), which report fund listing dates, allowing measurement of backfill periods.

[6] Longer truncation periods are very inefficient in terms of solving the backfill issue. For example, going from a 12 to 24 month truncation period ends up truncating as many non-backfilled returns as backfilled returns.

As a result, we document that using the ad hoc 12- or 24-month cutoff correction can lead to misleading conclusions about important empirical relations. Average returns are still significantly too high.[7] More importantly, these cutoffs still lead to biases in cross-sectional relations as well. While prior research finds that younger and smaller funds seem to outperform others,[8] we demonstrate that these results are likely due to the remaining backfill. Intuitively, the explanation for this finding is that funds are small and young prior to listing in a database, which is correlated with higher performance during the backfill period.

Additionally, we find that that fund characteristics such as fees or liquidity restrictions lose most of their return predictability after properly controlling for backfill. The explanation for this result is that funds with higher performance during their backfill periods tend to set higher fees and liquidity restrictions when they list to databases. Since these funds seem more attractive for investors, managers must believe that they can impose more onerous conditions. We confirm that, while funds' backfilled performance is related to their fee and liquidity characteristics, their true non-backfilled performance is not. Finally, we find that the inclusion of backfilled returns significantly biases fund alphas upward and fund market betas downward. On the other hand, some relations seem robust to backfill correction techniques, including the flow-performance relation and performance persistence.

Given these findings, we argue that the best adjustment for backfill bias involves using the "date added to the database" field (DADDB) as the cutoff point. HFR does contain listing dates, as should TASS. In essence, this date reveals the time at which the fund information becomes public, with certainty. Admittedly, the fund may have been available to outside investors before that date,

[7] Bhardwaj et al. (2014) show that using the truncation method still leads to average returns that are too high in the CTA industry.
[8] See, e.g., Getmansky (2012), Teo (2009), Yin (2016).

in which case some of the discarded track record could be valid.[9] Even so, we show that fund performance and size characteristics change dramatically at the time of listing. Thus, omitting some valid observations is a small price to pay for eliminating backfill bias completely.

Admittedly, our recommendation of using the DADDB field can be challenging for a number of reasons. TASS used to provide this field but unfortunately discontinued it after March 2011. For funds added after that date, the field is blank.[10] To our knowledge, this problem has been hitherto unrecognized. This is a major issue because most computer codes will automatically assume that there is no backfilled performance, which is incorrect.[11] More generally, many databases do not include these listing dates at all. This is becoming a widespread problem given the recent practice of merging several hedge fund databases in an attempt to expand the dataset. In this case, the ad hoc truncation rule seems like the only available correction.

To solve this issue, we create a novel technique to detect add dates when they are not available. The technique does not rely on fund returns, age, or size. Rather, our method relies on the fact that funds are typically given sequential identification numbers in chronological order when they are added to the database. Since funds must be alive at the time of listing, we group funds and use their overlapping return dates to estimate the add dates. Using HFR and TASS, we verify that our method generates dates that are close to the actual add dates and creates similar returns patterns around listing dates.

We then use our technique on the BarclayHedge database, which does not have a DADDB field, and find that our generated add dates have characteristics that match the actual HFR/TASS

---

[9] For example, Fung and Hsieh (2009) report that the date added to the database field for some funds in TASS actually refers to the merger of TASS with Tremont, which implies that not all the dates before this date are backfilled biased. On the other hand, Jorion and Schwarz (2014a) note that the decision to list to a second database is related to performance; thus, performance prior to any add date is biased upward.
[10] We also used snapshots from the last few years to calculate listing dates for TASS. However, this is not a practical approach as this requires numerous snapshots per year, for each year going back far in time. Even Patton et al. (2015) only use data back to 2007 when examining data revisions in multiple versions of TASS.
[11] For example, in SAS, “if date => DateAddedtoTass then output;” would retain all returns if the date was blank.

ones. We estimate many of the same empirical tests with all returns, truncated returns, and returns after our generated listing dates and find empirical biases that match those with TASS and HFR. Thus, this new method will allow for proper correction of the backfill bias even when using databases that are missing this field. These results also demonstrate that simply using a 12- or 24-month cutoff with BarclayHedge would create misleading empirical results and cause the appearance of inconsistent results across databases.

This paper provides several contributions to the literature. First, we evaluate whether the usual truncation approach properly controls for backfill bias for hedge funds using multiple databases. We find that performance measures are still substantially affected by the remaining backfill bias even when truncating 24 months of returns. This is because these truncations leave most backfilled returns in the database. Likewise, we document that cross-sectional relations, such as between size, age, and most fund characteristics and performance, are biased using the usual truncation method. We also show that the relation between fund terms, such as fees and liquidity restrictions, is actually due to funds setting stricter terms for funds with better backfilled performance. More generally, we show that results are generally consistent across databases when properly controlling for backfill. Finally, we provide a solution to both TASS's recent DADDB issue as well as provide a new novel technique to infer listing dates when not available.

This paper is structured as follows. Section II briefly reviews the literature and explains how backfill is generally dealt with. Section III describes the data, issues with add dates, and describes the distribution of backfill periods. Next, Section IV examines how empirical results are affected by various methods to deal with backfill. Given that backfill can create false results for age and size effects, we then describe the new method to generate a listing date when not available in Section V. Concluding comments are contained in Section VI.

II. Literature Review

Good quality data are essential for empirical research. This recognition motivated the development of the Center for Research in Securities Prices (CRSP) at the University of Chicago, which painstakingly started to construct a database of U.S. stock prices in 1960. This led to a revolution in empirical finance research. Later, in 1996, CRSP developed a survivor-bias-free database for U.S. mutual funds. This centralized compilation of data was made possible by the SEC rule that publicly available funds are required to report their performance. Because CRSP data are designed for academic research, considerable resources are expended to check data quality.[12]

In contrast, private investments such as hedge fund partnerships are not required to disclose their performance. Even so, they may voluntarily report to commercial databases, essentially for marketing purposes.[13] This has led to a plethora of competing commercial databases, with only limited overlap of funds.[14]

The availability of these databases has led to an expanding body of empirical research on hedge funds. Initially, the focus was on a single database, usually TASS. Recent research has pooled information across several databases in an attempt to increase coverage. Examples include Agarwal et al. (2009), Avramov et al. (2011), Joenvaara et al. (2016), and Agarwal et al. (2017),

[12] For example, the accuracy of the CRSP Stock Databases has been improved through the documentation of various data issues (e.g., Rosenberg and Houglet 1974; Bennin 1980; Shumway 1997; Canina et al. 1998; Shumway and Warther 1999.) Elton, Gruber, and Blake (2001) compare the coverage of CRSP Mutual Fund Database to that of Morningstar, finding a 26 basis point bias in early periods. Evans (2010) finds some mutual funds are incubated, leading to backfilled returns. Ljungqvist, Malloy, and Marston's (2009) discovery of changes in the I/B/E/S database has led to more reliable data for researchers. Schwarz and Potter (2016) find that CRSP's mutual fund portfolios are inaccuracy prior to 2008.

[13] Another issue is the quality of the reported returns. Patton et al. (2015) examine revisions of historical performance information within the same database, and find interesting systematic patterns.

[14] Jorion and Schwarz (2014a) examine the strategic decision by hedge fund managers to report simultaneously or sequentially across multiple databases.

among many others. The five most commonly used databases now include TASS, HFR, CISDM/Morningstar, BarclayHedge, and EurekaHedge.[15] Not all, however, provide the date added to the database field. As a result, backfill bias is generally addressed by truncation only, if at all.

As a reference, Table A-1 lists a selected sample of empirical research and describes how backfill is addressed. The table shows that awareness of the backfill issue has improved over time, with papers progressively using the date added to the database. In recent years, however, more papers have been combining several databases in order to increase the sample size. Because many of these do not report the listing date however, researchers use arbitrary cutoffs across all of their databases.

The truncation period is typically taken as 12 or 24 months. Early studies, such as Brown et al. (2001), Fung and Hsieh (2000) focus on the TASS database and report average incubation periods of 27 months and 12 to 15 months for CTA and hedge funds, respectively. As mentioned earlier, Malkiel and Saha (2005) estimated a "**backfill difference**" of around 7.3% per annum from 1995 to 2003. Using the DADDB field, this is estimated from the difference between the average performance of funds over the backfilled and non-backfilled periods. Agarwal et al. (2013) use 13F filings to compare returns imputed from reported long equity positions for hedge funds around the listing (added to the database) dates. They find a performance drop of around 7.2% per annum around the listing date, which is consistent with the Malkiel-Saha backfill difference. Aiken et al. (2013) also find that listing is associated with systematic performance changes.

More generally, "**backfill bias**" can be measured by taking the average of returns over the entire history, including backfill, minus that after the listing date. This bias is a function of the fund backfill difference Δ (e.g., 7%) multiplied by the ratio of the backfill period to total performance

[15] These databases have evolved over time. In 2005, Lipper acquired the TASS database from Tremont Capital, which itself had purchased TASS in 1999. Morningstar purchased the Altvest database in 2006, the MSCI database in 2008, and the CISDM database in 2010, originally called MAR.

period (T_b/T). Defining μ as the post-listing return and μ_T as the average over the entire period including backfill, backfill bias is generally measured as

$$\mu_T - \mu = \Delta\ (T_b/T) \tag{1}$$

so must be less than the backfill difference Δ.

Fung and Hsieh (2000) use a fixed truncation period of 12 months and report a backfill bias of 1.4%. Capocci et al. (2005) vary the truncation period from 12 to 60 months, and report a bias going up from 1.3% to 2.3%, respectively. The issue with a fixed truncation window, however, is that it ignores the distribution of backfill periods, so leaves many backfilled returns. In addition, these may be correlated with fund characteristics.

For perspective, it should be noted that backfill bias also arises with mutual funds. Mutual fund families can seed new funds without initially making their performance public. After a while, the fund may acquire a ticker symbol from the National Association of Securities Dealers (NASD), thus becoming public. Evans (2010) reports an incubation difference of 9.8% for domestic equity mutual funds, defined as the average performance difference between pre-listing incubated funds and non-incubated funds; on a risk-adjusted basis, this bias ranges from 1.4% to 3.5%. He suggests removing performance before the date of the ticker creation, which is similar to removing returns prior to the date added to the database. He also reports that 23% of these funds are incubated. The problem is worse for hedge funds, where incubation is more pervasive. Indeed, as shown in Figure 1, 60% of funds have a backfill period longer than one year.

III. Data

This study uses two widely-employed hedge fund databases, TASS and HFR. These do report the "date added to the database" (DADDB) field necessary for our analysis. We use the February 2016 version of the TASS database, which has 20,069 funds, and the February 2016 version of the HFR database, which has 23,396 funds. Both databases contain live and defunct funds since 1994, which eliminates survivorship bias after that date. After removing duplicates, the combined databases have 34,257 funds. We then impose the usual filters on our sample, eliminating funds of funds, as well as funds that report returns in non-dollar currencies, or gross of fees, or on a non-monthly basis.

A. Data Issues with Add Dates

HFR does not suffer any significant "add date" issues, although we do note that the first add dates are in May 1996, which is when HFR started tracking this field.[16] Thus, although HFR is survivorship bias free starting in 1994, using add dates will cause HFR to have no data prior to May 1996. TASS, unfortunately, has a major issue with the add date: The data vendor ceased updating the DADDB field since March 2011.[17] This is unfortunate, because researchers will probably erroneously assume that the backfill period is zero for funds listed after that date. TASS also has other issues with its add date field, which suggest that the information may not be quite as accurate as in HFR.[18]

For our analyses, we recreate add dates in TASS after 2011. Using monthly snapshots of TASS, we report the add date of a fund to TASS as the first month the fund appears in the sequences of database files. Using this methodology, we are able to obtain add dates for almost 6,000 funds for which this field is not reported. We double-check our methodology by looking at actual add dates from 2005 to 2011 and verify that our "reconstructed" add date and the reported

[16] 1,242 funds in HFR have an add date of May 1996. Thus, all returns prior to that point are labelled as backfilled. As we will note, almost all returns for those years are backfilled anyway.
[17] In private correspondence, Thomson Reuters, which sells the TASS database, has indicated that "this field is no longer maintained."
[18] The distribution of dates is not evenly clustered. Some of this is likely due to the merger of TASS and Tremont (see Aggarwal and Jorion (2010b)). For example, while the average number of funds added in 1999, 2000, and 2001 was approximately 20 funds per month, 203 funds were added in January 2001, 164 in September 2001, and 209 in December 2001. We also see strange behavior in 2007. In July, August, and September 2007, only 6 funds were added to TASS. This contrast with 566 funds added in September 2007.

TASS add date are almost always the same. Later, we will also examine the effect of this missing add date on empirical research using recent TASS data.

B. Distribution of Backfill Period

We first look at the distribution of backfill periods across our combined database, which is reported in Figure 1. This uses TASS data before March 2011 and our reconstructed series thereafter, as well as the add dates reported to HFR. If a fund is in both databases, we use the earlier of the two add dates. The "backfill period" is defined as the difference between the fund's first return date and the date added to the database.

The median backfill period is 16 months across our TASS and HFR sample. This probably explains the usual practice of truncating between 12 and 24 months of initial returns. This does not tell the entire story, however. The distribution has a long right tail. About 10% of funds have backfill periods longer than 6 years; 5% of funds have more than 10 years of backfill. Hence eliminating the first 12 or 24 months leaves a substantial bias in the dataset since leaving one fund with 10 years of backfill is equivalent to correcting 10 funds with one year of backfill.

Figure 1 also plots average returns over each fund's entire life in excess of the style average, sorted by backfill length. The bias increases with the length of the backfill period, as expected.[19] Beyond two years, the excess return is about 3% per annum.

Next, Table 1 reports the distribution of backfill periods using various cutoff periods, as well as across years. The first line in Panel A shows that both TASS and HFR funds have about 40% of backfilled returns, which is a very high fraction. Most funds, about 98%, have some backfill, however short. Note that the TASS reconstructed sample has a much higher fraction of backfilled returns than what is reported (43% vs. 35%). So, this will affect all studies using TASS data after

[19] The average is negative for funds with a backfill period below 12 months because the style average includes all funds with their backfilled periods.

March 2011. Next, each row shows the remaining fraction of backfill with various cutoff periods. The usual practice of eliminating the first 24 months results in a dataset with about 30% of the remaining returns still backfilled. This is because the longer truncation method is cutting some backfilled and non-backfilled returns at the same time. This demonstrates that the usual ad-hoc truncation method fails to purge the backfill bias.

<Insert Table 1 about here>

The second panel shows that backfill is a function of time. There is very little backfill in 2015, for example (about 5%). Going back in time systematically increases the extent of backfill, to 41% in 2005 and 93% in 1995. However, this should not be interpreted as an improvement in the data, such as funds reducing their backfill period in recent years. Instead, this is a purely mechanistic effect due to the passage of time. The high backfill rate in 1995 is due to funds that were allowed to add data during 1996 and subsequent years. This phenomenon has not yet happened in 2015.[20] This also means that estimates of the backfill bias increases with longer database coverage.[21]

IV. Empirical Results

In the prior section, we established that the usual truncation method does a poor job of removing backfill from databases. In this section, we examine how this affects empirical results.

A. Effect on Average Returns

First, we examine the effect of backfilling on average returns in Figure 2. Panel A plots the average returns relative to the add dates for the combined HFR and TASS databases. Returns are

[20] For example, currently 41% of 2005 returns are backfilled. However, if we look at the 2006 version of TASS, only 6% of 2005 returns were backfilled. This is similar to the 5% backfill rate for 2015 returns as of 2016.

[21] For instance, Joenvaara et al. (2016) compare the backfill bias across three databases (TASS, HFR, Eurekahedge) and find that for coverage periods of 18, 16.5, and 7.5 years, the bias is 4.2%, 3.2%, and 1.7%, respectively.

annualized and reported in excess of style returns, using all funds including backfilled periods. The effect is striking. Average returns plunge from around +3% before the listing date to around −3%. So, managers list their funds right after good performance, which then drops suddenly.

<Insert Figure 2 about here>

Panel B also shows the evolution of assets around the listing date. There is a slow drift down during the two years before to the listing date. The average fund size decreases from about $125 million to below $100 million. Given these funds are performing extremely well and thus should have strong investor interest, this confirms that the average fund has been hidden from view before listing. After the listing date, assets grow steadily, reflecting the good previous performance, which is now advertised. This panel also shows that fund AUM tends to be smaller during the backfill period, which has implications for tests of the performance vs. size relationship.

Table 2 then gives more detail on how this bias varies across time and database. Panel A reports results for TASS, with and without our correction, Panel B for HFR, and Panel C for the combined datasets. The table should be read from left to right, i.e., from no backfill adjustment to progressively better adjustments.[22]

<Insert Table 2 about here>

For the average over the entire sample, Panel A shows a systematic deterioration of performance going from left to right. With no backfill correction, the average return over the entire period for TASS is 10.16%; truncating the first 12/24/36 months changes this number to 9.11%/8.82%/8.76%. Using the reported add date drops this number further to 7.03%. Using our

[22] In this paper, since we are assessing the impact of backfill on empirical conclusions, we report results with equal-weighting for all funds. This is because almost all empirical tests in the literature (e.g. regressions) are equally weighted.

reconstructed date drops the number further to 5.66%. The total backfill bias is thus on the order of 10.16%−5.66%, or about 4.5% over this period, which is reported in the last column. All of returns using the truncation method are both economically and statistically higher than those using the add dates to remove backfill.

The difference widens going back in time. As explained for the previous table, this is a mechanistic effect due to the greater potential for backfill in earlier years since there has been more time for newer funds to backfill returns. For example, the difference is small for 2015. In contrast, the bias for 1995 is close to 13%. Some of this is due to funds listed before May 2011, but also to funds listed after. For example, the difference between the returns using the reported and reconstructed add dates was 13.98%-8.79%, or 5.19% in 1995. This is solely due to funds added after March 2011 that are backfilled all the way back to 1995. Hence, the omission of this field by TASS after 2011 has a significant effect on measured performance. There are also variations across years, e.g. during the 2007-2008 financial crisis, when there was more dispersion across fund returns.

Backfill bias also affects HFR, although to a lesser extent. Panel B shows a similar pattern of decreasing average returns from left to right, with a total backfill bias on the order of 10.34%−6.80%, or about 3.5%. For the combined databases, Panel C shows a drop of 10.26%–5.89%, or 4.5% using the first listing date for funds common to TASS and HFR. The bias worsens when using the second listing date. Since the fund becomes public on the first date, this second date effect reflects an increased marketing push after good performance, which suffers from mean-reversion, rather than conventional backfill bias, however.

B. Effect on Fund Return Prediction

Section A has shown that backfill bias significantly continues to affect measures of average returns, even after truncation of up to 36 months of returns. Another question is whether this also biases tests of relations between fund returns and indicators of interest. Table 3 considers tests of whether performance is related to size. Each quarter (year), we sort funds into quartiles based on their reported assets within the style. We then calculate the average style adjusted performance over the next quarter (year). These returns are averaged over our sample period, and tests of difference between the first and fourth quartiles performed using the Fama-MacBeth approach. Panels A, C, and E (B, D, and F) report quarterly (yearly) results for TASS, HFR, and our combined database, respectively. In each panel, we report results using all returns, various ad-hoc return cutoffs, and finally using the proper add dates cutoffs.

<Insert Table 3 about here>

All Panels show a significant impact of size on performance using all returns or the traditional cutoffs. When backfill is ignored, funds in the first quartile (large funds) tend to return significantly less than small funds. In Panel A, the difference is -0.77% per quarter, or about -3% per annum. As Figure 2 showed, however, this could be due to the combination of good performance and small AUM during the backfill period. Indeed, the effect slowly goes away when moving from the left side to the right side of the panels, which progressively eliminates the backfill period. Cutting off the first 24 months halves the size of the effect, but this is still statistically significant. Using the actual add date for cutoff actually changes the sign of the difference, which becomes insignificant. This finding generally occurs whether considering TASS, HFR, or the combined database, and for quarterly and annual horizons. Therefore, appropriately controlling for backfill largely eliminates the size effect and documents no diseconomies of scale in the hedge fund industry.

Next, Table 4 performs the same analysis controlling for age instead of size. The left column of Panel A shows that the oldest funds underperform the youngest funds by 1.08% per quarter, or about 4% annually, which is both economically and statistically significant. This is true even when cutting off the first 12- or 24- months of returns. Here again, however, this effect largely disappears when controlling for backfill bias.[23] This is because young funds tend to have a greater proportion of backfilled returns.

<Insert Table 4 about here>

The next effect we examine is the persistence in returns. This is central to the evaluation of alpha generation for hedge funds. The methodology is the same as before, using quartiles sorted on the past quarterly returns within styles.

<Insert Table 5 about here>

Panel A, for example, compares the subsequent 3-month returns of TASS funds sorted on their previous performance. The left column does not adjust for backfill. It shows that the best funds continue to outperform the worst funds by 1.93% per quarter, or about 6% per annum. A reasonable hypothesis is that these results might be driven by the backfill period, which includes a string of good returns that are built to appear persistent.[24]

Contrary to the age and size effect, however, we find that persistence of returns is not driven by backfilled data. There is no systematic drop in the performance going from the left to the right

---

[23] Aggarwal and Jorion (2010a) report that emerging (or young) managers tend to outperform others. This does not invalidate their results, however, because they use a sample of managers without backfilled data.

[24] Indeed, Jagannathan et al. (2010) argue that backfill bias could lead to spurious persistence. However, they still find persistence for returns only after the listing date.

of the table. Forecasting returns after the listed dated yield a dispersion of 1.83% per quarter (7.3% pa) between best and worst funds, which is still highly significant. Also note that this persistence decays over time. Indeed, the 12-month average return of 3.1% is much lower than the annualized 3-month return of 7.3%. These results confirm the persistence of performance reported in studies such as Baquero (2005), Kosowski et al. (2007), and Aggarwal and Jorion (2010a).[25] These results are also consistent with Jorion and Schwarz (2014a) who document that backfilled performance is informative about future non-backfilled performance.

The previous three tables presented univariate sorts. For completeness, Table 6 reports multivariate regressions of fund returns on prior returns, assets, and age. The table also adds other fund characteristics of considerable interest, including fund fees and liquidity variables.

<Insert Table 6 about here>

As usual, the left-most column presents results with all returns, including backfilled ones. We continue to find that future returns are strongly associated with previous returns, and negatively to size and age. The effect of fund characteristics is consistent with previous research. Higher returns seem associated with higher minimum investments, with higher management and incentive fees, with the presence of a high water mark, and with worse liquidity terms, including longer redemption notices and frequency.

Going from left to right, which properly eliminates backfill, wipes out most of these effects, however. Using 3-month returns in Panels A, C, and E, shows that the only variable that remains consistently significant is prior return. For that variable, the point estimates do not generally go down with less backfill. In contrast, the size and age effects largely go away, as in the univariate

[25] Baquero et al. (2005), for example, report evidence of persistence at a 12-month horizon, even after correcting for look-ahead bias.   Kosowski et al. (2007) report mild evidence of persistence using Ordinary Least Squares alphas but much stronger evidence in a Bayesian framework.   In another study, Aggarwal and Jorion (2010a) find that new funds that do not backfill generally have higher performance persistence than other funds, especially in earlier years.

sorts. We again note that simply cutting off 12 months of returns, or even 24, largely does not alter the conclusions as compared to using all returns. Using the add date, in contrast, leads to significantly different results.

Interestingly, we also find that the apparent relationship with fees and with liquidity variables goes away as well, e.g. across both samples in Panels E and F. This result is somewhat surprising as these variables are not obviously related to the backfill bias. However, managers maybe inclined to set stricter terms (i.e. higher fees and more liquidity restrictions) if the fund has outperformed in the backfill period. This is just like any product market. To test this explanation, we perform cross-sectional regressions of fees and liquidity restrictions on backfilled returns. Table 7 reports coefficient t-values.

<Insert Table 7 about here>

In the first panel, all coefficients are positive. This means that fund terms are systematically stricter when backfilled performance is better. When we combine the two databases, all variables but the subscription period are significant. Overall, these results demonstrate that backfill even biases the relations between fund characteristics and performance. Our findings again emphasize the importance of completely removing backfill, even for cross-sectional tests.

C. Effect on Fund Flows Prediction

Next, we examine the effect of backfill on the performance-fund flow relation. Table 8 presents regressions where the dependent variables are investor net flows over the next quarter or year. Independent variables include fund performance over the previous year,[26] previous flows, size, as well as various fund characteristics. Panel A, for example, shows that TASS quarterly

[26] To measure performance, we include two piecewise-linear variables. *Low Perf. Rank* is the minimum of the fund's performance rank within its style and 50%. *High Perf. Rank* is the maximum of the fund's performance rank minus 50% and zero. These two variables allow us to capture non-linear effects,

flows are positively related to previous performance, to previous fund and style flows, and negatively to assets.

<Insert Table 8 about here>

As before, the question is whether these relations are affected by backfill, which can be judged by moving from the left to the right side of the table. The table shows that the signs of these relationships are largely unaffected by backfill. Thus, flow relationships reported in empirical research seem robust to backfill.[27]

D. Effect on Fund Alphas

The stark difference in returns before and after the listing date shown in Figure 2 most likely reflect the portfolio managers' active decisions to list in a database after good performance. An alternative hypothesis is that this could reflect changes in the risk profile of the fund. Perhaps the higher returns are due to greater exposure to priced market risk factors. After listing, the fund could then be run more conservatively, leading to lower returns but also lower risk exposures.

To investigate this possibility, Table 9 presents the average alpha computed using two methods, from regressions on (1) the usual style index return, as well as (2) Carhart's (1997) 4-equity factors, which include the market, size, value, and momentum.[28]

<Insert Table 9 about here>

Panel A reports the alpha and beta estimated from the style index regressions. Contrary to the above hypothesis, the average style index beta, or slope coefficient, is lower during the backfill

[27] Jorion and Schwarz (2014a) show that investors use backfill information to make flow decisions. That will dampen the impact of any backfill bias on flow regressions. We do note, however, that the average flow is impacted by backfill. Without backfill the average flow is more than halved. However, this is largely an average effect as seen in the regressions.

[28] In untabulated results, we run the Fung and Hsieh (2001, 2004) model and find similar results to those reported.

period. This automatically generates alphas that are even higher during the backfill period. In the backfill period, the average beta is 0.78 only, and average alpha 0.79% per month, which is very high. The beta and alpha estimates change to 0.98 and -0.11% after the listing date.   So, the risk profile of funds regresses to the style average after listing, and the alpha drifts back down toward zero. The same pattern is observed in Panel B. Systematic risk (beta vs. the S&P 500 index) increases post backfill. Alpha also drops sharply, from 0.77% to -0.05%. To illustrate these effects, Figure 3 plots the average alphas and market betas around the listing dates, measured over the prior 24-month window. Beta drifts up, alpha goes down.

<Insert Figure 3 about here>

In summary, we found no evidence that higher returns during the backfill period are due to higher exposure to common market risk factors. Alphas are also higher, which reflects idiosyncratic risk and a large dose of luck.

V. Generating Listing Dates

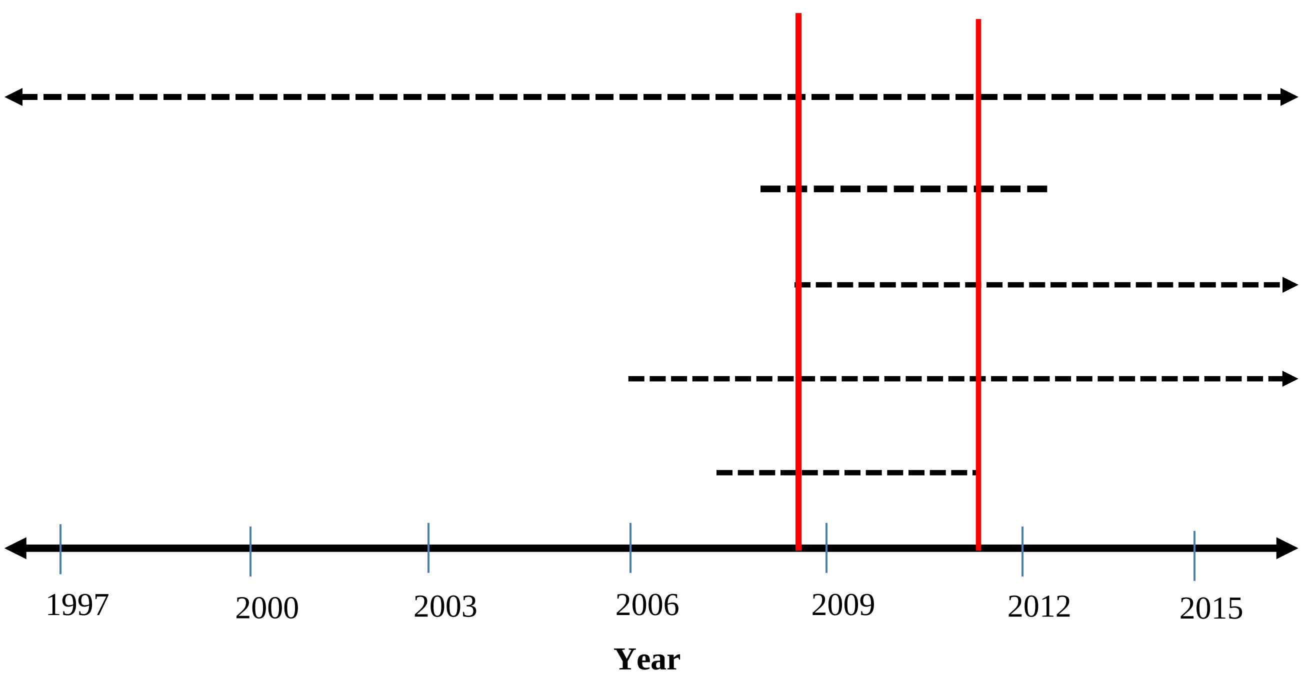
In the prior sections, we established that add dates are important for more precise empirical inferences concerning the hedge fund industry. However, this is easier said than done because many database vendors do not keep track of that date. Researchers likely do not have the ability to generate add dates using the consecutive snapshot method. Thus, in this section we propose a new method to generate the listing dates for databases that do not report this field.

One could try to infer the add date from the average performance drop reported around that date. Unfortunately, the noise in returns for individual funds does not make this practical. For example, the typical hedge fund in our sample has a standard deviation of monthly returns close to 4%. Thus, a performance drop from one month to the next of 0.7% such as reported in Figure 2 is common. Even if we group 100 funds together or lump six months of returns together for individual funds, the noise in the return series would still be too large to detect add dates with the required precision.

Instead, our method does not rely on fund performance. The approach relies only on the fact that funds are added largely chronologically to hedge fund databases. In other words, when a database vendor adds new funds to its database, the unique identifiers of all the new funds are sequential.[29]

To illustrate the intuition behind our method, the figure below displays the periods for which returns are available for a sample of five funds in the HFR database, taken as an illustrative example. These funds have sequential numbers (24860, 24861, 24863, 24864, and 24865). Our method assumes a common listing date for these funds.

Example of Return Periods for Five Sequential Funds



[29] Sequential means that the funds have consecutive numbers when added to the database. A proper sequence could include for instance 10, 20, 30, 40, 50, … as long as the vendor does not add later other funds numbered 11, 12, 13, and so on.

The funds exhibit wide differences in return periods, covering in total July 1996 to January 2016. What is important is not the range of returns for each fund, but rather the overlap of returns for all five funds. This is because we know for sure that, if these five funds were added at the same time, all the funds must have returns during the month in which they were added. In this particular example, the funds must have been added to HFR between July 2008 (which is the last start date across these funds) and January 2011 (which is the first end date), or between the red lines. One could assume either that all overlapping returns are backfilled or that none of the overlapping returns are backfilled.[30]

Based on the add date information from TASS and HFR, the shortest backfill period for a group of five funds is on average five months.[31] Thus, our method would infer that the listing date is around December 2008, i.e., five months after July 2008. In fact, the actual add dates for these five funds are near the end of November 2008, which is close. We use essentially the same method for all funds in our databases, making some adjustments. The overall procedure is as follows.

First, instead of grouping five funds together, we select 20 non-overlapping funds at a time (i.e., 1-20, 21-40, etc). Increasing the number of funds reduces the funds' return overlap, which increases the precision of our add date estimates. However, using too many, such as 500, is unreasonable because it would be inconsistent with the assumption that these were all listed during the same month. We use 20 as a compromise since this is the average number of funds added per month. Second, because some listings may be out of chronological order, or because the selected 20 funds may have add dates that span more than one month, we define the start date as the months with the most overlapping returns rather than requiring a month to have 20 overlapping returns.

[30] In our general application of the method, either of these assumptions led to similar conclusions.
[31] To derive the expected minimum amount of backfill, we used order statistics based on a negative binomial distribution, which resembles the empirical add date distribution, as well as a Monte Carlo simulation.

Finally, we choose the third month (or latest if less than three) after the start date as the add date, since this is the expected minimum backfill for 20 funds.[32]

We validate this approach using the actual listing dates in HFR. Figure 4 compares the distribution of actual add dates (Panel A) versus our generated add dates (Panel B).

<Insert Figure 4 about here>

Overall, we find similar patterns across the two panels.[33] The peak add dates occur during 2004 to 2008. We then further validate our method by comparing the returns around the generated and actual add dates in Figure 5, with TASS and HFR in Panels A and B, respectively. Both panels show the same large drop in returns around the actual and generated add dates.

<Insert Figure 5 about here>

As a last verification, we compare our generated add dates to the actual add dates of HFR and TASS. Table 10 shows the average difference and average absolute difference between these dates. We also report the percentage of funds where our generated add date is within three, six, and 12 months of the actual add date. As a baseline comparison, we report the same statistics with the inception date and the traditional 12 and 24 month cutoffs as the add dates. Results for HFR and TASS are in Panel A and B.

<Insert Table 10 about here>

[32] We tried 10 or 30 fund groups and found similar results. We also selected the first or last overlapping return and obtained similar results as well. One could think of many ways to make the procedure more complex and potentially more accurate. However, we want to ensure that the procedure is not over-fitted to any one database since we cannot observe the actual listing patterns of databases without add dates. More information about the procedure is described in Appendix A-2.

[33] HFR has no add dates prior to May 1996 as noted previously. Plots for TASS are similar, although TASS has a large number of add dates in certain months that seems unusual as noted in the Data Section.

Our method is much more accurate than using arbitrary cutoffs. In both databases, the method generates an add date within 3 months of the actual one for 52% of funds. This increases to about 67% and 82% when expanding the range to 6 and 12 months, respectively. As a comparison, using either cutoff length only gives about 10% of add dates within 3 months. Our method is also unbiased, with a systematically lower average difference. For HFR, where we are more confident of the accuracy of the add date field, our generated add date is on average within 3 months of the actual date.

We also examine the efficacy of our method. Looking at the Inception Date column, we see that without removing any backfill HFR would have 903,090 returns, of which 533,386 (59%) are non-backfilled and 501,532 (41%) are backfilled. Using our method, the database has 516,638 returns where 485,772 (94%) are non-backfilled and only 6% are backfilled. Even more importantly, our method retains almost all of the original non-backfilled returns (91%). Using the 24 month truncation method still leaves 30% of all returns backfilled while retaining only 83% of the original non-backfilled returns. Looking at TASS data leads to similar conclusions.

Finally, we apply our method to another database, BarclayHedge, which does not have an add date field, to demonstrate its usefulness. Figure 6 plots the returns around our generated add dates.

<Insert Figure 6 about here>

The pattern is similar to HFR and TASS, with a large drop in performance right after our generated listing dates.

Next, Table 11 evaluates the effect of the truncation rule on empirical conclusions using BarclayHedge. Panel A displays the percentage of backfilled returns retained using the truncation methods. Panel B shows the average returns using different cutoffs. The last panels describe the relation between performance and age as well as performance and size. These tests are run in a similar manner to those reported earlier in the paper.

<Insert Table 11 about here>

We find similar patterns using BarclayHedge and our generated listing dates as in HFR and TASS with the actual listing date. The usual method of truncating 12 or 24 returns only reduces the percentage of backfill returns by 30% at most. The backfill bias of 3.3%, which is economically and statistically significant, is similar to HFR and TASS. Finally, we also find that the relation between performance and age as well as size is seriously biased when ignoring backfill, or using ad hoc truncation corrections. Using our generated add dates, we find that these two relations disappear, which is similar to our findings with TASS and HFR.

Overall, we document that our method can be used to generate add dates for any database with sequential identification numbers. This should allow researchers to better control for backfill bias once and for all.

VI. Conclusion

Hedge funds are private investment vehicles and, as such, are basically banned from general advertising by the Securities and Exchange Commission. This includes public performance reporting, except through commercial hedge fund databases. As a result, the decision to report to one or several databases is voluntary. This creates several biases that cause major problems when trying to evaluate the performance of hedge funds. In particular, backfill is the bane of empirical research on hedge funds as it is difficult to correct for, especially since TASS now stopped reporting the listing date.

This paper provides a systematic assessment of the impact of ignoring backfill, or using ad hoc approximations, on the empirical performance of individual hedge funds. We show that backfill is systematically associated with abnormally high returns, mostly due to luck, and that the usual method of truncating a fixed number of months at the beginning of the return series fails to properly correct for the backfill problem. As expected, the bias mostly shows up when looking at variables that are related to the early life of a hedge fund return. In particular, the negative associations between fund return and both age and size largely disappear when properly controlling for backfill. A more subtle effect is that the relations between fund returns and fees and liquidity terms also disappear outside of the backfilled period. This is due to the fact they are correlated with backfilled performance.

Finally, we provide researchers with new tools that allow more precise backfill adjustments. First, we show that TASS stopped reporting the listing date after May 2011, but provide data that can help patch this problem. Second, we describe a new algorithm that seems to provide an excellent approximation to this listing date when unknown. Thus, researchers can now apply listing date information to any hedge fund database, which is invaluable when multiple databases are used.

References

Ackermann, Carl; McEnally, Richard; and David Ravenscraft. 1999. “The Performance of Hedge Funds: Risk, Return, and Incentives.” *Journal of Finance* 54: 833-874.

Agarwal, Vikas and Narayan Naik. 2000. “Multi-Period Performance Persistence Analysis of Hedge Funds.” *Journal of Financial and Quantitative Analysis* 35: 327—342.

Agarwal, Vikas; Daniel, Naveen; and Narayan Naik. 2009. “Role of Managerial Incentives and Discretion in Hedge Fund Performance.” *Journal of Finance*, 64: 2221–2256.

Agarwal, Vikas; Vyachelav Fos; and Wei Jiang. 2013. “Inferring Reporting Biases in Hedge Fund Databases from Hedge Fund Equity Holdings.” *Management Science*, 59: 1271–1289.

Agarwal, Vikas; Clifton Green; and Honglin Ren. 2017. “Alpha or Beta in the Eye of the Beholder: What Drives Hedge Fund Flows?” Forthcoming, *Journal of Financial Economics*.

Aggarwal, Raj and Philippe Jorion. 2010a. “The Performance of Emerging Hedge Funds and Managers.” *Journal of Financial Economics*, 96: 238–256.

Aggarwal, Raj and Philippe Jorion. 2010b. “Hidden Survivorship in Hedge Fund Returns.” *Financial Analysts Journal*, 66: 69–74.

Aiken, Adam; Christopher Clifford; and Jesse Ellis. 2013. “Out of the Dark: Hedge Fund Reporting Biases and Commercial Databases.” *Review of Financial Studies*, 26: 208–243.

Aragon, George; Bing Liang; and Hyuna Park. 2014. “Onshore and Offshore Hedge Funds: Are they Twins?” *Management Science*, 60: 74-91.

Avramov, Don; Kosowski, Robert; Naik, Narayan; and Melvyn Teo. 2011. “Hedge Funds, Managerial Skill, and Macroeconomic Variables.” *Journal of Financial Economics*, 99: 672–692.

Baquero, Guillermo; Jenke ter. Horst; and Marno Verbeek. 2005. “Survival, Look-Ahead Bias and the Persistence in Hedge Fund Performance.” *Journal of Financial and Quantitative Analysis*, 40: 493–517.

Bennin, Robert. 1980. “Error Rates in CRSP and COMPUSTAT: A Second Look.” *Journal of Finance*, 35: 1267–71.

Bhardwaj, Geetesh; Gorton Gary; and Geert Rouwenhorst. 2014. “Fooling Some of the People All of the Time: The Inefficient Performance and Persistence of Commodity Trading Advisors.” *Review of Financial Studies*, 27: 3099–3132.

Brown, Stephen; William Goetzmann; and James Park. 2001. “Careers and Survival: Competition and Risk in the Hedge Fund and CTA Industry.” *Journal of Finance*, 56: 1869–1886.

Canina, Linda; Roni Michaely; Richard Thaler; and Kent Womack. 1998. “Caveat Compounder: A Warning About using the Daily CRSP Equal-Weighted Index to Compute Long-run Excess Returns.” *Journal of Finance,* 53: 403–416.

Capocci, Daniel; Albert Corhay; and Georges Hübner. 2005. “Hedge Fund Performance and Persistence in Bull and Bear Markets.” *European Journal of Finance*, 11: 361-392.

Carhart, Mark. 1997. “On Persistence in Mutual Fund Performance.” *Journal of Finance,* 52: 57-82.

Elton, Edwin, Martin Gruber, and Christopher Blake. 2001. “A First Look at the Accuracy of the CRSP Mutual Fund Database and A Comparison of the CRSP and Morningstar Mutual Fund Databases.” *Journal of Finance,* 56: 2415–2430.

Evans, Richard. 2010. “The Incubation Bias.” *Journal of Finance,* 65: 1581-1611.

Fung, William and David Hsieh. 1997. “Empirical Characteristics of Dynamic Trading Strategies: the Case of Hedge Funds.” *Review of Financial Studies,* 10: 275–302.

Fung, William and David Hsieh. 2000. “Performance Characteristics of Hedge Funds and Commodity Funds: Natural vs. Spurious Biases.” *Journal of Financial and Quantitative Analysis,* 35: 291—307.

Fung, William and David Hsieh. 2001. “The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers.” *Review of Financial Studies,* 14: 313—341.

Fung, William and David Hsieh. 2004. “Hedge Fund Benchmarks: A Risk-Based Approach.” *Financial Analysts Journal*, 60: 65–80.

Fung, William and David Hsieh. 2009. “Measurement Biases in Hedge Fund Performance Data: An Update.” *Financial Analysts Journal*, 65: 36–38.

Getmansky, Mila. 2012. “The Life Cycle of Hedge Funds: Fund Flows, Size, Competition, and Performance.” *Quarterly Journal of Finance,* 2: 1— 53.

Getmansky, Mila; Lo, Andrew; and Igor Makarov. 2004. “An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns.” *Journal of Financial Economics,* 74: 529—609.

Jagannathan, Ravi; Alexey Malakhov; and Dmitry Novikov. 2010. “Do Hot Hands Persist Among Hedge Fund Managers? An Empirical Evaluation.” *Journal of Finance*, 65: 217–255.

Joenvaara, Juha; Robert Kosowski; and Pekka Tolonen. 2016. “Hedge Fund Performance: What Do We Know?” Imperial College Working Paper.

Jorion, Philippe and Christopher Schwarz. 2014a. “The Strategic Listing Decisions of Hedge Funds.” *Journal of Financial and Quantitative Analysis*, 49: 773–796.

Jorion, Philippe and Christopher Schwarz. 2014b. "The Delisting Bias in Hedge Fund Databases." *Journal of Alternative investments,* 16: 37–47.

Kosowski, Robert; Narayan Naik; and Melvyn Teo. 2007. "Do Hedge Funds Deliver Alpha? A Bayesian and Bootstrap Analysis." *Journal of Financial Economics*, 84: 229–264.

Liang, Bing. 2000. "Hedge Funds: The Living and the Dead." *Journal of Financial and Quantitative Analysis,* 35: 309–326.

Ljungqvist, Alexander; Christopher Malloy; and Felicia Marston. 2009. "Rewriting History." *Journal of Finance,* 64:1935–60.

Malkiel, Burton and Anatu Saha. 2005. "Hedge Funds: Risk and Return." *Financial Analysts Journal,* 61: 80—88.

Park, James. 1995. *Managed Futures As An investment Set*. Ph.D Dissertation, Columbia University.

Patton, Andrew; Ramadorai, Tarun; and Michael Steatfield. 2015. "Change You Can Believe In? Hedge Fund Data Revisions." *Journal of Finance,* 70: 963–999.

Rosenberg, Barr and Michel Houglet. 1974. "Error Rates in CRSP and Compustat Data Bases and their Implications." *Journal of Finance,* 29: 1303–1310.

Schwarz, Christopher and Mark Potter. 2016. "Revisiting Mutual Fund Portfolio Disclosure." *Review of Financial Studies,* 29: 3519–3544.

Shumway, Tyler. 1997. "The Delisting Bias in CRSP Data." *Journal of Finance,* 52: 327–40.

Teo, Melvyn. 2009. "Does Size Matter in the Hedge Fund Industry?" Singapore Management University Working Paper.

Yin, Chengdong. 2016. "The Optimal Size of Hedge Funds: Conflict between Investors and Fund Managers." *Journal of Finance,* 71: 1857–1894.

Figure 1: Distribution of Backfill Periods

This figure plots the cumulative distribution of backfill periods for funds in our sample based on the combined HFR and TASS databases. The backfill period is defined as the difference between the fund's first return date and the date added to the database. TASS add dates after May 2011 were reconstructed as described in the text. Backfill periods in months are on the horizontal axis. The frequency is on the left scale. The figure also shows average returns in excess of style returns over the fund's entire life, annualized with units on the right scale.

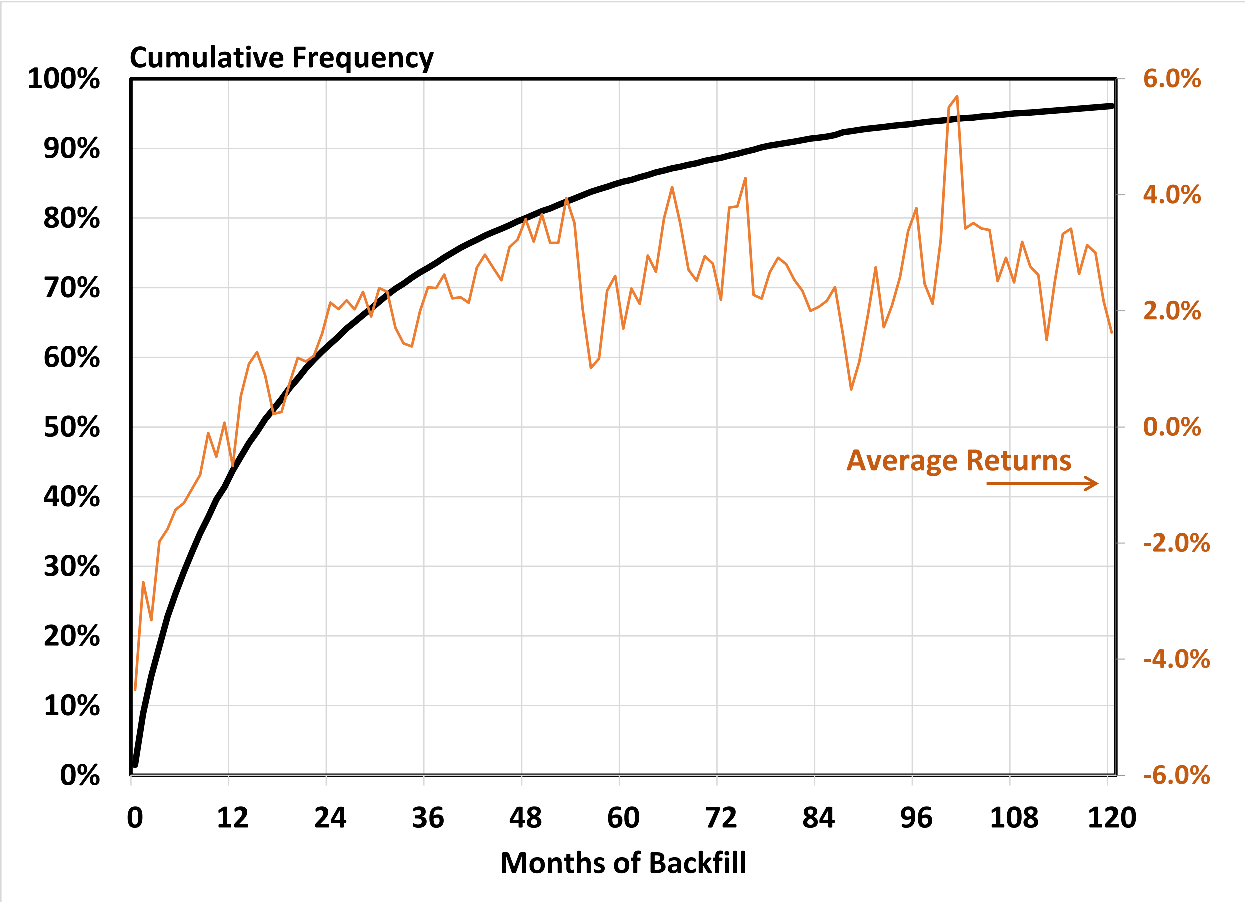


Table 1: Backfill using Various Cutoffs

This table describes the amount of backfill in our hedge fund databases. Panel A reports the percentage of returns (or funds) that are still backfilled using various return cutoff periods. Panel B reports the amount of backfill by calendar year using no cutoff.

Panel A: Backfill using Various Cutoffs

Cutoff	TASS Reported		TASS Reconstructed		HFR		Combined	
# months	Returns	Funds	Returns	Funds	Returns	Funds	Returns	Funds
None	35%	83%	43%	98%	41%	98%	40%	98%
3	33%	73%	41%	86%	38%	83%	38%	82%
6	32%	65%	39%	77%	37%	74%	36%	73%
12	29%	55%	37%	65%	34%	62%	33%	60%
18	27%	48%	34%	57%	32%	54%	31%	52%
24	26%	43%	32%	52%	30%	48%	29%	47%
30	24%	39%	31%	47%	29%	43%	28%	43%
36	23%	36%	29%	44%	27%	41%	26%	40%

Panel B: By Year

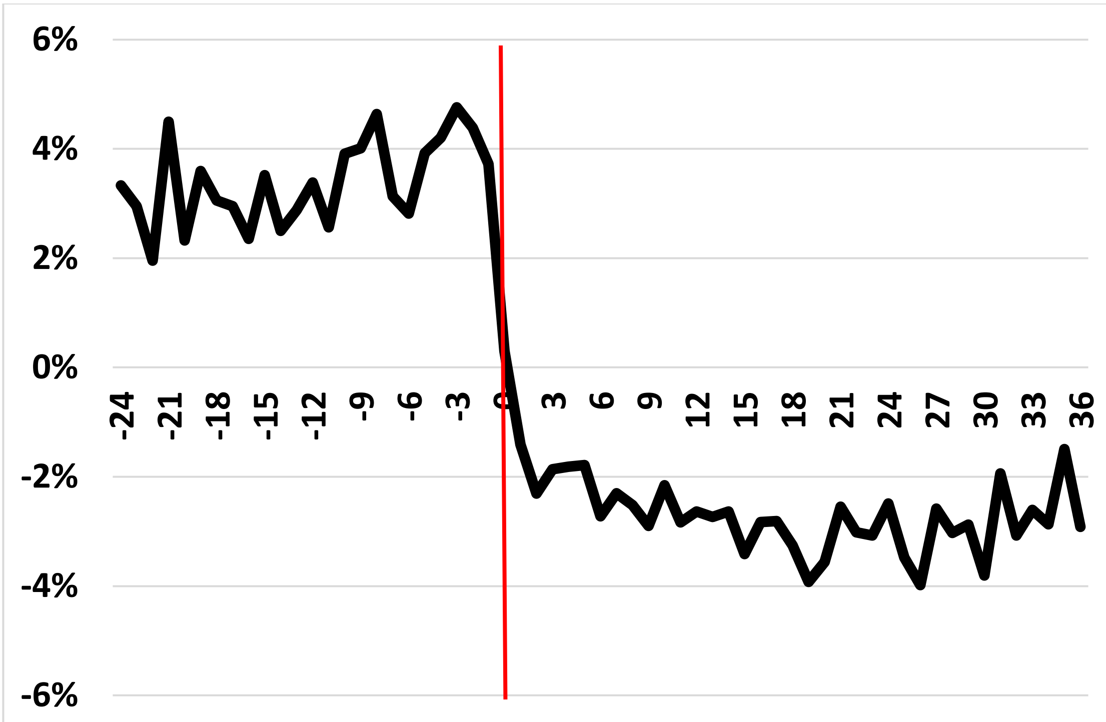
Year	TASS	HFR	Combined
1994	98%	100%	99%
1995	87%	100%	93%
1996	75%	64%	63%
1997	68%	41%	45%
1998	63%	43%	45%
1999	61%	47%	47%
2000	66%	51%	51%
2001	57%	54%	49%
2002	51%	51%	45%
2003	51%	52%	47%
2004	50%	48%	45%

Year	TASS	HFR	Combined
2005	47%	44%	41%
2006	46%	40%	39%
2007	47%	36%	38%
2008	34%	35%	35%
2009	32%	36%	35%
2010	27%	33%	31%
2011	20%	30%	28%
2012	16%	28%	26%
2013	13%	23%	22%
2014	9%	13%	13%
2015	2%	5%	5%

Figure 2: Returns and Assets around Databases Add Dates

Panel A plots the average returns in relation to the date the fund was added to the database. Panel B plots the average assets in millions of dollars around that date. Month 0 is the month the fund was added. Returns are style excess returns and annualized. The sample includes TASS and HFR, using the first add date for funds in common.

Panel A: Returns



Panel B: Assets

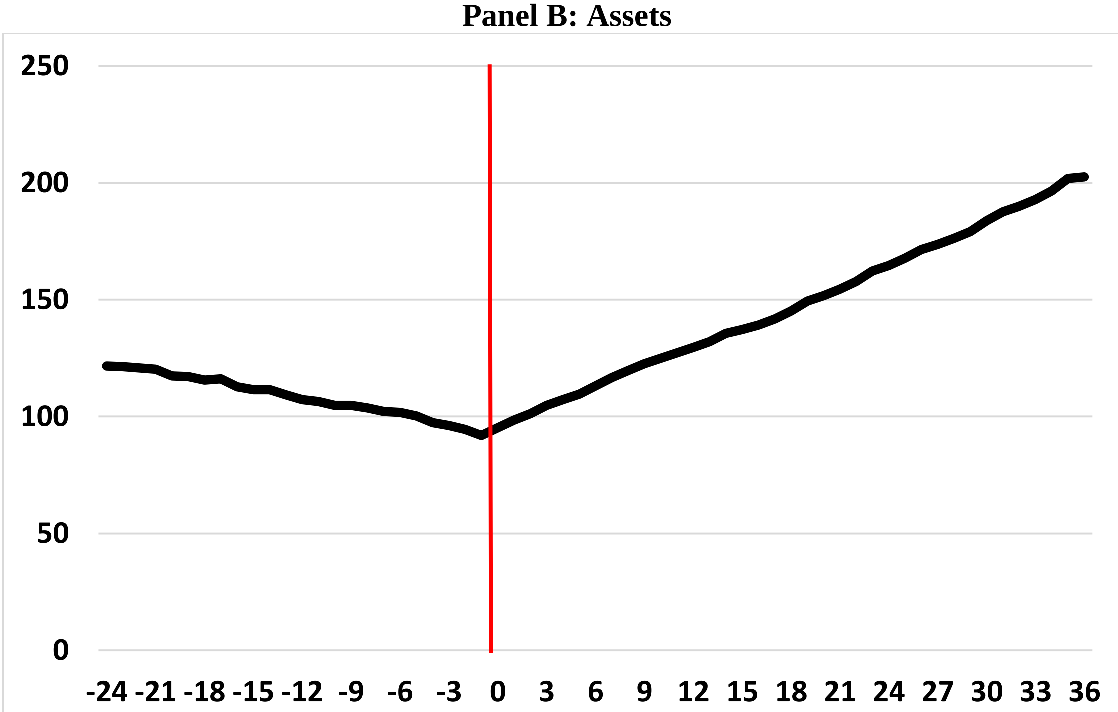


Table 2: Backfill and Average Returns

This table reports average returns each year using all returns, various return cutoffs, and using the add dates. In Panel A, for TASS, we display returns using the reported add dates as well as our constructed add dates. Bias is the difference between the average for all returns and for the reconstructed series. Panels B and C display returns for HFR and our combined dataset respectively. In the final row, we report the p-values of the difference between the average return in that column and the return after the listing date.

Panel A: TASS

Year	All Returns	Cut Offs			Add Dates		Bias
		1-year	2-years	3-years	Reported	Recon-structed	All – Recon.
1994	1.42%	-0.43%	-0.43%	-0.38%	-2.59%	-16.36%	17.78%
1995	21.54%	19.82%	19.80%	20.43%	13.98%	8.79%	12.75%
1996	20.83%	19.18%	17.81%	17.77%	14.12%	13.52%	7.31%
1997	19.60%	17.23%	16.38%	15.88%	13.96%	14.05%	5.55%
1998	5.76%	3.42%	3.01%	2.24%	-2.92%	-3.46%	9.22%
1999	28.01%	25.97%	24.67%	23.33%	25.57%	25.65%	2.36%
2000	10.87%	8.28%	8.15%	7.91%	1.96%	1.02%	9.84%
2001	7.55%	5.90%	5.15%	4.87%	2.35%	1.86%	5.69%
2002	3.77%	2.64%	1.93%	1.83%	1.09%	0.78%	3.00%
2003	20.58%	19.93%	20.36%	20.60%	18.70%	18.18%	2.39%
2004	9.20%	8.91%	8.58%	8.68%	7.97%	7.59%	1.61%
2005	10.03%	9.18%	8.87%	8.83%	8.38%	8.13%	1.90%
2006	14.17%	13.57%	13.15%	13.15%	12.66%	12.47%	1.71%
2007	16.24%	15.72%	15.64%	16.06%	10.93%	10.17%	6.07%
2008	-14.78%	-15.96%	-16.14%	-15.24%	-18.63%	-19.00%	4.22%
2009	21.98%	21.51%	21.68%	21.52%	20.86%	19.43%	2.54%
2010	11.67%	11.48%	11.15%	11.04%	11.25%	10.50%	1.17%
2011	-5.00%	-5.41%	-5.68%	-5.91%	-5.02%	-6.14%	1.13%
2012	7.14%	6.98%	7.00%	6.84%	7.14%	6.16%	0.98%
2013	10.55%	10.20%	10.83%	10.95%	10.53%	9.35%	1.20%
2014	2.45%	2.38%	2.41%	2.46%	2.45%	1.97%	0.48%
2015	-0.07%	-0.16%	-0.19%	-0.18%	-0.06%	-0.17%	0.10%
Avg.	10.16%	9.11%	8.82%	8.76%	7.03%	5.66%	4.5%
Cum.	767%	622%	588%	581%	406%	299%	468%
p-value of Diff.	0.00	0.00	0.00	0.00	0.04		

Panel B: HFR

Year	All Returns	Cut Offs			Add Dates	Bias
		1-year	2-years	3-years		
1996	21.94%	20.23%	18.62%	17.95%	8.96%	12.98%
1997	18.72%	17.31%	16.27%	15.80%	15.54%	3.18%
1998	6.61%	3.89%	3.90%	2.61%	2.28%	4.33%
1999	29.67%	28.44%	27.13%	26.00%	26.84%	2.83%
2000	12.15%	9.74%	8.83%	7.76%	3.20%	8.94%
2001	8.34%	6.63%	5.81%	5.56%	3.10%	5.24%
2002	3.55%	2.04%	1.36%	1.08%	-1.25%	4.80%
2003	21.07%	20.43%	20.79%	20.69%	18.02%	3.05%
2004	9.84%	9.28%	8.93%	9.04%	7.33%	2.51%
2005	9.82%	9.09%	8.55%	8.48%	7.60%	2.22%
2006	13.45%	12.67%	12.38%	12.34%	11.43%	2.01%
2007	13.14%	11.84%	10.88%	10.52%	10.17%	2.97%
2008	-15.27%	-16.92%	-17.81%	-18.14%	-19.28%	4.01%
2009	24.17%	23.22%	22.87%	22.98%	20.45%	3.72%
2010	12.06%	11.74%	11.32%	11.02%	10.10%	1.96%
2011	-2.84%	-3.41%	-3.94%	-4.06%	-4.85%	2.01%
2012	7.91%	7.66%	7.53%	7.29%	6.38%	1.53%
2013	10.11%	9.83%	10.03%	9.94%	8.82%	1.29%
2014	3.66%	3.31%	3.21%	3.10%	2.73%	0.92%
2015	-1.25%	-1.44%	-1.69%	-1.96%	-1.65%	0.40%
Avg.	10.34%	9.28%	8.75%	8.40%	6.80%	3.5%
Cum.	657%	541%	491%	460%	342%	215%
p-value of Diff.	0.00	0.00	0.00	0.00		

Panel C: HFR + TASS Combined

Year	All Returns	Cut Offs			Add Dates	2nd Add Date
		1-year	2-years	3-years		
1994	3.50%	1.88%	1.66%	1.45%	-16.36%	-17.18%
1995	21.05%	20.47%	20.19%	20.02%	6.27%	8.03%
1996	20.98%	19.29%	17.90%	17.51%	14.52%	12.65%
1997	18.70%	17.03%	15.86%	15.31%	15.21%	13.36%
1998	6.27%	3.77%	3.61%	2.67%	1.04%	-1.46%
1999	27.85%	26.21%	24.94%	23.64%	25.45%	24.16%
2000	11.41%	9.05%	8.52%	7.80%	3.35%	0.27%
2001	7.84%	6.30%	5.66%	5.42%	2.87%	1.52%
2002	3.74%	2.48%	1.86%	1.69%	0.04%	-0.77%
2003	20.53%	19.83%	20.26%	20.38%	17.83%	17.14%
2004	9.49%	9.03%	8.72%	8.81%	7.47%	6.80%
2005	9.73%	8.97%	8.46%	8.40%	7.76%	7.06%
2006	13.60%	12.91%	12.55%	12.58%	11.97%	11.29%
2007	14.75%	13.69%	13.17%	13.40%	10.38%	9.35%
2008	-14.68%	-15.99%	-16.56%	-16.38%	-18.89%	-19.48%
2009	22.88%	22.09%	21.86%	21.92%	19.52%	18.82%
2010	11.86%	11.65%	11.34%	11.04%	10.41%	10.07%
2011	-3.59%	-4.11%	-4.51%	-4.75%	-5.40%	-5.74%
2012	7.57%	7.25%	7.16%	6.98%	6.19%	5.93%
2013	9.77%	9.53%	9.80%	9.73%	8.66%	8.41%
2014	3.39%	3.12%	3.05%	2.98%	2.59%	2.52%
2015	-1.01%	-1.17%	-1.34%	-1.57%	-1.34%	-1.36%
Avg.	10.26%	9.24%	8.82%	8.59%	5.89%	5.06%
Cum.	686%	541%	489%	463%	215%	165%
p-value of Diff.	0.00	0.00	0.01	0.01		0.00

Table 3: Size and Performance Relation

This table reports returns for various size quartiles. Each quarter (year), we sort funds into quartiles based on their reported assets within the style. We then calculate the average style adjusted performance over the next quarter (year). These returns are averaged over our sample period. Finally, we calculate the difference between the largest and smallest funds using Fama-MacBeth (1973) to calculate p-values. Panels A, C, and E report quarterly results for TASS, HFR, and our combined database respectively. Panels B, D, and F report yearly results for TASS, HFR, and our combined database respectively. In each panel, we report results using all returns, various fixed period return cutoffs, and add dates cutoffs.

Panel A: TASS 3-Month Returns

	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Largest	-0.34%	-0.22%	-0.19%	-0.10%	0.26%
2nd	-0.15%	-0.04%	0.01%	-0.06%	-0.14%
3rd	0.08%	0.04%	0.01%	0.07%	-0.18%
Smallest	0.43%	0.24%	0.19%	0.09%	-0.08%
Diff	-0.77%	-0.46%	-0.38%	-0.19%	0.34%
p-value	0.00	0.00	0.00	0.27	0.36

Panel B: TASS 12-Month Returns

	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Largest	-1.61%	-1.34%	-1.08%	-0.84%	0.07%
2nd	-0.38%	-0.34%	-0.17%	-0.27%	0.62%
3rd	0.27%	0.31%	0.27%	0.10%	-0.64%
Smallest	2.05%	1.62%	1.19%	1.21%	-0.49%
Diff	-3.66%	-2.96%	-2.27%	-2.05%	0.56%
p-value	0.00	0.00	0.00	0.01	0.69

Panel C: HFR 3-Month Returns

	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Largest	-0.36%	-0.25%	-0.20%	-0.13%	-0.01%
2nd	-0.20%	-0.07%	-0.01%	-0.07%	0.01%
3rd	0.08%	0.05%	0.02%	0.03%	0.02%
Smallest	0.52%	0.30%	0.22%	0.20%	-0.03%
Diff	-0.88%	-0.55 %	-0.42%	-0.33%	0.02%
p-value	0.00	0.00	0.01	0.02	0.82

Panel D: HFR 12-Month Returns

	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Largest	-1.40%	-1.18%	-1.02%	-0.87%	-0.61%
2nd	-0.79%	-0.51%	-0.38%	-0.56%	-0.48%
3rd	0.36%	0.33%	0.10%	-0.07%	0.43%
Smallest	2.18%	1.66%	1.61%	1.88%	0.92%
Diff	-3.58%	-2.84%	-2.63%	-2.75%	-1.53%
p-value	0.00	0.00	0.01	0.01	0.10

Panel E: Combined 3-Month Returns

	All Returns	Cut Offs			Add Dates	2nd Add Date
		12 months	24 months	36 months		
Largest	-0.45%	-0.31%	-0.26%	-0.20%	0.12%	0.09%
2nd	-0.20%	-0.07%	-0.01%	-0.08%	-0.13%	-0.13%
3rd	0.10%	0.04%	0.02%	0.09%	-0.02%	-0.30%
Smallest	0.57%	0.36%	0.27%	0.20%	0.09%	0.22%
Diff	-1.02%	-0.67%	-0.53%	-0.40%	-0.03%	-0.13%
p-value	0.00	0.00	0.00	0.01	0.94	0.76

Panel F: Combined 12-Month Returns

	All Returns	Cut Offs			Add Dates	2nd Add Date
		12 months	24 months	36 months		
Largest	-1.86%	-1.49%	-1.28%	-1.09%	-1.00%	-0.31%
2nd	-0.96%	-0.81%	-0.75%	-0.61%	1.57%	0.13%
3rd	0.62%	0.80%	0.45%	0.08%	0.62%	1.57%
Smallest	2.58%	1.82%	1.86%	1.90%	-1.01%	-0.49%
Diff	-4.44%	-3.32%	-3.14%	-2.99%	-0.01%	-0.18%
p-value	0.00	0.00	0.01	0.01	0.98	0.92

Table 4: Age and Performance Relation

This table reports returns for various age quartiles. Each quarter (year), we sort funds into quartiles based on their age within the style. We then calculate the average style adjusted performance over the next quarter (year). We average these returns over our sample period. Finally, we calculate the difference between the oldest and smallest funds, using Fama-MacBeth (1973) to calculate p-values. Panels A, C, and E report quarterly youngest funds for TASS, HFR, and our combined database, respectively. Panels B, D, and F report yearly results for TASS, HFR, and our combined database, respectively. In each panel, we report results using all returns, various fixed period return cutoffs, and add dates cutoffs.

Panel A: TASS 3-Month Returns

	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Oldest	-0.39%	-0.22%	-0.21%	-0.19%	0.09%
2nd	-0.11%	0.06%	-0.07%	-0.04%	-0.18%
3rd	-0.16%	-0.04%	0.11%	0.11%	-0.15%
Youngest	0.69%	0.21%	0.18%	0.12%	0.15%
Diff	-1.08%	-0.43%	-0.39%	-0.31%	-0.06%
p-value	0.00	0.00	0.00	0.02	0.84

Panel B: TASS 12-Month Returns

	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Oldest	-1.00%	-1.05%	-0.79%	-0.73%	0.07%
2nd	-0.06%	0.03%	-0.02%	0.01%	2.00%
3rd	0.09%	0.06%	0.23%	-0.17%	-0.02%
Youngest	1.05%	1.10%	0.67%	0.98%	-1.43%
Diff	-2.05%	-2.15%	-1.46%	-1.61%	-1.50%
p-value	0.00	0.00	0.01	0.02	0.17

Panel C: HFR 3-Month Returns

	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Oldest	-0.39%	-0.21%	-0.12%	-0.07%	0.01%
2nd	-0.26%	-0.11%	0.01%	0.01%	0.02%
3rd	-0.03%	0.02%	-0.09%	0.06%	-0.08%
Youngest	0.76%	0.35%	0.24%	0.02%	0.05%
Diff	-1.15%	-0.56%	-0.36%	-0.09%	-0.04%
p-value	0.00	0.00	0.00	0.44	0.76

Panel D: HFR 12-Month Returns

	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Oldest	-1.16%	-0.82%	-0.52%	-0.76%	0.24%
2nd	-0.19%	0.09%	0.09%	-0.18%	0.34%
3rd	-0.02%	0.06%	0.46%	0.03%	-0.15%
Youngest	1.61%	0.90%	0.06%	0.85%	0.20%
Diff	-2.76%	-1.72%	-0.58%	-1.61%	-0.44%
p-value	0.00	0.00	0.36	0.03	0.42

Panel E: Combined 3-Month Returns

	All Returns	Cut Offs			Add Dates	2nd Add Date
		12 months	24 months	36 months		
Oldest	-0.39%	-0.18%	-0.13%	-0.08%	0.16%	0.24%
2nd	-0.25%	-0.06%	-0.04%	-0.08%	-0.30%	-0.28%
3rd	-0.09%	-0.04%	0.01%	0.17%	-0.07%	-0.37%
Youngest	0.77%	0.30%	0.17%	0.00%	0.14%	0.38%
Diff	-1.16%	-0.48%	-0.30%	-0.08%	-0.02%	-0.14%
p-value	0.00	0.00	0.01	0.43	0.57	0.77

Panel F: Combined 12-Month Returns

	All Returns	Cut Offs			Add Dates	2nd Add Date
		12 months	24 months	36 months		
Oldest	-1.02%	-0.72%	-0.40%	-0.73%	-0.07%	0.81%
2nd	-0.24%	-0.15%	-0.15%	-0.01%	-0.05%	1.24%
3rd	-0.13%	-0.06%	0.35%	-0.22%	0.55%	-2.23%
Youngest	1.47%	1.01%	0.24%	1.05%	-2.21%	-1.09%
Diff	-2.49%	-1.73%	-0.64%	-1.78%	-2.14%	1.90%
p-value	0.00	0.00	0.33	0.05	0.23	0.08

Table 5: Performance Persistence

This table reports returns for various past performance quartiles. Each quarter (year), we sort funds into quartiles based on prior quarter performance within the style. We then calculate the average style adjusted performance over the next quarter (year). We average these returns over our sample period. Finally, we calculate the difference between the best and worst performing funds, using Fama-MacBeth (1973) to calculate p-values. Panels A, C, and E report quarterly results for TASS, HFR, and our combined database, respectively. Panels B, D, and F report yearly results for TASS, HFR, and our combined database, respectively. In each panel, we report results using all returns, various fixed period return cutoffs, and add dates cutoffs.

Panel A: TASS 3-Month Returns

	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Highest	1.23%	0.99%	0.90%	0.80%	1.07%
2nd	-0.02%	0.06%	0.04%	-0.02%	0.06%
3rd	-0.57%	-0.48%	-0.43%	-0.38%	-0.50%
Lowest	-0.70%	-0.62%	-0.57%	-0.45%	-0.76%
Diff	1.93%	1.61%	1.47%	1.25%	1.83%
p-value	0.00	0.00	0.00	0.00	0.00

Panel B: TASS 12-Month Return

	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Highest	2.13%	1.78%	1.55%	1.48%	2.10%
2nd	-0.38%	-0.28%	-0.13%	-0.28%	-1.78%
3rd	-1.46%	-1.10%	-1.10%	-1.18%	-1.93%
Lowest	-0.68%	-0.80%	-0.74%	-0.32%	-1.02%
Diff	2.81%	2.58%	2.29%	1.80%	3.12%
p-value	0.21	0.24	0.30	0.41	0.19

Panel C: HFR 3-Month Returns

	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Highest	1.28%	1.13%	1.06%	0.95%	1.07%
2nd	-0.07%	-0.05%	-0.00%	-0.03%	0.04%
3rd	-0.54%	-0.45%	-0.42%	-0.36%	-0.40%
Lowest	-0.75%	-0.70%	-0.72%	-0.64%	-0.82%
Diff	2.03%	1.83%	1.78%	1.59%	1.89%
p-value	0.00	0.00	0.00	0.00	0.00

Panel D: HFR 12-Month Returns

	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Highest	2.47%	1.86%	1.60%	1.41%	1.28%
2nd	-0.85%	-0.75%	-0.84%	-0.63%	-0.73%
3rd	-1.71%	-1.30%	-1.05%	-1.19%	-1.13%
Lowest	-0.23%	-0.07%	0.04%	0.20%	0.33%
Diff	2.70%	1.93%	1.64%	1.21%	0.95%
p-value	0.37	0.51	0.59	0.69	0.76

Panel E: Combined 3-Month Returns

	All Returns	Cut Offs			Add Dates	2nd Add Date
		12 months	24 months	36 months		
Highest	1.31%	1.10%	1.01%	0.88%	1.07%	1.26%
2nd	-0.00%	0.02%	0.05%	0.00%	-0.01%	-0.18%
3rd	-0.55%	-0.46%	-0.45%	-0.37%	-0.46%	-0.43%
Lowest	-0.78%	-0.71%	-0.66%	-0.56%	-0.64%	-0.88%
Diff	2.09%	1.81%	1.67%	1.44%	1.71%	2.14%
p-value	0.00	0.00	0.00	0.00	0.00	0.00

Panel F: Combined 12-Month Return

	All Returns	Cut Offs			Add Dates	2nd Add Date
		12 months	24 months	36 months		
Highest	2.29%	1.84%	1.64%	1.58%	1.58%	2.74%
2nd	-0.78%	-0.75%	-0.68%	-0.80%	-1.17%	-1.77%
3rd	-1.52%	-1.10%	-0.96%	-0.86%	-1.51%	-3.08%
Lowest	-0.33%	-0.27%	-0.23%	-0.15%	-0.59%	-1.45%
Diff	2.62%	2.11%	1.87%	1.73%	2.17%	4.19%
p-value	0.38	0.48	0.53	0.58	0.51	0.23

Table 6: Fund Characteristics and Return Prediction

This table reports results from regressions where the dependent variable is the return in quarter (year) *t*. Independent variables include prior returns, size, age, as well as various fund characteristics. Overall coefficients and t-values are computed using Fama-MacBeth (1973). Panels A, C, and E report quarterly results for TASS, HFR, and our combined database, respectively. Panels B, D, and F report yearly results for TASS, HFR, and our combined database respectively. In each panel, we report results using all returns, various fixed period return cutoffs, and add dates cutoffs.

Panel A: TASS 3-Month Returns

	All Returns		12 months cut		24 months cut		36 months cut		Add Date	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Prior Return	0.079	3.27***	0.055	2.23**	0.041	1.59	0.028	1.01	0.092	2.10**
Log Assets	-0.002	-5.58***	-0.001	-3.38***	-0.001	-2.55**	-0.001	-1.72*	0.000	-0.18
Age	0.000	-1.87*	0.000	-1.11	0.000	-0.79	0.000	-1.32	0.000	-0.03
Min. Invt.	0.002	4.10***	0.001	3.00***	0.001	3.06***	0.001	2.36**	0.000	0.28
Mfee	0.224	3.17***	0.203	3.03***	0.197	2.70***	0.217	2.85***	0.045	0.43
Ifee	0.024	2.80***	0.017	1.83*	0.014	1.42	0.015	1.23	0.007	0.51
HWM	0.004	3.19***	0.003	2.94***	0.002	1.52	0.002	1.54	0.003	1.40
Red. Notice	0.017	2.53**	0.021	2.98***	0.020	2.29**	0.017	1.73*	0.048	1.33
Sub. Freq.	-0.006	-0.73	-0.002	-0.24	-0.010	-1.11	-0.007	-0.69	0.012	0.60
Red. Freq.	0.011	1.23	0.009	1.00	0.014	1.56	0.010	1.09	0.011	0.63
Lockup	0.002	1.72*	0.001	1.18	0.001	1.28	0.001	0.76	-0.002	-0.45
N	1289		1096		917		761		860	
Adj. R-sq.	5.75%		5.70%		5.88%		6.04%		6.54%	

Panel B: TASS 12-Month Returns

	All Returns		12 months cut		24 months cut		36 months cut		Add Date	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Prior Return	0.070	1.32	0.062	1.11	0.050	0.90	0.038	0.67	0.085	1.26
Log Assets	-0.008	-3.67***	-0.006	-2.96***	-0.005	-2.73**	-0.005	-2.73**	-0.004	-2.02*
Age	0.000	-0.57	-0.001	-1.97**	-0.001	-1.38	-0.001	-1.20	0.000	0.18
Min. Invt.	0.007	3.15***	0.006	3.15***	0.007	3.04***	0.005	2.10**	0.007	2.93***
Mfee	1.104	3.37***	0.957	2.77**	0.907	2.63**	0.717	1.92*	0.648	2.59**
Ifee	0.090	1.80*	0.062	1.21	0.068	1.31	0.048	0.92	0.078	0.93
HWM	0.013	2.21**	0.012	1.73*	0.008	1.28	0.010	1.96*	0.007	1.28
Red. Notice	0.074	2.57**	0.048	1.27	0.025	0.58	0.064	0.91	0.082	2.24**
Sub. Freq.	0.010	0.23	-0.001	-0.02	-0.026	-0.51	-0.067	-1.22	0.039	1.10
Red. Freq.	0.022	0.57	0.047	1.02	0.060	1.27	0.060	1.28	-0.013	-0.41
Lockup	0.004	0.76	0.001	0.22	0.002	0.41	0.002	0.28	0.013	1.26
N	1047		877		730		605		654	
Adj. R-sq.	7.47%		7.30%		7.04%		6.63%		8.38%	

Panel C: HFR 3-Month Returns

	All Returns		12 months cut		24 months cut		36 months cut		Add Date	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Prior Return	0.086	3.71***	0.069	2.77***	0.070	2.72***	0.068	2.51**	0.081	2.73***
Log Assets	-0.003	-6.83***	-0.002	-4.59***	-0.002	-3.84***	-0.001	-3.79***	-0.001	-1.46
Age	0.000	-2.01**	-0.000	-0.81	0.000	0.30	0.000	1.19	0.000	0.17
Min. Invt.	0.001	4.12***	0.001	2.54**	0.001	2.48**	0.001	2.15**	0.001	1.66
Mfee	0.002	2.17**	0.002	2.27**	0.001	1.38*	0.001	1.25	0.001	1.00
Ifee	0.000	2.14**	0.000	0.70	0.000	0.65	0.000	0.70	0.000	0.12
HWM	0.001	1.51	0.002	2.37**	0.002	2.11**	0.002	1.27	0.001	1.19
Red. Notice	0.012	1.61	0.010	1.20	0.008	1.05	0.010	1.09	0.008	0.11
Sub. Freq.	0.001	0.76	0.003	1.76*	0.003	1.45	0.000	0.13	0.005	2.40**
Red. Freq.	0.002	1.09	0.001	0.31	-0.001	-0.31	-0.001	-0.38	0.001	0.46
Lockup	0.038	2.34**	0.043	2.00**	0.045	2.14**	0.035	1.61	0.034	1.33
N	2260		1917		1605		1335		1604	
Adj. R-sq.	5.58%		5.64%		5.64%		6.01%		6.39%	

Panel D: HFR 12-Month Returns

	All Returns		12 months cut		24 months cut		36 months cut		Add Date	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Prior Return	0.043	0.78	0.038	0.68	0.016	0.29	0.012	0.22	0.025	0.43
Log Assets	-0.011	-4.37***	-0.009	-3.91***	-0.008	-3.47***	-0.008	-3.09***	-0.005	-2.51**
Age	0.000	0.03	0.000	0.32	0.000	0.71	-0.000	-0.29	0.000	0.08
Min. Invt.	0.003	2.09**	0.003	1.99*	0.003	1.83*	0.001	1.06	0.000	0.17
Mfee	0.009	1.76*	0.005	1.23	0.005	1.19	0.006	1.29	0.009	1.71
Ifee	0.001	1.76*	0.001	1.20	0.000	0.95	0.000	0.62	0.001	1.16
HWM	0.011	2.13**	0.013	1.98*	0.012	1.52	0.011	1.34	0.012	1.44
Red. Notice	0.035	0.77	0.001	0.01	0.003	0.05	0.022	0.40	-0.034	-0.65
Sub. Freq.	0.010	1.50	0.007	1.19	0.006	0.45	-0.002	-0.32	0.014	1.61
Red. Freq.	0.005	0.58	0.004	0.45	0.007	0.75	0.005	0.38	-0.002	-0.20
Lockup	0.204	1.40	0.271	1.67	0.124	0.92	0.134	1.36	0.157	1.34
N	1827		1531		1276		1058		1178	
Adj. R-sq.	7.06%		6.60%		6.16%		5.37%		6.02%	

Panel E: Combined 3-Month Return

	All Returns		12 months cut		24 months cut		36 months cut		1st Add Date		2nd Add Date	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Prior Return	0.091	4.03***	0.070	2.91***	0.062	2.50**	0.051	1.95*	0.089	1.76*	0.164	2.08**
Log Assets	-0.002	-6.23***	-0.002	-4.50***	-0.001	-3.61***	-0.001	-3.26***	-0.006	-0.88	0.000	0.39
Age	-0.000	-2.70***	-0.000	-0.72	-0.000	-0.11	0.000	0.27	-0.002	-0.64	0.000	-0.19
Min. Invt.	0.001	3.44***	0.001	2.71***	0.001	2.43**	0.001	2.02**	0.001	1.08	0.000	0.89
Mfee	0.002	2.10**	0.001	2.02**	0.001	1.25	0.001	1.16	0.000	0.08	0.001	0.61
Ifee	0.000	1.74*	0.000	0.43	0.000	0.30	0.000	0.54	-0.000	-0.95	0.000	-0.77
HWM	0.003	3.67***	0.004	4.49***	0.004	3.87***	0.003	2.31**	-0.001	-0.14	0.002	1.41
Red. Notice	0.011	1.74*	0.008	1.21	0.003	0.46	0.004	0.47	0.045	1.46	0.015	0.76
Sub. Freq.	0.003	1.45	0.005	2.09**	0.004	1.83*	0.002	0.62	0.023	1.45	-0.023	-0.71
Red. Freq.	0.003	1.35	0.002	0.71	0.001	0.45	0.001	0.53	-0.009	-1.05	0.003	0.30
Lockup	0.060	3.59***	0.066	6.16***	0.082	3.94***	0.080	3.69***	0.049	0.42	-0.153	-0.70
N	2817		2395		2010		1676		1946		1675	
Adj. R-sq.	5.60%		5.53%		5.61%		5.93%		6.56%		7.01%	

Panel F: Combined 12-Month Return

	All Returns		12 months cut		24 months cut		36 months cut		1st Add Date		2nd Add Date	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Prior Return	0.059	1.12	0.057	1.09	0.050	0.95	0.043	0.80	0.170	1.33	0.051	0.83
Log Assets	-0.010	-4.20***	-0.008	-3.51***	-0.007	-3.01***	-0.006	-2.80***	0.008	0.59	-0.004	-2.19**
Age	-0.000	-0.76	-0.000	-0.77	-0.000	-0.42	-0.001	-1.38	-0.007	-0.99	0.000	0.27
Min. Invt.	0.003	1.84*	0.003	1.76*	0.003	1.70	0.002	1.21	-0.004	-0.55	0.002	0.88
Mfee	0.007	1.53	0.005	1.18	0.004	1.05	0.004	0.86	0.027	1.20	0.005	1.15
Ifee	0.000	0.94	0.000	0.67	0.000	0.86	0.000	0.37	-0.001	-0.64	0.001	1.08
HWM	0.015	2.76***	0.015	2.21**	0.011	1.74*	0.013	2.00*	-0.014	-0.53	0.012	1.67
Red. Notice	0.024	0.60	-0.018	-0.41	-0.026	-0.53	-0.003	-0.07	0.016	0.33	0.063	1.20
Sub. Freq.	0.014	1.92*	0.007	0.97	0.005	0.64	0.003	0.59	-0.002	-0.05	-0.012	-0.51
Red. Freq.	0.011	1.28	0.010	1.32	0.014	1.69	0.012	1.06	0.139	1.08	0.021	1.04
Lockup	0.317	2.24**	0.417	2.57**	0.344	2.57**	0.329	2.89***	-2.068	-0.89	0.337	2.01*
N	2288		1923		1608		1340		1419		1211	
Adj. R-sq.	7.65%		7.33%		6.88%		6.65%		6.73`%		7.44%	

Table 7: Relation between Backfill Performance and Fund Terms

This table reports the cross-sectional relation between fund terms and backfill performance. We regress each variable separately against the average excess monthly return during the backfill period. We display the t-value of the slope coefficient, for the TASS, HFR, and our combined TASS/HFR dataset.

	TASS	HFR	Combined
	Fees		
Management Fee	0.23	3.26***	2.57***
Incentive Fee	3.38***	6.28***	8.24***
	Liquidity Restrictions		
Minimum Investment	2.41**	-0.53	2.68***
Notice Period	4.17***	4.30***	6.21***
Redemption Freq.	4.53***	2.56***	3.63***
Lockup Period	3.36***	1.94*	2.86***
Subscription Freq.	0.85	0.87	1.62

Table 8: Investor Net Flows and Fund Characteristics

This table reports results from regressions where the dependent variable is net investor flows in quarter (year) *t*. Independent variables include prior size, flows, and various fund characteristics. We also include two piecewise performance variables for returns over the past year, which should capture non-linear effects. *Low Perf. Rank* is the minimum of the fund's performance rank within its style and 50%. *High Perf. Rank* is the maximum of the fund's performance rank minus 50% and zero. Coefficients and t-values are computed using Fama-MacBeth (1973). Panels A, C, and E report quarterly results for TASS, HFR, and our combined database respectively. Panels B, D, and F report yearly results. In each panel, we report results using all returns, various fixed periods return cutoffs, and add dates cutoffs.

Panel A: TASS Quarterly Flows

	All Returns		12 months cut		24 months cut		36 months cut		Add Date	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Low Perf	0.188	15.48***	0.144	15.84***	0.124	12.27***	0.102	9.15***	0.161	11.58***
High Perf	0.129	10.44***	0.097	9.70***	0.076	7.67***	0.079	6.51***	0.098	7.37***
Prior Flows	0.051	7.28***	0.067	8.97***	0.077	9.63***	0.089	9.63***	0.102	8.52***
Style Flow	0.923	57.74***	0.927	56.67***	0.947	56.87***	0.935	49.36***	0.940	60.45***
Assets	-0.024	-16.53***	-0.011	-11.92***	-0.008	-9.12***	-0.006	-6.80***	-0.011	-8.58***
Min. Invt.	0.012	10.93***	0.006	6.80***	0.004	4.73***	0.003	3.51***	0.006	3.51***
Ifee	0.467	2.42**	0.035	0.25	-0.056	-0.42	-0.084	-0.56	-0.006	-0.03
Mfee	-0.001	-0.06	0.006	0.41	-0.002	-0.14	-0.021	-1.29	0.006	0.24
HWM	0.022	6.34***	0.014	5.56***	0.010	4.23***	0.010	4.49***	0.015	2.42**
Red. Notice	0.140	7.06***	0.077	4.65***	0.063	3.44***	0.059	3.46***	0.148	2.71***
Sub. Freq.	-0.152	-5.59***	-0.066	-3.00***	-0.062	-2.69***	-0.049	-2.08**	-0.064	-2.38**
Red. Freq.	-0.043	-2.31**	-0.008	-0.47	0.014	0.73	0.005	0.26	-0.009	-0.42
Lockup	-0.002	-0.81	0.001	0.43	0.000	0.12	0.002	0.52	0.003	0.65
N	1210		1026		857		711		811	
Adj R-sq	8.98%		8.00%		8.06%		8.67%		9.86%	

Panel B: TASS Annual Flows

	All Returns		12 months cut		24 months cut		36 months cut		Add Date	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Low Perf	0.889	6.66***	0.716	6.73***	0.599	6.51***	0.558	7.41***	0.397	3.01***
High Perf	0.614	5.23***	0.525	4.62***	0.514	5.73***	0.456	6.73***	0.698	4.96***
Prior Flows	0.018	3.49***	0.039	3.32***	0.065	4.11***	0.085	3.47***	0.055	3.67***
Style Flow	0.923	30.60***	0.972	25.38***	0.884	21.85***	0.915	14.25***	0.943	35.42***
Assets	-0.119	-9.14***	-0.086	-8.17***	-0.067	-7.27***	-0.057	-6.54***	-0.088	-5.74***
Min. Invt.	0.056	7.26***	0.043	5.93***	0.037	5.66***	0.032	3.98***	0.053	4.04***
Ifee	0.407	0.24	-0.567	-0.41	-0.910	-0.67	-0.885	-0.86	-1.421	-0.79
Mfee	0.026	0.18	-0.065	-0.56	-0.043	-0.30	-0.145	-0.94	-0.013	-0.10
HWM	0.067	4.31***	0.071	4.37***	0.070	4.58***	0.084	4.57***	0.046	1.43
Red. Notice	0.368	3.15***	0.367	3.59***	0.361	3.10***	0.371	3.62***	0.350	2.44**
Sub. Freq.	-0.586	-3.12***	-0.434	-2.61**	-0.377	-2.20**	-0.272	-1.85*	-0.240	-1.18
Red. Freq.	-0.222	-2.56**	-0.072	-0.60	-0.124	-0.85	-0.137	-0.87	-0.103	-0.90
Lockup	0.018	0.50	-0.031	-2.00*	-0.028	-1.77*	-0.042	-2.72**	0.094	0.84
N	924		773		642		532		579	
Adj R-sq	13.34%		13.50%		13.68%		14.01%		15.88%	

Panel C: HFR Quarterly Flows

	All Returns		12 months cut		24 months cut		36 months cut		Add Date	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Low Perf	0.178	15.60***	0.139	16.93***	0.124	15.26***	0.115	13.06***	0.142	14.08***
High Perf	0.120	10.56***	0.088	10.49***	0.078	9.92***	0.068	8.28***	0.108	10.21***
Prior Flows	0.033	7.37***	0.042	7.96***	0.042	7.98***	0.048	8.86***	0.053	8.54***
Style Flow	0.891	76.01***	0.926	94.05***	0.937	101.36***	0.942	89.94***	0.934	93.64***
Assets	-0.022	-16.62***	-0.012	-13.72***	-0.009	-10.51***	-0.007	-9.02***	-0.013	-11.37***
Min. Invt.	0.007	9.50***	0.004	7.87***	0.003	7.01***	0.003	5.79***	0.005	6.26***
Ifee	-0.001	-0.65	-0.000	-0.58	-0.000	-1.95*	-0.000	-1.97*	-0.000	-0.41
Mfee	0.007	4.01***	0.003	1.71*	0.003	0.23	0.000	0.04	0.005	3.16***
HWM	0.003	0.86	0.002	0.69	0.000	0.11	-0.001	-0.33	0.007	1.89*
Red. Notice	0.123	6.80***	0.055	4.28***	0.030	2.31**	0.026	2.18**	0.079	5.00***
Sub. Freq.	-0.019	-6.56***	-0.012	-4.94***	-0.010	-5.38***	-0.009	-3.46***	-0.012	-3.33***
Red. Freq.	-0.030	-6.59***	-0.015	-3.64***	-0.007	-1.88*	-0.007	-1.63	-0.019	-3.67***
Lockup	-0.093	-2.38**	-0.040	-1.34	-0.055	-1.52	-0.072	-2.08**	-0.055	-1.48
N	2436		2070		1737		1448		1571	
Adj R-sq	7.62%		6.98%		6.96%		7.48%		7.42%	

Panel D: HFR Annual Flows

	All Returns		12 months cut		24 months cut		36 months cut		Add Date	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Low Perf	0.907	10.70***	0.771	8.89***	0.674	9.70***	0.556	11.65***	0.549	7.96***
High Perf	0.582	6.81***	0.464	5.87***	0.491	6.86***	0.501	7.79***	0.643	7.80***
Prior Flows	0.015	3.02***	0.027	3.03***	0.030	3.58***	0.038	3.41***	0.014	2.98***
Style Flow	0.858	39.61***	0.912	46.92***	0.896	34.34***	0.946	20.85***	0.934	36.11***
Assets	-0.124	-10.52***	-0.093	-9.95***	-0.074	-8.60***	-0.062	-6.83***	-0.073	-7.08***
Min. Invt.	0.042	7.53***	0.030	6.59***	0.024	6.22***	0.022	6.18***	0.027	4.86***
Ifee	-0.002	-1.31	-0.003	-3.31***	-0.002	-2.22**	-0.002	-1.44	-0.000	-0.44
Mfee	0.015	1.22	0.006	0.68	0.005	0.42	-0.004	-0.40	0.022	1.54
HWM	0.025	0.90	0.014	0.47	-0.005	-0.23	-0.017	-0.75	0.021	0.79
Red. Notice	0.411	3.19***	0.272	2.69**	0.306	2.24**	0.246	1.76*	0.163	2.16**
Sub. Freq.	-0.086	-2.79**	-0.109	-3.54***	-0.084	-1.38	-0.117	-2.08**	-0.068	-5.14***
Red. Freq.	-0.153	-4.40***	-0.104	-4.09***	-0.075	-2.59**	-0.041	-1.75*	-0.105	-4.35***
Lockup	-0.476	-1.77*	-0.060	-0.15	-0.489	-1.51	-0.877	-3.33***	-0.328	-0.96
N	1745		1460		1215		1007		1152	
Adj R-sq	12.92%		13.24%		13.77%		16.10%		12.39%	

Panel E: Combined Quarterly Flows

	All Returns		12 months cut		24 months cut		36 months cut		1st Add Date		2nd Add Date	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Low Perf	0.184	17.24***	0.147	17.67***	0.130	15.19***	0.122	14.40***	0.161	14.60***	0.174	6.58***
High Perf	0.125	11.67***	0.088	10.67***	0.074	8.52***	0.068	6.94***	0.108	8.95***	0.072	2.55**
Prior Flow	0.030	6.91***	0.037	6.75***	0.041	6.79***	0.049	7.36***	0.048	8.05***	0.050	5.74***
Style Flow	0.949	68.79***	0.936	64.14***	0.934	55.45***	0.953	58.67***	0.944	68.61***	0.942	34.60***
Assets	-0.024	-17.31***	-0.013	-14.32***	-0.010	-11.10***	-0.008	-9.39***	-0.013	-12.48***	-0.013	-7.92***
Min. Invt.	0.008	11.55***	0.005	10.02***	0.004	8.96***	0.003	7.35***	0.005	6.45***	0.005	3.08***
Ifee	-0.000	-0.74	-0.000	-1.49	-0.000	-2.63**	-0.000	-3.61***	-0.000	-1.69	-0.001	-1.96*
Mfee	0.005	3.10***	0.002	1.27	0.000	0.17	-0.000	-0.12	0.003	1.98*	-0.000	-0.03
HWM	0.014	5.04***	0.011	4.54***	0.007	3.17***	0.007	2.98***	0.018	4.89***	0.013	3.76***
Red. Notice	0.110	7.23***	0.068	3.31***	0.011	0.96	-0.001	-0.09	0.069	3.87***	0.253	1.97*
Sub. Freq.	-0.013	-3.17***	-0.007	-1.33	-0.010	-2.36**	-0.009	-1.74*	-0.029	-3.04**	0.011	0.13
Red. Freq.	-0.029	-6.48***	-0.011	-2.96***	-0.001	-0.20	0.001	0.37	-0.016	-1.73*	-0.051	-0.54
Lockup	-0.036	-0.84	-0.001	-0.02	0.033	0.75	0.016	0.39	0.071	0.53	0.084	0.27
N	2773		2354		1974		1645		1936		1642	
Adj R-sq	6.45%		5.39%		5.23%		5.38%		6.16%		5.70%	

Panel F: Combined Annual Flows

	All Returns		12 months cut		24 months cut		36 months cut		1st Add Date		2nd Add Date	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Low Perf	0.927	9.93***	0.758	9.99***	0.660	9.76***	0.564	11.87***	0.622	8.65***	0.537	6.49***
High Perf	0.594	5.87***	0.523	6.02***	0.527	6.27***	0.510	7.29***	0.627	7.11***	0.624	5.88***
Prior Flow	0.013	3.42***	0.028	2.56**	0.034	3.18***	0.051	2.96***	0.015	3.35***	0.019	2.50**
Style Flow	0.896	34.74***	0.924	45.08***	0.889	25.01***	0.952	28.08***	0.918	34.03***	0.929	43.16***
Assets	-0.126	-10.10***	-0.093	-8.93***	-0.074	-8.39***	-0.050	-8.23***	-0.086	-8.32***	-0.075	-7.36***
Min. Invt.	0.043	8.23***	0.032	7.37***	0.027	7.07***	0.021	7.84***	0.035	4.83***	0.039	4.51***
Ifee	-0.002	-1.71	-0.003	-3.66***	-0.003	-2.92***	-0.002	-2.14**	-0.002	-1.20	-0.003	-1.98*
Mfee	0.020	1.64	0.012	1.13	0.007	0.66	-0.005	-0.058	0.015	1.04	0.019	1.29
HWM	0.051	2.46**	0.038	2.04*	0.019	1.20	0.006	0.34	0.055	2.05*	0.064	2.11**
Red. Notice	0.228	2.58**	0.131	1.71	0.149	1.35	0.095	0.84	0.316	1.69	0.240	1.45
Sub. Freq.	-0.064	-2.76***	-0.081	-3.09***	-0.010	-2.08*	-0.107	-1.88*	-0.090	-2.89***	-0.129	-2.35**
Red. Freq.	-0.010	-3.45***	-0.051	-2.39**	-0.032	-1.24	-0.023	-1.08	-0.088	-3.99***	-0.013	-0.19***
Lockup	-0.021	-0.08	0.102	0.34	-0.091	-0.31	-0.283	-1.03	-0.329	0.87	1.164	0.88
N	2182		1831		1530		1273		1448		1169	
Adj R-sq	11.42%		11.51%		11.07%		11.86%		10.23%		12.49%	

Table 9: Alpha and Beta Comparison

This table presents fund alpha estimates, in percent per month, using two different factor models. We only run this analysis on our combined dataset. Panel A uses an index model where the only factor is the style index return, taken from the cross-sectional average of all funds. Panel B uses the Carhart 4-factor model. We report the average alphas and average values of the factor loadings. Alphas and betas are estimated once using all available fund returns. (i.e. "All" uses all returns to estimate beta and alpha.) Results are reported for all return period, as well as backfilled and post-backfilled returns. We require at least 12 returns for a fund to be included in our results

Panel A: Style Index Factor Model

	All	Backfill	Post Backfill
Alpha	0.206%	0.793%	-0.110%
Index Beta	0.89	0.78	0.98
Adjusted R2	25.8%	23.1%	28.8%
Residual Std. Dev.	3.53%	3.55%	3.15%

Panel B: 4-Factor Carhart

	All	Backfill	Post Backfill
Alpha	0.263%	0.770%	-0.049%
Mkt_RF	0.306	0.293	0.324
SMB	0.061	0.067	0.056
HML	-0.018	0.018	-0.025
MOM	0.017	0.029	0.001
Adjusted R2	22.7%	22.3%	24.8%
Residual Std. Dev.	3.47%	3.38%	3.07%

Figure 3: Alphas and Market Betas around Listing Dates

This figure plots the alpha and market beta using the 4-factor Carhart model around the date added to the database, measured over the prior 24-month period. Alphas are represented by the solid black lines and left axis; dotted lines and the right axis represent the average market beta. Month 0 is the month the fund was added.

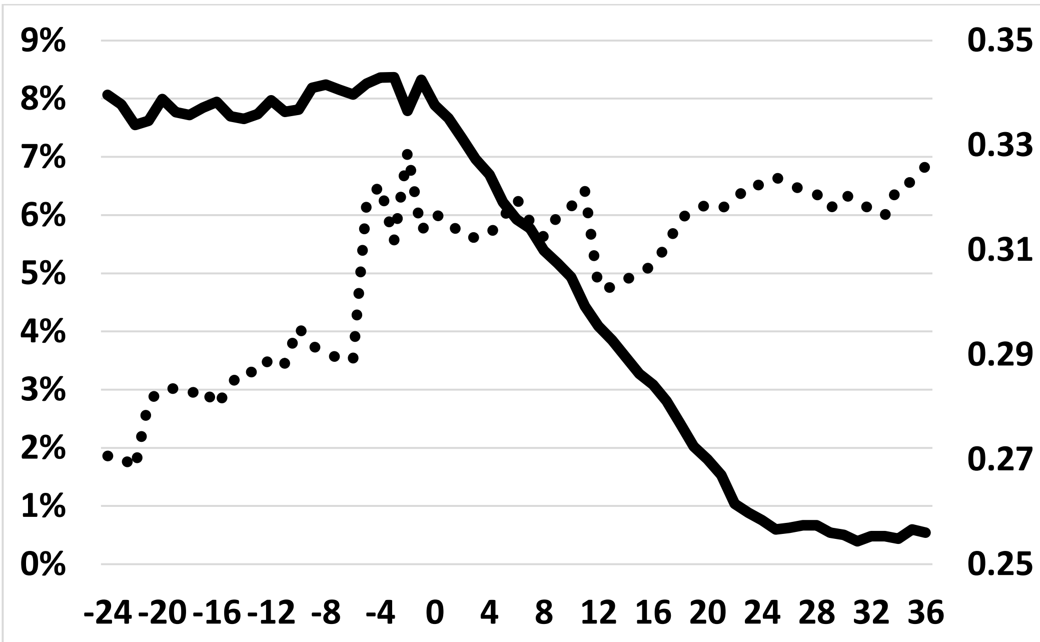
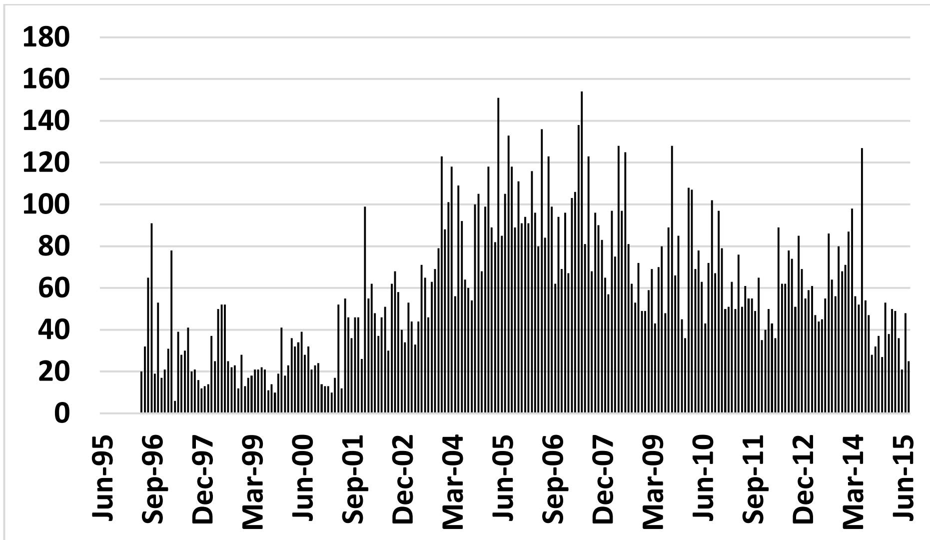


Figure 4: Comparison of Actual and Generated Add Dates for HFR

In this figure, we display the distribution of add dates across time using the actual add dates from HFR (Panel A) and the add dates generated by our algorithm (Panel B).

Panel A: Actual



Panel B: Generated

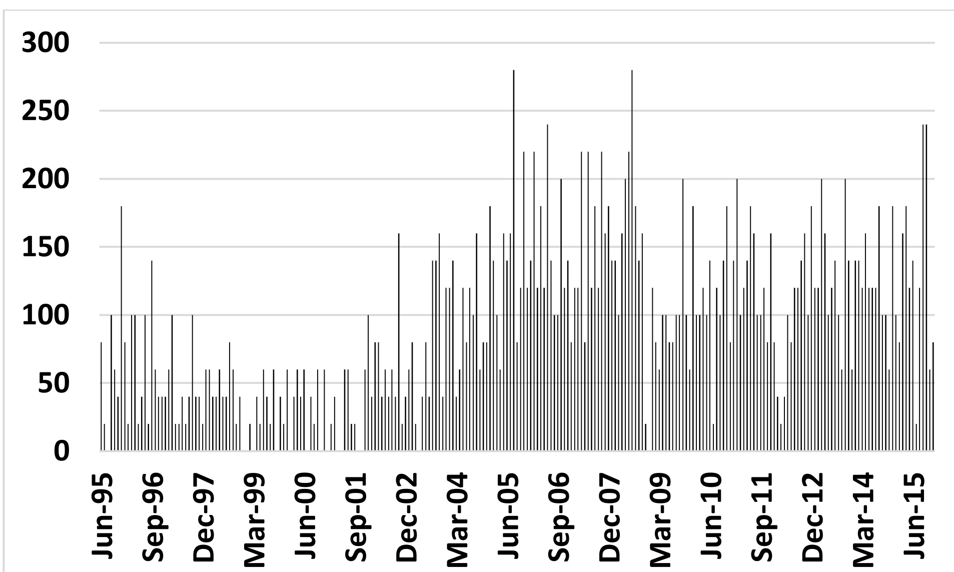
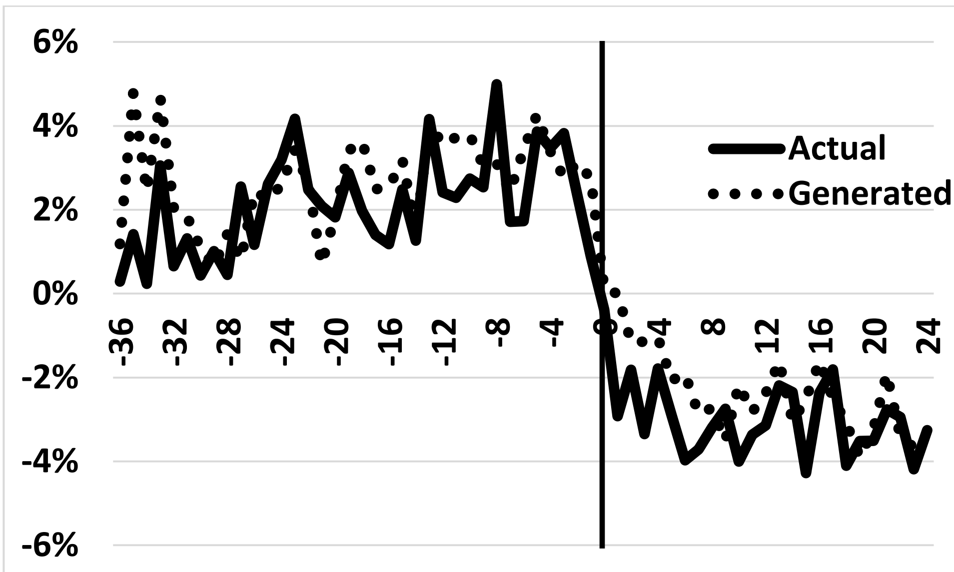


Figure 5: Returns around Generated Add Dates

This figure plots average returns relative to the date the fund was added to the database. Month 0 is the month the fund was added. Returns are style excess returns. In Panels A and B, we plot the returns using TASS and HFR data, respectively. The analysis is reported using the "actual" date reported in the databases, and the "generated" date using our algorithm.

Panel A: TASS



Panel B: HFR

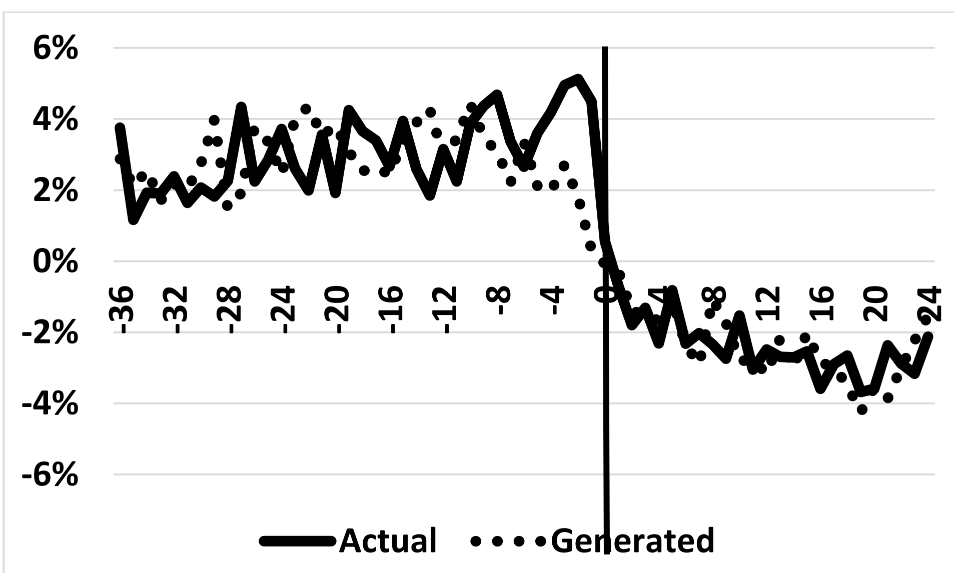


Table 10: Generated Add Date Accuracy

This table compares the accuracy of our generated add dates compared to the actual ones. We compute the average difference and average absolute difference in months. We also show the percentage of our generated add dates that are within three, six, and 12 months of the actual add date. We also report the percentage of backfilled returns retained using our add dates as well as the total number of observations and number of non-backfilled observations. For comparison, we report the same statistics for the inception date and the standard 12- and 24-month cutoffs. Panel A and B report results for HFR and TASS, respectively.

Panel A: HFR

	Generated	Inception	12-month Cutoff	24-month Cutoff
Average Difference	2.55	30.32	18.28	6.25
Average Abs(Difference)	6.72	30.32	24.49	24.95
Percent within 3 months	49%	16%	12%	7%
Percent within 6 months	69%	27%	25%	14%
Percent within 12 months	88%	42%	59%	30%
Percent of Backfilled	6%	41%	34%	30%
Total Num. of Obs.	516,638	903,090	763,508	637,018
Non-Backfilled Obs.	484,772	533,386	501,532	444,215

Panel B: TASS

	Generated	Inception	12-month Cutoff	24-month Cutoff
Average Difference	7.94	33.21	21.18	9.14
Average Abs(Difference)	9.84	33.21	26.27	25.18
Percent within 3 months	55%	9%	13%	8%
Percent within 6 months	65%	19%	26%	15%
Percent within 12 months	77%	34%	53%	32%
Percent of Backfilled	9%	43%	37%	32%
Total Num. of Obs.	386,114	616,315	520,472	433,311
Non-Backfilled Obs.	342,457	355,641	336,983	298,988

Figure 6: Performance around Generated Add Dates for BarclayHedge

This figure plots excess returns around generated add dates for BarclayHedge, which does not have an add date field. Month 0 is the month the fund was added. Returns are style excess returns and annualized.

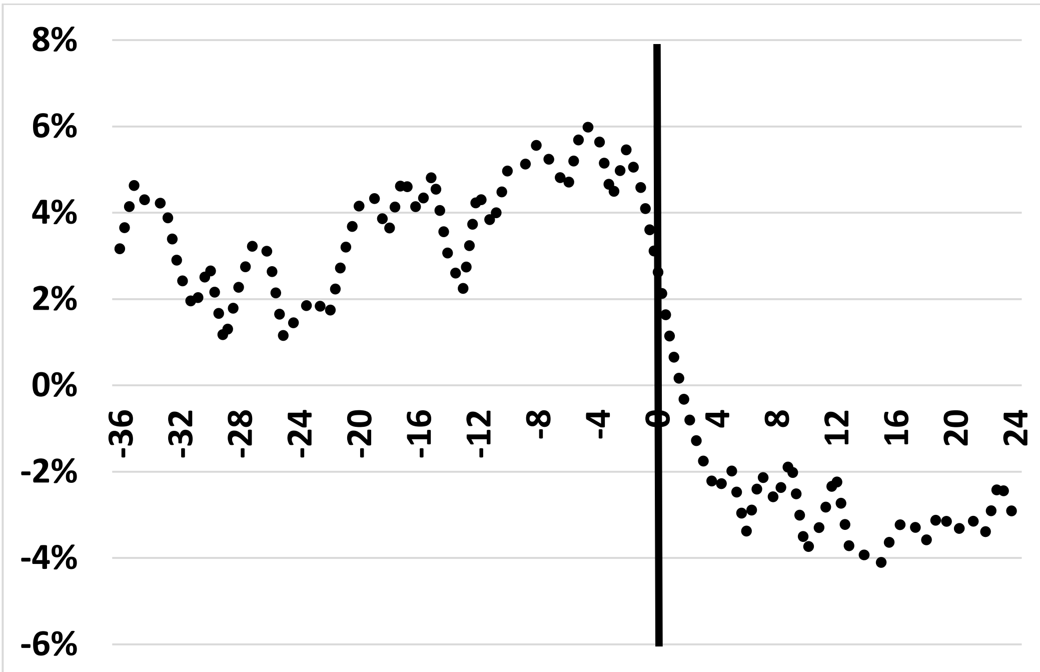


Table 11: Impact of Backfill on the BarclayHedge Database

In this table, we examine the impact of backfill on the BarclayHedge database. Panel A reports the percentage of backfilled returns and funds with backfilled returns based on various truncation lengths. Panel B reports average returns using various truncation lengths as well as our generated add dates. Bias is defined as the difference between average returns using all returns and only those after our generated add date. In Panel C (D), we report the relation between performance and size (age).

Panel A: Amount of Backfill

# months	Returns	Funds
None	38%	95%
3	35%	77%
6	33%	69%
12	30%	57%
18	28%	50%
24	26%	44%
30	24%	39%
36	22%	36%

Panel B: Average Performance

	All Returns	Cut Offs			Add Dates	Bias
		1-year	2-years	3-years		
Avg.	10.70%	9.39%	8.88%	8.24%	7.07%	3.30%
Cum.	716%	564%	513%	459%	3.73%	343%
p-value of Diff.	0.00	0.00	0.00	0.00		

Panel C: Size and Performance

Quarterly Results					
	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Highest	-0.52%	-0.33%	-0.22%	-0.13%	-0.02%
2nd	-0.18%	0.01%	0.02%	-0.01%	0.14%
3rd	0.02%	0.02%	0.03%	0.05%	-0.05%
Lowest	0.70%	0.33%	0.18%	0.09%	-0.08%
Diff	1.22%	0.66%	0.40%	0.22%	-0.06%
p-value	0.00	0.00	0.01	0.16	0.75

Annual Results					
	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Highest	-2.04%	-1.54%	-1.20%	-0.99%	-0.66%
2^{nd}	-0.43%	-0.30%	-0.52%	-0.37%	-0.05%
3^{rd}	0.64%	0.54%	0.68%	0.54%	0.17%
Lowest	2.24%	1.62%	1.32%	1.02%	0.82%
Diff	4.28%	3.16%	2.52%	2.01%	1.48%
p-value	0.00	0.00	0.00	0.03	0.10

Panel D: Age and Performance

Quarterly Results					
	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Highest	-0.58%	-0.32%	-0.15%	-0.09%	0.09%
2^{nd}	-0.50%	-0.36%	-0.27%	-0.23%	-0.23%
3^{rd}	0.25%	0.11%	-0.02%	0.00%	-0.14%
Lowest	0.84%	0.57%	0.44%	0.32%	0.27%
Diff	1.42%	0.89%	0.59%	0.41%	0.18%
p-value	0.00	0.00	0.00	0.00	0.25

Annual Results					
	All Returns	Cut Offs			Add Dates
		12 months	24 months	36 months	
Highest	-1.29%	-0.66%	-0.45%	-0.35%	1.01%
2^{nd}	-1.48%	-1.15%	-0.71%	-0.90%	-1.05%
3^{rd}	0.32%	-0.03%	-0.10%	0.40%	-0.96%
Lowest	2.28%	1.67%	1.10%	0.71%	0.66%
Diff	3.57%	2.33%	1.55%	1.06%	-0.35%
p-value	0.00	0.00	0.01	0.14	0.67

APPENDIX A-1

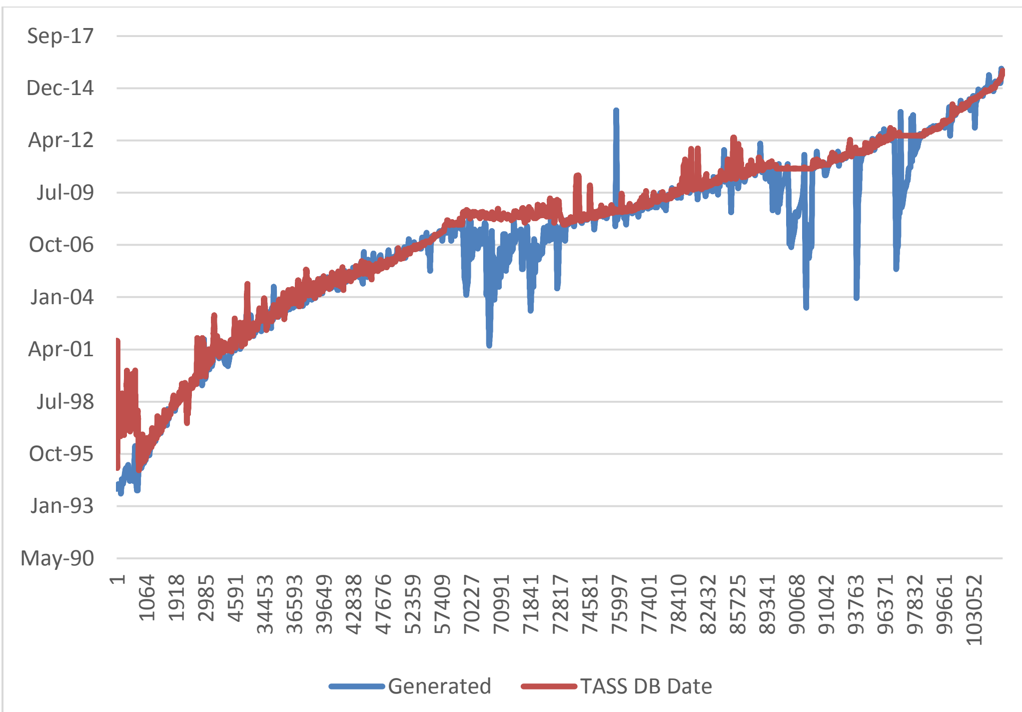
Table A-1: Illustrative Examples of How Empirical Research Deals with Backfill

Paper	Database / Approach
Fung, W. and D. Hsieh, "Empirical characteristics of dynamic trading strategies: The case of hedge funds," *Review of Financial Studies* (1997).	TASS / No returns removed
Ackermann, C.; R. McEnally; and D. Ravenscraft. "The Performance of Hedge Funds: Risk, Return, and Incentives." *Journal of Finance* (1999).	MAR, HFR / Main tests have no returns removed. Removed first 24 returns for a robustness check
Liang, B. "Hedge Funds: The Living and the Dead." *Journal of Financial and Quantitative Analysis* (2000).	TASS, HFR / No returns removed
Agarwal, V. and N. Naik, "Multi-Period Performance Persistence Analysis of Hedge Funds," *Journal of Financial and Quantitative Analysis* (2000).	HFR / No returns removed
Fung, W. and D. Hsieh, "The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers," *Review of Financial Studies* (2001).	TASS / No returns removed
Getmansky, M.; A. Lo; and I. Makarov, "An econometric model of serial correlation and illiquidity in hedge fund returns," *Journal of Financial Economics* (2004).	TASS / No returns removed
Baquero, G.; J. Horst; and M. Verbeek. "Survival, Look-Ahead Bias and the Persistence in Hedge Fund Performance." *Journal of Financial and Quantitative Analysis* (2005).	TASS / No returns removed
Kosowski, R.; N. Naik; and M. Teo. "Do Hedge Funds Deliver Alpha? A Bayesian and Bootstrap Analysis." *Journal of Financial Economics* (2007).	TASS, HFR, CISDM, MSCI / Robustness with 12 months removed
Agarwal, V.; N. Daniel; and N. Naik. "Role of Managerial Incentives and Discretion in Hedge Fund Performance." *Journal of Finance* (2009).	TASS, HFR, CISDM, MSCI / Main tests have no returns removed. Removed first 24 returns for a robustness check
Jagannathan, R.; Malakhov A.; and D. Novikov. "Do Hot Hands Persist Among Hedge Fund Managers? An Empirical Evaluation." *Journal of Finance* (2010).	HFR / Use added to database date
Aggarwal, R. and P. Jorion. "The Performance of Emerging Hedge Funds and Managers." *Journal of Financial Economics* (2010).	TASS / Use added to database date
Avramov D.; Kosowski R; Naik N.; and M. Teo, "Hedge funds, managerial skill, and macroeconomic variables." *Journal of Financial Economics* (2011).	TASS, HFR, CISDM, MSCI / 12 months removed
Aragon, G; Liang, B.; and H. Park. "Onshore and Offshore Hedge Funds: Are They Twins?" *Management Science* (2014).	TASS / Use added to database date
Jorion, P. and C. Schwarz. "The Strategic Listing Decisions of Hedge Funds." *Journal of Financial and Quantitative Analysis* (2014a).	TASS, HFR / Use added to database date
Joenvaara, J.; R. Kosowski; and P. Tolonen. "Hedge Fund Performance: What Do We Know?" *Imperial College WP* (2016).	TASS, HFR, BarclayHedge, EurekaHedge, Morningstar / 12 months removed

APPENDIX A-2

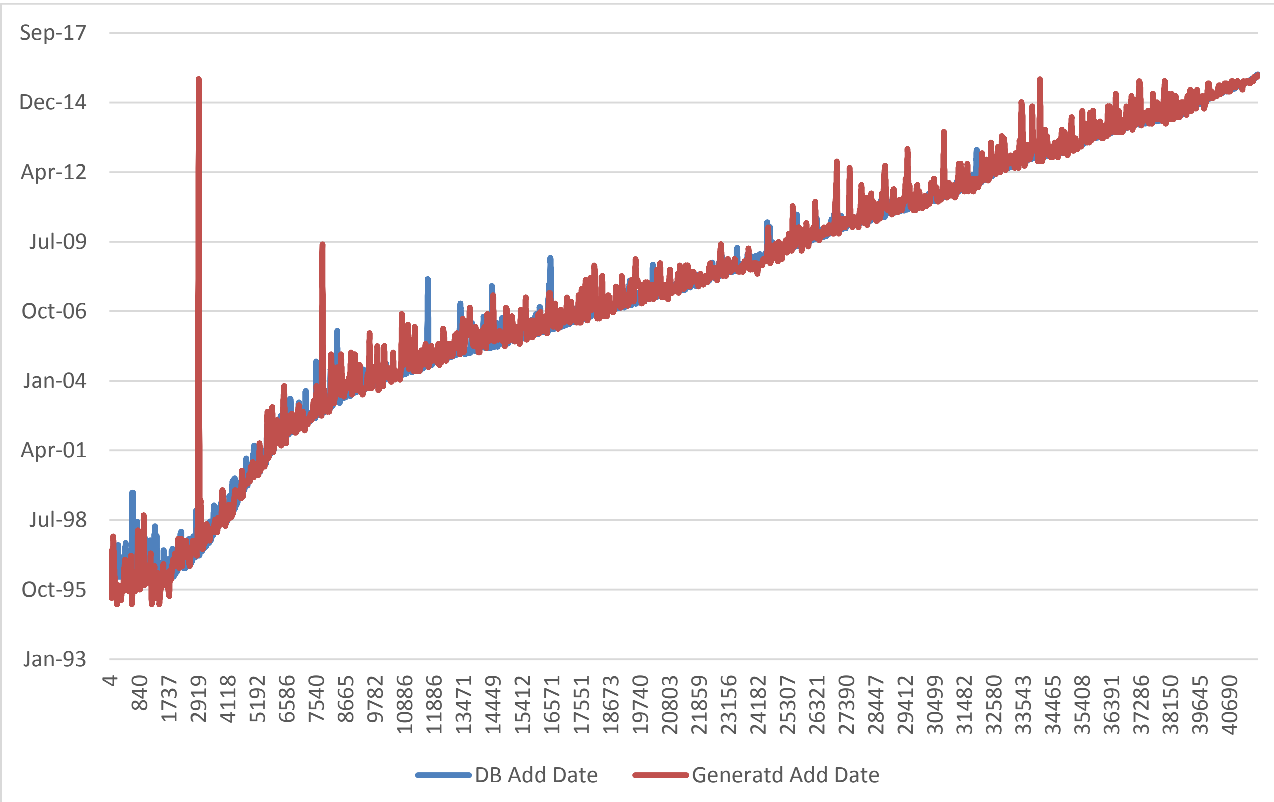
Table A-2: Add Date Generating Information

This appendix describes the detailed procedure, or code used to generate add dates for HFR in SAS. With simple modifications, this code can be used for any hedge fund database. In the following figures, we compare the generated add dates for TASS and HFR to the actual add dates. The x-axis represents the product code; the y-axis shows the 20-fund moving average of add dates. Here is the TASS graph:



As noted in the main body of the paper, funds are largely listed in chronological order and therefore the add dates we generate are fairly accurate and also in chronological order. The main problems with add dates are in areas were TASS has some issues. For example, starting around product number 70,000, TASS lists approximately 3,000 funds with the same add date. These are likely not the true add dates. Thus our method does not match TASS here. The same issue appears near product numbers 90,000 and 96,000.

Next, consider HFR add dates. These are much more smooth. As a result, our add date generation is smoother as well:



As with TASS, products are generally listed in chronological order. Except for one large miss for a group of 20 funds near product number 3,000, our generated add dates are generally very accurate.

SAS Code

```
/* HFR */

/* get a list of funds from the characteristic file*/
data hfrchartmp;
      set hfrdata.hfrfullfeb16;

      keep hfr_fund_id;
run;

/* make sure they are sorted by product number*/
proc sort data=hfrchartmp;
      by hfr_fund_id;
run;

/* label funds from the first to the last fund in chronological order
      Necessary since product numbers aren't 1,2,3,etc.*/
data hfrchar;
      set hfrchartmp;

      listnumber = _n_;
run;

/* get performance data. Remove any bad observations */
data hfrperf;
      set hfrdata.hfrperformancefeb16;

      keep hfr_fund_id date;

      if date ~= . and performance ~= . then output;
run;

/* sort to merge with fund number list */

proc sort data=hfrperf;
      by hfr_fund_id date;
run;

/* combine performance data and fund number list.
      Remove any funds without performance data */
data together;
      merge hfrperf hfrchar;
      by hfr_fund_id;

      if date ~= . then output;
run;


%macro getadddates;

/* select data for this group of 20 funds. */
data selectdata;
      set together;

      if  listnumber > ((&counter - 1)*20) and listnumber <= &counter*20 then
output;
run;
```

```
/* generate list of 20 funds in this batch to merge later with add date*/
data fundlist;
      set selectdata;
      by hfr_fund_id;

      keep mergeme hfr_fund_id;

      mergeme = 1;

      if first.hfr_fund_id then output;
run;

/* sort data by date */
proc sort data=selectdata;
      by date;
run;

/* count observations per month */
proc univariate data=selectdata noprint;
      by date;
      var hfr_fund_id;
      output out=countrtns n=n;
run;

/* put most overlap funds at top and sort in chrono order */
proc sort data=countrtns;
      by descending n date;
run;

/* select first 3 (or less) months with max overlap */
data maxoverlapmonths;
      set countrtns;

      retain tempnum;

      mergeme = 1;

      if _n_ = 1 then tempnum = n;

      if tempnum = n and _n_ < 4 then output;
run;

/* keep last observation for add date */
data adddateforgroup;
      set maxoverlapmonths  end=last;

      keep mergeme HFRcomputedadddate;

      format HFRcomputedadddate mmddyy10.;

      HFRcomputedadddate = date;

      if last then output;
run;

/* merge add date with list of 20 funds in batch */
data cutoffs;
```

```
      merge fundlist adddateforgroup;
      by mergeme;

      drop mergeme;
run;
%mend;


%macro killbackfill;

/* set batch counter to 1, run macro*/
data tempfile;
      CALL SYMPUT('counter',1);
run;

%getadddates

/* create final list file w/ first batch*/
data finallist;
      set cutoffs;
run;

/* run macro up to 1170. Final number should be
      number funds divided by 20 rounded up*/
%do i=2 %to 1170;

      data tempfile;
            CALL SYMPUT('counter',&i);
      run;

      %getadddates

      data finallist;
            set finallist cutoffs;
      run;
%end;
%mend;

/* turn off log */
filename junk dummy;
proc printto log=junk; run;

/* run macro */

%killbackfill


/* turn on logging again*/
proc printto; run;

/* work.finallist has add dates*/
```