Hedge Fund Return Misreporting: Incentives and Effects

Petri Jylhä*

February 15, 2011

Abstract

We study the motives and effects of return misreporting in the hedge fund industry, and find the following results. First, misreporting is most prevalent in young funds, in funds that have strong flow-performance relation, and during months of positive capital flows. These empirical findings are consistent with a simple model where hedge fund managers have two motives to misreport returns: attraction of larger capital flows in the future and wealth transfer from the new investors to the old investors in the fund. Second, under mild conditions, return misreporting decreases the estimates of funds' risks and increases estimates of risk-adjusted returns. Ex post estimates of volatility are lower and estimates of alpha and autocorrelation higher for funds with lower quality of reported returns.

JEL Classification: G23.

Keywords: hedge funds, return misreporting, incentives, performance measurement.

---

*Aalto University (former Helsinki School of Economics). P.O.Box 21220 00076 Aalto Finland, petri.jylha@aalto.fi. I thank Reena Aggarwal, Joni Kokkonen, Peter Mackay, Martin Ruckes, Henri Servaes, Matti Suominen, and the participants at Nordic Finance Network 2010 workshop for helpful comments. I also thank Finnish Foundation for Advancement of Securities Markets and OP-Pohjola Reseacrh Foundation for financial support.

1 Introduction

The hedge fund industry has grown at a ferocious pace over the past 15 years with the most recent estimates putting the total assets under management at about 2 trillion dollars. At the same time, hedge fund related frauds have become more common and concerns over the veracity of reported returns more strident. Recent research shows that hedge funds do misreport their returns and estimates the wealth transfer due to misreported returns to be between 1 and 2 billion dollars between the years 1994 and 2005 (Bollen and Pool, 2009).

Bollen and Pool (2009) show that the distribution of hedge fund returns displays a discontinuity around zero: The amount of small positive returns far exceeds the amount of small negative returns. They show that this is a result of deliberate misreporting rather than skill, similar discontinuity in underlying assets or strategies, or database biases. They argue that reporting a small positive return instead of a small negative one has a positive effect on the fund's capital flows. They also show that no misreporting takes place right before an audit and that misreporting is more prevalent in the population of funds for which it is easier to misvalue assets (e.g. funds investing in distressed securities). Bollen and Pool (2009) concentrate on identifying the phenomenon of misreporting and some fixed characteristics that govern the possibility of a fund to misreport its returns. In this paper we extend their work and contribute to the literature in three ways.

First, we extend the analysis of Bollen and Pool (2009) regarding the motives for misreporting by presenting a simple model of a hedge fund manager's return reporting problem and by testing the model predictions empirically. Two motives to misreport returns arise from the model: attraction of future capital flows and wealth transfer. Attraction of future flows is already mentioned as the motive for misreporting by Bollen and Pool (2009). In addition, by overstating returns, and hence asset values, during periods of positive capital flows a fund manager can overcharge new investors for their shares resulting in a wealth transfer from new investors to old investors. Likewise, understating returns and asset values during negative capital flows will result in too little being paid out to the leaving investor and a value transfer to the remaining ones.

Comparing the discontinuities in the distributions of various sub-samples of hedge fund returns, we find empirical evidence of both of these motives. Hedge funds with strongly performance-dependent capital flows misreport returns more than those with weaker flow-performance relation. Also, misreporting is more prevalent during months of capital inflows than during months of outflows.

The second contribution of the paper deals with the effects of misreporting on ex post performance measurement. Under relatively mild conditions misreporting induces a negative bias to risk estimates such as volatility and beta and a positive bias to alpha and return autocorrelation estimates. We show this to be the case empirically. Using various measures for the quality of reported returns, we find that lower return volatility, higher alpha, and higher return autocorrelation are related to misreporting.

Third, in order to formally test whether the prevalence of misreporting differs between two sub-samples of return data, we extend the methodological framework developed by Bollen and Pool (2009). The original framework focuses on identifying discontinuities in hedge fund return distributions that the authors argue and show to be a result of deliberate return misreporting. We introduce formal tests for the existence of such discontinuity in one sample of return data and for the difference in discontinuities in two samples of return data. This extended framework may prove useful in other applications as well. One example could be testing for differences in earnings management in various sub-samples of corporate earnings data.

In addition to Bollen and Pool (2009), this paper is related to a number of papers examining misreporting, return smoothing, and other suspicious patterns in hedge fund returns. Getmansky, Lo, and Makarov (2004) document large positive autocorrelations in hedge fund returns and provide innocuous explanations such as time-varying expected returns, time-varying leverage, fee structures, and illiquidity of assets. In an earlier paper, Bollen and Pool (2008) study conditional serial correlations of hedge fund returns and find some evidence that hedge fund managers delay reporting losses but fully report gains. Asness, Krail, and Liew (2001) find that hedge fund returns are smoother during low benchmark returns, consistent with more return misreporting during months of low actual returns. Agarwal, Daniel, and Naik (2007) document that hedge fund returns in December are significantly higher than during other months. This 'Santa effect' is more pronounced for funds with higher incentives, as measured by the 'delta' of the fund manager's compensation contract. Bollen and Pool (2010) show that suspicious patterns in hedge fund returns can be used to predict fraud. Finally, Cumming and Dai (2010) use a simplified version of the Bollen and Pool (2009) framework to provide evidence that the regulatory environment affects hedge funds' return misreporting.

Misreporting by hedge funds is also studied from different angles by two interesting recent papers. Using unique data on due diligence reports, Brown, Goetzmann, Liang, and Schwarz (2010) find that a large fraction of hedge funds use internal pricing which might also facilitate misreporting of returns. Cici, Kempf, and Puetz (2010) study the SEC filings and show that hedge fund advisors intentionally misprice their stock positions providing direct evidence that hedge funds do misreport asset values.

The rest of the paper is organized as follows. Sections 2 and 3 contain the theoretical part of the paper with the former describing fund managers' incentives to misreport returns and the latter focusing on the effects of misreporting. Section 4 describes the data and methodology, Sections 5 and 6 present results and robustness checks, respectively, and Section 7 concludes the paper.

2 When Should Hedge Fund Managers Misreport?

Here, we consider the return reporting problem of a risk neutral hedge fund manager in a model with discrete and finite time horizon. Figure 1 illustrates the timing of the model. At time $t = 0$ the hedge fund is launched with A_0 units of assets under management. We simplify by normalizing A_0 to one. The fund manager holds an ω_0 share of the fund and outside investors hold the rest, i.e. $1 - \omega_0$. Each period, the fund's return, r_t, is a random draw from a normal distribution with μ_r mean and σ_r^2 variance. Note that the true returns are completely independent of any action taken by the manager.

[Insert Figure 1 here]

The events at times $t = 1$ and $t = 2$ happen sequentially as follows. First, outside investors submit their subscriptions or redemptions after which the fund manager observes the return on the fund's existing assets. Following the observation of the flows and actual returns, the fund manager chooses what return he will report to outside investors. The capital flow at time $t = 1$, f_1 is exogenous and random but flow at time $t = 2$ is a deterministic function of the return reported at $t = 1$: $f_2 = b_0 + b_r\tilde{r}_1$, where $\tilde{r}_1 = r_1 + m$ is the reported return at $t = 1$, m is the amount of misreporting, and $b_r > 0$ measures the strength of the fund's flow-performance relation. Such positive dependence of flows on past returns is documented in e.g. Agarwal, Daniel, and Naik (2004), Getmansky (2005), Fung, Hsieh, Naik, and Ramadorai (2008), and Wang and Zheng (2008)).

At time $t = 3$ a final return is realized. The fund managers charges a management fee equal to ϕA_3 from the fund, where A_3 is the fund's assets under management at time $t = 3$. The net-of-fees assets are then divided among the fund manager and outside investors according to their ownership in the fund. The management fee and the ownership stake in the fund represent the only sources of consumable income for the fund manager.

In this setting, the fund manager only has discretion over what returns to report in times $t = 1$ and $t = 2$. Let us further assume that any misvaluation of assets resulting from misreported returns at $t = 1$ must be corrected at $t = 2$. Thus, the fund manager's problem reduces to choosing the amount of misreporting, m at $t = 1$. Misreporting returns, i.e. choosing $m \neq 0$, has three distinct effects on the fund manager's expected income.

First, reporting higher than realized returns has a positive effect on the $t = 2$ flows which in turn increases the size of the fund at $t = 3$ and the management fee charged by the manager. The monetary increase in management fee resulting from misreporting by m is equal to[1]

$$\phi b_r m(1 + r_3). \tag{1}$$

[1] See Appendix A for the derivation of the results.

Note that r_2 and r_3 are unknown by the time the manager is making his misreporting decision. Since $b_r > 0$, overstating returns will always have a positive effect on future flows and management fees. This flow effect of fund performance is what Bollen and Pool (2009) argue is the main motive return misreporting.

Second, overstating returns while flows are positive results in a transfer of wealth from the new investors to the existing ones, including the fund manager. This is because the new investors overpay for their share in the fund when assets are overvalued. Similarly, understating the fund's asset value when flows are negative leads to a wealth transfer from those investors who are leaving the fund to those who stay. The $t = 3$ net-of-fees value of this wealth transfer to the fund manager is equal to

$$\frac{m f_1 \omega_0 (1 + r_2)(1 + r_3)(1 - \phi)}{1 + r_1 + m + f_1}. \tag{2}$$

It is obvious from (2) that whenever misreporting (m) and capital flows (f_1) have the same sign, the fund manager is extracting value from the subscribing or redeeming investors.

This effect of capital flows inducing misreporting with the same sign can also be seen as a form of an anti-dilution levy. Flows, especially large ones, may result in significant trading costs that are borne by all investors in the fund, not just the subscribing or redeeming investors. To compensate for such dilution of fund value, some mutual funds charge (or reserve the right to charge at fund manager's discretion) the subscribing or redeeming investor an anti-dilution levy which is remitted to the fund.[2] Though not completely unheard-of, such practice is rare in the hedge fund industry.[3] However, in face of larger redemptions or subscriptions, fund managers may be inclined to collect such levy by misreporting asset values. Misreporting returns in the same direction as capital flows are can then be seem as a kind of an undisclosed anti-dilution levy.

Finally, the fund manager might be caught for misreporting. We assume that this happens with a probability $1 - \exp(-\kappa m^2)$, where κ is the parameter that governs the

[2]Note that management fees, performance fees, subscription fees, and redemption fees are all paid to the fund manager whereas the anti-dilution levy is paid to the fund itself.

[3]For example, in July 2009 the London based hedge fund company Polar Capital Partners imposed an redemption fee on its Forager fund to protect staying investors in face of large outflows from the fund.

likelihood of being caught. If the manager is caught for misreporting, he is subjected to a fixed penalty, c. The expected penalty of being caught for misreporting by m is then

$$\left(1 - e^{-\kappa m^2}\right) c \tag{3}$$

We have now outlined the three channels through which misreporting affects the fund manager's consumption. Being risk neutral, the manager chooses m^* that maximizes the expected value of the sum of the three effects:

$$m^* = \underset{m}{\operatorname{argmax}} \left\{ \phi b_r m(1+\mu) + \frac{m f_1 \omega_0 (1+\mu)^2 (1-\phi)}{1 + r_1 + m + f_1} - \left(1 - e^{-\kappa m^2}\right) c \right\}. \tag{4}$$

The first order condition of (4) is

$$\phi b_r (1+\mu) + \frac{f_1 \omega_0 (1+\mu)^2 (1-\phi)}{1 + r_1 + m^* + f_1} - \frac{m^* f_1 \omega_0 (1+\mu)^2 (1-\phi)}{(1 + r_1 + m^* + f_1)^2} - \frac{2c\kappa m^*}{e^{\kappa m^{*2}}} = 0. \tag{5}$$

No simple closed form solution exists for (5).[4] To examine the fund manager's optimal strategy, we fix the model parameters and solve numerically for m^*. Figure 2 presents the optimal levels of misreporting for varying levels of capital flows (f_1), flow performance relations (b_r), capture likelihoods (κ), and penalties (c).[5]

Generally, greater flows result in a greater amount of misreporting. The only exception to this occurs when the fund manager does not have an ownership stake in the fund ($\omega_0 = 0$) or ignores the value transfer effects. In such a case, due to linear flow-performance relation, the optimal amount of misreporting is positive and equal for all values of capital flows. A fund manager that does not charge a management fee ($\phi = 0$) misreports only to steal from the subscribing or redeeming investors. Hence, he overstates returns during positive flows, understates returns

[4] This is mainly due to the $m^* e^{-\kappa m^{*2}}$ term. Other alternative definitions of the penalty function, such as a simple quadratic one, would help to resolve this particular problem but still result in a complicated closed form solution and potentially multiple solutions and imaginary solutions.

[5] We do not study the effect of true return, r_1, on misreporting. This is because our methods of identifying misreporting are based either on discontinuity in the distribution of reported returns around zero return or individual funds' full return history. Hence, we cannot empirically test the prevalence of misreporting at different levels of true returns. Theoretically, it might be interesting to examine how different forms of flow-performance relations and fund managers' preferences affect the relation between true returns and misreporting. However, such examination is beyond the scope of this paper.

during negative flows, and reports truthfully during zero flows. These results provide for an empirical identification strategy. Any misreporting we observe during zero flows must be driven by the fund managers' wish to attract future flows. Further, if we observe misreporting to be larger during positive flows than it is during zero flows, the difference is driven by value transfer motives. Same is true for smaller misreporting during negative flows.

Stronger flow performance relation (i.e. higher b_r) leads to higher misreporting in a nearly linear fashion. This we can also test empirically by identifying funds whose future flows depend on past performance particularly strongly or weakly and comparing the misreporting propensity in these two sub-samples of funds.

Likelihood of being caught (governed by κ) and the penalty imposed for captured misreporter (c) serve as substitutes to each other decreasing optimal misreporting towards zero. Also this prediction can be tested empirically by comparing misreporting between funds that differ in their likelihood of being caught for misreporting or in the penalty they are likely to suffer if caught misreporting. The results in Bollen and Pool (2009) that there is no misreporting right before an audit and that funds with more transparent assets misreport less are manifestations of higher capture probability decreasing misreporting.

3 Effects of Misreporting

We now abstract from the three-period setup above to study the effects of misreporting on the measures of risk and risk adjusted returns. Let us assume a simple return generating process and see how misreporting biases estimates of return volatility, autocorrelation, as well as fund betas and alphas. The true returns on a fund (r_t) are generated by an exposure to a risk factor (λ_t) and an idiosyncratic term (ε_t): $r_t = \alpha + \beta\lambda_t + \varepsilon_t$, where $\lambda_t \sim N(\mu_\lambda, \sigma^2_\lambda)$ and $\varepsilon_t \sim N(0, \sigma^2_\varepsilon)$. Since λ_t and ε_t are independent random draws, the true returns follow a normal distribution with $\mu_r = \alpha + \beta_\lambda\mu_\lambda$ mean, $\sigma^2_r = \beta^2\sigma^2_\lambda + \sigma^2_\varepsilon$ variance, and zero autocorrelation.[6]

Again, the reported returns are the sum of true returns and misreporting: $\tilde{r}_t =$

[6] Let us also assume that $\beta \geq 0$ and $\mu_\lambda > 0$, i.e. the fund has a non-negative exposure to a risk factor with positive expected return.

$r_t + m_t$, where m_t is the amount of misreporting. The amount of misreporting is not necessarily i.i.d. but can depend on its own history as well as current and past returns. Actually, it is quite reasonable to assume that true returns and misreporting are negatively correlated, i.e. that fund managers overstate returns when true returns are low, and eliminate the thus generated misvaluation when true returns are high (Bollen and Pool, 2008, 2009). Negative correlation between m_t and r_t could also imply that the simple correlation between m_t and λ_t is negative, i.e. funds misreport positively during low factor returns.[7] Evidence in Asness, Krail, and Liew (2001) suggests that more misreporting happens during low benchmark returns. It is also reasonable to assume the standard deviation of misreporting (σ_m) is rather low compared to the standard deviation of the true returns (σ_r)

To rule out Ponzi schemes, we require that in the long run the expected cumulative amount of misreporting must be zero, $E(\sum_{t=0}^{\infty} m_t) = 0$. First of all, this condition implies that the unconditional expectation of m_t is zero and the expected value of reported returns is equal to the expected value of true returns, $\mu_{\tilde{r}} = \mu_r$. Second, the no-Ponzi constraint implies that m_t is negatively autocorrelated as following an overstating of returns the fund manager needs to correct for misvaluation by understating returns in a later period. If m_t is negatively autocorrelated and negatively correlated with r_t, we should expect the simple correlation between m_t and r_{t-1} to be positive.

We are now in a position to compare the return volatility, autocorrelation, beta, and alpha estimates based on the reported returns to the true ones based on the true returns. The standard deviation of reported returns is less than that of the true returns when correlation between the true return and the amount of misreporting is sufficiently low:[8]

$$Cor(r_t, m_t) < -\frac{\sigma_m}{2\sigma_r}. \tag{6}$$

Naturally, if misreporting is high during periods of low returns and vice versa this shrinks the distribution of reported returns and lowers the estimated standard devi-

[7] Negative correlation between m_t and r_t, however does not necessarily lead to a negative correlation between m_t and λ_t. One example of such a case is if fund manager overstates returns only when true return is low due to a low idiosyncratic return.

[8] Appendix B contains the derivation of the results in this section.

ation.

If m_t is negatively autocorrelated and positively correlated with r_{t-1} the autocorrelation of reported returns is positively biased when

$$\frac{Cor(m_t, r_{t-1})}{Cor(m_t, m_{t-1})} \quad < \quad -\frac{\sigma_m}{\sigma_r}. \tag{7}$$

This condition holds as long as there is sufficient amount of positive correlation between m_t and r_{t-1} and the standard deviation of misreporting is sufficiently low compared to the standard deviation of returns.[9]

Misreporting also affects estimated factor exposures and abnormal returns. This, naturally, happens when the amount of misreporting is correlated with the factor returns (λ_t). Beta estimate is negatively and alpha estimate positively biased when

$$Cor(m_t, \lambda_t) < 0. \tag{8}$$

When fund managers overstate returns during low factor returns this results in too low beta estimates and too high alpha estimates. Above we argued that negative correlation between m_t and λ_t is the most likely case.

The most likely case of misreporting, i.e. one where returns are overstated during low true returns and understated during high true returns, now leads an econometrician to infer to fund as having lower than actual risk (negatively biased variance and beta estimates), generate higher than actual abnormal return (positively biased alpha estimate), and generate persistent returns (positively biased autocorrelation).[10] We will test these implication in Section 5.2 by comparing ex post measures of risk and return to our fund level measures of misreporting.

[9] Appendix B also describes the conditions for the more unlikely cases where the correlations are both positive, both negative, and when $Cor(m_t, m_{t-1}) > 0$ and $Cor(m_t, r_{t-1}) < 0$.

[10] Note that the same biases can be generated by other sources of return smoothing such as illiquidity exposure (Getmansky, Lo, and Makarov, 2004).

4 Data and Methodology

4.1 Data

The data used in this study are from TASS database and contains monthly return and assets under management (AUM) figures for a total of 9,714 active and inactive hedge funds from March 1977 through December 2008. Following Bollen and Pool (2009), we exclude observations with zero returns and consecutive observations with return of 0.0001. Further, we exclude observations where the fund's AUM does not change from the previous month as in these cases the reported AUM is unlikely to represent the true value. After these eliminations we are left with 352,709 fund-month observations.[11]

In addition to the return data, we need data on the monthly flows of new investments into the funds. The net flow of capital (f_t) is calculated for each fund each month (t) based on the funds' self reported AUM (A_t) and returns (r_t) as

$$f_t = \frac{A_t - A_{t-1}(1 + r_t)}{A_{t-1}}. \tag{9}$$

4.2 Methodology

We use two differing methodologies to measure misreporting. First, we follow and extend the pooled distribution method used by Bollen and Pool (2009). This method is useful and efficient for identifying misreporting in large samples containing at least tens of thousands of return observations. However, this method is not suitable for measuring misreporting at individual fund level due to the low number of observations per fund. Hence, in the second part of the empirical section, we use several fund level measures of return suspiciousness to identify misreporting.

4.2.1 Pooled Distributions

Our methodology to identify return misreporting in large samples is based on Bollen and Pool (2009). In simple terms, the idea is to compare empirical return distribution of a large pool of hedge funds to a fitted continuous nonparametric distribution.

[11] For a more detailed description of the TASS database, see e.g. Getmansky (2005).

If the empirical distribution exhibits discontinuities not reproduced by the fitted distribution, there is a reason to suspect return misreporting. The most likely candidate for the point of discontinuity is around zero return. Similar methodology has been used by e.g. Burgstahler and Dichev (1997) and Degeorge, Patel, and Zeckhauser (1999) to study earnings management by corporations.

To study return misreporting, we first construct the pooled distribution, a histogram, of hedge fund returns. An important step here is to choose an appropriate bin width. Too narrow bins may bias us to identify discontinuities where there are none whereas too wide bins prevent us from identifying existing discontinuities. Following Silverman (1986) and Bollen and Pool (2009) we set the bin width equal to

$$\alpha 1.364 \min\left(\sigma, \frac{Q_3 - Q_1}{1.340}\right) N^{-\frac{1}{5}}, \tag{10}$$

where σ is the standard deviation of observed returns, Q_3 and Q_1 are the third and first quartiles of the returns, N is the number of observations and α is a scaling parameter set equal to 0.776 corresponding to a normal distribution. The top panel of Figure 3 presents the return histogram of all funds in the TASS database. The highlighted bins surrounding zero do exhibit some degree of discontinuity: The bin left of zero has far less observations than the bin right of zero.

[Insert Figure 3 here]

Next, we fit a nonparametric continuous distribution to the histogram. Still following Bollen and Pool (2009), we employ a Gaussian kernel with a density estimate at point r defined as

$$\hat{f}(r; h) = \frac{1}{Nh} \sum_{j=1}^{N} \phi\left(\frac{x_j - r}{h}\right), \tag{11}$$

where h is the chosen bandwidth of the kernel, N is the number of observations, ϕ is the density of standard normal distribution, and x_j are the pooled return data. Since we are using a Gaussian kernel and assumed normal distribution above when setting the bin width, the optimal bandwidth (h) is equal to the optimal bin width in (10). The solid line in the top panel of Figure 3 plots the fitted Gaussian kernel density for the pooled return histogram of all TASS funds. The kernel distribution fits the actual rather well except for the fact that it cannot fully reproduce the discontinuity around zero return.

Bollen and Pool (2009) us a z test to investigate difference between the empirical and fitted densities in any given bin. Since such a test indicates that the empirical density in the bin left (right) of zero is significantly lower (higher) than the fitted density they conclude that the return distribution is discontinuous around zero.[12] However, since we are interested in the differences in the discontinuities between sub-samples of return data, we develop a single number measures for the degree of discontinuity and the difference in discontinuities. Further, we also develop formal tests to examine the statistical significance of the discontinuity and difference in discontinuities.

Our measure of discontinuity is based on the observed and fitted densities in the two bins neighboring zero return. According to the DeMoivre-Laplace theorem the number of observations in bin k is asymptotically normally distributed with mean Np_k and variance $Np_k(1-p_k)$. The probability p_k of an observation residing in bin k is given by the definite integral of the estimated kernel density along the bounds of the bin (r_k^L to r_k^U):

$$p_k = \int_{r_k^L}^{r_k^U} \hat{f}(r;h)dr. \tag{12}$$

Due to the discontinuity, there are $X_{+1} - Np_{+1}$ more observations in bin just right of zero than is expected based on the kernel density estimate. Likewise, the are $Np_{-1} - X_{-1}$ too few observations just left of zero. The total amount of 'misplaced' observations, hence, is $X_{+1} - Np_{+1} + Np_{-1} - X_{-1}$. Since this number is dependent on the sample size, we divide it by the expected number of observations in the two bins. Hence, our measure of discontinuity, DC, is given by

$$DC = \frac{X_{+1} - X_{-1} - N(p_{+1} - p_{-1})}{N(p_{+1} + p_{-1})}. \tag{13}$$

Under no misreporting and a continuous return distribution, the relative amount of

[12] The lower panel of Figure 3 plots the z test statistics for the difference between observed and fitted densities for each of the bins in the histogram.

misplaced observations has expected value of zero and variance equal to

$$D^2(DC) = \frac{[p_{+1} + p_{-1} - (p_{+1} - p_{-1})^2]}{N(p_{+1} + p_{-1})^2}. \tag{14}$$

Hence, the z statistic for testing the existence of discontinuity, z_{DC}, is given by

$$z_{DC} = \frac{DC}{\sqrt{D^2(DC)}} \sim \mathcal{N}(0,1). \tag{15}$$

The full sample has a DC of 7.2%. The associated z_{DC}-statistic is equal to 10.9 which is statistically significant at any reasonable level and a clear evidence that the return distribution is discontinuous around zero return.

Next, to investigate differences in the prevalence of misreporting, we compare the DC figures of different subsets of hedge fund return data. Under the assumption of discontinuity being equal in the two sub-samples i and j, the difference between the DC figures, $DDC = DC^{(i)} - DC^{(j)}$, follows a normal distribution with zero mean and variance equal to the sum of the variances of the individual DC figures. Hence, our asymptotic test statistic for the difference in discontinuities is given by

$$z^a_{DDC} = \frac{DC^{(i)} - DC^{(j)}}{\sqrt{D^2(DC^{(i)}) + D^2(DC^{(j)})}} \sim \mathcal{N}(0,1). \tag{16}$$

We also employ a more flexible, simulation based, approach to test whether the discontinuity is stronger in one sub-sample than the other. We draw 5,000 random pairs of sub-samples whose sample sizes correspond to those we are testing. We calculate the differences in the degree of discontinuity for each of the pairs and use the mean, $\overline{DDC^s}$, and variance, $D^2(DDC^s)$, of simulated DDC's to form a simulated test statistic for the difference in discontinuity:

$$z^s_{DDC} = \frac{DDC - \overline{DDC^s}}{\sqrt{D^2(DDC^s)}} \sim \mathcal{N}(0,1). \tag{17}$$

We use the tests of differences in discontinuities when testing empirically the predictions laid out in Section 2 regarding when fund managers should misreport returns. We split the full sample of hedge fund returns in sub-samples according to capital flow, strength of flow-performance relation, and risk of being caught misreporting and test whether the discontinuities in return distributions differ among the sub-samples.

4.2.2 Return Quality

The pooled distribution method described above is useful for detecting discontinuities, i.e. misreporting, in large samples of returns. However, due to low number of return observations per fund, the pooled distribution method cannot be used to detect misreporting at the level of individual funds. To properly test the effect of misreporting on ex post performance metrics, we use three different fund level measures for return quality: kink in the return distribution, fraction of unique return figures, and correlation with benchmark returns. These measures are based on those presented in Bollen and Pool (2010) with the exception that we use continuous measures of return quality whereas Bollen and Pool (2010) use binary 'performance flags' of statistically significant return suspiciousness. In Section 5.2 below, we regress fund level measures of risks and returns on these three measures of return quality.

The first measure, *kink*, is a simplified version of the pooled distribution method described above. For each individual fund, we create a histogram of the full time series of returns using (10) to calculate the optimal bin size. Rather than fitting a kernel distribution, we next make the simplifying assumption that the density in the bin left of zero should be approximately equal to the average of the densities of the neighboring bins. Denoting by X_{-2}, X_{-1}, and X_{+1} the amount of observations in the two bins left of zero and the bin right of zero, we calculate *kink* as

$$kink = \min\left(\frac{X_{-1} - \frac{1}{2}(X_{-2} + X_{+1})}{\frac{1}{2}(X_{-2} + X_{+1})}, 0\right). \tag{18}$$

This measure tells how many fewer observations are there in the bin left of zero than would be expected if the density of that bin was equal to the average of neighboring bins' densities. This number is then divided by the average number of observations in bins -2 and $+1$ to arrive at a relative measure. The truncation at zero is made to ensure that we do not infer a fund with more than expected amount of observation in bin -1 to have high quality returns.[13]

Our second measure, *unique*, is simply the number of unique return observations divided by the total number of observations. This is motivated by Straumann (2008) who finds that hedge fund data exhibits too few unique returns. A low fraction of unique returns is suspicious given that the returns are mostly reported with two decimals of a percent.

Our third, and final, measure of return quality is correlation with other asset classes as misreporting might decrease estimates of correlation between fund returns and the risk factors (see Section 3 above). We calculate our return suspiciousness measure 'correlation' as the adjusted R-squared of the optimal factor model for each fund. Starting from a model with ten factors we exclude factors one-by-one to maximize the Akaike Information Criterion (AIC) of the model. Optimal factor model then is the one with highest AIC.[14,15]

All three return quality measures are defined so that a high value signals high quality of returns and low value signal more suspicious returns and a higher likelihood of misreporting. Table 1 shows the discontinuity and difference in discontinuity measures for sub-samples of low (bottom quartile) and high (top quartile) return quality funds.

[Insert Table 1 here]

Low value of *kink* is associated with a much stronger discontinuity in the return distribution than high *kink*. This result indicates that *kink* as a fund level measure of return quality is related to misreporting on the sub-sample level. The same also holds for *unique*, but not for *correlation*. The discontinuities in return distributions are very actually similar for high and low *correlation* funds. This indicates that *kink* and *unique* may be better measures of return quality than *correlation*. However, we

[13] The truncation does not have a material quantitative or qualitative effect on the results.

[14] The factors are are the three Fama and French (1993) equity risk factors, the five Fung and Hsieh (2001) trend following factors, and two bond market factors (return on Barclays US Aggregate index, and the return difference of Barclays US Corporate AAA and Barclays US Corporate BAA indexes). Data for the Fung and Hsieh (2001) factors are available at http://faculty.fuqua.duke.edu/ dah7/DataLibrary/TF-FAC.xls.

[15] Note that low factor model fit is not necessarily indication of misreporting but may also reflect fund manager's skill and informational advantages to pursue profitable idiosyncratic trading strategies. Actually Titman and Tiu (2008) and Sun, Wang, and Zheng (2009) find that hedge fund performance is negatively related to factor model R-squared. However, Bollen and Pool (2010) show that low model fit is associated with hedge fund frauds.

choose to use *correlation* as a measure for return quality as Bollen and Pool (2010) find low correlation to be associated with an elevated risk of fraud.

5 Results

We first examine the empirical validity of the predictions laid out in Section 2 on when hedge fund managers should misreport returns. Second, we study whether misreporting has an impact on fund level performance metrics as is predicted in Section 3.

5.1 When Do Hedge Fund Managers Misreport?

Capital flow. The first prediction of the solution to the simple misreporting problem is that the fund managers should misreport more during positive capital flows than during negative ones. Overstating returns, and hence asset values, during months of positive flows leads the new investor to overpay for her share in the fund. However, overstating returns and asset values during negative net flows results in the fund overpaying the redeeming investor. Hence fund managers should overstate returns more during months of positive net flow and overstate less, or even understate, during negative net flow months.

We test the empirical validity of this prediction by splitting the sample into three sub-samples. The first sub-sample contains the fund-month observations where net capital flow is significantly negative (less than -1%), the second sub-sample contains those where net flow is practically zero (between -1% and 1%), while the third contains observations with significantly positive net flows (over 1%). This split is not constant and a fund may belong to one group in one month and another group in the next month. Figure 4 presents the return histograms and Table 2 presents the test statistics for the three sub-samples.

[Insert Figure 4 and Table 2 here]

First of all, the amount of discontinuity in the zero capital flow sub-sample, 7.9%, establishes the level of misreporting that is not driven by the wealth transfer considerations. This level is statistically significant (z_{DC} statistic equal to 8.3) and of same magnitude that we find for the full sample, 7.2%.

The discontinuity is weakest, 5.0%, for the negative flow sub-sample and strongest, 9.8%, for the positive flow sub-sample. The degree of discontinuity is significantly smaller in the negative flows sub-sample than in the other two indicating that hedge fund managers reduce misreporting when faced with negative capital flows. We argue that this is done to attenuate the negative wealth effects that overstating returns during capital outflows has on the remaining shareholders in the fund. The difference in the discontinuities between the positive flow and zero flow sub-samples is of the predicted sign (discontinuity is stronger during month of positive flow) but, using a one-sided test, only significant at a 10% level (the asymptotic and simulated z_{DCC} statistics are equal to 1.4 and 1.5, respectively). This provides some, although weak, evidence that some hedge fund managers use misreported returns and asset values as means to extract wealth from new fund investors.

Flow-performance relation. The second implication of the model is that fund managers who face more strongly performance-dependent flows should misreport more. We use two differing proxies for the performance-dependence of capital flows: a regression based strength of flow-performance relation and fund age.

Our regression based measure for the fund-specific strength of the flow-performance relationship is measured by b_r in

$$f_t = b_0 + b_r r_{t-1} + \varepsilon_t, \tag{19}$$

where f_t is the net capital flow to the fund during month t and r_{t-1} is the fund's return during month $t-1$. Funds with a top (bottom) quartile b_r coefficient are classified as having a strong (weak) flow-performance relation.[16] Figure 5 presents the return histograms and Table 3 gives the test statistics for the two groups of funds.

[Insert Figure 5 and Table 3 here]

Reported returns of the funds in both categories exhibit statistically significant discontinuity. The value of our DC measure is 5.7% for funds with weak flow-

[16] 24 months of data are required to estimate the flow-performance relation for a fund, a requirement that is met by 4,912 out of 9,714 funds.

performance relation and almost twice as large, 10.4%, for the strong flow-performance relation funds. The difference in the discontinuities is statistically significantly different from zero with asymptotic and simulated z_{DDC} statistics equal to 2.7 and 2.9, respectively. This result is in line with the theoretical prediction that stronger performance-dependence of flows leads the fund managers to misreport returns more often to improve future capital flows.[17]

Our second proxy for the performance-dependence of future capital flows is the fund age. Manager of a newly launched fund is in a similar situation as a young employee in Holmstrom (1999) who should extract higher effort in order to build reputation as the market has diffuse priors of her type. Similarly, the manager of a newly launched fund needs to build a successful track record for his fund to signal high ability and attract new investors. Agarwal, Daniel, and Naik (2004) show that flows into younger hedge funds are, indeed, more sensitive to recent performance than flows in to older funds.[18] Hence, we argue that fund age presents itself as a suitable proxy for performance-dependence of flows in a broader sense than b_r above.

We divide the hedge fund return data into two sub-samples: young funds and old funds. A fund is considered to be young when it has been less than three years since its inception. Funds with more than three years since inception are considered old.[19,20] Anecdotal evidence suggests that many institutional investors require funds to have a three-year track record before considering investment in the fund. Hence the first three years of the fund's life is the most crucial time for building an attractive performance history. Figure 6 presents the return histograms and Table 4 presents the test statistics for the young and old fund sub-samples.

[Insert Figure 6 and Table 4 here]

Again, both groups display significant amount of discontinuity: 10.3% for young funds and 6.1% for the old funds. The difference in discontinuities is also statistically very significant with associated z_{DDC} statistics equal to 3.4 and 3.8. Interestingly, the

[17]The strength of flow-performance relation may vary systematically across fund styles. Hence, we repeat the analysis by ranking funds within style categories instead of across all funds. The results are quantitatively and qualitatively unchanged from those presented in Figure 5 and Table 3.

[18]This is also true for the mutual fund industry, see Chevalier and Ellison (1997).

[19]Note that the split into young and old funds is not fixed. A fund belongs to the young group for the first three years of operations and is then moved to the old group.

[20]The results are quantitatively and qualitatively the same if we used other cut-offs, such as two or four years.

level of discontinuity in the young funds sub-sample is very close to that in the strong flow-performance sub-sample in Table 3 above. Also, DC for old funds is close to that of the funds with weakly performance-dependent flows. These results provide further empirical support to our theoretical prediction that the performance-dependence of capital flows positively affects hedge fund managers' propensity to misreport returns.

Risk of Being Caught. Third, and quite expectedly, our model predicts that managers misreport less when likelihood of being caught or the penalty of being caught is higher. Bollen and Pool (2009) show that there is no discontinuity in return distribution during months prior to audit whereas significant discontinuity prevails during other months, and that funds with more transparent assets exhibit less discontinuity than funds with opaque assets. Both of these findings support the hypothesis that likelihood of being caught misreporting lowers the propensity to misreport.

We extend these analyses using the fund domicile as a proxy for both the likelihood of being caught and the expected penalty if caught. We divide the fund domiciles in two categories: offshore financial centers (OFCs) and non-OFCs.[21] We argue that being domiciled in an OFC may offer fund managers a more lax legal and regulatory environment to report untruthful returns. Figure 7 presents the return histograms and Table 5 presents the test statistics for the OFC and non-OFC hedge funds.

[Insert Figure 7 and Table 5 here]

Contrary to our expectations, discontinuities in the two domicile categories are surprisingly similar: DC measure is equal to 7.9% for OFC funds and 7.8% for non-OFC funds. The z_{DDC} statistics for the difference are practically equal to zero. Hence, it seems that domicile in a non-OFC country does not provide for sufficient risk of being caught or expected penalty to hinder hedge fund managers from misreporting returns.[22]

[21] In accordance with IMF staff assessment (`http://www.imf.org/external/np/ofca/ofca.asp`), we classify the following hedge fund domiciles as OFCs: Andorra, Anguilla, Bahamas, Bermuda, Cayman Islands, Cook Islands, Gibraltar, Guernsey, Isle of Man, Jersey, Liechtenstein, Malaysia, Netherlands Antilles, Samoa, and British Virgin Islands.

[22] US domiciled funds represent 53% of the non-OFC funds and 72% of the non-OFC return observations. Respectively, Cayman Islands is by far the largest OFC domicile corresponding to 55% of funds and return observations.

5.2 Effects of Misreporting

In Section 3 above, we show that under fairly mild conditions return misreporting will induce a negative bias to ex post risk estimates and hence a positive bias to ex post risk-adjusted return estimates. We study the validity of these predictions by regressing fund level measure of risks and risk-adjusted returns on the three measures of return quality introduced in Section 4.2.2. Note that low return quality measure is associated with higher likelihood of misreporting.

First, Table 6 gives the results of regressing fund return volatility and autocorrelation on the measures of return quality. Return volatility has a highly significant positive association with return quality. This provides support for our hypothesis that funds who misreport returns, i.e. are more likely to have lower return quality, exhibit lower ex post volatility of reported returns.

[Insert Table 6 here]

Also, in line with our prediction, return autocorrelation is negatively and significantly related to two measures of quality: *kink* and *unique*. However, contrary to what we expected, return autocorrelation and correlation with other asset classes are positively and very significantly related. Hence the evidence here provides partial support for our prediction that misreporting induces a positive bias in the autocorrelation estimates.

Table 7 presents the results of regressing hedge fund alphas on the measures of return quality. In Panel A the dependent variable is the alpha from a ten factor model while in Panel B the dependent variable is the alpha from a parsimonious model where non-significant factors are excluded to maximize model AIC. [23,24]

[Insert Table 7 here]

Two of the return quality measures, *kink* and *correlation*, are significantly negatively associated with fund alphas while the coefficient of *unique* is not different from

[23]The factors are are the three Fama and French (1993) equity risk factors, the five Fung and Hsieh (2001) trend following factors, and two bond market factors (return on Barclays US Aggregate index, and the return difference of Barclays US Corporate AAA and Barclays US Corporate BAA indexes). Data for the Fung and Hsieh (2001) factors are available at http://faculty.fuqua.duke.edu/ dah7/DataLibrary/TF-FAC.xls.

[24]Since the factors are USD based, we only run the regressions for funds that report their returns in USD to be sure that the factor loadings are appropriately measured.

zero except in the model where the parsimonious model alpha is regressed on *unique* only. Altogether, these results support our prediction that misreporting biases alpha estimates positively.

All the results presented in this section are also of economically significant magnitude. A one standard deviation decrease in *kink*, for example, will decrease volatility estimate by 0.32% (which corresponds to 11% of the cross sectional standard deviation of the estimated volatilities), increase autocorrelation by 0.034 (14% of cross sectional variation) and increase alpha by 0.65% (7% of cross sectional variation). These effects are sizeable enough to affect investment decisions based on ex post performance measures.

6 Robustness checks

Derolles and Gourieroux (2009) show that the discontinuity around zero in the hedge fund return distributions can be due to performance fees charged by the funds. The performance fee is usually charged as a fixed percentage of fund's positive return over a high-water mark. In most cases the high-water mark is defined as the fund's highest historical net asset value (NAV). Such a performance fee scheme shrinks the return distribution asymmetrically by moving the positive returns towards zero while having no effect on the negative returns. This asymmetric shrinking of return distributions may then result in a discontinuity in observed return distributions which might not be captured by a continuous non-parametric distribution.

To check that our results are not driven by performance fees, we repeat our analyses for a sub-sample that only contains those fund-month observations where the beginning-of-month NAV is strictly below the highest historical NAV of the fund. In these cases the potential performance fees would not be applied to the full positive return, but only to that part for which the NAV exceeds its historical maximum. Hence the return distributions will not be shrunk at zero and any discontinuity around zero return will not be due to performance fees.

Table 8 reproduces the analyses above for the sub-sample of 211,457 observations that should be clear of any performance fee induced discontinuity around zero return.

[Table 8 here]

According to Table 8 our results are not driven by the performance fees charged by the hedge funds. The full sample exhibits very strong discontinuity showing that there really is misreporting as Bollen and Pool (2009) argue. Also, most of the results for the sub-samples are quantitatively and qualitatively very similar to those presented above. However, there are two exceptions. First, though discontinuity is increasing in capital flows, the discontinuity in the negative flows sub-sample is not statistically significantly smaller than in the zero and positive flow sub-samples. Second, the discontinuity is somewhat larger for the non-OFC domiciled funds than for those domiciled in OFCs. The statistical significance of this result, however, is weak.

7 Conclusions

In this paper we extend the literature on hedge fund return misreporting in three ways. First, we deepen the analysis of the fund managers' motives to misreport returns. We show that misreporting is more prevalent in funds whose capital flows are strongly dependent on past performance, in young funds, and during months of capital outflows. All these findings are consistent with a simple model where fund manager misreports to attract future flows and to extract wealth from subscribing and redeeming investors. Second, we show that misreporting affects ex post performance measurement. As we predict, misreporting is associated with lower estimates of the riskiness of the fund and higher estimates of risk adjusted returns and return persistence. Third, we extend the methodology of identifying misreporting and testing for differences in misreporting in different sub-samples of return data.

Our results bear significance for investors and academics alike. Misreporting biases ex post performance measures and makes the misreporting funds appear as more attractive investments than they are in reality. Also, misreported returns, and asset values, result in a wealth transfer between trading and non-trading investors in the fund.[25] Investors, both trading and non-trading, should be aware of this effect and its impact on their wealth. Interestingly, we also find that fund domicile does not affect

[25]Trading investors are those who place either subscriptions or redemptions to the fund during the month in question. Non-trading investors are then those that do not change their investment in the fund.

the misreporting propensity: Funds domiciled in overseas financial centers (OFC) misreport as much as those domiciled in non-OFCs. This makes return misreporting relevant also to those investors who, mistakenly, rely on the stricter regulation of onshore funds to prevent such practices.

Our results also suggest that any analysis of the relation between hedge fund returns, flows, age, and flow-performance relation may be biased by the fact that fund managers' propensity to manage reported returns are affected by the flows, low-performance relation and age. Finally, our relatively simple methodology to identify differences in misreporting may prove useful in other applications as well. One example of such is the study of differences in earnings management in various sub-samples of corporate earnings data.

Appendix A: Derivation of Results in Section 2

Management fee. We first investigate how misreporting at $t = 1$ affects the management fee charged at $t = 3$. Under truthfully reported returns the $t = 2$ flow would be equal to $b_0 + b_r r_1$, whereas under misreporting it is $b_0 + b_r(r_1 + m)$. The difference between the two is equal to $b_r m$. The change in $t = 2$ flow will change the management fee by

$$\phi b_r m(1 + r_3). \tag{20}$$

Wealth transfer. Misreporting during non-zero capital flows leads to transfer of wealth between the new and old shareholders in the case of inflows and between leaving and remaining shareholders in the case of outflows. We analyze each of these cases separately and show that the wealth transfer is the same in both cases.

In the case of an inflow ($f_1 > 0$) and under fair reporting, the new shareholder should get

$$\frac{f_1}{A_1 + f_1} = \frac{f_1}{1 + r_1 + f_1} \tag{21}$$

stake of the fund's assets. Due to misreporting, she gets a stake equal to

$$\frac{f_1}{1 + r_1 + m + f_1}. \tag{22}$$

The difference between these stakes is equal to

$$\frac{f_1}{1 + r_1 + f_1} - \frac{f_1}{1 + r_1 + m + f_1} = \frac{m f_1}{(1 + r_1 + f_1)(1 + r_1 + m + f_1)} \tag{23}$$

and the monetary value of the difference (wealth transfer from new shareholder to old ones) is

$$
\begin{aligned}
& \left(\frac{mf_1}{(1+r_1+f_1)(1+r_1+m+f_1)}\right)(A_1+f_1) && (24)\\
= & \left(\frac{mf_1}{(1+r_1+f_1)(1+r_1+m+f_1)}\right)(1+r_1+f_1) && (25)\\
= & \frac{mf_1}{1+r_1+m+f_1}. && (26)
\end{aligned}
$$

The manager's share of this monetary wealth transfer will grow to a net-of-fees value of

$$\frac{mf_1}{1+r_1+m+f_1}\omega_0(1+r_2)(1+r_3)(1-\phi) \tag{27}$$

by time $t = 3$.

In the case of an outflow ($f_1 < 0$), the leaving shareholders should, under fair reporting, sell

$$-\frac{f_1}{1+r_1} \tag{28}$$

shares in the fund, leaving a total of

$$1+\frac{f_1}{1+r_1} \tag{29}$$

shares outstanding. Due to misreported asset values, they end up selling

$$-\frac{f_1}{1+r_1+m} \tag{30}$$

shares with the number of outstanding shares equal to

$$1+\frac{f_1}{1+r_1+m}. \tag{31}$$

The manager owns ω_0 shares. Hence, under truthful reporting, his ownership stake in the fund would be

$$\frac{\omega_0}{1 + \frac{f_1}{1+r_1}} \tag{32}$$

whereas it is

$$\frac{\omega_0}{1 + \frac{f_1}{1+r_1+m}} \tag{33}$$

under misreported asset values. The difference of the two ownership shares is equal to

$$\frac{\omega_0}{1 + \frac{f_1}{1+r_1+m}} - \frac{\omega_0}{1 + \frac{f_1}{1+r_1}} \quad = \quad \frac{mf_1}{(1+r_1+f_1)(1+r_1+m+f_1)}\omega_0 \tag{34}$$

The total value of the fund at $t = 1$ is $1 + r_1 + f_1$. Hence, the monetary value of the difference in manager's ownership at $t = 1$ is

$$\frac{mf_1}{(1+r_1+f_1)(1+r_1+m+f_1)}\omega_0(1+r_1+f_1) \quad = \quad \frac{mf_1}{1+r_1+m+f_1}\omega_0 \tag{35}$$

which will grow to net-of-fees value of

$$\frac{mf_1}{1+r_1+m+f_1}\omega_0(1+r_2)(1+r_3)(1-\phi) \tag{36}$$

at time $t = 3$. This is equal to the wealth transfer in the case of inflows (27).

Appendix B: Derivation of Results in Section 3

Return volatility. The variance of $\tilde{r}_t$ is equal to

$$\sigma_{\tilde{r}}^2 = \sigma_r^2 + \sigma_m^2 + 2Cor(r_t, m_t)\sigma_r\sigma_m, \tag{37}$$

which is less than σ_r^2 when

$$Cor(r_t, m_t) < -\frac{\sigma_m}{2\sigma_r}. \tag{38}$$

Autocorrelation. The first order autocorrelation of $\tilde{r}_t$ is

$$\rho_{\tilde{r}} = \frac{Cov(\tilde{r}_t, \tilde{r}_{t-1})}{\sigma_{\tilde{r}}^2} \tag{39}$$

$$= \frac{Cor(m_t, r_{t-1})\sigma_r\sigma_m + Cor(m_t, m_{t-1})\sigma_m^2}{\sigma_r^2 + \sigma_m^2 + 2Cov(r_t, m_t)}, \tag{40}$$

where we get from (39) to (40) by the assumption of i.i.d. true returns. If $Cor(m_t, m_{t-1})$ and $Cor(m_t, r_{t-1})$ are both negative the autocorrelation is also negative. The autocorrelation is positive (i.e. greater than that of the true returns) when both $Cor(m_t, m_{t-1})$ and $Cor(m_t, r_{t-1})$ are positive. If $Cor(m_t, m_{t-1}) < 0$ and $Cor(m_t, r_{t-1}) > 0$ the autocorrelation is positive when

$$\frac{Cor(m_t, r_{t-1})}{Cor(m_t, m_{t-1})} > -\frac{\sigma_m}{\sigma_r}, \tag{41}$$

and if $Cor(m_t, m_{t-1}) < 0$ and $Cor(m_t, r_{t-1}) > 0$ the autocorrelation is positive when

$$\frac{Cor(m_t, r_{t-1})}{Cor(m_t, m_{t-1})} < -\frac{\sigma_m}{\sigma_r}. \tag{42}$$

Alpha and beta. The estimated factor exposure based on the reported returns is equal to

$$\tilde{\beta} = \frac{Cov(\tilde{r}_t, \lambda_t)}{\sigma_\lambda^2} \tag{43}$$

$$= \frac{Cor(r_t, \lambda_t)\sigma_r + Cor(m_t, \lambda_t)\sigma_m}{\sigma_\lambda} \tag{44}$$

$$= \beta + \frac{Cor(m_t, \lambda_t)\sigma_m}{\sigma_\lambda}. \tag{45}$$

The estimated alpha of the fund then becomes

$$\tilde{\alpha} = \mu_{\tilde{r}} - \tilde{\beta}\mu_\lambda \tag{46}$$

$$= \alpha - \frac{Cor(m_t, \lambda_t)\sigma_m}{\sigma_\lambda}\mu_\lambda. \tag{47}$$

Beta is negatively alpha positively biased when

$$Cor(m_t, \lambda_t) < 0. \tag{48}$$

References

Agarwal, V., N. D. Daniel, and N. Y. Naik (2004): "Flows, Performance, and Managerial Incentives in Hedge Funds," Available at SSRN: http://ssrn.com/abstract=424369.

——— (2007): "Why Is Santa So Kind to Hedge Funds? The December Return Puzzle!," Available at SSRN: http://ssrn.com/abstract=891169.

Asness, C., R. Krail, and J. Liew (2001): "Do Hedge Funds Hedge," *Journal of Portfolio Management*, 28(1), 6–19.

Bollen, N. P., and V. K. Pool (2008): "Conditional Return Smoothing in the Hedge Fund Industry," *Journal of Financial and Quantitative Analysis*, 43(2), 267–298.

——— (2009): "Do Hedge Fund Managers Misreport Returns? Evidence from the Pooled Distributions," *Journal of Finance*, 64(5), 2257–2288.

——— (2010): "Predicting Hedge Fund Fraud with Performance Flags," Available at SSRN: http://ssrn.com/abstract=1569626.

Brown, S. J., W. N. Goetzmann, B. Liang, and C. Schwarz (2010): "Trust and Delegation," Available at SSRN: http://ssrn.com/abstract=1456414.

Burgstahler, D., and I. Dichev (1997): "Earnings Management to Avoid Earnings Decreases and Losses," *Journal of Accounting and Economics*, 24(1), 99–126.

Chevalier, J., and G. Ellison (1997): "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy*, 105(6), 1167–1199.

Cici, G., A. Kempf, and A. Puetz (2010): "Caught in the Act: How Hedge Funds Manipulate Their Equity Positions," Available at SSRN: http://ssrn.com/abstract=1664461.

Cumming, D., and N. Dai (2010): "Hedge Fund Regulation and Misreported Returns," *European Financial Management*, Forthcoming.

Degeorge, F., J. Patel, and R. Zeckhauser (1999): "Earnings management to Exceed Thresholds," *Journal of Business*, 72(1), 1–33.

Derolles, S., and C. Gourieroux (2009): "The Effects of Management and Provision Accounts on Hedge Fund Returns," Available at SSRN: http://ssrn.com/abstract=1485289.

Fama, E. F., and K. R. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33(1), 3–56.

Fung, W., and D. A. Hsieh (2001): "The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers," *Review of Financial Studies*, 14(2), 313–341.

Fung, W., D. A. Hsieh, N. Y. Naik, and T. Ramadorai (2008): "Hedge Funds: Performance, Risk, and Capital Formation," *Journal of Finance*, 63(4), 1777–1803.

Getmansky, M. (2005): "The Life Cycle of Hedge Funds: Fund Flows, Size and Performance," Working paper.

Getmansky, M., A. W. Lo, and I. Makarov (2004): "An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns," *Journal of Financial Economics*, 74(3), 529–609.

Holmstrom, B. (1999): "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66(1), 169–182.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.

Straumann, D. (2008): "Measuring the Quality of Hedge Fund Data," *RiskMetrics Journal*, pp. 65–93.

Sun, Z., A. Wang, and L. Zheng (2009): "The Road Less Traveled: Strategy Distinctiveness and Hedge Fund Performance," Available at SSRN: http://ssrn.com/abstract=1337424.

Titman, S., and C. I. Tiu (2008): "Do the Best Hedge Funds Hedge?," Available at SSRN: http://ssrn.com/abstract=1107627.

Wang, A., and L. Zheng (2008): “Aggregate Hedge Fund Flows and Asset Returns,” Available at SSRN: http://ssrn.com/abstract=1236502.

Table 1: **Return quality and return distributions.** This table presents the return distribution discontinuities for hedge funds with different levels of return quality. *Kink* measures return quality as the difference between realized and expected amount of return observations just left of zero return. *Unique* is the proportional amount of unique return observations. *Correlation* is the R^2 of a parsimonious factor model on the funds returns. *Discontinuity* measures the degree of discontinuity around zero return. *Difference in discontinuity* presents the differences in the discontinuity measure across the different sub-samples. Asymptotic and simulated z test statistics are given in parentheses and brackets, respectively. Figures significant at a 5% significance level are bolded.

	Discontinuity		Differences in
	Low	High	discontinuity
Kink	**25.13%**	**-3.67%**	**28.80%**
	(20.05)	(-3.14)	(16.81)
			[17.22]
Unique	**9.61%**	**4.25%**	**5.36%**
	(9.05)	(3.33)	(3.23)
			[3.58]
Correlation	**9.04%**	**8.94%**	0.10%
	(7.84)	(6.93)	(0.06)
			[0.13]

Table 2: **Capital flows and return distributions.** This table presents the return distribution discontinuities for three pools of monthly hedge fund return observations: those with a negative capital flow, those with zero capital flow, and those with a positive capital flow. *Discontinuity* measures the degree of discontinuity around zero return. *Difference in discontinuity* presents the differences in the discontinuity measure across the different sub-samples. Asymptotic and simulated z test statistics are given in parentheses and brackets, respectively. Figures significant at a 5% significance level are bolded.

	Negative	Zero	Positive
Discontinuity	**5.03%**	**7.86%**	**9.80%**
	(4.48)	(8.27)	(9.40)
Differences in discontinuity			
wrt. zero	**-2.84%**		
	(-1.99)		
	[-2.00]		
wrt. positive	**-4.77%**	-1.93%	
	(-3.11)	(-1.37)	
	[-3.44]	[-1.52]	
Observations	83,789	144,515	124,405
Bin width	0.28%	0.23%	0.22%

Table 3: **Flow-performance relation and return distributions.** This table presents the return distribution discontinuities for two pools of monthly hedge fund return observations: those with a weak flow-performance relation and those with a strong flow-performance relation. The strength of flow-performance relation is measured by the impact of past returns on capital flows. Flow-performance relation in the bottom (top) quartile is classified as weak (strong). *Discontinuity* measures the degree of discontinuity around zero return. *Difference in discontinuity* presents the differences in the discontinuity measure across the different sub-samples. Asymptotic and simulated z test statistics are given in parentheses and brackets, respectively. Figures significant at a 5% significance level are bolded.

	Weak	Strong
Discontinuity	**5.72%**	**10.39%**
	(4.64)	(8.49)
Differences in discontinuity		
wrt. strong	**-4.68%**	
	(-2.69)	
	[-2.89]	
Observations	74,620	74,717
Bin width	0.33%	0.35%

Table 4: **Fund age and return distributions.** This table presents the return distribution discontinuities for two pools of monthly hedge fund return observations: those of young funds and those old funds. A fund is classified as young (old) when it has been less (more) than three years since its launch. *Discontinuity* measures the degree of discontinuity around zero return. *Difference in discontinuity* presents the differences in the discontinuity measure across the different sub-samples. Asymptotic and simulated z test statistics are given in parentheses and brackets, respectively. Figures significant at a 5% significance level are bolded.

	Young	Old
Discontinuity	**10.33%**	**6.07%**
	(11.19)	(7.26)
Differences in discontinuity		
wrt. old	**4.26%**	
	(3.42)	
	[3.80]	
Observations	159,742	192,967
Bin width	0.21%	0.22%

Table 5: **Fund domicile and return distributions.** This table presents the return distribution discontinuities for two pools of monthly hedge fund return observations: those of funds domiciled in an overseas financial center (OFC) and those of funds domiciled in a non-OFC. *Discontinuity* measures the degree of discontinuity around zero return. *Difference in discontinuity* presents the differences in the discontinuity measure across the different sub-samples. Asymptotic and simulated z test statistics are given in parentheses and brackets, respectively. Figures significant at a 5% significance level are bolded.

	OFC	Non-Ofc
Discontinuity	**7.93%**	**7.77%**
	(9.00)	(8.92)
Differences in discontinuity		
wrt. non-OFC	0.16%	
	(0.12)	
	[0.01]	
Observations	169,815	182,894
Bin width	0.23%	0.21%

Table 6: **Return quality, volatility, and autocorrelation.** This table presents the results of cross-sectional regressions of estimates of hedge fund return volatilities and autocorrelations on measures of return data quality. *Kink* measures the degree to which the density of small negative returns differs from from an expected density. *Unique* is the number of unique return figures reported divided by the total number of reported returns. *Correlation* is the adjusted R^2 of a parsimonious factor model. For each measure, a lower value indicates more suspicious reported returns. The dependent variable is the return standard deviation in Panel A and first-order return autocorrelation in Panel B. Fund category dummies are included in all specifications. *Incremental* R^2 gives the difference between the adjusted R^2 of the model and that of a model containing only the fund category dummies. Heteroscedasticity consistent t-statistics are reported in parenthesis, and figures significant at a 5% significance level are bolded. The sample includes all hedge funds and the sample size in all models is 8,113.

PANEL A: Return volatility	(1)	(2)	(3)	(4)
Kink	**1.33**			**1.25**
	(10.70)			(10.13)
Unique		**5.14**		**4.16**
		(14.75)		(12.18)
Correlation			**1.48**	**1.33**
			(8.54)	(7.62)
Adjusted R^2	0.666	0.667	0.666	0.674
Incremental R^2	0.004	0.005	0.004	0.012

PANEL B: Return autocorrelation	(1)	(2)	(3)	(4)
Kink	**-0.14**			**-0.13**
	(-9.71)			(-8.92)
Unique		**-0.10**		**-0.16**
		(-2.39)		(-3.77)
Correlation			**0.24**	**0.24**
			(17.81)	(17.85)
Adjusted R^2	0.504	0.494	0.516	0.527
Incremental R^2	0.011	0.000	0.023	0.034

Table 7: **Return quality and alphas.** This table presents the results of cross-sectional regressions of hedge fund alphas on measures of return data quality. *Kink* measures the degree to which the density of small negative returns differs from from an expected density. *Unique* is the number of unique return figures reported divided by the total number of reported returns. *Correlation* is the adjusted R^2 of a parsimonious factor model. For each measure, a lower value indicates more suspicious reported returns. The dependent variable in Panel A is the intercept term in a 10-factor model where the factors are the three Fama and French (1993) equity risk factors, the five Fung and Hsieh (2001) trend following factors, and two bond market factors (return on Barclays US Aggregate index, and the return difference of Barclays US Corporate AAA and Barclays US Corporate BAA indexes). In Panel B, the dependent variable is the intercept term in a parsimonious factor model resulting from exclusion of non-significant factors to achieve the highest Akaike Information Criterion. Fund category dummies are included in all specifications. *Incremental* R^2 gives the difference between the adjusted R^2 of the model and that of a model containing only the fund category dummies. Heteroscedasticity consistent t-statistics are reported in parenthesis, and figures significant at a 5% significance level are bolded. The sample includes all hedge funds reporting returns in USD and the sample size in all models is 5,772.

PANEL A: Alpha (full factor model)				
	(1)	(2)	(3)	(4)
Kink	**-0.27**			**-0.27**
	(-4.46)			(-4.53)
Unique		-0.10		0.15
		(-0.85)		(1.25)
Correlation			**-0.37**	**-0.38**
			(-5.90)	(-5.96)
Adjusted R^2	0.212	0.209	0.213	0.216
Incremental R^2	0.003	0.000	0.004	0.007

PANEL B: Alpha (parsimonious factor model)				
	(1)	(2)	(3)	(4)
Kink	**-0.26**			**-0.26**
	(-4.58)			(-4.57)
Unique		**-0.22**		0.02
		(-1.99)		(0.19)
Correlation			**-0.38**	**-0.38**
			(-6.17)	(-6.15)
Adjusted R^2	0.201	0.198	0.203	0.206
Incremental R^2	0.003	0.000	0.005	0.008

Table 8: **Robustness checks.** This table reproduces the the z-statistics presented in Tables 2-5 for a sub-sample containing only those fund-month observations where the beginning-of-month net asset value is strictly below the fund's highest historical asset value. This sub-sample should be clear of potential performance fee induced discontinuity around zero return (Derolles and Gourieroux, 2009). Figures significant at a 5% significance level are bolded.

	DC	z_{DC}	Obs.	Bin width
All funds				
All funds	**4.88%**	6.35	211,457	0.25%
Capital flow				
Negative	**5.11%**	4.11	61,214	0.34%
Zero	**5.46%**	5.05	93,368	0.29%
Positive	**6.67%**	5.09	56,875	0.31%
Flow-performance relation				
Weak	**3.97%**	2.83	48,604	0.41%
Strong	**8.85%**	6.24	46.071	0.45%
Fund age				
Young	**7.85%**	7.06	85,924	0.28%
Old	**3.30%**	3.49	125,533	0.29%
Fund domicile				
OFC	**4.43%**	4.39	107,501	0.27%
Non-OFC	**6.75%**	6.59	103,956	0.30%

Figure 1: **Timing of the model.** This figure presents the timing of the model. The decision variable to be chosen by the fund manager is the return reported at $t = 1$.

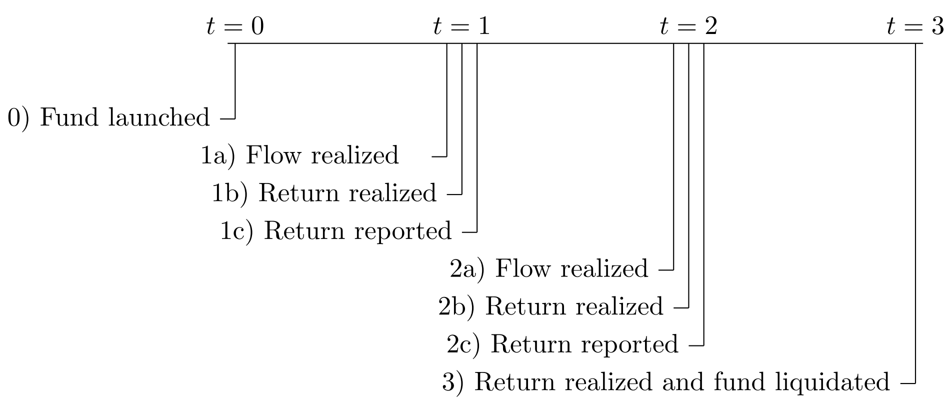

Figure 2: **Comparative statics.** This graphs presents the optimal levels of misreporting (m^*) for varying levels of capital flows (f_1), flow performance relations (b_r), capture likelihoods (κ), and penalties (c)

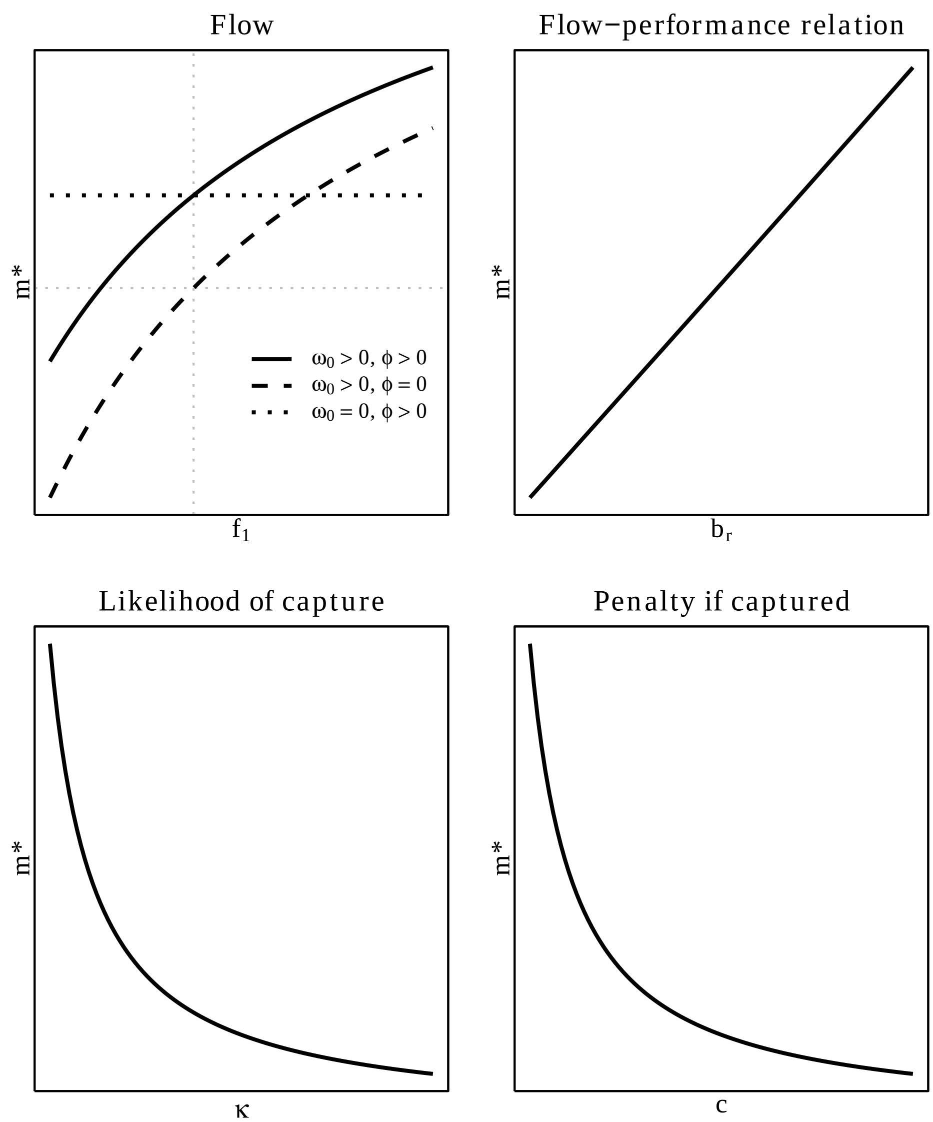

Figure 3: **Return distribution, full sample.** This figure gives the distribution monthly returns on all hedge funds. The top panel gives the histogram of returns with fitted Gaussian kernel density estimate given by the solid line. The bottom panel gives the standard normal test statistics measuring the statistical significance of the difference between the actual density and estimated density in each bin. The dash line gives the 1% critical values. Bins neighboring zero return are highlighted.

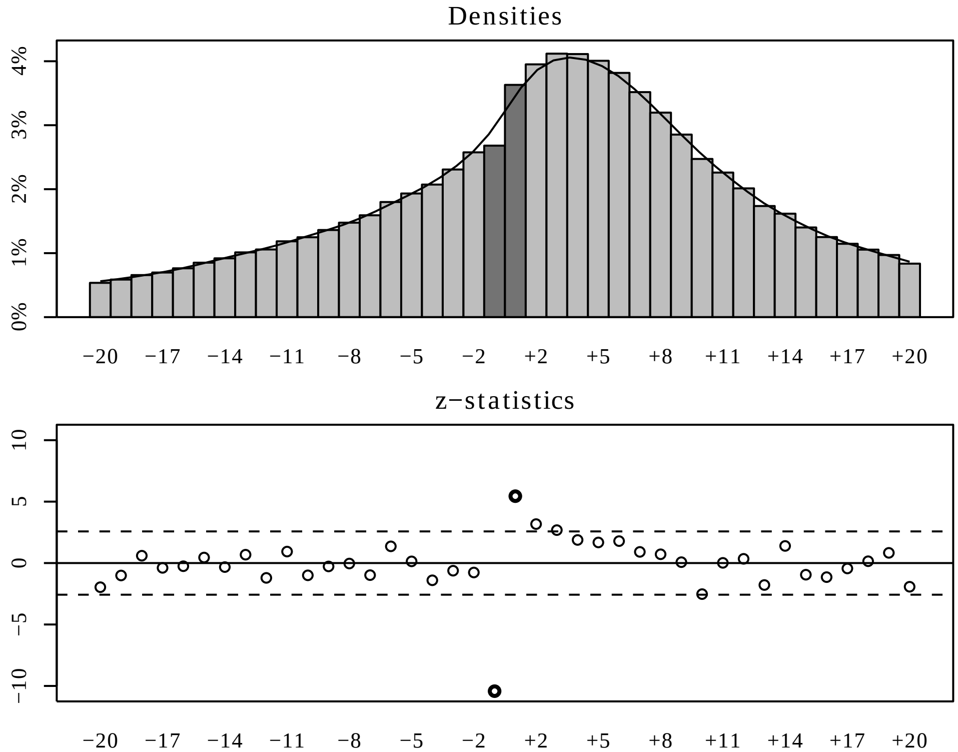

Figure 4: **Capital flows and return distributions.** This figure gives the distribution monthly returns on hedge funds in three sub-samples according to capital flows. The top panel gives the distribution for fund-month observations where capital flow is negative, the middle panel gives the distribution for observations where capital flow is zero, and the bottom panel gives the distribution for positive flow observations. The left panel gives the histogram of returns with fitted Gaussian kernel density estimate given by the solid line. The right panel gives the standard normal test statistics measuring the statistical significance of the difference between the actual density and estimated density in each bin. The dash line gives the 1% critical values. Bins neighboring zero return are highlighted.

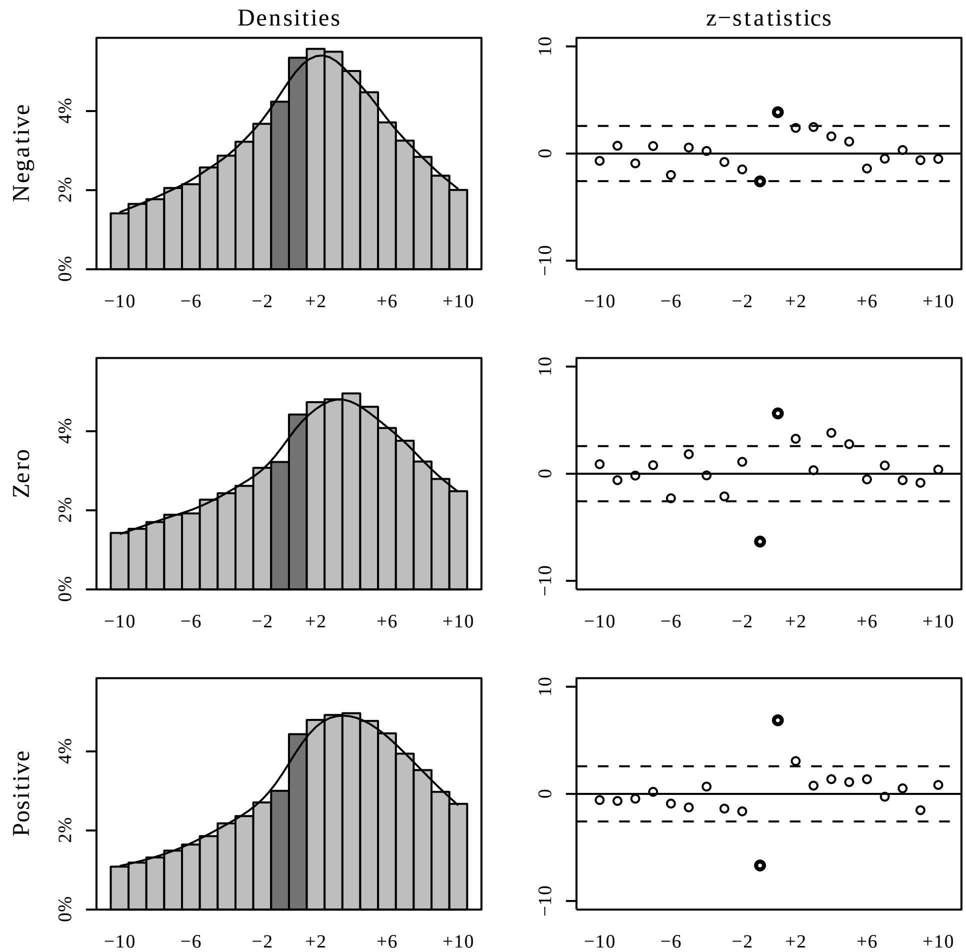

Figure 5: **Flow-performance relation and return distributions.** This figure gives the distribution monthly returns on hedge funds in two sub-samples according to flow-performance relation of the funds. The top panel gives the distribution for the funds with weak flow-performance relation and the bottom panel gives the distribution for strong flow-performance relation funds. The left panel gives the histogram of returns with fitted Gaussian kernel density estimate given by the solid line. The right panel gives the standard normal test statistics measuring the statistical significance of the difference between the actual density and estimated density in each bin. The dash line gives the 1% critical values. Bins neighboring zero return are highlighted.

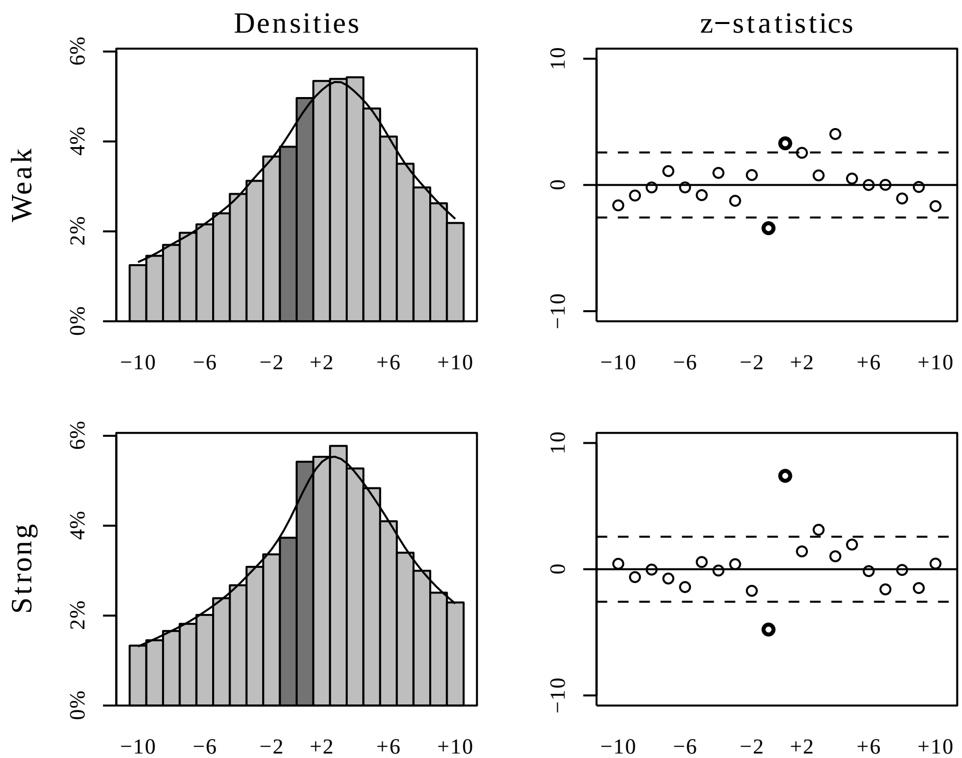

Figure 6: **Fund age and return distributions.** This figure gives the distribution monthly returns on hedge funds in two sub-samples according to fund age. The top panel gives the distribution for young funds (less than three years since inception) and the bottom panel gives the distribution for old funds (more than three years since inception). The left panel gives the histogram of returns with fitted Gaussian kernel density estimate given by the solid line. The right panel gives the standard normal test statistics measuring the statistical significance of the difference between the actual density and estimated density in each bin. The dash line gives the 1% critical values. Bins neighboring zero return are highlighted.

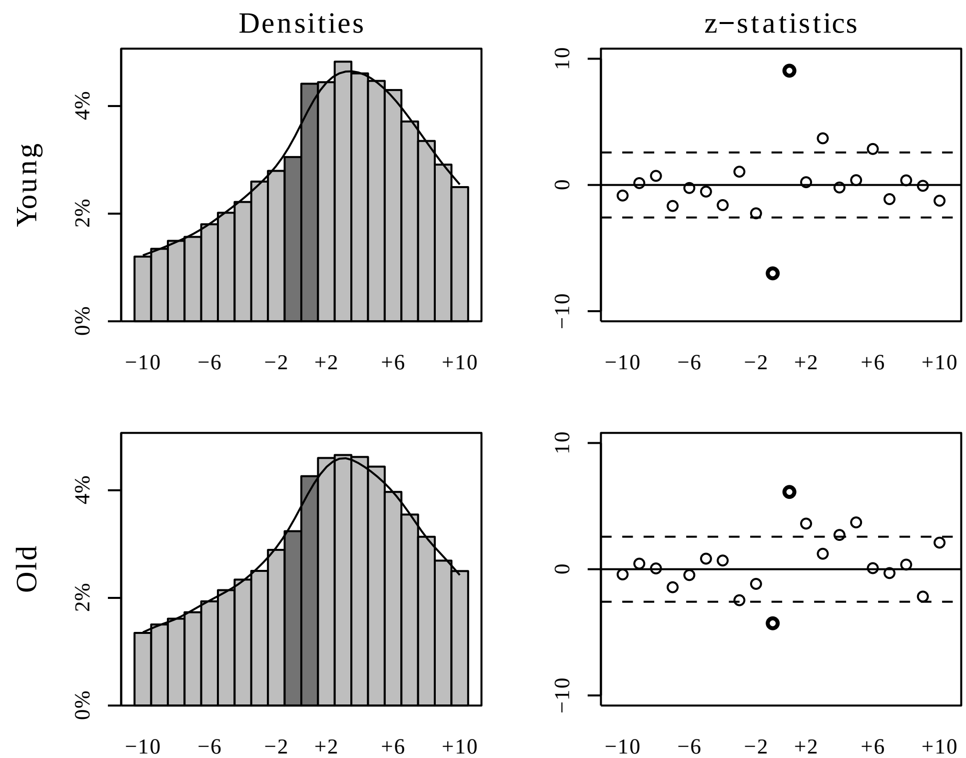

Figure 7: **Fund domicile and return distributions.** This figure gives the distribution monthly returns on hedge funds in two sub-samples according to fund domicile. The top panel gives the distribution for funds domiciled in overseas financial centers (OFC) and the bottom panel gives the distribution for fund domiciled in non-OFCs. The left panel gives the histogram of returns with fitted Gaussian kernel density estimate given by the solid line. The right panel gives the standard normal test statistics measuring the statistical significance of the difference between the actual density and estimated density in each bin. The dash line gives the 1% critical values. Bins neighboring zero return are highlighted.