BIAS IN HEDGE FUND DATA

Nandita Das
Bloomsburg University
and
David L. Muething
Buckley Muething Capital Management Co.
Bethlehem, PA 18018

ABSTRACT

The study of performance persistence in the hedge fund industry is a recent phenomenon. Bias is very closely linked to the issue of performance persistence. The direction of the bias is not clear even for traditional investments like mutual funds. We document the bias in performance data for all classifications of hedge funds in the ZCM/Hedge database. We also document the performance of complete portfolio, surviving portfolio, disappeared funds, and new funds. Hedge funds disappear from the database primarily due to poor performance. Hedge funds have a superior performance when they enter the database. Two kinds of biases act in opposite direction. The overall direction of the bias depends on the relative magnitude of these two biases.

INTRODUCTION

The study of performance persistence in the hedge fund industry is a recent phenomenon. Bias is closely linked to the issue of performance persistence; the direction of the bias is not clear even for traditional investments like mutual funds. For hedge funds, the issue becomes more complicated, because it is possible that hedge funds disappear from the database for various reasons. There is no way to track the disappeared funds. There is no regulatory authority that collects data for hedge funds, nor are the hedge funds required to report their performance. Therefore, it is necessary to estimate bias to better measure performance and to get an idea of the relative performance. Hedge funds have different performance characteristics depending upon their investment strategy. To measure performance accurately it is important to estimate bias for different categories.

LITERATURE REVIEW

Survivorship bias is the effect of considering only the performance of funds that are present in the database at a given time. Since investors are interested only in the funds that are available to them, most databases do not provide the performance of the defunct funds. Performance studies that use only surviving funds will result in biased measures. Much work has been done in providing estimates of survivorship bias for traditional investments. All these studies have documented an upward bias in measures of performance.

For bond funds, Blake, Elton, and Gruber (1993) find an upward survivorship bias of 27 basis points per annum. For equity funds, various researchers have come up with different estimates of survivorship bias. Grinblatt and Titman (1989) come up with several estimates of survivorship bias ranging from 10 to 30 basis points. Brown and Goetzmann (1994) estimate an upward bias of 80 basis points; Malkiel (1994) reports an upward bias of 150 basis points, whereas Carhart (1994) estimates an upward bias of 300 to 500 basis points. Elton, Gruber, and Blake (1996) estimate upward survivorship bias to range from 4 basis points to 97 basis points per annum depending upon the length of study. This apparent difference in estimates of survivorship bias for the same investment class is attributable to differences in methodology used and on the length of the study period.

For Commodity Trading Funds (CTAs), researchers have estimated an upward survivorship bias to range from 350 to 470 basis points. Schneeweis, Spurgin, and McCarthy (1996) estimate survivorship bias to be 120 basis points per annum. Fung and Hsieh (1997) estimate survivorship bias as the difference between the equal-weighted portfolio of existing CTAs and CTAs that have survived the complete study period, and find the bias to be 29 basis points per month, or 348 basis points per year. Diz (1999) studies the performance characteristics of surviving and non-surviving CTAs and find that survivors generated higher returns even on a risk-adjusted basis.

The estimates of survivorship bias for hedge funds range from 16 basis points to 300 basis points.

All the studies in hedge fund survivorship bias have mostly concentrated on estimating bias for the complete database, except Liang (2000). The databases used for all these studies are different. The period of study also differs, as does the procedure of calculating the survivorship bias. The commonality in all these studies is the conclusion that hedge fund survivorship study is different from other survivorship studies.

Hedge funds presumably disappear from the database for two opposite reasons. Some funds disappear because of poor performance and others disappear because they no longer need new money or are closed to the general investors who have access to the database. This is because, unlike traditional investments, hedge funds are not required to disclose any information, and they report to databases with the sole purpose of attracting new investors. There is bias due to the disappearance of funds from the database. There could be bias also in the performance because funds presumably enter the database with positive performance experience. These three factors that lead to bias in the performance measurement could counteract each other depending on the magnitude and direction of bias. We use the term ‘bias’ instead of ‘survivorship bias’ since survivorship bias in mutual funds presumably leads to an upward bias in performance measurement.

We estimate bias for each class and category of hedge funds and for the complete database, including and excluding fund-of funds. Our study is comprehensive and allows for easy comparison of the results with the results obtained by other researchers. We vary the period under study to observe the impact of the length of the study period on bias.

DATA ORGANIZATION AND METHODOLOGY

Three primary databases are popular among academic researchers and the investment industry. Providers of these databases offer different services to the industry. Hedge Fund Research (HFR) database contains more equity-based hedge funds. ZCM/Hedge provides comprehensive coverage of global alternative investment (particularly hedge funds marketplace). TASS is the information and research subsidiary of Tremont Advisers, Inc. TASS classifies managed futures as hedge funds.

We use the ZCM/Hedge database. The ZCM/Hedge database classifies hedge funds into four general classes and eight broad categories of investment styles. The classes are ‘onshore’ hedge fund (HF-US), ‘offshore’ hedge fund (HF-NON), ‘onshore’ fund-of-funds (FOF-US), and ‘offshore’ fund-of-funds (FOF-NON). ZCM/Hedge database categories are shown in Figure 1.

Data Organization

The ZCM/Hedge database provides monthly returns for all the funds. We select a study period from January 1994 and December 2000. ZCM/Hedge data has 180,180 observations of monthly returns for 2,797 funds. A study period dataset from January 1994 to December 2000 is constructed from the available dataset. The dataset is further divided into two parts; from January 1997 to December 2000, and from January 1998 to December 2000.

For the seven-year study period (1994 to 2000), 109,272 observations of monthly return data are available. A total of 74,657 observations of monthly return data are available for the four-year study period (1997 to 2000), and 57,691 observations of monthly return data are available for the three-year study period (1998 to 2000). These four-year and three-year datasets are used to observe the effect of the study period on bias.

Portfolio Construction

The bias can be calculated in two different ways, as in Brown et al. (1999). A surviving portfolio in a particular month during the study period consists of all funds that have reported return until the end of the study period. For example, for the study period that ends in December 2000, a surviving portfolio for each of the previous month will have only those funds that have reported return up to December 2000; the start date of these funds could vary. A surviving portfolio takes into consideration the new money coming in, but does not consider the funds that disappear from the database during the study period. An observed portfolio consists of funds that are in the database for that particular month irrespective of their start and end date. A complete portfolio consists of funds that have reported returns for the complete study period. By definition, a complete portfolio consists of the same funds each month during the study period.
Figure 2 illustrates the portfolio construction process. For the month of October 2000, the surviving portfolio consists of B, C, D, E, and F funds; the observed portfolio consists of B, C, D, E, F, and G funds; and the complete portfolio consists of funds C, D, E, and F. The complete portfolio consists of funds C, D, E, and F.

We calculate equal-weighted and value-weighted portfolio returns. Bias is calculated in two ways: the difference between the return of the surviving portfolio and the observed portfolio (SP-OP), and the difference between the return of the complete portfolio and the observed portfolio (CP-OP). The return data is available for each month, so bias is calculated on a monthly basis and is reported as average monthly bias for each year of study. Bias is calculated for each category, each class, total database including fund-of-funds, and total database excluding fund-of-funds.

**PERFORMANCE CHARACTERISTICS OF DIFFERENT PORTFOLIOS**

We consider both after-fee returns and before-fee returns. A before-fee return is more robust than the after-fee-return, because of the vagaries of the fee structure and the complexities of calculation. In general, hedge funds charge two types of fees: an asset management fee and an incentive fee. The asset management fee is based on amount of the assets in the fund, usually 1%, or 2% per year.

The incentive fee or the “carried interest” is the hedge fund manager’s share in a fund’s profit. Usually this is 20 percent and is paid annually in the United States. Two other important features of a hedge fund fee structure are the hurdle rate and the high water mark. The ZCM/Hedge database provides information on annual fee structure for each of the hedge funds. Subtracting 1/12th of the stated percent fee from the monthly return approximates the administrative fee. Both the hurdle rate and the high water mark feature are considered for computing the incentive fee.

**Observed Portfolio**

Table 1 provides the summary statistics (Before-fee return statistics are not reported here for the sake of brevity) for the categories, the classes of hedge funds and total hedge funds (both including and excluding fund-of-funds), for the study period of 1994 to 2000. The returns of hedge funds vary from a maximum return of 903% (Global US Category) to a minimum return of -99.99%. The maximum volatility of 14% of monthly returns is also for the Global US category.

The category *Sector* has outperformed all other categories on a risk-return basis, followed by the category *Event Driven* for both the study periods of 84 and 48 months.

**Surviving and Complete Portfolio**

We studied the performance characteristics of surviving, and complete portfolio. It appears that the new funds have better performance than the old funds (existing in the database at the beginning of the study period) that disappear. The median return for the surviving portfolio is above the mean return 51% of the time, indicating that probably there are some very poor and very good performers in the categories, pulling the mean towards their performance. No conclusion can be made as to the number of good versus poor performers as it is the magnitude of performance that will affect the mean and median measures.

The surviving portfolio has lower maximum return and higher minimum return, leading to the conclusion that probably the surviving portfolio has less variability in returns compared to the observed portfolio. This is intuitively correct if we relate dispersion in return to the risk taken by the hedge fund managers. The riskier the hedge fund, the greater the chance of its disappearance if the risk ultimately leads to poor performance. Before confirming this conclusion, it is important to see how the disappeared funds perform prior to their disappearance.

The return characteristics vary between the surviving and complete portfolio. This is expected since the two portfolios are constructed differently. If attrition and entry rate have any impact on hedge fund returns, then it is expected that the return characteristics will vary. The rate of entry of new hedge funds will affect the return of the surviving portfolio; whereas, by definition no new funds are included in the complete portfolio. The median return for the complete portfolio is above its mean return 72% of the time, indicating that probably there are some very poor performers in the categories, pulling the mean towards their performance. The minimum return of the complete portfolio is in most cases higher than the minimum return of the observed portfolio.

**BIAS STUDY RESULTS**

We calculate bias for each class, category, and total dataset separately. The bias is calculated both for after-fee return and before-fee return and for equal-weighted and value-weighted returns. The bias is calculated based on the complete and observed
portfolios. The complete portfolio is the portfolio of hedge funds that have survived for the entire study period. By definition, the number of funds in the portfolio remains fixed for each year of study for the complete portfolio. The surviving portfolio consists of funds that have survived until the end of the study period (December 2000), irrespective of the starting date. The observed portfolio consists of all hedge funds that are in the database when the monthly return is calculated.

We also calculate bias as the difference between the surviving portfolio and the observed portfolio (SP-OP). The surviving portfolio considers new funds coming in, but funds that drop out before the end of the study period are not part of the portfolio. In constructing the portfolio, there is an implied assumption that funds that come in anytime during the study period will continue in the database. For example, if a fund enters the database in November 2000 it is considered a survivor even though it has an age of only one month. It is possible that the fund will survive, but it is equally possible that it will not survive if we change the end date of the study period. Calculating survivorship bias, as the difference in the surviving and observed portfolio is useful to understand performance characteristics of incoming funds.

Table 2 shows the bias calculated as the difference in portfolio returns of the observed and complete portfolios and as the difference in portfolio returns of the observed and surviving portfolios for 84-month study period. It also shows the corresponding t-statistics.

The Wilcoxon Signed Rank Test is used to calculate the t-statistic. The significant t-statistics are highlighted. In general, all the classes appear to have a significant t-statistic for bias calculated as SP-OP, and for total hedge funds both including and excluding fund-of-funds.

The direction of the bias differs according to category. If bias is a negative number, then the average performance of survived new funds and disappeared funds (new and old) is greater than the average performance of funds that have survived for the complete study period. It is important to analyze the performance of survived new and disappeared funds. Hedge funds drop out of the database for two entirely opposite reasons, poor performance and probably also because of limitations in the arbitrage opportunities in the investment strategy. If funds drop out because of poor performance, this would, in general impart an upward bias in performance measures, if there is no impact from survived new funds. The analysis becomes more complicated when survived new funds also have an effect on the bias.

Table 2 shows that seven out of nine categories show a negative bias (calculated as CP-OP) for an equal-weighted method of bias computation while six out of nine categories show a negative bias for a value-weighted method of bias computation. Poor performers drop out from the database more frequently than good performers. The bias results are different for different hedge fund categories and for different ways of calculating bias (equal-weighted versus value-weighted portfolio). The bias (calculated as CP-OP) result varies from +0.13% to -0.44% for different categories for the 84 month study period, and the overall bias of hedge funds excluding fund-of-funds is -0.11% per month using value-weighted method. The corresponding results for the 48-month and 36-month study periods are +0.33% to -0.44% and +0.4% to -0.14%, respectively. The bias result of +156 to -528 basis points per year for different categories is different from the results obtained by other researchers. It could be because of the database and the methodology for bias calculation.

We calculate bias on a monthly basis and then average bias is calculated for the study period. This gives a better estimate because the monthly calculations are a good representation of the inflow and outflow of funds from the database, and of the monthly fund return data. Fund-of-funds have a slightly lower bias than hedge funds alone. Bias results for hedge funds, both including fund-of funds and excluding fund-of-funds are not significantly different. Composite bias for all categories can be inaccurate as the bias results of different categories may cancel out. The result for the complete hedge-fund database is provided for comparison with the results obtained by other researchers.

Table 3 compares the category and class bias results (Value-weighted results are available from the authors) for different study periods of 84, 48, and 36 months. As the length of study period increases, the CP-OP bias becomes more negative. This is because the complete portfolio (CP) depends on the length of the study period. The number of funds in the CP portfolio decreases as the length of study period increases. Presumably, these funds have lower risk and possibly lower returns leading to a negative bias. The magnitude of the change in bias result as a factor of study period length is much smaller for SP-OP bias than that of CP-OP bias. It appears that SP-OP is a more stable measure of bias, specifically with the varying lengths of the study period. However, as
mentioned earlier this measure of bias has its limitation because of the way the surviving portfolio is constructed. The advantage of this measure of bias is in its robustness not its accuracy, compared to the CP-OP measure.

**PERFORMANCE CHARACTERISTICS OF DISAPPEARED AND NEW FUNDS**

The performance of funds that disappeared during the study period is analyzed for 36 months prior to their disappearance. In hedge funds, unlike mutual funds, survivorship bias does not necessarily mean that only poor performers have dropped from the database. The performance of new funds that entered the database during the study period is analyzed for thirty-six months after they enter the database.

**Cumulative Average Excess Return (CAER) Of Disappeared funds**

We analyze the performance of disappeared funds to help predict any pattern using the cumulative average excess return. For each class and category, we calculate excess return for the disappeared funds. Excess return is the difference between the fund return and the category return for that month. We do this calculation going back 35 periods for each disappeared fund from the last available monthly return. We calculate average excess return for the class or category by taking the arithmetic average of excess return of all disappeared funds in that category for a particular month. Cumulative average excess return is calculated by linking the average excess returns. The following set of equations represents the process of calculating CAER for disappeared funds.

\[ \varepsilon_{it} = R_{it} - \bar{E}[R_{it}] \] (1)

\[ \text{AER}_t = \frac{\sum_{i=1}^{n} \varepsilon_{it}}{n} \] (2)

\[ \text{CAER}_{t+m} = \left\{ 0 \prod_{j=-m}^{0} (1 + \text{AER}_{t+j}) \right\} - 1 \] (3)

where:

- \( \varepsilon_{it} \) is the excess return of \( i \)-th hedge fund for time \( t \),
- \( R_{it} \) is the monthly return of \( i \)-th hedge fund for time \( t \),
- \( \bar{E}[R_{it}] \) is the category/class return for time \( t \),
- \( \text{AER}_t \) is category/class average excess return for time \( t \),
- \( n \) is the number of funds in the category/class,

\( \text{CAER}_{t+m} \) is category/class cumulative average excess return.

The results of the average and cumulative average excess return of the disappeared funds for representative category, class are shown in Figure 3. All of the nine categories show a strong pattern of a downward trend, indicating that the categories had poor performers as the portfolio of disappeared funds. In the analysis of class of hedge funds, all the four groups show a downward trend and so does the total hedge funds, indicating that the portfolio of disappeared hedge funds consist of mostly poor performers. This is the same conclusion reached by analyzing the return characteristics of the observed and the complete portfolio.

Individual funds drop out for various reasons. Possibly, funds disappear from the database because they perform poorly or also because they no longer wish or need to advertise in the database. Hedge funds report to the database for the sole purpose of attracting new investors. They stop reporting if the fund-manager no longer needs new investors. This could be for a couple of reasons. There may be no arbitrage opportunity that would increase performance for the investment strategy followed by the fund-manager, or the fund manager may decide that it is better to confine to the present number of investors and obtain the necessary fund from the existing investors. Yet, it seems that the majority of the funds disappear due to poor performance. Thus having ruled out the conjecture that, funds disappear from the database because of lack of arbitrage opportunities, it is safe to conclude that the average return of disappeared funds is less than the average return of the funds that have survived for the complete study period. This does not explain why the bias results are negative. If the disappeared funds have low average return, the funds that contributed to the negative bias could be due to the performance of the survived new funds.

**Cumulative Average Excess Return (CAER) Of New Funds**

We analyze the performance of the funds that entered the database after the first month of the start of the study period, namely the new funds, to help predict the pattern in the cumulative average excess return. For each class and category, excess return is calculated for the new funds. This calculation is done going forward 35 periods for each new fund from the start date of the fund. Equation 1 and equation 2 above represent excess return and average excess
return for a category/class. Equation 4 below gives the cumulative average excess return for new funds.

$$CAER_{t+m} = \left( \prod_{j=0}^{m} \left( 1 + AER_{t+j} \right) \right) - 1$$

The results of the average and cumulative average excess return of the new funds for representative category, class, and total hedge fund are shown in Figure 4. Eight out of the nine categories show a strong pattern of an upward trend, indicating that these categories had good performers in the portfolio of new funds.

In the analysis of class of hedge funds, all four groups show an upward trend and so does the total hedge funds, indicating that the new portfolio of hedge funds consist of mostly good performers. There are two caveats to this result. There is an implied assumption that the start date of the fund return data in the database coincides with the date of the entry of the fund in the database. That is, there is no back filing of any fund performance data. Secondly, even if the assumption of no back filing is valid, there is no way of knowing the actual age of the hedge fund, that is, the actual start date of the hedge fund.

It can be concluded that hedge funds’ performance is good for the first 36 months (from the date of registration in the database). It is very likely that funds register after a certain time lag from the date of the inception of the fund. Initially, the hedge fund managers probably manage without having the need to advertise in the database for attracting funds. When these funds reach a particular level of activity, they register in the database to attract more investors. However, it should be noted that not all funds reach this level of activity. In fact, some funds disappear in the process and their existence is never recorded. This should not distort the results much because the bias results are not based on the total hedge fund industry, but are specific to the database under study.

**SUMMARY AND CONCLUSION**

The performance of hedge funds is analyzed for each category and class of the ZCM/Hedge database. The performance of fund-of-funds is inferior to that of other hedge funds.

We estimate bias for different categories, class, and the complete database. For the 84-month study period, the bias result using the value-weighted method varies from +0.13% to -0.44% for different categories and the overall bias of hedge funds excluding fund-of-funds is -0.11% per month. The corresponding results for the 48 month (1997-2000) and 36 month (1998-2000) study periods are from +0.33% to -0.44% and from +0.4% to -0.14% respectively. This result of +156 to -528 basis points per year for different categories is different from the results obtained by other researchers.

The return characteristic of the complete and complete-complement portfolio is compared to that of the observed portfolio. To understand the characteristic of disappeared funds, the cumulative average excess return of the portfolio of disappeared funds is analyzed. All the categories, class and the total hedge funds show a strong pattern of a downward trend, indicating that they had poor performers as the portfolio of disappeared funds.

The average return of disappeared funds is less than the average return of the funds that have survived for the complete study period. If the disappeared funds have low average returns, the funds that contributed to the negative bias could be the performance of the survived new funds. It can be concluded that the new portfolio of hedge funds consists of mostly good performers. However, caution is appropriate here because of the voluntary reporting of hedge funds and possible back filing of return data in the database. It is probably safe to conclude that the hedge funds perform well for the first 36 months from the date of their registration in the database. It is very likely that there is a time lag between the inception of a fund and its registration in the database. If back filing is minimal or negligible, voluntary reporting should not distort the results much since the bias results are not based on the total hedge fund industry, but are specific to the database under study. This study uses the ZCM/Hedge database assuming that the database is a good representation of the hedge fund industry.

Bias study is carried out for three different lengths of study period. The results show that as the length of the study period increases bias calculated as the difference in return between the complete and the observed portfolio (CP-OP) becomes more negative. The analysis is also done for different study periods using the bias measure as the difference in return between the surviving portfolio and the observed portfolio (SP-OP). The study finds SP-OP to be a more stable measure of bias, specifically with varying lengths of study period although it has its limitation because of the way the surviving portfolio is constructed.
FIGURE 1: ZCM/Hedge Classification of Hedge Funds

1. Event Driven
   - 1a. Risk Arbitrage
   - 1b. Distressed Securities

2. Fund of Funds
   - 2a. Global International
   - 2b. Global Regional Established
   - 2c. Global Regional Emerging

3. Diversified
   - 3a. International
   - 3b. Regional Established
   - 3c. Regional Emerging

4. Niche
   - 4a. Long Short
   - 4b. Arbitrage Convertible

5. Global
   - 5a. Long/Short
   - 5b. Arbitrage Stocks

6. Macro
   - 6a. Arbitrage Bonds

7. Long Only/ Leveraged
   - 7a. Event Driven
   - 7b. Global International
   - 7c. Global Regional Established
   - 7d. Global Regional Emerging
   - 7e. Global US
   - 7f. Global Macro
   - 7g. US
   - 7h. Market Neutral
   - 7i. Long/Short
   - 7j. Arbitrage Convertible
   - 7k. Arbitrage Stocks
   - 7l. Arbitrage Bonds

8. Market Neutral

9. Sector

10. Short Sellers

FIGURE 2: Portfolio Construction for the Study of Bias

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Category/Class/Total</th>
<th>Mean Monthly Return (%)</th>
<th>Median Monthly Return (%)</th>
<th>Max Monthly Return (%)</th>
<th>Min Monthly Return (%)</th>
<th>Std. Dev Monthly Return (%)</th>
</tr>
</thead>
</table>

Panel A. Category

- Event Driven
- Global International
- Global Regional Established
- Global Regional Emerging
- Global US
- Global Macro
- US
- Market Neutral
- Sector
- Short Sellers

Panel B. Class

- HF-US
- HF-NON
- FOF-US
- FOF-NON

Panel C. Total

- Excluding FOF
- Including FOF
FIGURE 4: Average and Cumulative Average Excess Return of New Funds

For each category of new funds, the figure shows the average and cumulative average excess return over time. The graphs display the return (%) and time period, illustrating the performance of different fund categories.
**ENDNOTES**

1. A hurdle rate is a compensation feature, whereby fund performance must exceed that of a commonly recognized market index or rate for a given period in order for the manager to earn incentive compensation. When a fund’s performance exceeds an established hurdle rate, incentive compensation is determined as a percentage of the calculated excess. The treasury-bill rate is often used as the hurdle rate.

2. A high water mark is a feature within an incentive compensation structure whereby losses are accumulated in a ‘loss recovery account’ against which future gains are applied in calculating incentive compensation. This feature ensures that incentive compensation is only derived from net positive performance on a cumulative basis.

**ACKNOWLEDGEMENTS**

The authors are thankful to The Foundation for Managed Derivatives Research for providing a research grant and, to Mr. Richard E. Oberuc of LaPorte Asset Allocation System for providing the ZCM/Hedge database.

**REFERENCES**


