Performance evaluation and the potential biases in fund manager return databases

History may judge us all but when it comes to the databases that can provide a clue to past fund manager performance, **ANDREW AINSWORTH**, **DAVID GALLAGHER** and **PETER GARDNER** discover that there is the potential for bias.

Performance measurement is critical to our understanding fund manager ability. However, one important aspect of the performance evaluation process relies on the construction and maintenance of databases containing historical fund manager performance records. Recent evidence in the US suggests that biases in fund data can lead to significant problems in accurately quantifying the universe of funds' performance. In particular, survivorship bias and incubation bias represent some of the most problematic issues for researchers.

Our study provides an opportunity to assess the integrity of one historical fund performance database, and therefore allow for accurate inferences regarding managerial ability. We find minor evidence that both survivorship and backfill biases are present in a highly regarded and widely used database. However, this does not lead to a directional bias in the database which would otherwise lead to inconsistent inferences being drawn on managerial ability.

The most predominant bias we observe is due to alterations in fund histories for months where a return was previously reported (i.e. a record change), although it does not lead to a systematic bias. We also explore the reported returns between newly added and pre-existing fund managers, and find that newly added managers generally outperform extant managers. This may give rise to a selection bias, but this is expected to exist in databases of this nature.

There are a variety of potential biases documented in performance evaluation studies:

- Survivorship bias arises when a manager's return series is removed from a database following liquidation or a merger with another entity. If these fund managers have generally underperformed their peers, the exclusion of their return histories would otherwise inflate the average returns of the surviving managers.
- Backfilling (or instant history) bias arises where new managers selectively report the (superior) returns achieved in periods prior to the date of their actual inclusion in the database. Incubation is one explanation for the backfilling bias.
- Incubation bias occurs where managers operate a number of funds as private and discrete portfolios, and then subsequently decide to publicly offer only the best performing fund(s), thus generating a superior return history that is then backfilled within a database.
- A selection bias would be present if investment managers in the sample did not accurately represent the population of fund managers.

SURVIVORSHIP AND BACKFILL BIASES

The data employed in this study is sourced from the Mercer Investment Consulting Manager Performance

Andrew B. Ainsworth[1]
Australian School of Business, The University of New South Wales

David R. Gallagher PhD
Associate Professor of Finance; Director, Centre for Research in Finance, Australian School of Business, The University of New South Wales

Peter Gardner
Australian School of Business, The University of New South Wales

Analytics (MPA) database for specialist Australian active equity managers. As asset consultants, Mercer use the database as an input into their process of short-listing fund managers for their superannuation clients that are likely to yield superior performance in the future. As such, the MPA database does not include all possible managers in the investment universe, and it is likely that some form of selection bias may exist (although this is expected to be small).

Four MPA database snapshots at approximately three-year intervals are used to undertake the analysis — downloads for the periods ending 31 December 1996, 30 September 1999, 31 December 2002 and 30 September 2005. The selection of these particular dates was based on limited availability of data, with Vanguard Investments Australia providing the earliest data snapshot. We must also point out that Mercer removed the equity component of diversified funds from the specialist funds database in April 2000. As this is not a survivorship issue we added these funds back to the specialist funds to comprise our sample.

Where a specialist fund was used as the equity component of a diversified fund, we removed that fund to avoid including its return series twice. Mercer provided data on the equity component of the diversified funds that were removed from the MPA database in April 2000. Passive and socially responsible managers are excluded from the analysis.[2]

Table 1 documents the entry and exit of managers between consecutive download series. There were in excess of 20 new managers added to the database between each snapshot. In terms of non-surviving funds, six funds ceased reporting based on the four periods used in our analysis. There were 22 instances of backfilling between the four datasets under examination. It must be noted that from April 1999, Mercer Investment Consulting altered its process on the inclusion of new funds' return histories, limiting it to three months for new entrants in their survey.

TABLE 1 FUND ENTRY AND EXIT

	Dec 96	Sept 99	Dec 02	Sept 05
Survivors	n/a	65	94	120
Exits	n/a	3	1	2
Entries	n/a	30	28	21
Total	68	95	122	141
Backfilled	n/a	16	4	2

n/a = not applicable.

Table 2 presents the annualised average and median manager returns, as well as the percentage of funds that outperformed the market index for the period between 1992 and 2002. A preliminary examination of Table 2 reveals that all three statistics change between survey periods, though the sign of the change is not consistent. For example, if we look at 1992, the average return decreased from over 0.7 percent in December 1996 to about –0.1 percent in September 2005. Conversely, for 1996, the average return increased from 16.6 percent in the December 1996 dataset to 17.7 percent in September 2005, with changes of 0.9, zero and 0.2 percent in the intervening snapshots.

Three potential explanations for the differences in the values reported in Table 2 are survivorship bias, the backfilling of historical returns (for months where a fund did not previously report returns) and a record change bias (where a fund's previously reported returns are altered).

TABLE 2 ANNUAL MANAGER RETURNS AND PERFORMANCE

Data download date	Dec 1996	Sept 1999	Dec 2002	Sept 2005
Panel A: Average manager return				
1992	0.75%	0.19%	0.09%	-0.07%
1993	47.42%	47.40%	47.31%	47.61%
1994	-7.47%	-7.41%	-7.45%	-7.37%
1995	19.85%	19.93%	19.89%	19.94%
1996	16.62%	17.49%	17.49%	17.67%
1997	n/a	14.20%	14.19%	13.93%
1998	n/a	11.85%	11.89%	11.39%
1999	n/a	n/a	19.40%	19.27%
2000	n/a	n/a	8.51%	8.66%
2001	n/a	n/a	12.09%	12.10%
2002	n/a	n/a	-8.57%	-8.22%
Panel B: Median manager return				
1992	-0.17%	-0.47%	-0.47%	-0.74%
1993	46.28%	46.28%	46.11%	46.34%
1994	-7.60%	-7.62%	-7.63%	-7.63%
1995	19.75%	19.57%	19.57%	19.60%
1996	15.99%	16.11%	16.09%	16.04%
1997	n/a	13.61%	13.53%	13.46%
1998	n/a	12.17%	12.21%	12.04%
1999	n/a	n/a	17.82%	17.83%
2000	n/a	n/a	8.93%	8.90%
2001	n/a	n/a	11.12%	11.25%
2002	n/a	n/a	-8.43%	-8.11%
Panel C: Average percent of managers out-performing				
1992	61.8%	60.0%	60.2%	59.4%
1993	54.1%	54.7%	52.7%	55.1%
1994	51.2%	54.6%	55.7%	55.0%
1995	50.3%	47.2%	49.1%	47.5%
1996	60.0%	60.0%	63.0%	60.7%
1997	n/a	56.9%	58.3%	56.3%
1998	n/a	50.2%	51.6%	49.6%
1999	n/a	n/a	56.9%	57.2%
2000	n/a	n/a	63.0%	64.0%
2001	n/a	n/a	54.8%	56.1%
2002	n/a	n/a	53.9%	55.4%

n/a = not applicable.

Table 3 documents the size of these three biases for each year from 1992 to 2002.[3] A positive number indicates that the exclusion of non-surviving funds and backfilling of returns has created an upward bias in returns over time. In 1996, the average manager return has increased by 105 basis points as a result of these three biases. At the other extreme, the minimum total bias is –82 basis points in 1992. As such, the bias that does exist in the database is somewhat directionless, and does not consistently inflate the returns of active managers over time. In other words, the few funds that exit are not necessarily the worst performers. Those funds that have their return histories backfilled are not strictly backfilling good performance. And more importantly, survivorship and backfill bias do not explain the majority of the total returns difference across time, with more than half of the absolute deviations a result of changes in fund records between surveys. Also, the backfill bias is substantially smaller in the latter years of our sample, consistent with Mercer's policy change that only three months of historical returns were to be reported by newly added funds after April 1999.

TABLE 3 ANNUAL DATABASE BIASES

	Backfill bias	Survivorship bias	Record change	Total
1992	-16	-10	-56	-82
1993	56	35	-72	19
1994	-13	12	11	10
1995	7	-1	3	9
1996	55	9	41	105
1997	1	-1	-26	-26
1998	-3	-1	-42	-46
1999	-5	0	-8	-13
2000	-1	0	16	16
2001	2	0	-1	1
2002	2	15	18	35

From the analysis above, we can conclude that the presence of survivorship, backfilling and record change biases in the MPA database generally inflates the average returns of active equity managers by only three basis points per annum between 1992 and 2002. This is equivalent to one-fifth of a basis point per month. However, based on our four snapshots, the survivorship and backfill biases are less prevalent in the latter part of our sample. The five years to December 1996 reveals an average upward bias of 27 basis points per annum for these two biases, compared to less than two basis points per year for the six years to December 2002. This is to be expected though, as there is limited opportunity for managers to backfill returns relative to the early part of our sample following Mercer's policy change. In terms of the record change bias, this averages –14 basis points in the first half of the sample versus –7 basis points in the second half.

There is one caveat that must be noted however. The results we find are, to a certain extent, dependent on the download periods employed. As such, they should be interpreted as indicative. The number of non-surviving managers may differ between different download periods. Also, we are unable to perfectly capture the backfill bias given our fixed time event periods.

SAMPLE SELECTION, NEW FUND AND INCUBATION BIASES

An issue of similar importance to the biases discussed above is whether the recently added managers are reflective of the entire universe of fund managers, or if a selection bias exists that inflates the performance of active equity managers. Intuitively, a selection bias would be present, as managers that do not perform well have little incentive to seek inclusion in databases, as well as the difficulty of achieving acceptance from intermediaries in the market.

We also explore the issue of fund incubation, as the newly added funds may have generated a superior return history prior to their public offer that may not be othrwise sustained following their inclusion in the database. For example, it is widely accepted that successful boutiques perform very strongly in the formative years, and once cash inflows become significant, alpha generation becomes more challenging than previously.

In order to ascertain the magnitude of this bias, fund managers are partitioned as either newly added (denoted 'new') or pre-existing ('old'). For the more recent dataset 'new' is defined as those managers that reported in the September 2005 sample but were not included in the December 2002 sample. 'Old' managers are those that were present in both samples. As such, the period over which we can examine differences in manager returns is limited to those months between January 2003 and September 2005, inclusive. Similar classifications are undertaken based on the December 2002, September 1999 and December 1996 return series to provide additional samples to investigate.

To maximise the degrees of freedom, the returns for each manager type are pooled together over time. The 1999 sample aggregates returns between January 1997 and September 1999. For the 2002 sample, this period is from October 1999 to December 2002 and for the 2005 sample, it is between January 2003 and September 2005.

Analysis of a t-test for a difference in means for new and old managers is presented in Table 4. New managers in the 1999 sample significantly outperform old managers by 14 basis points per month. The addition of new managers to the database leads to a 4 basis point increase in total average return per month, or 48 basis points per year.

The 2002 sample of new managers achieve monthly excess returns that are 16 basis points higher than the old managers, on average, with a reported p-value of 1 percent. If we compare the average monthly excess return for old managers with that of the new and old managers combined, we can see that the inclusion of new managers in the sample increases the average monthly excess return by 2.6 basis points per month or 0.3 percent on an annualised basis. Similarly with the 2005 dataset, the difference between the two manager groups is statistically significant with a 13 basis point difference in excess returns per month.

TABLE 4 MONTHLY RETURN DIFFERENCE BETWEEN NEW AND OLD MANAGERS

	Average excess return	p-value	No. monthly returns
Panel A: September 1999 dataset			
New managers	0.23%	0.00	689
Old managers	0.09%	0.00	1796
Difference	0.14%	0.01	n/a
Combined (new and old)	0.13%	0.00	2485
Panel B: December 2002 dataset			
New managers	0.26%	0.00	529
Old managers	0.10%	0.00	2631
Difference	0.16%	0.01	n/a
Combined (new and old)	0.13%	0.00	3160
Panel C: September 2005 dataset			
New managers	0.21%	0.00	366
Old managers	0.09%	0.00	2616
Difference	0.13%	0.01	n/a
Combined (new and old)	0.10%	0.00	2982

The inclusion of new managers inflates excess returns by about 1.6 basis points per month or 0.19 percent per year. Irrespective of the impact of new managers, active managers still outperform the market on average for the time periods examined, although we have not accounted for the costs, taxes or risks incurred by the fund manager (see Table 4).

To shed further light on the performance differential, we compare returns between new and old managers over three different time periods between the start of the sample in January 1992 and the end of the sample, be it September 2005, December 2002 or September 1999. These are all relative to each manager's first reported return in the MPA database on or after January 1992. The "one year after inception" refers to the first 12 months of returns that the manager reported to the database. The earliest this could be is the 12 months to December 1992, though the time period will generally differ across managers. The "period after initial year" is from the 13th month that manager reported until the final month reported (this may be the end of the sample or the earlier if the manager has ceased reporting).

The time period and number of monthly observations will differ for each manager. The one year to the end of sample is the final 12 months of each sample (September 1999, December 2002 or September 2005). Managers had to be in existence for 12, 24 or the final 12 months of the sample to be included in the one year after inception, period after

TABLE 5 AVERAGE MONTHLY MANAGER PERFORMANCE BY NEW–OLD PARTITION

	Excess return				Alpha				Scaled excess return				No. funds	
	New	Old	Diff	p-val	New	Old	Diff	p-val	New	Old	Diff	p-val	New	Old
Panel A: September 1999 dataset														
One year after inception	0.23%	0.18%	0.05%	0.50	0.26%	0.19%	0.07%	0.38	0.39	0.10	0.28	0.23	21	64
Period after initial year	0.13%	0.08%	0.05%	0.31	0.14%	0.08%	0.06%	0.26	0.18	0.12	0.06	0.60	17	63
One year to September 1999	0.37%	0.25%	0.11%	0.15	0.42%	0.27%	0.15%	0.08	0.41	0.21	0.20	0.16	18	47
Panel B: December 2002 dataset														
One year after inception	0.50%	0.18%	0.32%	0.00	0.49%	0.20%	0.29%	0.00	0.32	0.18	0.15	0.21	18	92
Period after initial year	0.21%	0.09%	0.12%	0.17	0.23%	0.09%	0.14%	0.11	0.17	0.12	0.06	0.65	10	90
One year to December 2002	0.21%	-0.01%	0.22%	0.08	0.26%	0.02%	0.24%	0.05	0.22	0.04	0.18	0.06	17	62
Panel C: September 2005 dataset														
One year after inception	0.17%	0.21%	-0.04%	0.71	0.16%	0.22%	-0.07%	0.52	0.31	0.17	0.15	0.25	15	115
Period after initial year	0.29%	0.08%	0.21%	0.00	0.05%	0.09%	-0.04%	0.64	0.16	0.12	0.04	0.57	6	110
One year to September 2005	0.20%	0.10%	0.10%	0.14	0.06%	-0.01%	0.07%	0.20	0.09	0.04	0.05	0.13	15	69

initial year and one year to end of sample groups, respectively. In addition to the excess return performance measure, we also use the intercept from a single factor model:

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \varepsilon_{i,t} \quad (1)$$

where $r_{i,t}$ is the return of manager i at time t and $r_{m,t}$ is the return on the S&P/ASX 200 Accumulation Index at time t. All returns are measured in excess of the 90-day bank-accepted bill rate on a monthly basis. α_i measures risk-adjusted performance and is the intercept of the regression. β_i is an estimate of manager i's systematic risk. We also report the excess return scaled by the market return.

Table 5 presents mixed results on the difference between new and old managers over the various performance measures. For both the excess return and alpha performance measures, there are instances of statistically different performance between the two groups of managers over all three time periods. As such we cannot draw any firm conclusions regarding fund performance and length of returns history.

However, when comparing returns over the same time period, namely the 12 months to the end of each respective sample, we can see that new funds outperform old funds across all three performance measures. In some instances, these average monthly differences are quite substantial. For example new managers achieve an alpha of 0.26 percent in 2002, whereas old mangers are only able to generate alpha of 2 basis points.

There are two problems that arise from the risk-adjustment process that are likely to limit the statistical significance of our results. Firstly, we have to estimate alphas using a minimum of 12 monthly observations of managers' returns, thus precluding the use of those managers with a limited returns history and shrinking our sample size. Secondly, we are unable to pool managers across the three periods due to the fact that a manager deemed as 'new' in 1999 will be reclassified as 'old' in 2002. This reduces the degrees of freedom with which to conduct our tests and adversely impacts the statistical significance of our results, especially in light of the statistical difference between the returns of new and old managers documented in Table 4.

An alternative approach to examining cross-sectional manager performance is to test whether the initial one or three years of performance differ from that achieved subsequently for each manager over time. To be included in this analysis, managers needed to be in existence for at least 24 months to compute the one year average or 72 months to be included in the three year analysis. Table 6 contains the results of this exercise utilising excess returns.

Despite the majority of manager returns having an insignificant difference between the two periods, 9 percent of managers did perform significantly better in their first year of reported operations versus the subsequent performance. In terms of the three-year performance sub-group, 15 percent of managers had excess returns that were significantly higher in their early years.

In contrast only 2 percent of managers performed significantly worse in their first year, relative to their post-first year performance. Two percent of managers had a negative and significant difference between their first three years of returns and those achieved subsequently.

Ignoring statistical significance momentarily, 64 percent of managers performed better in their first year and 72 percent of managers reaped higher returns in the first three years relative to that achieved subsequently. This lends support to the idea that some positive bias does exist as a result of adding new managers to the database. The outperformance by new managers could be a result of fund incubation. An alternative explanation from Chan, Faff, Gallagher and Looi (2005) is that pre-existing, and more likely larger, funds suffer performance erosion from both flow-induced purchases and attempts to avoid market impact costs. Newly added managers are likely to be smaller in size than pre-existing managers and are therefore more likely to avoid these two adverse factors.

TABLE 6 DIFFERENCE IN EXCESS RETURN ACROSS TIME FOR EACH MANAGER

	One year		Three years	
	No. managers	**Percent**	**No. managers**	**Percent**
Positive & Significant	11	9.4%	13	15.1%
Negative & Significant	2	1.7%	2	2.3%
Positive & Insignificant	64	54.7%	49	57.0%
Negative & Insignificant	40	34.2%	22	25.6%
Total	117		86	

CONCLUSION

Performance measurement is critical to our understanding fund manager ability, in terms of the selection, monitoring and review of funds management institutions. One important aspect of the performance evaluation process relies on the construction and maintenance of databases containing historical fund manager performance records. Surprisingly, in Australia little published work exists on this topic, such that we are able to achieve a better understanding of the integrity of historical fund performance databases.

Using a case study approach, our research explores the extent to which certain biases are present in a well used and highly respected institutional fund performance database, sourced from Mercer Investment Consulting. While we find some minor evidence of survivorship bias and backfill bias affecting returns in the early years of the database, these biases have become less substantial in the later years of our sample, and this result can be explained by the enhancements executed by Mercer Investment Consulting.

Changes in fund records comprise the majority of the difference in returns between surveys. This leads us to conclude that, while survivorship and backfill bias are present, their impact is likely to be minimal. However, we find evidence that newly added managers perform better than pre-existing managers, and that this selection bias is likely to enhance the overall ability of active equity managers as a group relative to the original universe of managers. However, this is somewhat expected, as the Mercer database is not static in time and has needed to adapt to capture new players in the market. That said, further investigation of the performance differential between these two groups might be an avenue for future research. In addition, the usefulness of the Mercer database to clients indeed relies on the universe of funds included to have up-to-date information, including the capability of relatively recent starters, such as boutiques. Overall, our study confirms that the Mercer database is well-managed and the significance of database biases is indeed small in nature.

References

Chan, H.W.H, Faff, R.W., Gallagher, D.R. and Looi, A., (2005), 'Fund Size, Fund Flow, Transaction Costs and Performance: Size Matters!', *Working paper*, University of New South Wales and Monash University.

Frontier Investment Consulting, (2001), 'Australian Equity Managers — Just How Good are They?', *Frontier Line*, No.13.

Notes

[1] Corresponding author: andrew.ainsworth@student.unsw.edu.au. This research was funded through an ARC Linkage Grant (LP0561160) involving Vanguard Investments Australia and SIRCA. The authors also thank Vanguard Investments for research support and Mercer Investment Consulting for the use of the MPA database to undertake our analysis. We also thank Eric Smith, Scott Lawrence, Russell Clarke and David Carruthers for helpful comments and suggestions.

[2] Our sample is similar to that of Frontier Investment Consulting (2001). However there are key differences between the two datasets. Our sample is comprised of four snapshots, rather than quarterly surveys and we include only active managers. Our data is sourced electronically, whereas the Frontier sample is from hard-copy records. As such, the data may differ between the original hard copy surveys and our electronic copies if retrospective changes to the database have been implemented. We also add back the equity component of balanced funds, which is not mentioned in the Frontier study.

[3] The Appendix contains two alternative sets of bias estimates. Firstly, we report results using a sample that excludes the diversified funds that Mercer removes from the MPA database. We also present findings using median returns rather than average returns. J

TABLE A1
ANNUAL DATABASE BIASES: EXCLUDING DIVERSIFIED FUNDS

	Backfill Bias	Survivorship Bias	Record Change	Total
1992	-16	103	-55	31
1993	56	166	-84	137
1994	-13	86	4	77
1995	7	34	-31	11
1996	55	144	34	233
1997	1	23	-43	-19
1998	-4	21	-39	-22
1999	-7	0	-20	-27
2000	-2	0	1	0
2001	2	0	-26	-24
2002	2	19	9	30

TABLE A2
ANNUAL DATABASE BIASES: MEDIAN MANAGER

	Backfill Bias	Survivorship Bias	Record Change	Total
1992	-12	-13	-32	-58
1993	-15	12	9	6
1994	-20	15	2	-3
1995	-7	-2	-7	-15
1996	2	17	-15	5
1997	-6	-3	-7	-16
1998	6	-5	-12	-12
1999	-1	0	2	1
2000	4	0	-7	-4
2001	5	0	8	13
2002	2	2	27	32

JASSA ARTICLES

CONTRIBUTIONS

If you are interested in contributing to JASSA, please contact Rosemary McLauchlan on (+61 2) 8248 7617 or email **JASSAcontent@finsia.edu.au** to discuss your ideas.

Download the current JASSA Author Guidelines from JASSA Online page at: **www.finsia.edu.au/members**

ARCHIVED ARTICLES: ONLINE

Articles published in JASSA are available to members online.

Search the online archives back to 1998. Full text of articles is available back to mid-1999.

Find JASSA online index at: **www.finsia.edu.au/members/members_JASSA_Online.asp.**

ARCHIVED ARTICLES: IN PRINT

Authors may order reprints of their published articles.

For details of costs and quantities contact the publisher: Hardie Grant Magazines, 85 High St Prahran, Victoria, 3181. Phone: (03) 8520 6444.

CORRECTION

Benchmarking Discriminatory Power of Credit Risk Rating Models

D.H. Liyana Arachchige, JASSA, Journal of FINSIA, Issue 1, Autumn 2007, pp 6–12

Due to an oversight, the last equation on page 9 is incorrect in the above paper. The author unreservedly apologises for this error and asks the readers to note the following changes. The **Z** values given in each row of Table 2 in the paper should be multiplied by the factor, $2(\boldsymbol{g}+\boldsymbol{b})/\boldsymbol{g}$ using the values in the corresponding rows for the columns labelled 'Goods ($\boldsymbol{g}$)' and 'Bads ($\boldsymbol{b}$)' in Table 1. This change leads to three more failures for the last four **Z** values in Table 2. Figure 2 is also slightly affected and the second paragraph on page 10 should be read accordingly. The corrected version of the paper can be obtained by writing to the author's email address given in the paper. The following formula encompasses the above changes.

$$Z = \frac{(2(g+b)/g)(A-A')}{\sqrt{V(G)+V(G')-2Cov(G,G')}}$$

where g and b are respectively total number of non-defaults and total number of defaults.