Performance Persistence in Mutual Funds: An Independent Assessment of the Studies Prepared by Charles River Associates for the Investment Management Association

Professor David Blake PhD*

Birkbeck College

University of London

London W1T 1LL, UK

and

Professor Allan Timmermann PhD*

University of California, San Diego

9500 Gilman Drive

La Jolla, CA 92093-0508, USA

Prepared for the Financial Services Authority

April 2003

This report is available at:

http://www.pensions-institute.org/reports/performance_persistence.pdf,

* The authors thank Russ Wermers for conversations related to performance persistence. The authors will not enter into any correspondence with members of the public concerning this report.

Table of Contents

Background and Terms of Reference **1**
Executive Summary **2**
1. An Introduction to Risk, Return and Performance Measures **4**
1.1 Risk and return 4
1.2 Performance measures 6
1.3 Implications for investment decision making 8
2. A Review of Previous Studies Prepared for the FSA on Performance Persistence **9**
3. A Review of the CRA Studies of Performance Persistence in Mutual Funds **12**
3.1 The first CRA report (CRA1) 12
3.2 The second CRA report (CRA2) 13
4. An Assessment of the CRA Studies **15**
4.1 The empirical evidence 15
4.2 Raw returns versus risk-adjusted performance measures 18
4.3 The case for using raw returns 18
4.4 Risk and persistence in raw returns 20
4.5 The case for adjusting for risk 21
4.6 Should performance be risk-adjusted? 23
4.7 Persistence and market efficiency 24
5. An Assessment of CRA's Statistical Procedures and Hypothesis Tests **26**
6. Weaknesses in CRA's Analysis and Conclusions **30**
6.1 The use of raw unadjusted returns 30
6.2 The robustness of the statistical analysis 30
6.3 Ignoring the differential effect of charges on the performance of active and passive strategies 30
6.4 The exclusion of tracker funds from the analysis 31
6.5 Failure to report evidence on the performance of switching strategies 32
6.6 Failure to identify the sources of performance persistence 34
6.7 Failure to analyse flows and sizes of funds 35
6.8 Dealing with these weaknesses 35
7. Can (and If So How) Performance Information Inform Investors' Decisions? **37**
8. Conclusions **39**
Glossary **43**
References **46**
Appendix: About the Authors of this Report **49**
Professor David Blake 49
Professor Allan Timmermann 50

Background and Terms of Reference

1. In 2002, Charles River Associates produced for the Investment Management Association (IMA) a report on the persistence of performance in mutual funds (*Performance Persistence in UK Equity Funds – An Empirical Analysis*, October 2002). This report differed from previous studies in focusing on raw returns data. It found greater evidence of persistence in raw data than the previous studies had found in risk-adjusted returns.

2. The IMA has asked the Financial Services Authority (FSA) to consider whether CRA's report justifies a change in policy about the FSA's Comparative Tables. The policy has been that it would be inappropriate for the FSA to include past performance information in these tables since:

    - risk-adjusted performance data are the data that ought to matter to consumers;
    - there is very little persistence in the risk-adjusted performance of mutual funds;
    - information on past performance is in any event available to those who want it;
    - provision of this information by the regulator might attach to it an importance that it does not have;
    - in particular, it might lead consumers to pick funds on the basis of past performance, rather than cost (and the FSA has included cost in the Comparative Tables because it believes that on average cost is the most reliable available indicator of fund performance).

3. The FSA is genuinely open to important new research and has therefore commissioned a short analysis from independent consultants. The terms of reference were to consider:

    - the merits of CRA's approach and methodology, including, in particular, the advantages and disadvantages of using performance data that have not been adjusted for risk;
    - the reliability of CRA's findings, taking into account, for example, the statistical tests used and the important choices made in generating the contingency tables (such as the treatment of dead funds);
    - weaknesses in the CRA's analysis and conclusions;
    - the merits of the part of the FSA's policy position that is set out in the final two bullets in 2. above.

Executive Summary

1. In 2002 Charles River Associates produced for the Investment Management Association two studies that provide valuable contributions to the ongoing debate about performance persistence. These studies reviewed the extensive empirical literature on mutual fund performance and also conducted an empirical analysis of the performance of a large sample of UK unit trusts.
2. The second of the two studies (*Performance Persistence in UK Equity Funds – An Empirical Analysis*, October 2002) concluded that there is strong evidence of persistence in the raw return performance of UK unit trusts at a variety of investment horizons. It argued that this finding was robust to a number of factors, including charges, survivorship bias, investment horizon and the assumptions underlying the statistical analysis.
3. We do not believe that CRA's decision to conduct their analysis of performance persistence using raw returns is justified. The absence of risk adjustment in the performance study means that performance persistence figures are likely to divide mutual funds according to their levels of risk exposure and not according to the degree of fund manager skill or value-added. High-risk funds are more likely to be top performers (particularly in the long run), while low-risk funds are more likely to be among the worst performers. Performance figures based on raw returns are likely to induce investors to hold more mutual funds with high risk and fewer mutual funds with low risk, regardless of whether the returns generated by these funds are justified by their level of risk exposure.
4. We do not think that the publication of persistence measures that induce investors to seek more risk are in the best interests of investors. No skill on the part of a fund manager is required to select a high-risk strategy. To the extent that superior performance in raw returns merely reflects higher levels of risk, it can be replicated at a lower cost (since active fund management fees can be avoided) by using a geared or leveraged investment strategy based on tracker funds. If the purpose of performance league tables is to enable investors to select the most skilfully run mutual funds, performance should therefore be measured and reported on a risk-adjusted basis.
5. CRA's choice of raw returns as the basis for their analysis also means that the chosen measure of performance is likely to be strongly correlated across funds. Such dependencies across mutual funds' raw returns make it difficult to assess the statistical

significance of the persistence results reported by CRA.

6. Further empirical work is therefore required. We suggest that an analysis of persistence in risk-adjusted performance be carried out on the basis of the comprehensive dataset used in the CRA reports. This would provide a valuable complement to the findings reported in the existing CRA studies. We also recommend modifying parts of the statistical analysis conducted by CRA.
7. Studies in the empirical literature on persistence in mutual fund performance provide fairly strong evidence of persistence in negative performance, but much weaker evidence of persistence in positive performance.
8. If mutual funds are left to advertise past performance figures on their own and are free to choose the way in which they do so, this is unlikely to bring as much attention to past underperformance as would seem appropriate given its persistence. Superior performance, though less persistent, is likely to be more aggressively promoted through voluntary advertisement than underperformance. Hence, we believe it is important for both existing and potential investors in mutual funds to have an independent outlet that reports persistent underperformance on a risk-adjusted basis where this exists.
9. We therefore believe that there is a reasonable case for arguing that risk-adjusted past-performance data should be included in the FSA's Comparative Tables. This is not because of the traditional argument over whether superior performance might or might not persist, which we regard as inconclusive, but rather because of the evidence that inferior performance seems to persist. We believe it is important for investors to have easy access to reliable information on underperforming funds so they can modify their investment strategies accordingly. Additional research needs to be undertaken on the most suitable risk-adjustment procedure as well as the most suitable investment horizon over which past performance is reported.

1. An Introduction to Risk, Return and Performance Measures

1.1 Risk and return

1.1. Investment is the sacrifice of certain current consumption for generally uncertain future consumption. All investors will want to be compensated for the time that their funds are tied up: in addition to a real rate of return they will expect a term (or horizon or liquidity) premium which depends on the length of their investment horizon.

1.2. Most investors are also risk averse, many of them extremely so, and such investors will want to be compensated for the risk (i.e., potential reduction in the value of the initial sum invested) to which their investment is subjected over the investment horizon: they will therefore expect an additional risk premium. Only risk-neutral investors will make investment decisions by assessing the expected returns on investments without considering the risk. Risk-seeking investors will also be interested in the risk of their investments but clearly in a rather different way from risk-averse investors.

1.3. Risk can be hard to quantify and assessing attitudes to risk even more so. Nevertheless, finance theory tells us that risk and risk aversion are critical factors in the determination of most investors' portfolios.

1.4. It is difficult to provide a single best summary measure of the risk attached to a particular investment such as a mutual fund because the best measure of risk depends on whether the mutual fund is held in isolation or is part of a larger portfolio of assets. In the latter case, risk can actually be lower than the sum of the risks attached to the investments held separately. This is because a portfolio generally contains investments with offsetting risks: when one investment is doing well, another might be doing badly and vice versa, and so the overall risk from combining the two investments can be reduced. The process of combining investments in a way that reduces overall risk is called diversification.

1.5. Standard deviation is one commonly used measure of risk. It is useful both when considering an asset held in isolation and when considering a portfolio as a whole. In

the case of a single asset, standard deviation measures the dispersion of an asset's returns around its mean or average value. It is equal to the square root of the variance of the asset's returns. The variance of the returns on a portfolio comprising two assets is equal to the weighted sum of the variances of the two assets' returns plus twice the covariance between these returns. Again the standard deviation is the square root of the variance.

1.6. Hence, while standard deviation may be an appropriate measure of risk for a single asset, the presence of covariance terms means that when two or more assets are held in combination, the individual assets' standard deviations are no longer sufficient measures of the risk of the combined asset holdings.

1.7. A relatively simple measure of risk that accounts for an asset's correlation with other assets in a larger portfolio is provided by the classic Capital Asset Pricing Model (CAPM) devised by the Nobel Prize winning economist William Sharpe (1963). According to this model, the expected return on a risky asset at time t, $E(R_t)$, equals the risk-free rate at time t, R_{ft}, plus the asset's risk premium (which is equal to the product of the asset's beta (β) and the expected excess return on the market portfolio, $E(R_{mt} - R_{ft})$):

$$E(R_t) = R_{ft} + \beta E(R_{mt} - R_{ft}).$$

The expected excess return on the market portfolio is also known as the market risk premium and β is a measure of the risky asset's sensitivity to movements in the market risk premium (i.e., it is a measure of the market risk of the asset).

1.8. According to the CAPM, assets with higher betas can expect to earn a higher rate of return. This is only in expectation (that is, 'on average'), however. The downside is that such assets will also experience more volatile returns so that, over any given investment horizon, they also stand a higher chance of losing more money whenever the market return ends up being less than the risk-free rate.

1.9. More recently, models based on multiple risk-factors, besides exposure to the market, have been proposed. The so-called Fama-French factors, namely book-to-market ratio

and company size, seem to capture systematic differences in returns (Fama and French (1992)). If these factors are viewed as systematic risk-factors, it would also be necessary to adjust asset returns for their exposure to such factors. Momentum is another risk-factor adopted in many empirical studies, c.f. Jegadeesh and Titman (1993) and Carhart (1997).

1.10. Even though it is difficult to establish a single, comprehensive measure of risk that has the same relevance to all investors irrespective of their other asset holdings, this does not mean that risk indicators should not be reported. Taking into account some measure of risk, however imperfect, is clearly preferable to ignoring risk altogether.

1.2 Performance measures

1.11. Should superior or poor performance always count for the same? Theory would suggest the answer is no. If a mutual fund underperforms on average, but nevertheless manages to outperform during recessions or during periods when the rest of the market is down, then it provides a form of insurance to those who invest in it. Likewise, superior performance during times when the market is up by a significant amount is not as valuable as if it occurred in a down market.

1.12. Managerial skill cannot be assessed simply by reference to a mutual fund's raw returns. Even very good fund managers are likely to experience negative returns during periods where the market goes down by a significant amount. Likewise, even bad managers are likely to experience high returns when markets go through a rapid rise. Performance must therefore be measured in relation to a benchmark that reflects the returns on other comparable investments.

1.13. An unskilled fund manager can increase the fund's average raw return over time simply by increasing its risk exposure. The simplest way to do this is through gearing or leverage, that is, borrowing money to invest in additional risky assets. However, this comes at the expense of increasing the fund's risk as measured by its beta. Only skilled fund managers can improve on the fund's ratio of expected excess return to market (i.e., beta) risk (known as the Treynor (1965) ratio).

1.14. A frequently used measure of the skill of a fund manager is Jensen (1969)'s alpha (α). This is the expected return a fund manager earns over-and-above the return expected from the CAPM:

$$E(R_t) = \alpha + R_{ft} + \beta E(R_{mt} - R_{ft}).$$

A positive α indicates a genuinely skilled fund manager whose decisions add value to the fund, while a negative α indicates a poorly performing manager whose decisions subtract value from the fund. Jensen's α can also be measured in relation to a multi-factor performance benchmark.

1.15. Alpha is a measure of risk-adjusted performance since it controls for the fund's exposure to risk. Hence, if a particular mutual fund earns a high average return not because the manager is skilled but simply because the manager took on a high level of risk (i.e., chose a portfolio of assets with high betas), this will not show up in Jensen's alpha. In contrast, if raw returns are used as a performance measure, even a poor fund manager with a negative value-added ($\alpha < 0$) could appear to outperform by choosing a high-risk (high β) strategy. This is because on average, over long periods of time, the mean excess return on the market portfolio, i.e., the market risk premium ($R_{mt} - R_{ft}$), tends to be positive.

1.16. If a mutual fund holding represents a small part of an investor's total portfolio, then alpha is an appropriate measure of the mutual fund's performance, since the contribution to overall portfolio risk depends on how much market risk the mutual fund takes on. It is thus necessary to control for market risk exposure. An alternative to alpha which gives a similar if not identical ranking of mutual funds is the Treynor ratio.

1.17. If, on the other hand, the mutual fund holding comprises a large part or indeed all of the investor's total portfolio, then a more appropriate measure of risk is the standard deviation of the fund's returns which includes both the market risk and the fund-specific (or idiosyncratic) risk. Performance measures such as the Sharpe (1966) ratio (the ratio of the mean excess return to the standard deviation of portfolio returns) can be used in this situation.

1.3 Implications for investment decision making

1.18. The most important investment decision an investor makes is likely to be his or her overall strategic asset allocation (SAA), i.e., how much of total wealth to invest in broadly defined asset classes, such as UK and international stocks, bonds, property and cash. The SAA should be based on a careful consideration of both available investment opportunities and key characteristics defining the current stage in the investor's life cycle (such as existing and future family commitments, the anticipated length of the remaining working life, and projected earnings over this period). The SAA also depends on the level of risk that is acceptable to the investor. The selection of specific mutual funds should therefore be seen in the context of and in relation to the SAA.

1.19. One way to conceptualise the investment process is to assume that investors have a risk budget that allows them to accept a certain level of total risk across all their investment holdings. If they spend this risk budget on high-risk, actively-managed mutual funds that do not add any value on a risk-adjusted basis, then they are unlikely to outperform a simple passive strategy invested in tracker funds with the same overall level of risk. Of course, in practice investors will trade off risk against expected returns in their portfolio decisions.

1.20. With this brief introduction to risk and return in investment portfolios, we now turn first to review the current debate on performance persistence promoted by the FSA and then to review and assess Charles River Associates' contribution to the debate.

2. A Review of Previous Studies Prepared for the FSA on Performance Persistence

2.1. Three studies prepared for the FSA are particularly relevant to the current debate about past performance:

- Bacon and Woodrow (1999), *Comparative Tables*, September (B&W);
- Rhodes, M (2000), *Past Imperfect? The Performance of UK Equity Managed Funds*, Occasional Paper Series No. 9, Financial Services Authority, August;
- Financial Services Authority (2001), *Report of the Task Force on Past Performance*, September.

Since the CRA reports were partly prepared as a response to these studies, we briefly document their key arguments and findings.

2.2. In considering the use of past performance information, B&W concluded: '*B&W suggest that to provide retail investors with information that would enable them to make informed choices between competing investment products, it would be necessary to present qualitative as well as quantitative data on a like for like basis and for this to be capable of external, objective verification. Further, the overwhelming emphasis should be on the former [covering knowledge of the investment philosophy and process applied by fund management organisations, an awareness of the management structure within which they operate and the portfolio construction and risk control systems in place across the range of funds under management], with the latter used merely as a validation. In the absence of such comparative information we suggest that performance data taken in isolation will provide no reliable guide to future success and could be very misleading. We recommend that past performance is rejected as an indicator for Comparative Tables…In the context of Comparative Tables, B&W recommend that for, unit-linked products, emphasis is placed on the Risk & Return Key Indicator*' (p. 76).

2.3. The B&W study also found that in its comparison of international comparative finance information, past performance is '*often adjusted for risk before ranking and assigning a star rating*' (p 124).

2.4. Rhodes (2000) argued as follows:

- It is skills that are important to consumers and it is skills that are being tested: '*Persistent investment performance implies that some fund managers are able consistently to outperform their peers. This implies that the funds' managers must either have access to information that is not widespread or make use of available information in a better and speedier way than most other managers. This will allow them to trade stocks, so that net of the costs of purchasing and selling stock, they achieve a high relative return. Claims of persistently high relative performance must therefore relate to more skilled fund management, which might be found in better research or trading…[A]s markets become more efficient it will become more difficult for any fund manager consistently to beat her peers, to any significant degree. The gains to be made from conducting ever more thorough research will diminish. Therefore theoretically at least, it seems unlikely that a given fund manager could maintain a meaningful outperformance of her peers for a long period of time*' (p 7).
- *'[F]irms are able to publish past performance figures on their own account. So the decision by the FSA not to publish this comparative information will not remove performance information from the market'* (p.8).
- *'[A] fund which exposes the investor to a relatively low level of risk might still be attractive even if the expected rate of return is low. A measure of relative performance should therefore account for the level of risk to which the investor was exposed'* (p.10).

2.5. The FSA Task Force on Past Performance (2001) made the following points:

- In choosing which funds to purchase, consumers looked at: past performance, brand reputation and editorials in the financial press. '*All respondents* [to the survey of consumers it commissioned] *felt the balance of risk–return within the advertising was disproportionately in favour of 'return', with the risk warnings not overt enough and meaningful information on the risk of losing money virtually non-existent'* (p.9).
- *'Both firms and* [advertising] *agencies* [which were also surveyed] *believed that consumers use past performance and brand to select funds, with independent*

endorsement (via press editorial) also playing a key role. This perceived demand for past performance, and firms' experience that response rates decline if performance was omitted from the advertising, encouraged the current wide use of the information' (p10).

- *'All respondents agreed that past performance and brand was key to the choice of investment product. The less sophisticated investors placed great faith in the brand, although it also remained important for the more knowledgeable consumer. Past performance was seen as fundamental, and if not present would be sought elsewhere by many. Consumers were using performance figures as a reasonable indicator of future performance and benchmarking their (minimum) expectations against perceived savings account rates.*' (p11).

2.6. In summary, these reports argue that information on past performance is of little value, that retail customers cannot effectively exploit this information, and that the provision of this information could potentially mislead investors. As such they provide a significant challenge to those who want to see past performance information included in the FSA's Comparative Tables. However, it should also be noted that the reports concentrate their attention almost exclusively on above-average performance persistence and have little to say on below-average performance persistence.

3. A Review of the CRA Studies of Performance Persistence in Mutual Funds

3.1. In 2002 Charles River Associates produced two studies for the Investment Management Association (formerly the Association of Unit Trusts and Investment Funds) on performance persistence in UK equity mutual funds (the global term for unit trusts).

3.1 The first CRA report (CRA1)

3.2. The first CRA report (*Performance Persistence in UK Equity Funds – A Literature Review*, January 2002) reviewed the existing academic literature on the performance persistence of mutual funds, most of which relates to UK and US studies of equity funds. We refer to this as the CRA1 report. This report was critical of the earlier reports prepared for the FSA. It found the following:

- Many studies both for the UK and US have found strong evidence of persistence amongst poorly performing funds ('losers repeat') and mixed evidence for persistence amongst top performing funds ('some winners can repeat for a while at least').
- Where present, the persistence of superior performance lasted between 1 and 3 years.
- Most studies were interested in testing the skills of the fund manager over time, rather than assessing persistence from the consumer's (i.e., investor's) perspective.
- Performance persistence may have fallen over time as securities markets become more efficient.
- Superior performing funds attracted inflows of new money, but poorly performing funds did not experience outflows of funds at the same rate.

3.3. The report concluded that:

- '*Discounting the value of information on persistently poorly performing funds throws up a significant regulatory risk…[C]onsumers invest and spend a disproportionate amount of time considering top funds where past performance may be a weak predictor. At the same time, consumers tend to keep their losing funds,*

when in fact they should not. Therefore, there may be a valid role for regulation in the use of past performance information' (p.1)...'*With this level of importance being placed on past performance, academic papers have called for regulation to improve its use rather than constrain it*' (p. 25).

- '*If consumers were discouraged from taking account of past performance: management incentives would be weakened and underperformance could become more prevalent; and the flow of funds into poorly performing funds is likely to increase*' (p. 30).
- '*Indeed the FSA approach appears to believe only charges or performance can be used, ignoring the possibility that charges and performance are jointly important*' (p. 30).
- What is important to consumers is *whether* performance persistence exists, not *why* it exists (p. 4).

3.2 The second CRA report (CRA2)

3.4. The second CRA report (*Performance Persistence in UK Equity Funds – An Empirical Analysis*, October 2002) presented the results of a new study of performance persistence using a newly compiled database of UK equity-based unit trusts between 1981 and 2001 that included both 'live' funds (those surviving until the end of the period) and 'dead' funds (those that closed down or merged with other funds before the end of the period). The unit trusts are taken from four sectors:

- UK Smaller Companies
- UK Equity Income
- UK Equity & Bond Income
- UK All Companies.

3.5. The main findings of the CRA2 study were (p. 1):

- Performance broadly persisted in UK equity-based unit trusts between 1981 and 2001.
- Based on this information, it is possible for retail customers (and their advisers) to use this performance information to aid their investment decision making.

- The importance of persistence depends on both the time horizon and the sector in which the fund is invested. Performance is strongly significant only in the short-term for funds in the UK Equity & Bond Income and UK Smaller Companies sectors. However, it is significant over all time horizons for the UK All Companies and UK Equity Income sectors.
- Choosing a top quartile fund as opposed to a bottom quartile fund will, on average, add to an investor's potential return. The cumulative return for funds that are in the top quartile exceeds the returns for funds that are in the bottom quartile over the majority of time horizon and sector combinations.
- The results are not sensitive to charges. The persistence of unit trusts is not countered by charges and actually may increase under certain assumptions when charges are included. In addition, there is no relationship between charges and returns: unit trusts with higher charges (and therefore able to pay for more expensive fund managers) do not generate higher returns on average.
- Based on this data set, the report concludes that consumers (and their advisers) can use past information as a beneficial part of their investment decision making process.

4. An Assessment of the CRA Studies

4.1. Before commenting on the specific empirical study conducted by CRA, it is worth making some general points. Clear-cut and decisive empirical results can often be difficult to find in the performance literature due to problems with poor quality of data, the absence of complete data on transactions costs and the effects of survivorship bias due to the closure of underperforming funds.

4.2. Furthermore, it is also difficult to draw precise statistical inferences on mutual fund performance due to the low signal-to-noise ratio, i.e., the relatively small value of alpha (the average risk-adjusted performance of the fund) relative to the level of specific risk found for many funds. A long data sample is therefore often needed in practice to verify statistically whether abnormal performance is genuine or spurious. In view of these points, CRA should be commended for having constructed such a long database based on a large number of funds, including dead ones. It is also a valuable addition to the debate that CRA has managed to acquire and utilise data on charges.

4.1 The empirical evidence

4.3. The key empirical evidence of the CRA2 study is presented in a series of four-by-four contingency tables containing so-called transition probabilities between assessment and investment periods of different lengths (CRA2: Tables 4-11). These are the probabilities that a fund whose historical performance fell in a particular quartile during the assessment period will fall in a given quartile during the investment period based on its future performance. The transition probabilities will add up to 100% across the rows of the tables. The benchmark case where past performance contains no information over future performance corresponds to a transition probability of 25%. Transition probabilities above 27% are bold faced in the CRA tables.[1]

[1] We find it difficult to understand why such numbers are bold-faced. It is clear from the tables in Section 5 of CRA2 that 27% is not sufficient to establish statistical significance. Furthermore, as the number of funds is different in the four sectors, the same critical cut-off point for assessing significance cannot be used for the four sectors.

4.4. The tables cover the four unit trust sectors shown at para. 3.5 above, first with annual charges only deducted, and then with annual and initial charges deducted. Empirical results are reported for many different horizons. The general conclusion drawn by CRA is that there is strong evidence supporting performance persistence at the short horizon of 12-24 months, whereas the only systematic evidence of performance persistence at longer horizons is for UK Equity Income funds and UK All Companies funds.

4.5. However, the results often differ significantly across horizons. For instance, Table 1 shows the following pattern in the persistence of top quartile UK Smaller Companies funds as the investment horizon increases from 12 to 84 months:

Table 1 – Transition Probabilities: Top Quartile, UK Smaller Companies	
Assessment period – Investment period	**Top-to-top quartile transition probabilities (%)**
12 months back – 12 months forward	39.0
24 months back – 24 months forward	30.3
36 months back – 36 months forward	26.4
48 months back – 48 months forward	22.8
60 months back – 60 months forward	27.0
72 months back – 72 months forward	32.2
84 months back – 84 months forward	23.4
Source: Table 4, CRA2 (with annual charges deducted)	

What should an investor conclude from these results? Is there evidence that past performance persists at the 12, 24 and 72 month horizons where the estimates all appear to be much higher than 25%, but not at the other horizons? This would seem an odd conclusion. Clearly a joint test of persistence across all the different horizons is needed to address this type of question. No joint tests are reported in the CRA report, however, which makes it difficult to interpret some of the results.

4.6. Similarly, although an investment strategy based on the top quartile of funds is generally found to outperform investments in funds in the other quartiles at the 60-month investment horizon, it is only really for UK Equity Income that there appears to be a

higher return associated with investing in top-quartile funds, c.f. Figure 14 in CRA2. If so, why should this be the case? These observations are not intended as criticisms of CRA's analysis, but they do serve to remind us of how difficult it can be to get clear-cut results on performance persistence even with a sample as large as that utilised by CRA.

4.7. Some of the empirical findings are also surprising in the light of the existing empirical literature on performance persistence. Table 2 shows the following pattern in the persistence of top and bottom quartile UK All Companies funds as the investment horizon increases from 12 to 84 months:

Table 2 – Transition Probabilities: Top and Bottom Quartiles, UK All Companies

Assessment period – Investment period	**Bottom-to-bottom quartile transition probabilities (%)**	**Top-to-top quartile transition probabilities (%)**
12 months back – 12 months forward	34.9	37.6
24 months back – 24 months forward	30.3	29.4
36 months back – 36 months forward	26.4	29.1
48 months back – 48 months forward	25.0	29.7
60 months back – 60 months forward	26.1	31.5
72 months back – 72 months forward	27.1	33.0
84 months back – 84 months forward	27.6	38.7
Source: Table 11, CRA2 (with annual and initial charges deducted)		

One striking finding from this table is that the top-to-top quartile transition probabilities are, with one exception, larger than the bottom-to-bottom quartile transition probabilities. This is somewhat surprising since persistence generally has been found to be stronger among the bottom than among the top performers. This difference could well reflect the absence of risk adjustment: most studies adjust for risk, CRA did not. If the spread in the risk exposures of the top funds is larger than the spread in the risk exposures of the bottom funds, this would tend to generate more persistence among top funds than among bottom funds. The pattern is not so clear-cut in the other tables, but there are a surprisingly large number of cases for which it holds.

4.2 Raw returns versus risk-adjusted performance measures

4.8. The most surprising feature of the CRA2 study is that the authors do not adjust for risk. In their first report, CRA notes: '*Not surprisingly, often studies find persistence in risk-adjusted returns more elusive than persistence in unadjusted returns. Accordingly, we take care not to reject the possibility of persistence if the study has simply found insignificant persistence in risk-adjusted returns*' (CRA1: p. 18). We turn now to a discussion of the advantages and disadvantages of using raw returns or risk-adjusted performance measures.

4.3 The case for using raw returns

4.9. In their empirical analysis of performance persistence, CRA consider raw returns adjusted for annual fund management fees and transaction costs. The argument they put forward for using raw returns is that '*A consumer would be making a decision based solely on nominal returns, as this is how performance information is displayed*' (CRA2: p. 2). Later on, the report states that '*.. given that reliable risk models employ two to four risk factors and the calculation is beyond the average investor, we felt any results would be rejected on grounds of impracticality of calculation and use*' (CRA2: p. 5).

4.10. This argument is not convincing: most consumers would not be able to calculate for themselves the performance information provided about new cars, but no one would suggest that this information should not be calculated and published.

4.11. Further, the fact that performance is often reported in the form of raw, nominal returns is not a convincing argument for conducting a statistical analysis of performance if there are strong arguments from finance theory for adjusting for risk before comparing returns. Associated with the decision on which characteristics of a mutual fund to include in the FSA's Comparative Tables ought to be a programme of information and education for lay investors on what finance theory and empirical research tells us about the past performance of mutual funds. It is essential that the FSA's decision is based on sound theoretical and empirical arguments irrespective of whether investors historically made choices that were not based on the best available information.

4.12. Another argument put forward by CRA for using raw returns is that by comparing mutual fund returns within certain sectors, risk is already accounted for. Their justification for this argument is that risk varies considerably across investment sectors (UK Small Companies, UK Equity Income, UK Equity & Bond Income, UK All Companies), but not so much within each of the sectors.

4.13. How valid is this point? Figure 8 in CRA2 shows that market risk exposures, as captured by CAPM betas, vary significantly even across funds within the same sector. The dominant ranges are:

- UK All Companies: 0.7 – 1.1;
- UK Smaller Companies: 0.5 – 0.95;
- UK Equity & Bond Income: 0.1 – 0.85;
- UK Equity Income: 0.5 – 1.0.

We may conclude from this that using sector raw returns does not adequately control for market risk.

4.14. As we shall see, this variation in risk exposures within sectors has a number of consequences for the CRA analysis. As the investment horizon gets longer, the higher historical mean return on equities compared with that on cash and bonds means that a fund with a high beta will increasingly have a higher chance of outperforming in consecutive multi-year periods (and vice versa for funds with relatively smaller holdings in equities).

4.15. The only potentially persuasive argument for using raw returns is that it is a 'model-free' approach that does not involve taking a stand on which particular model to use for risk-adjustment purposes or on how to estimate the parameters of the risk-adjustment model. However, we would regard this as a defeatist argument. As we have already stated, risk is hard to quantify and interpret, but that does not mean information on risk and risk-adjusted performance data should not be published. Substantial progress has been made in the finance profession towards establishing practical procedures that account for risk when measuring performance.

4.4 Risk and persistence in raw returns

4.16. If past performance is measured in terms of raw returns, a key question is what generates persistence. Possible sources of persistence in raw return figures include differences in risk exposures, fees and fund manager skills.

4.17. The CRA report acknowledges that if persistence measures are based on raw returns, then they are likely to be dominated by differences in funds' risk exposures: '*If, as theory suggests, higher risk funds should out-perform low risk funds then, on average, forming a portfolio from funds that have performed well in the past is likely to result in a higher risk profile than the average and consequently higher returns in the future than the average*' (CRA2: p. 14).

4.18. From a theoretical perspective, the CRA2 report argues that the main source of performance persistence is differences in risk exposures and that this is consistent with market efficiency. The report states that: '*Orthodox finance theory suggests that persistence should be expected in raw non-risk adjusted returns, unless the risks carried by the funds change unpredictably over time or if all funds were exposed to the same risk*' (CRA2: p. 4). This statement is a rather convoluted way of saying that firms that take on more risk will, on average, end up earning higher returns.

4.19. Funds with high risk exposures are thus more likely, particularly over the longer term, to have the highest return, so high-risk funds would be over-represented in the top quartile (not necessarily because they were well managed). In contrast, low-risk funds are likely to be over-represented in the bottom quartile (not necessarily because they were poorly managed). Persistence figures, if based on raw returns, would thus induce investors to hold high-risk funds and to avoid low-risk funds, irrespective of the skills of the fund manager.

4.20. Such differences in risk exposures across funds are likely to be significant in explaining persistence in raw returns, c.f. Figure 8 in CRA2. It is difficult to say how important the other factors are. To answer this question properly, a careful analysis would need to be

conducted linking risk exposures to the evidence of differential skills amongst fund managers.

4.5 The case for adjusting for risk

4.21. The argument for adjusting performance figures for risk can perhaps best be understood in the context of a simple example. Suppose that an investor is considering whether to buy into a UK All Companies fund (denoted Fund 1 below) with a beta of unity and an alpha of zero or to buy a fund from the same sector (denoted Fund 2 below) with a beta of 0.5 and an alpha of 0.1 per month. The expected returns on the two funds on the basis of the CAPM are:

$$\text{Fund 1:} \qquad E(R_{1t}) = R_{ft} + E(R_{mt} - R_{ft})$$
$$\text{Fund 2:} \qquad E(R_{2t}) = 0.1 + R_{ft} + 0.5E(R_{mt} - R_{ft})$$

Provided that the market risk premium (the expected excess return on the market portfolio) exceeds 0.2, the expected return on Fund 1 exceeds that of Fund 2. However, this is purely a consequence of Fund 1's higher beta and ignores the fact that the level of risk is higher on the first fund.

4.22. In order to make investments in the two funds directly comparable, consider the following two investment strategies:

Investor 1: 75% in T-bills + 25% in Fund 1
Investor 2: 50% in T-bills + 50% in Fund 2.

The risk of the two strategies is the same since both have a beta of 0.25. However, the expected returns on the strategies of the two investors are:

$$\text{Investor 1:} \quad 0.75R_{ft} + 0.25E(R_{mt})$$
$$\text{Investor 2:} \quad 0.05 + 0.75R_{ft} + 0.25E(R_{mt})$$

Investor 2 expects a higher performance of 5 basis points per month or 60 basis points

per annum, even though this investor chose the fund (Fund 2) that is likely to have the worse raw return performance.

4.23. This simple example illustrates the possibility that basing performance on raw returns can result in sub-optimal investment decisions. If risk adjustment is considered, Fund 2 comes out on top, as it should given that only the manager of Fund 2 has genuine skills. However, by using raw returns, Fund 1 appears to produce the best performance.

4.24. Having demonstrated the importance of adjusting for risk, there remains the practical difficulty of choosing the most appropriate common risk-factors to adjust for. Empirical research (notably by Fama and French (1992, 1993)) suggests that exposure to the market index does not explain a great deal of the variation in performance across equities. Instead, the so-called Fama-French factors, book-to-market ratio and company size, seem to capture systematic differences in returns much better.[2] Similarly, funds within each sector are also likely to differ in terms of their 'styles', i.e. their relative holdings in small/large stocks and in growth/value stocks. Such style-factors have been found to be closely correlated with fund performance in the US, c.f. Carhart (1997) and Wermers (2002). If all these factors are viewed as risk-factors, it would be necessary to adjust returns for their exposure to them.

4.25. Clearly additional research is required in order to address the question of how best to produce risk-adjusted performance figures for UK unit trusts. The problem is by no means insurmountable, however, as demonstrated by the fact that companies such as Lipper, Standard & Poor and Morningstar in the US have designed procedures for accounting for risk when assessing mutual fund performance.

[2] These or related factors have also been found to be relevant to UK equities, c.f. Miles and Timmermann (1996). In the context of mutual funds they relate to the underlying investments held by the mutual fund.

4.6 Should performance be risk-adjusted?

4.26. We are surprised that the CRA empirical study did not report persistence results both for raw and for risk-adjusted returns. This would have allowed readers of the report to decide for themselves which set of results were more relevant.

4.27. Performance in raw returns depends on risk exposure, persistence by the underlying risk factors, charges and the fund manager's skills. Even if the objective were to study persistence in raw returns, a decomposition of this into risk, charges and skill would be of interest in order to further our understanding of the sources responsible for persistence in returns. This is an important consideration when assessing how likely it is that performance will persist in future periods.

4.28. Furthermore, even if the objective is forecasting raw return performance, using risk-adjusted performance as a predictor may do a better job. In fact, this is what Gruber (1996: p.796) reports in his Presidential Address to the American Finance Association: '*The results indicate clearly that the four index [risk adjustment] model does a better job than past raw returns of forecasting future raw returns.*'

4.29. We do not think it is appropriate to use past performance figures based on raw returns to induce investors to maximise their risks. Past performance figures, in so far as they are useful in guiding retail investors' decisions, should help to identify the best talent in the fund management industry as opposed to selecting the riskiest strategies within a given sector.

4.30. There are good reasons why funds with higher risks on average pay higher returns. Although, on average, such funds tend to pay higher returns than low-risk funds, sometimes they pay a much lower return, as witnessed by the performance of the stock market over the last three years. This represents a real loss to investors. The higher average return is a compensation for bearing risk (i.e., it is a risk premium) and should not be seen as a free bonus that makes high-risk funds more attractive than low-risk funds. Risk-averse investors are not generally or, indeed, necessarily better off by investing in high-risk, high-expected-return mutual funds.

4.7 Persistence and market efficiency

4.31. Much of the theoretical debate on mutual fund performance persistence has been conducted with reference to the efficient market hypothesis. For example, as CRA2 notes, the efficient market hypothesis has some powerful implications: '*The efficient market hypothesis implies that, on a risk-adjusted basis, it is impossible to generate superior returns consistently*' (CRA2: p. 14).

4.32. There are at least two ways in which competitive pressure might rule out consistently superior performance. First, suppose that mutual fund managers do not differ in terms of their abilities. Grossman and Stiglitz (1980) show that while mutual funds can outperform on a gross-return basis, net of costs (including research costs) they will not be able to do so in a strongly efficient market. Of course, if some fund managers are simply more skilled than others, superior performance need not be matched by higher research costs, and investors who are smart (or lucky) enough to pick a skilled fund manager might end up with an above-average return on a risk-adjusted basis.

4.33. However, suppose that genuinely skilled fund managers exist but that skill is a scarce resource while there are a large number of investors ensuring that investment capital is abundant. In this second scenario, it is highly likely that skilled fund managers will simply charge higher fees, so that, net of fees, investors will end up being indifferent between investing in funds managed by skilled or unskilled managers, at least in so far as the latter do not underperform. Economic models, such as that proposed by Berk and Green (2002), predict little or no persistence, because mutual fund stars capture the full rent from their skills through higher management fees.

4.34. On the other hand, while the efficient market hypothesis in its strongest form rules out consistently superior performance, it does not rule out the possibility that underperforming funds may continue to exist in the market place. Rational decision-making by <u>all</u> investors is required to rule out this situation. Nevertheless, as markets become more informationally efficient, persistently underperforming funds should eventually disappear as investors withdraw their money from such funds. However, this

assumes that information on past underperformers is readily available to investors, a condition that does not appear to be satisfied in practice at the present time.

4.35. While competition among informed investors may be strong enough to compete away most of the surplus generated by genuinely skilled fund managers, the same mechanism is unlikely to operate at the other end of the performance and skill spectrum which is dominated by less knowledgeable investors displaying considerable inertia in their behaviour.

5. An Assessment of CRA's Statistical Procedures and Hypothesis Tests

5.1. Apart from basing its analysis on raw returns, the CRA2 empirical study also suffers from some drawbacks in respect of the statistical procedures used.

5.2. Most seriously, the CRA2 study calculates standard errors on the basis of the assumption that the individual funds' returns are independently distributed not only through time but also across funds.[3]

5.3. For example, for the case underlying CRA2's Figure 24, the binomial distribution is based on 96 funds followed over 20 years of which 96 x 20 x 0.25 = 480 observations (n = 480) correspond to top-quartile performance. Under the assumption of no performance persistence and independence across funds, the transition probability equals 25% (p = 0.25), so the expected number of funds with top-quartile performance and the associated standard deviations and critical values are:

$$\text{E (No. of funds)} = n \times p = 480 \times 0.25 = 120$$
$$\text{Standard deviation (No. of funds)} = \sqrt{n \times p \times (1-p)} = \sqrt{480 \times 0.25 \times 0.75} = 9.49$$
$$95\%\ \text{critical value} = 120 + 1.645 \times 9.49 = 136.$$

This figure of 136 is equal to 28.33% of 480. It is the critical value cited in the CRA2 report on page 43. It is based on the assumption of independence in the draws across funds. This assumption can be justified as a first approximation when the common risk factors have been subtracted from fund returns. However, it is much harder to justify when the analysis is based on raw returns. Raw returns are likely to be strongly correlated across funds since they are driven by common risk factors such as the market return, book-to-market ratio, firm size and momentum.

[3] The statistical analysis in section 5 of CRA2 explains how the analysis was conducted. Footnote 6 (CRA2: p.43) explains how the binomial probabilities are computed.

5.4. As a simple illustration of what can go wrong as a result of failing to account for dependence across the funds' performance, consider the extreme case of having only four different groups of funds. This case would arise if the funds are fully diversified and only differ in terms of the four different (but constant) betas which the funds' managers choose for their portfolios. The funds are sorted into low, medium-low, medium-high and high beta risk groups. The performance of all funds within the same risk group is identical, but that performance differs across the four risk groups.

5.5. Rather than having 96 independent fund observations each year, as implied in CRA's Figure 24, there would only be four independent observations (risk groups) per year. The resulting number of independent risk group observations would be 4 x 20 = 80 with 4 x 20 x 0.25 = 20 of these in the top quartile. Repeating the above analysis for such risk groups or 'blocks of funds', we have:

$$\text{E (No. of groups)} = n \times p = 20 \times 0.25 = 5$$
$$\text{Standard deviation (No. of groups)} = \sqrt{n \times p \times (1-p)} = \sqrt{20 \times 0.25 \times 0.75} = 1.94$$
$$95\% \text{ critical value} = 5 + 1.645 \times 1.94 = 8.2.$$

Each risk group contains 24 funds, so the corresponding values for the underlying number of funds can be found by scaling the above numbers by 24:

$$\text{E (No. of funds)} = 120$$
$$\text{Standard deviation (No. of funds)} = 46.5$$
$$95\% \text{ critical value} = 120 + 1.645 \times 46.5 = 196.5.$$

Dividing 196.5 by 480, gives 40.93% as the critical value that needs to be exceeded in order for the transition probability to be regarded as statistically significantly different from 25% at the 95% confidence level. This percentage is far higher than that computed under the assumption of independence between funds (28.33%) and very few of the transition probability estimates in Tables 4-11 in CRA2 exceed 40%.

5.6. Neither of these extreme cases (independence or perfect dependence group-wise) will be consistent with the data and the truth will lie somewhere in between. However, these

calculations simply demonstrate how important it can be to account for the potential effect of dependence across funds. The statistical results are difficult to interpret (and can lead to false inferences) unless cross-sectional dependence is taken into consideration in the raw return performance analysis.

5.7. Our simple illustration also helps to explain what would otherwise be a very surprising finding, namely how much the transition probabilities differ even for very similar holding periods. For example, for UK Smaller Companies, the top quartile transition probability drops from 32.2% at the 72-month horizon to 23.4% at the 84 month horizon despite the considerable overlap between these two investment periods (see Table 1 above).

5.8. The large variations in some of the reported transition probabilities would appear to support the conjecture that dependence in the raw return performance across funds and across overlapping holding periods leads to a higher level of statistical uncertainty than is assumed in the statistical analysis of the CRA study.

5.9. The CRA2 report also considers the effect of overlaps in the Monte Carlo simulations it documents. It is clearly appropriate to address this point. Indeed, it is clear from CRA2's Table 15 and other tables that accounting for overlap can have a big effect on the statistical analysis. For example, for the longest horizon considered, 84 months or seven years, the sample used by CRA effectively involves only three non-overlapping observations. This is important, since there may well have been considerable persistence in some of the return factors for substantial periods of the sample: e.g., momentum (or the past performance risk-factor) is likely to have been important in the 1990s as documented in US studies (e.g., Carhart (1997)). Given the many types of assumptions that one can make in designing the Monte Carlo simulation, it would be desirable to conduct a series of sensitivity analyses for the overlap analysis, but the CRA2 study fails to report these.

5.10. Although contingency tables provide an easy way to comprehend and report persistence in performance, it would be useful to supplement them with a statistical analysis that is not dependent on forming quartiles of funds. One disadvantage of an approach based on

quartiles is that they can induce a degree of discreteness into the financial decision-making process that might not be desirable. For example, if such information were included in the Comparative Tables, a fund that only just avoids falling into the bottom quartile might not experience large cash outflows, while the neighbouring fund that is the top performer in the fourth quartile might experience significant cash outflows despite the near-identical performance of the two funds.

5.11. Simply reporting the extent of past underperformance on a risk-adjusted basis might serve investors better, particularly if an empirical analysis were able to find a statistically significant relationship between the level of past performance and future performance. This would suggest using regression analysis to estimate risk-adjusted performance that could then be reported for each individual fund (as opposed to simply reporting the quartile to which the fund's risk-adjusted performance belongs).

6. Weaknesses in CRA's Analysis and Conclusions

6.1. In this section, we summarise the weaknesses in the CRA empirical study. For the sake of completeness, we begin by summarising the weaknesses we have already identified in the study.

6.1 The use of raw unadjusted returns

6.2. The first problem with the CRA study is that it has not conclusively established that there are significant gains to be earned from past performance information, particularly once risk-adjusted performance measures are used, as we have advocated should be done. Measuring performance based on raw returns is likely to induce investors to pursue riskier strategies than if performance measures were recorded on a risk-adjusted basis. Such an investment strategy will, in general, be sub-optimal for risk-averse investors.

6.2 The robustness of the statistical analysis

6.3. The second problem with the CRA study is that the robustness of the statistical significance of some of its findings has not been tested. In particular:

- joint tests of the statistical significance of persistence across the different assessment and investment horizons have not been performed; and
- the likely dependence of performance across the funds has not been taken into account.

6.3 Ignoring the differential effect of charges on the performance of active and passive strategies

6.4. The CRA empirical study does account for annual charges. Furthermore, the report finds that performance persistence actually increases after accounting for initial charges. In the case of the worst performing funds in the bottom quartile, this is because a new initial charge is incurred when a fund dies. Funds with poor performance are more likely to close down. As such funds are over-represented in the bottom quartile, the initial charge will subtract more from the returns of funds in this quartile than in the

other quartiles. This will tend to enhance the persistence of the poor performance of these funds. This is another good reason for investors to avoid funds with poor past performance.

6.5. When a fund dies, the CRA study assumes that the residual fund value is reinvested according to one of three rules: equally across all surviving funds, equally across all surviving funds in the same quartile, or in surviving funds in the same fund management group. Investments in new funds are associated with a new initial charge.

6.6. While these scenarios seem reasonable to us, the third problem with the CRA study is that it does not compare the effect of initial charges on realised performance of a strategy of re-investing in top-quartile funds with a strategy of investing in tracker funds that are much less likely to close down and hence more likely to result in initial charges being incurred only once. This could be an important omission since initial and annual charges appear to affect cumulative returns quite significantly and such charges will be much lower for tracker funds. If an active strategy of investing in top-quartile funds does not consistently outperform an investment in a tracker fund net of annual and initial charges, it would make the active investment strategy considerably less attractive.

6.4 The exclusion of tracker funds from the analysis

6.7. The CRA2 study should be commended for using a large sample of funds including dead funds and accounting for the effect of survivorship bias. Survivorship bias (the effect on reported returns of excluding funds that have died some time during the sample period) can lead to important distortions in the statistical analysis. Our earlier work (Lunde, Timmermann and Blake (1999)) found that underperforming unit trusts do eventually merge with more successful unit trusts, but that on average it takes some time for this to occur. Overall, our study found that about 40% of funds over the period 1972-1995 were eventually wound up or merged. The most frequently occurring (modal) duration for a fund was 4.25 years (51 months), but the average duration was about 16 years. Across the whole unit trust industry, the average return on funds that survived the whole period was 13.7% per annum, while the average return on funds that

were wound up or merged during the period was 11.3% per annum. This implies that a typical unit trust investor might find himself locked into an underperforming trust that is eventually wound up or merged into a more successful trust, experiencing an underperformance of 2.4 percentage point per annum over a 16 year period. This translates into a fund value that is 19% lower after 16 years than a fund that is not wound up or merged.

6.8. The fourth problem with the CRA analysis is that it excludes trackers from the sample. It is difficult to conjecture how this affects the analysis of persistence. The exclusion of trackers is likely to have the effect of increasing both the relative out-performance in the top quartile and the relative under-performance in the bottom quartile, since the trackers are most likely to be concentrated around the median. Their exclusion will therefore pull some top-quartile funds into the second quartile and some bottom-quartile funds into the third quartile, and thereby leave the top and bottom quartiles with the more outlying funds. It would be important to see how the inclusion of tracker funds affects the results since investment in tracker funds is, of course, one of the key options open to retail investors.

6.5 Failure to report evidence on the performance of switching strategies

6.9. A key objective of a persistence analysis is to shed light on the potential economic gains investors would get from using past performance figures. We agree with the CRA2 report when, on page 13-14, it argues that consumers care about cumulative returns and that even transitory persistence generates a permanent gain to the value of the fund.

6.10. Such gains can be measured through the return associated with investment strategies based on performance persistence information. The CRA2 report considers a range of investment strategies, but concentrates on studying top-quartile or bottom-quartile investment strategies.

6.11. Although results on performance persistence may differ depending on sample period, type of mutual fund, method of risk adjustment, treatment of charges and so forth, there is considerable evidence in the empirical literature of strong negative persistence. In

contrast, there is only weak evidence of positive persistence, mainly at the short (1-3 year) horizon. We do not intend to provide a comprehensive survey of the performance literature, but will mention here a few studies that support this point of view.

6.12. Hendricks, Patel and Zeckhauser (1993: p.93) argue that '*Portfolios of recent poor performers do significantly worse than standard benchmarks; those of recent top performers do better, though not significantly so.*' They conclude (p.122), '*Icy hands, the evil counterpart of hot hands, also shows up in our sample: funds that perform poorly in the most recent year continue to be inferior performers in the near term. Indeed, they are more inferior than hot hands are superior. While there is little support for funds that are sustained superior performers, we do identify some funds that are sustained underperformers*.'

6.13. Similarly, Gruber (1996) finds that the underperformance of the worst decile of funds is much higher than the superior performance of the best funds. He asks the question (p. 807): '*Why do we see any money remain in funds that are predicted to do poorly and in fact do perform poorly?*'.

6.14. Carhart (1997: p.80) concludes that '*the spread in mean return unexplained by common factors and investment costs is concentrated in strong underperformance by the bottom decile relative to the remaining sample*'. He later states (p.80) that '*Although the top-decile mutual funds earn back their investment costs, most funds underperform by about the magnitude of their investment expenses. The bottom-decile funds, however, underperform by about twice their reported investment costs*.'

6.15. All this evidence indicates that the largest gains from past performance information are likely to arise from inducing investors to avoid underperforming funds. However, the fifth problem with the CRA study is that it does not shed light on the economic returns from a strategy of moving funds away from underperforming funds and into funds with the best past performance.

6.6 Failure to identify the sources of performance persistence

6.16. The CRA2 study argued that what is important to consumers is *whether* performance persistence exists, not *why* it exists (p. 4). As a consequence of this view, the report does not identify whether the source of persistence is due to risk, momentum, size or other characteristics.

6.17. We disagree with this point of view. Clearly it is of interest to investors to know whether the source of raw return persistence is skill, risk profile or charges. Any measure of persistence will invariably be surrounded by statistical uncertainty. The source of past persistence thus provides important clues as to whether or not it will continue in future periods and whether the optimal strategy for the investor is to continue to invest in a fund with high historical returns. The sixth problem with the CRA study is that it makes no effort to identify the sources of performance persistence.

6.18. Suppose that it is found that the fund manager's skills account for a large share of performance persistence. This would raise a number of interesting questions such as why the fund manager does not simply increase charges. Furthermore, little is known about the factors responsible for superior investment performance. Is it the individual fund manager (e.g. Peter Lynch), the 'team' running the fund or the fund family that creates abnormal performance? If superior performance is due to the individual fund manager and this manager ceases to run the fund, then this would be vital information to disclose. Also, how involved is a fund manager required to be in order for investors to treat a particular fund as being under his/her management?

6.19. On the other hand, if we find the source of raw return persistence is differences in risk exposures, this would lead to a different conclusion. If fund manager skill is not important and only the risk assumed by the fund manager matters (which seems to be the premise underlying the CRA analysis) then investors do not need active fund managers at all. They could create any desired risk exposure themselves much more cheaply using a combination of borrowing and low-cost trackers. For example an investor with £1000 to invest could create a fund with a beta of 1.2 by borrowing £200 and investing £1200 in a tracker. This would be cheaper than buying an actively

managed fund with a beta of 1.2. Therefore, it is important to conduct a more detailed analysis to shed light on the sources of performance persistence.

6.7 Failure to analyse flows and sizes of funds

6.20. An analysis of the investment flows into and out of mutual funds would be helpful to assess potential welfare gains from publicizing past performance information. US studies have found that 'money is smart' and tends to flow into funds with good past and (to some extent) good future performance. The extent to which similar results hold for the UK would be helpful in providing a better understanding of UK investors' actual behaviour. This would set a benchmark for how measures of past performance and risk should be conveyed to the public.

6.21. Similarly, the size of investment holdings in the worst-performing mutual funds needs to be calculated so that an estimate of the likely gains from reporting of underperformance can be obtained. The seventh problem with the CRA study is that it failed to conduct a flow-of-funds and size analysis.

6.8 Dealing with these weaknesses

6.22. These weaknesses could be rectified along the following lines:

- The statistical significance of many of the findings in the CRA2 report could be revisited along the lines proposed in our report. The effects of cross-sectional dependencies and of including the tracker funds on the results in the contingency tables are of particular interest as is the way that data overlaps are accounted for in the analysis of the multi-year holding returns.
- Similarly, it would be possible to conduct an analysis of the significance of the cumulative return differentials and of the differential effects of charges on various investment strategies (active versus passive, bottom- to top-quartile switching strategies), again computed on a risk-adjusted basis.
- It would also be possible to conduct an analysis of the sources of performance persistence.
- All this could be done with CRA's existing dataset. Finally, it would be desirable

to conduct a flow-of-funds and size analysis to see if UK mutual fund money is as smart as US mutual fund money, although this may not be possible with CRA's current dataset.

7. Can (and If So How) Performance Information Inform Investors' Decisions?

7.1. As we have stated above, many studies in the empirical literature on mutual fund performance find that underperformance persists. A potentially significant benefit from reporting persistence is that funds, if required to disclose their performance over a fixed recent period, would be given a stronger incentive not to underperform.

7.2. Currently, there is an important asymmetry in how funds promote their performance figures. The advertisement of funds' past performance is voluntary although it is regulated and a certain amount of standardised information must be reported in footnotes. In practice, this means that superior past-performance records are actively promoted, while poorly performing funds are far less likely to advertise their past performance. This gives an incentive for mutual fund houses to engage in risk taking across the individual funds in their stable (known as families in the literature). Superior performers (the lucky ones?) can be heavily advertised, while inferior performers (the unlucky ones?) do not have to be actively publicised. It is important to design a reporting system that cannot be 'gamed' in this way, c.f. Nanda, Wang and Zheng (2002).

7.3. One strategy that avoids this distortion would be to require that inferior as well as superior risk-adjusted past-performance figures be reported in the Comparative Tables. As a consequence and as a clear benefit to investors, fund management houses might also become more aggressive in closing down or merging underperforming funds.

7.4. In the absence of readily available figures on inferior past performance, the flow of funds into poorly performing funds is likely to increase, as CRA argue. This would tend to weaken the incentives of fund managers and underperformance could become more widespread.

7.5. On the downside, if past performance figures were more easily available to the public, it might be argued that fund managers would be less willing to take on risk on account of the possibility of being penalised for underperformance. This could lead to the type of

herding behaviour observed in the management of UK pension funds, c.f. Blake, Lehmann and Timmermann (2002).

7.6. This effect may be important, but we believe that the most effective way of providing appropriate managerial incentives is through performance-related fees as proposed in Blake and Timmermann (2002). This would provide appropriate awards to skilled fund managers, while penalising poor fund managers and at the same time reducing the incentive for losing fund managers to take excessive risks with their clients' funds in an attempt to improve their relative performance, a pattern of behaviour found in Brown, Harlow and Starks (1996).

7.7. Yet none of this would detract from the benefit to investors of having an easily accessible and independently verified source of past-performance data on a risk-adjusted basis.

8. Conclusions

8.1. A key finding of CRA's empirical analysis of the performance persistence of UK mutual funds is that: '*the importance of persistence depends on both the time horizon and the sector in which the fund is invested: performance is strongly significant only in the short-term for funds in the Equity & Bond Income and Smaller Companies sectors; however, it is significant over all time horizons for the UK All Companies and UK Equity Income sectors*' (CRA2, p.1).

8.2. We have argued in our report that this result could well be explained by differences in risk in the short term and by the risk premium in the long term and need not reflect any intrinsic superior investment skills by the fund managers involved.

8.3. Further, the investment strategy which follows from this and which CRA recommends can be summarised succinctly as: '*invest in high risk funds and hope for the best, since if you hold the investment for long enough it generally works out*'. This strategy, because it in effect discounts any notion of fund manager skill, has a number of implications:

- Investors do not need to pay for active fund management, since they can replicate the high-risk exposures of active fund managers more cheaply using a combination of home-made leverage and tracking funds.
- There is no need for the sector allocation of unit trusts. If the objective is to have high returns, why divide unit trusts into sectors? Why not group all the funds together and select those from the highest quartile of past performance?

8.4. A second key finding of CRA's empirical analysis is that poor fund managers do systematically underperform. One of the authors of this report has been persuaded to change his mind as a result of the increasing amount of empirical evidence on systematic underperformance. Blake and Board (2000, p. 560) concluded:

- '*Economic theory [in the form of the efficient markets hypothesis], supported by the demonstrated inability of funds to sustain consistent performance over extended periods, makes the use of past investment returns to project future performance hard*

to support.

- *If [fund managers] wish to improve value added in the [fund management] industry, they should compete on the basis of their charges, rather than on the basis of their past performance or their promised future performance*'.

These remarks were made in relation to consistent superior performance, but the efficient markets hypothesis does not rule out consistent inferior performance, the presence of which requires the irrationality of just a small number of investors.

8.5. The findings in the empirical literature on mutual fund performance have persuaded us that there is a good case for the FSA publishing risk-adjusted past-performance data in the Comparative Tables, not because we believe that superior risk-adjusted performance can be sustained over long periods (the CRA reports have not shown this), but because poorly managed funds can be exposed more quickly. The benefit to investors is that such funds can then be closed down or merged with more successful funds more rapidly and at lower cost to investors than if they sought to achieve this result themselves by switching into better performing funds and incurring the switching costs.

8.6. Our position mirrors the FSA's current view that costs and charges are very important determinants of mutual funds' net performance. In fact, underperforming funds are exactly those funds that do not have any fund manager ability, but do have high charges and costs.

8.7. We are not persuaded by the argument that important information should not be published just because investors might 'misuse' it. While we acknowledge that the FSA has an obligation to protect the investor, it is not clear that investors are protected from a failure to expose the performance of poor fund managers.

8.8. Although it is unclear whether risk-adjusted superior performance is sustainable over long investment horizons, we do believe that the empirical evidence shows that poor investment performance is persistent and that by preventing publication of past-performance data an important piece of information is lost. The worst outcome it seems to us is to leave the decision on what past-performance information to publish to the

fund managers themselves, since they are far more likely to publish information on the best performing funds on a non-risk-adjusted basis. It is far better for the FSA to publish past performance data on a risk-adjusted basis for all funds and to give a 'health warning' along the lines that the mass of existing empirical research shows that '*losers generally repeat, while winners do not necessarily repeat*'. Only the FSA can do this, and only the FSA can reduce the risks of any potential 'misuse' of information through an education programme. The situation is similar to the use of performance information published on schools and universities by their 'regulator'.

8.9. We have also expressed concerns about how certain aspects of the statistical analysis was conducted, in particular how dependencies across funds' raw returns were dealt with. CRA2 (p.3) states that it '*provides new evidence based on statistically and economically robust analysis*'. However, a robust analysis would need to consider the effect of correlations in performance across funds. At this stage, irrespective of whether or not the performance figures adjust for risk, the statistical results reported by CRA must be viewed as preliminary.

8.10. Although we have advocated the use of risk-adjusted past-performance measures, one uncertainty remaining is deciding on the best method for risk adjustment in the context of UK mutual funds. We discussed a range of possibilities, based on US studies, and these could be investigated using the data set used by CRA. This research is necessary to assess whether factors such as investment 'style' are as important to UK funds as it has been found to be for US funds.

8.11. The CRA reports also raise a question mark over the FSA's permissible growth rates for projecting future fund values. Although '*the principal aim is to show the impact of charges on growth*' (FSA (2001: p.12)), they are used by investors to provide an idea of '*what you might get back*'. But the critical problem with them is that they do not take account of future risk. We believe that the best way of dealing with this problem is to present key percentiles (such as 20%, 50% and 80%) of the distribution of possible returns over a relevant investment horizon.

8.12. The CRA reports have made a valuable contribution to the debate and in doing so have presented the FSA with an ideal opportunity to design a coherent framework for measuring and reporting information on:

- charges
- past performance on a risk-adjusted basis, and
- key percentiles of the distribution of possible returns on the basis of the risk exposure assumed by the fund manager.

Glossary

The terms below are explained in more detail in Blake (2000).

Alpha – a measure of the return to active fund management and hence fund manager skill. Defined as the difference between the actual return realised by the fund manager and the expected return from the Capital Asset Pricing Model, given the level of market risk (beta) assumed by the fund manager. Also known as Jensen's alpha.

Beta – a measure of the market risk of an investment. It measures the degree to which the return on an investment moves in line with the return on the market (a beta of zero indicates that an investment does not move at all in response to a move by the market as a whole, while a beta of unity indicates that the investment and the market move up and down in tandem).

Capital Asset Pricing Model – a model for determining the expected return on a risky asset as the sum of two components, the risk-free rate (all risky assets should expect to earn at least the risk-free rate) and the asset's risk premium (which is defined as the product of the asset's beta and the market risk premium).

Contingency tables – two-dimensional tables reporting transition probabilities.

Correlation – a rescaled measure of covariance: varies between –1 (the two variables move in exactly opposite directions), through zero (the movements in the two variables are completely unrelated) to +1 (the two variables move exactly in line with each other). Correlation is defined as the ratio of the covariance to the product of the standard deviations of the two variables.

Covariance – the degree to which two variables move together over time: positive if the two variables move in the same direction and negative if they move in opposite directions. The covariance of the returns on two assets is defined as the average of the cross products of the deviations of the actual returns on the two asset from the mean returns on the two assets over a specified investment horizon.

Critical value – value above which statistical significance at a specified level is established; e.g., the 95% critical value in the case of the normal distribution is equal to the mean of the distribution plus 1.645 times the standard deviation of the distribution.

Diversification – combining investments together in a portfolio with the aim of reducing the aggregate level of risk by exploiting the offsetting risks contained in the individual assets

held in the portfolio.

Efficient markets hypothesis – information is so rapidly incorporated into the prices of traded securities that trading strategies based on the use of public information cannot systematically generate positive excess returns after adjusting for risk and transactions costs.

Excess return – return on an investment in excess of the risk-free return.

Fund family – the range of mutual funds offered by the same fund management house.

Gearing – using borrowed funds to invest in risky assets.

Leverage – same as gearing.

Market efficiency – the current market values of investments incorporate all relevant information about those investments, so that there are no unexploited profit opportunities from using any particular piece of information to undertake transactions in a particular security.

Market portfolio – the portfolio of assets comprising all the assets in the economy in direct proportion to their market values. It is the most diversified portfolio of all and so is used in finance theory as a benchmark for comparing the returns on other less well-diversified portfolios. Tracker funds seek to replicate the market portfolio.

Market risk premium – the difference between the expected return on the market portfolio and the risk-free rate (i.e, the expected excess return on the market).

Mean return – the average return over a number of years.

Momentum – an asset is performing well (or badly) in one year and this good (or bad) performance spills over to the next year and possibly to later years.

Mutual fund – an entity that invests in the securities issued by quoted companies.

Performance persistence – a fund that performs well (or badly) in one year also performs well (or badly) in subsequent years, due to fund manager skill, momentum or level of market risk assumed by the fund manager.

Regression analysis – a method of estimating the parameters of the statistical relationship between a set of variables that is based on minimising the sum of the squares of the residuals (i.e., unexplained part) of the regression equation.

Risk – the degree to which the value of an investment can deviate from its initial value (*total risk*) or the degree to which the value of an investment can deviate from its initial value in response to a movement by the market as a whole (*market risk*). Total risk is equal to market risk plus specific (or idiosyncratic) risk. *Specific risk* is the risk embodied within an investment that is unrelated to movements in the market: in principle it can be diversified

away.

Risk aversion – a dislike of risk: risk-averse investors need a risk premium to compensate them for bearing risk.

Risk-free rate – the return on a risk-free asset such as Treasury bills.

Risk premium – the additional return on an asset in excess of the risk-free rate needed to compensate investors for bearing the market risk of the asset.

Sharpe ratio – ratio of excess return on an investment to its standard deviation.

Standard deviation – a measure of the total risk of an investment, usually explained in terms of the 2-in-3 rule: in two years out of every three, the actual return on an investment will lie within one standard deviation of the mean return on the investment. Standard deviation is defined as the square root of the variance.

Star – a highly skilled fund manager.

Survivorship bias – the effect on reported returns of excluding funds that have died some time during the sample period.

Term premium – the additional return needed to induce investors to forgo access to their funds for a particular period of time.

Transition probability – the probability of being in a particular state (e.g., the fourth quartile of investment performance) in one period, conditional on being in a particular state (e.g., the first quartile of investment performance) in the previous period.

Treynor ratio – ratio of excess return on an investment to its beta.

Unit trust – same as mutual fund.

Variance – square of the standard deviation. The variance is defined as the average of the squared deviations of the actual returns on an asset from the mean return on the asset over a specified investment horizon.

References

Bacon and Woodrow (1999), *Comparative Tables*, September.

Berk, J. and Green, R.C. (2002), *Mutual Fund Flows and Performance in Rational Markets*, NBER Working Paper W9275.

Blake, D. (2000), *Financial Market Analysis*, Wiley, Chichester.

Blake, D., and Board, J. (2000), Measuring Value Added in the Pensions Industry, *Geneva Papers on Risk Insurance: Issues and Practice*, 25, 539-567.

Blake, D., and Timmermann, A. (2002), Performance Benchmarks for Institutional Investors: Measuring, Monitoring and Modifying Investment Behaviour, in Knight, J. and Satchell, S. (Eds) *Performance Measurement in Finance: Firms, Funds and Managers*, Butterworth-Heinemann, Oxford.

Blake, D., Lehmann, B., and Timmermann, A. (2002), Performance Clustering and Incentives in the UK Pension Fund Industry, *Journal of Asset Management*, 3, 173-194.

Brown, K.C., Harlow, W.V., and Starks, L.T. (1996), Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry, *Journal of Finance*, 51, 85-110.

Carhart, M. (1997), On Persistence in Mutual Fund Performance, *Journal of Finance*, 52, 57-82.

Charles River Associates (2002a), *Performance Persistence in UK Equity Funds – A Literature Review* (prepared by T Giles, T Wilsdon and T Worboys), January.

Charles River Associates (2002b), *Performance Persistence in UK Equity Funds – An Empirical Analysis* (prepared by T Giles, T Wilsdon and T Worboys), October.

Fama, E., and French, K. (1992), The Cross-Section of Expected Stock Returns, *Journal of Finance*, 47, 427-465.

Fama, E., and French, K. (1993), Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics*, 33, 3-56.

Financial Services Authority (2001), *Report of the Task force on Past Performance*, September.

Grossman, S.J., and Stiglitz, J. (1980), On the Impossibility of Informationally Efficient Markets, *American Economic Review*, 70, 393-408.

Gruber, M.J. (1996), Another Puzzle: The Growth in Actively Managed Funds, *Journal of Finance* 51, 783-810.

Hendricks, D., Patel, J., and Zeckhauser, R. (1993), Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, *Journal of Finance* 48, 93-130.

Jegadeesh, N., and Titman, S. (1993), Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance*, 48, 65-91.

Jensen, M. (1969), Risk, the Pricing of Capital Assets and the Evaluation of Investment Portfolios, *Journal of Business*, 42, 167-247.

Lunde, A., Timmermann, A., and Blake, D. (1999), The Hazards of Mutual Fund Underperformance: A Cox Regression Analysis, *Journal of Empirical Finance*, 6, 121-152.

Miles, D. and Timmermann, A. (1996), Variation in Expected Stock Returns: Evidence on the Pricing of Equities from a Cross-Section of UK Companies, *Economica*, 63, 369-382.

Nanda, V., Wang Z., and Zheng. L. (2002), *Family Values and the Star Phenomenon*, Working Paper, University of Michigan Business School.

Rhodes, M (2000), *Past Imperfect? The Performance of UK Equity Managed Funds*, Occasional Paper Series No. 9, Financial Services Authority, August.

Sharpe, W. (1963), Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance*, 19, 425-442.

Sharpe, W. (1966), Mutual Fund Performance, *Journal of Business*, 39, 119-138.

Treynor, J. (1965), How to Rate Management of Investment Funds, *Harvard Business Review*, January-February, 63-75.

Wermers, R., 2002, *Is Money Really 'Smart'? New Evidence on the Relation Between Mutual Fund Flows, Manager Behavior, and Performance Persistence*. Mimeo, University of Maryland.

Appendix: About the Authors of this Report

Professor David Blake

Dr David Blake is Professor of Financial Economics at Birkbeck College in the University of London and Chairman of Square Mile Consultants, a training and research consultancy. Senior Research Associate, Financial Markets Group, London School of Economics. Senior Consultant, UBS Pensions Research Centre, London School of Economics. Research Associate, Centre for Risk & Insurance Studies, University of Nottingham Business School. Formerly Director of the Securities Industry Programme at City University Business School and Research Fellow at both the London Business School and the London School of Economics. Consultant to many organisations, including UBS Global Asset Management, Merrill Lynch, Deutsche Bank, Paribas Capital Markets, McKinsey & Co., Financial Research Services Ltd, James Capel, Schroders, Browne Jacobson Solicitors, Artemis Trustees, Taunton Cider Company, Hill and Knowlton, the Independent Television Companies Association, the Office of Fair Trading, UNESCO and the World Bank. External Examiner for the Wholesale Markets Brokers Association Diploma. Registered as an Approved Person with the Financial Services Authority.

David Blake was a student at the London School of Economics in the 1970s and early 1980s, gaining his PhD on UK pension fund investment behaviour in 1986. His research interests include the modelling of asset demands and financial innovations, the investment behaviour and performance of pension funds and mutual funds, and pension plan design. He has published in major economics and finance journals in all these fields. He is author of ***Financial Market Analysis*** published by Wiley in 2000, ***A Short Course of Economics*** published by McGraw Hill in 1993, ***Modelling Pension Fund Investment Behaviour*** and ***Issues in Pension Funding*** both published by Routledge in 1992, and ***Pension Schemes and Pension Funds in the United Kingdom*** published by Oxford University Press in 2003. Professor Blake gave evidence to the Goode Committee in March 1993 and was a member of the Retirement Income Working Party which published its report ***Improving Security and Flexibility in Retirement*** in March 2000.

In June 1996, he established the Pensions Institute at Birkbeck College. The Pensions Institute undertakes high quality research on all pension-related issues and publishes details of its research activities on the internet (http://www.pensions-institute.org). The Pensions Institute was the first academic research centre devoted exclusively to studying pensions matters to be established outside the US.

David Blake regularly speaks at international conferences on pensions and pension fund management in the UK, Europe, the Far East, Australia and the Americas.

Professor Allan Timmermann

Dr Allan Timmermann is a professor of Economics at University of California at San Diego. Previously, he was a lecturer of financial economics at Birkbeck College and a Professor of Finance at London School of Economics. In the Spring of 2003, he will be the Kaiser visiting Professor at Stanford University. He is a fellow of the Center for Research in Economic Policy. Dr. Timmermann holds a Master's degree in economics from the London School of Economics and earned his PhD from the University of Cambridge in 1992.

Dr. Timmermann has given lectures and invited talks to academics and practitioner audiences on topics in finance, economics and statistical methods in the UK, Europe, Australia and the US. His research has been supported by the National Science Foundation (US), Banco Swisse Italiano (BSI) and UK Inquire. Dr. Timmermann has consulted for several companies and institutions including the European Central Bank, the International Monetary Foundation (IMF), the Board of the Federal Reserve of the USA, Barclays Global Investors and Oppenheim.

Allan Timmermann is an Associate Editor of the Journal of Asset Management, the American Statistical Association's Journal of Business and Economic Statistics, Journal of Economic Dynamics and Control and a Departmental Editor of the Journal of Forecasting. He has published papers on asset prices and the performance of UK unit trusts and pension funds in major international academic journals.