Occasional Paper Series ◆ 9

Past Imperfect?

The performance of UK equity managed funds

Mark Rhodes

August 2000

FSA
Financial
Services
Authority

FSA OCCASIONAL PAPERS IN FINANCIAL REGULATION

Foreword

The FSA is committed to encouraging debate among academics, practitioners and policy-makers in all aspects of financial regulation. To facilitate this, the FSA is publishing a series of occasional papers in financial regulation, extending across economics and other disciplines.

These papers will cover such topics as the rationale for regulation, the costs and benefits of various aspects of regulation, and the structure and development of the financial services industry. Since their main purpose will be to stimulate interest and debate, we welcome the opportunity to publish controversial and challenging material and to publish papers that may have been presented or published elsewhere.

The main criterion for acceptance of papers, which will be independently refereed, is that they should make substantial contributions to knowledge and understanding in the area of financial regulation. We will encourage contributions from external authors as well as from within the FSA. In either case the papers will express the views of the author and not necessarily those of the FSA. Comments on these papers are welcomed and should addressed to the series editors.

Authors wishing to contribute to this series should contact Clive Briault or Paul Johnson at:

The Financial Services Authority
25 The North Colonnade
Canary Wharf
London
E14 5HS

Telephone: (0)20 7676 3100 or 3120

e-mail: clive.briault@fsa.gov.uk or paul.johnson@fsa.gov.uk

FSA Occasional Papers are available on the FSA website www.fsa.gov.uk

Past Imperfect?

The performance of UK equity managed funds

Mark Rhodes

FSA Occasional Paper

Biographical Note

Mark Rhodes is an associate in the Economics of Financial Regulation Department (Central Policy) at the Financial Services Authority.

Contents

Executive Summary

This paper presents evidence on whether information on the past investment performance of unit trusts can be useful to retail investors. The conclusions are important for investors who may be unsure of the value of information on past investment performance when choosing a fund. The issue is also important to the FSA in considering what information to publish in its comparative information tables.

A number of researchers have examined whether past investment performance repeats. The conclusion from an examination this literature is that repeat performance (if there is any) is both small in the size of effect and short lived. Repeat performance (persistency) is most evident for smaller poorly performing funds but the degree of persistency remains low. The conclusion from studies of UK unit trusts, and the more reliable of US studies, is that retail investors could not usefully exploit information on past performance.

Some of the literature suggests that persistency is even less evident for recent years. Evidence is provided on whether this may be true for funds in the largest four UK sectors. Both charts and descriptive statistics suggest insignificant persistency since London moved to a system of electronic trading. This is tested further by using a new measure of persistency. Relative to most other studies this new measure examines persistency over a longer time horizon. The approach has greater relevance to the practical choice faced by consumers. The results from using this new measure support earlier conclusions. There is no evidence that investment performance persisted over recent years.

To conclude, information on past investment performance would not, in general, be useful to retail investors. Findings from the literature and our own study do not commend the inclusion of information on past performance in the FSA's comparative information tables.

Introduction

If the past investment performance of managed funds is significantly persistent and this provides useful information about future performance then consumers would be foolish not to make use of such information in choosing where to invest. Firms would also have an incentive to put appropriate resources into maintaining and advertising their out performance.

However if there is no significant link between past investment performance and future outcomes consumers should not take heed of any information on past performance. To the extent that they nevertheless use such information they could be led into error and give inadequate weight to the features of the product that do matter. In addition if there is inertia in behaviour, so consumers buy into funds with good recent investment performance but do not sell when their performance worsens, then a further problem emerges. Funds which have been 'lucky' in the past will have an incentive to advertise their past performance, even though this provides the consumer with no useful information.

This issue has become more prominent for the FSA in the context of what information should be contained in the Comparative Information Tables that it is committed to publishing.

This paper is intended to present the available evidence on whether past investment performance persists. The next section outlines some theoretical issues concerning the likelihood of performance persistency in different market situations. The paper then discusses some of the empirical literature that has been accumulating over the last few decades and then provides an overview of UK data on persistence. The final section uses a novel approach to the measurement of persistence, putting it in the context of consumer preferences and presents some formal tests of persistence.

1 Theoretical issues

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 or speedier way than most other managers. This will allow them to trade stock 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 skillful fund management, which might be found in better research or trading.

The research conducted by fund managers is important in promoting efficient capital markets. This ensures that firms can raise capital at reasonable prices and that investors are appropriately compensated for the risks of their investments. If capital markets are fully efficient[1] then stock prices will adjust very quickly to reflect all new information on the value of the firm. So as markets become more efficient it will be 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 out-performance of her peers for a long period of time. To quote Bernstein (1998) 'Many smart people fail to get rich because people not so smart soon follow in their footsteps and smother the advantage their strategy was designed to create' (pp299). A corollary is that persistently good performance by one or more funds is likely to indicate relatively inefficient capital markets. It would imply that available information is consistently used too slowly or poorly by other managers.

Since it is costly to purchase and sell stock, managers could also persistently underperform their peers if they over-trade. If this happened it would provide a reason for informing customers about past investment performance. So if consumers are aware of this under-performance they can avoid the fund concerned.

1 Fama (1970) discusses market efficiency and conditions under which it might be met.

The question is whether any persistency in past performance is sufficiently material to include in the FSA's comparative information tables. Note that firms are able to publish past performance figures on their own account. So a decision by the FSA not to publish this comparative information will not remove performance information from the market.

2. Literature Review

A number of commercial organisations and academics have examined the performance of managed funds and whether relative performance persists. This literature is useful in allowing us to examine a range of evidence. However, although there are studies of both unit trusts and pension funds in the UK these are not as numerous as those conducted on US funds. Studies of US funds are reported here because the potential for a fund manager to outperform her peers is, as discussed above, influenced by the extent to which the market is inefficient and fund managers have different skills. Arguably, US equity markets are not very different to those in the UK. It is probable, then, that the conclusions of US studies will 'broadly' apply to the UK. The US literature is therefore used as a cross-check on the findings of studies in the UK.

Examining persistency in the performance of managed funds is a difficult process because there is a great deal of noise in the data. The returns that one might expect from a unit trust have a significant random element, capable of swamping any systematic differences in the skills of fund managers. Some funds may perform persistently well even over relatively long periods of time. However if enough funds are examined to start with this outcome might be found even if relative performance is truly random. It is not possible to reject the claim that, in the case of any individual fund, such persistence is the result of skill rather than luck. What can be examined, however, is whether the general pattern of relative performance is random.

UK fund managers

There is a burgeoning interest amongst academics and others in the question of whether investment performance repeats for UK fund managers. A simple analysis of whether relative performance persists was undertaken by WM Company (1999), in comparing active and passive fund management. They examined funds in the top 25% of performers over five year periods in the UK Growth and Income Sector for a period of 20 years (1979 to 1998). The proportion of these funds that remained within the top quartile (on a five year rolling basis) was then examined. This proportion declined. When there was no longer any overlap between the sample periods the general result was that there were no more funds still in the top quartile than would be expected by chance. The conclusion is that good performance is of no use in predicting the future.

This approach was formalised (partly) by an examination of the probability of a fund moving from one quartile to another (funds in quartile 1 being the top performers and in quartile 4 the worst). For example one quarter of funds would be ranked in quartile 1 over an initial period, say 1984 to 88. The proportion of these funds that were then found in each individual quartile for a subsequent period, say 1989 to 93 was then recorded. If relative performance were random then one would expect the quartile in which a fund starts to have no impact on a subsequent (non-overlapping) period. The chances of each subsequent ranking (1-4) will be the same *if relative performance does not persist*.

WM found that funds which are initially ranked in quartile 1 show no more than a random chance of being similarly ranked in a subsequent period[2]. In fact they found that a fund ranked in quartile 4 showed a stronger probability of subsequently being ranked top. The conclusion was that historic top quartile performance was not a good guide to picking funds. One of the tables is reproduced below[3]. WM (1999) summarised the findings in the following way 'Over the entire period of study the probability of selecting a top quartile performer based on historic top quartile performance was no better than would be expected by chance' (pp 28).

---

2 The exception was funds in the top quartile over 1989 / 93, of which 40% were still in the top quartile over the period 1994 /98. This methodology is not very robust, but the general result provides an indication that persistency is not evident.

3 These are the aggregate results for funds from the UK Growth and Income sector for the initial and subsequent periods of [1979 – 83 against 1984 –88], [1984 – 88 against 1989 – 93], [1989 – 93 against 1994 – 98]

Table 1: Past and Subsequent performance rankings of UK managed funds.

	Quartile positions in period 2				
	%	1	2	3	4
Quartile positions in period 1	1	22.2	22.2	33.3	22.2
	2	19.1	34.0	17.0	29.8
	3	13.0	17.4	32.6	37.0
	4	43.8	25.0	16.7	14.6

Source: WM Company

The direct examination of persistency adopted by WM may not be the most accurate or reliable measure. A rational investor will be interested in the risk that he is adopting when placing his funds with a particular manager. For example, 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. Measures might also be biased if they fail to account for the different performance of funds that have been closed as a result of poor performance. Funds which, by chance alone, perform poorly may be closed because customers choose not to invest with them. If there are costs to setting up a new fund the manager may decide on closure rather than committing to further advertising expenditure or fixed administration expenses[4].

Lunde et al (1999) model the decision to close a fund and examine the effects of fund closure on persistency. They find that performance does affect the decision to close a fund and note that this implies, although it does not prove, persistently poor performance by funds. They ask the reasonable question - otherwise why close them down? Industrial economics can provide a further explanation for this:

If performance is random but fund managers can attract poorly informed customers through advertising then there may be an incentive for a fund management group to offer lots of brands. By chance some funds will perform well. It will thus prove possible to attract a lot of new custom by advertising these funds heavily. In addition, if there are fixed (sunk) costs to setting up a fund then brand proliferation will act as an

4 This suggests that a fund may need to achieve a certain size or be supported for a certain length of time before it becomes viable.

entry deterrent. A new fund manager would have to offer a number of new managed funds in order to capture enough of the market. Each fund would also have to be supported for long enough for a performance record to be built up. A result of this is that fewer new firms than expected would enter the market to provide active fund management services. Existing firms may offer a higher than expected number of new funds to replace those that, by chance, have a poor performance record. (One might expect a high proportion of new entry that does occur to concentrate on index trackers, for which the issue of past performance is less relevant).

When Lunde et al (1999) included all funds in the study, non-survivors as well as those still around, they did find evidence of persistency. This is attributed to persistently poor investment performance by funds that were closed. The methodology they used is well suited to illustrating the problem of bias introduced by excluding information on non-surviving funds. Examining only surviving funds prevents comparison with funds that may have performed poorly before being closed. However the methodology, similar to that of WM Company (1999), may provide less reliable conclusions on persistency[5].

Quigley and Sinquefield (1998) focus more closely on whether past performance is a good predictor of the future. They take monthly returns on all UK equity unit trusts that were in existence at any time over the period 1978- 97. The authors form equal-weighted portfolios of funds based on their performance ranking, by decile, over an initial year. The performance of each portfolio over the following year is then examined. Funds are included in the analysis of each subsequent period until the point they 'die' and then the portfolio is re-weighted for the rest of the year. The authors repeat this process across the entire sample. So the top 10% of funds in 1978 are formed into a portfolio and the performance of this portfolio in 1979 is then calculated. For 1979 a different top 10% of funds is then found and so on.

The authors take the average, subsequent, performance across all of the years of these top 10% portfolios (so the performance figures are calculated for the years 1979 to 1997). A similar average is calculated for the bottom 10% portfolios. There is an apparently substantial difference in the average performance figures between these top and bottom sets of funds of 3.54% points. However the authors find that, as an investment strategy, reforming one's portfolio each year with last year's top 10% of funds would

5 Lunde et al (1999) concentrate on the factors leading to the closure of a fund and provide some useful insights. However more robust methodologies exist for examining persistency.

not produce higher returns than a simple buy and hold strategy. The costs of reforming the portfolios each year (on average selling 80% of the funds that were held in the previous period) would be too high.

The authors also provide a more rigorous analysis. They examine the monthly returns to each of the 10 portfolios formed by deciles relative to the market as a whole (over the whole period 1979-97). Two alternative adjustments for risk are made, one following a basic market model and the other also adjusting for size and value factors. In neither case is the subsequent performance of high decile portfolios found to be significantly different from the average. However low decile portfolios are found to have persistently poor subsequent annual performance with either type of risk adjustment made[6]. To quote the authors 'Losers repeat, winners do not' (pp7).

Fletcher (1992) took a sample of UK unit trusts and applied a further adjustment for risk and return. His approach, based on Arbitrage Pricing Theory, allows for a wider range of factors to influence the expected returns to assets and hence funds than under other approaches. Using this different approach Fletcher (1992) found that using past performance to pick funds was not able to offer any advantages in generating high relative subsequent returns across the sample period.

The performance of other types of UK fund manager has also been examined, particularly pension fund managers. Blake et al (1998) examined over 300 institutionally managed UK pension funds. Conclusions on persistency are rather more difficult to draw. The market is more concentrated and observed differences in performance are typically very small. Differences found by Blake et al (1998) were largely the result of long term decisions on investment policy (the mix of assets chosen) rather than decisions that were taken on timing or the particular security in which to invest[7]. Blake et al (1998) argue that this provides 'more robust empirical support *in claiming* the absence of extensive attempts at active management' by institutional pension fund managers (pp21 my italics). In other words pension fund managers do not adopt a very active trading strategy, with the result that there is little economic difference between the returns that they achieve.

6 This is the only result that is statistically significant. Further results are reported when explicit expenses are added back into the analysis and interested readers are referred to the paper for a discussion of these.

7 We might characterise investors in unit trusts as making their own decision about the mix of assets, an example being the decision on the amounts to place in a unit trust rather than cash.

Brown et al (1997) found 'limited evidence of persistence in pension fund performance' (pp 176) in a sample of some 600 pension funds. The findings are similar to Blake et al (1998) in that house style is cited as the main source of persistency in performance, analogous to the earlier findings on asset allocation. The conclusion we may draw on UK pension fund managers is therefore that although there is some evidence of persistently good performance, this is largely explained by factors other than shorter term stock picking or market timing (i.e. skill)[8]. The spread between good and bad funds is also very low in comparison with that for unit trusts.

Studies of US fund managers

This section provides an indication of the findings of studies carried out in the US. This is likely to be informative in developing our understanding of persistency in UK unit trusts. The stock markets in both countries are well developed and fund managers are given similar tasks. It seems unlikely that there are considerable differences between the skills of UK fund managers without a similar situation emerging in the US. The exception would be if the UK market is considerably less efficient than the US, although this again does not seem likely. Analysing persistency in the US mutual fund sector is likely to give more robust results, there being more funds and therefore more data on which to perform statistical tests. A discussion of this literature will therefore provide a useful cross-check of our conclusions for the UK.

There are a number of US studies where persistency in performance is found to be both economically and statistically significant i.e. potential investors can gain by picking funds that have performed relatively well in the past. Grinblatt and Titman (1992) found that there was exploitable persistency in the performance of mutual funds, which could not be explained entirely by differences in fees and transactions costs. The implication of this study of two non-overlapping five year periods (over 1975 to 1984) was that fund managers exhibited skill in the timing of investment. Hendricks et al (1993) found that persistency existed over the short term, a phenomenon they christened 'Hot Hands' in the case of good performers. However this result should be interpreted with more care following the work of Carhart (1997) who argues that this is a *momentum* effect and largely the result of luck rather than skill (this explanation is

8 Objectives are, nominally, different for firms managing more than one UK unit trust fund (i.e. UK Equity Growth and UK Smaller Companies). Drawing a comparison between funds, rather than managers, is therefore appropriate in the case of unit trusts but returns may have to be adjusted for risk.

expanded upon further below). Goetzman and Ibottson (1994) find that winners repeat (that there is persistent good performance) although a later paper by Brown and Goetzman (1995) finds that most persistence is due to poor performance.

In a more recent paper Elton et al (1996) make the strongest claim that performance, good or bad, persists and that it is economically significant for the investor. This paper suggests that most of the literature is incorrectly adjusting for risk in tracking the subsequent performance of portfolios of funds[9]. The authors prefer to track funds individually and then to aggregate performance. This methodology finds evidence for persistency into the longer term, as well as strengthening the results of others who find that short term (one year) persistency exists. They find that momentum effects alone are insufficient to explain short-term persistency. The conclusion is that past performance can provide useful information in picking mutual funds. However the reliability of this result is questionable because of the possibility that bias is introduced by their treatment of non-surviving funds[10].

There are a number of more recent studies that contrast with the above findings. In a study of US funds Carhart (1997) finds from his analysis that 'The results do not support the existence of skilled or informed mutual fund managers'(pp57). His paper adopts a methodology later followed by the Quigley and Sinquefield (1998) paper discussed above. Portfolios of mutual funds are formed from the ranking, by decile, of funds' performance in the previous year. The performance of these portfolios over the following year is then examined. As with Quigley and Sinquefield (1998) portfolios are re-formed for every year for which there are data. Carhart (1997) finds that there is a one year persistence in the investment performance of mutual funds. This is attributed to luck, rather than being taken as evidence that some fund mangers are more skilled than others. By chance a mutual fund may hold stock that produces a high return in a given year. It is unlikely that these better returns cease at whatever year-end one might choose. So there is a momentum effect, whereby mutual funds which by chance

9 An example is given that the composition of the portfolio formed by the top decile of funds will change over time. The riskiness of the portfolio relative to the market will also therefore change, with the effect that adjusting the performance of the portfolio for risk (assumed constant over time) is inaccurate.

10 The standard way in which managed funds 'die' in both the US and the UK is through merger. Elton et al (1996) find that all funds in their sample die in this way and claim that tracking them through mergers removes survivor bias. This approach does not account for the fact that the fund's manager changes and possibly also the investment strategy. The result is that the nature of bias may change but is not removed.

happened to be holding this good performing stock appear to perform relatively well in the following year too. Although the managers of other mutual funds may decide that it is a good idea to buy the stock concerned, there will be costs incurred in them doing so which would reduce their observed performance. So the result is the presence of some short term persistence in investment performance. But it would be costly to switch funds to a different fund manager on an annual basis and so an important question is whether these short run effects, lucky or not, carry over to a longer period of time.

Consequently Carhart (1997) also studies persistency in investment performance over the longer term. To do this the performance of each fund is measured over 3 years. Adjustments are made to the performance measure for the riskiness of the fund relative to the market and for four other factors: expense ratio, turnover, size and load fees[11]. The adjustment is made so that a measure of the different skill of the fund managers is obtained, rather than some measure affected by other influences not directly related to such skill. This adjusted performance measure is then used to rank funds by decile. Portfolios of these funds are then formed as in the analysis of the short run described above. The performance of these portfolios of funds is then examined over the following 3 years. The measured subsequent performance of the portfolios is also adjusted for risk and the four factors mentioned above[12]. This process is repeated for each 3 year period. Investment performance is found to persist for the top and especially the bottom decile portfolios.

Recall the discussion of the momentum effect mentioned earlier. On average the performance of the top decile portfolios is considerably better than the rest. One would therefore expect the momentum effect to result in considerably better average subsequent performance for these top decile portfolios. However Carhart (1997) finds only a small difference in the subsequent performance of these portfolios (although the difference is statistically significant). The implication is that an insufficient amount of short term persistence carried into the longer term to be of benefit to investors. So an investor would not benefit by using past performance to try to pick a winner. Significant and persistent poor performance by bottom decile funds remains.

11 See Fama & McBeth (1973) or Carhart (1997) for a discussion.

12 If the adjustments for performance miss something important out then the effect might be that persistency is thought to be significant where this is not truly the case. Carhart (1997) downplays the importance of these later results because of this possible problem.

These funds typically have a very small capitalisation, so this form of persistency relates to only a small fraction of the total market.

The adjustments made to the performance measures provide some further interesting results. Expense ratios are found (on average) to reduce performance a little more than one for one. So a higher charging fund provides a slightly worse performance as well. Also every round trip trade, selling some stock and then buying different shares to the same value, reduces performance, on the amount traded, by 95 basis points[13]. The study concludes that there is only weak evidence that fund managers possess different levels of stock picking skill but expenses and trading costs have a demonstrable effect. Carhart (1997) summarises the paper in stating 'While the popular press will no doubt continue to glamorise the best performing mutual fund managers, the mundane explanations of strategy and investment costs account for almost all of the important predictability in mutual fund returns' (pp 81).

Brown and Goetzman (1995) examine year on year persistency for US mutual funds over the period 1976 to 1988. They analyse the year on year persistency in investment performance separately for each possible span of two years. Persistence is found to be strongly related to the time period under examination. For example some years in the sample show reversals[14]. This finding implies that any relationship between the past and the future is fragile. The paper does find that investors can 'beat the pack' by using past performance to pick funds. However much of the persistency is again attributed to poor performers.

Phelps and Detzel (1997) also find that the time period over which one examines performance is likely to affect conclusions over persistency. They find that persistence disappears when the more recent past is examined, or when more comprehensive adjustments for risk are made. The paper provides a succinct critique of much of the literature which has found evidence of persistency. They demonstrate that unexpected changes in general economic conditions, rather than fund manager skill, can be used to explain earlier results. Some weak evidence is provided that funds with high expense ratios should be avoided. However the conclusion is that past performance is of no use in picking tomorrow's winners or avoiding losers. Sauer (1997) also fails to find persis-

13 This is a measure of the effects of turnover and only decile 1 funds are found to cover the costs of trading (just).

14 Higher initial performance in 1980 and 1987 is associated with low relative returns in each of the following years.

tence in mutual fund performance, once he has accounted for the different investment objectives of the funds. He finds that prior period performance is of little value in the selection of mutual funds.

Summary of discussion on the literature

There is only weak evidence that the relative performance of managed funds repeats. When investment performance is found to be persistent this occurs only over relatively short periods of time. Poorly performing funds are most persistent and this result appears to be largely responsible for finding any evidence of persistency at all. However these poorly performing funds are small and represent only a tiny fraction of the total market. So in practice the value of information on past performance may be limited. An important further fact is that only a small portion of good or bad performance ever persists into a subsequent period. As a result only a few of the academic studies (in the US) have found a procedure to use past performance that would have been of benefit to customers in practice, especially when account is taken of switching costs. The weight of evidence is therefore that past performance is not of use to retail investors choosing a fund.

Some of the studies in the US have found that any persistency is even weaker in recent years. Studies of the UK have not addressed this possibility. If the contemporary experience is of no persistency in performance then the use of older information (which may exhibit limited persistency) may present a misleading picture for investors.

3. Data Description

Overview - Unit Trusts in UK Sectors

This section provides a direct look at those unit trusts into which the majority of UK investors place their funds. The data are described and a number of graphs chart the

actual and relative performance of funds and present a picture of the riskiness, or volatility, of different types of fund. There is also a simple examination of persistency.

For the purposes of this paper data have been acquired on all unit trusts that were in existence at any point over the years 1980 to 1998 inclusive and that were in one of four sectors. These sectors are UK Equity Growth, UK Growth and Income, UK Income and UK Smaller Companies. Index tracker funds largely fall into the UK Growth and Income sector and can be identified separately. Performance is measured bid to bid with income re-invested. Throughout the figures for fund performance are net of trading costs but gross of explicit costs such as annual management charges or the cost of advice. The data, obtained from S&P's Micropal, report the monthly return to funds. The table below provides some summary statistics. Note that the number of unit trusts in existence normally increases with time.

Table 2: Number of funds and average return

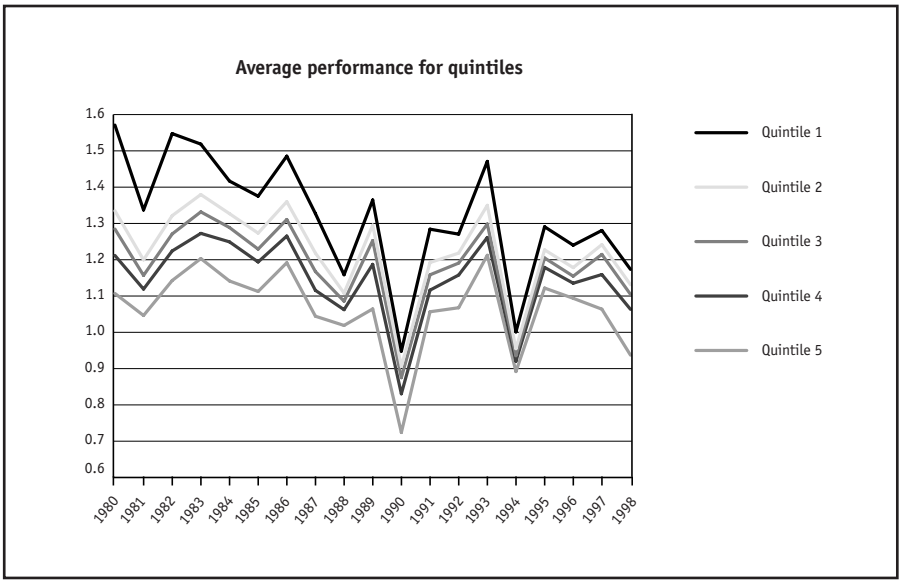
Year	1981	1984	1987	1990	1993	1996	1998
Number of funds	202	232	345	437	437	458	439
Average Return	17%	28%	17%	-14%	31%	16%	8%

Source: S&P Micropal

For each year in the sample the performance of all funds that existed from January to December was examined. Funds were ranked according to quintile and the (unweighed) average performance within each quintile was calculated. In 1983 the average performance of funds in Quintile 5 was 20% nominal growth. This is recorded as a return of 1.2 in Chart 1. Similarly if funds within a quintile on average lost 20% the performance would be recorded as 0.8. The full results are plotted in Chart 1. As might be expected the performance of each quintile follows a similar pattern over time. The appearance is of generally high returns achieved during the early 1980s[15]. Some trends do appear within the data, most noticeably there appears to be a convergence, at least in the absolute difference, between quintiles from the 1980's to the end of the 1990's.

15 These high returns are verified from other sources, see for example Quigley and Sinquefield (1998).

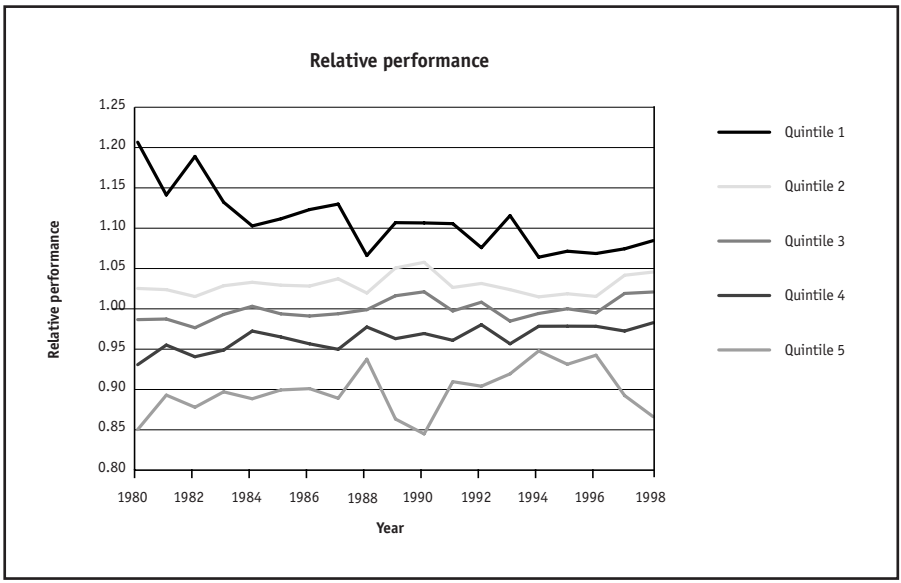
Chart 1



Trends may be easier to see if the relative performance is plotted rather than the actual. This is found in Chart 2. The performance of each fund was divided by the (unweighed) average performance across all funds in existence over the year concerned. So a recorded performance of 1 means that the fund (or quintile) matched the average outcome across all funds. The performance of funds in the top quintile (Qn1) appears to converge on the average, as suggested by Chart 1. It is more difficult to draw a conclusion about funds in quintile 5 (Qn5). Note that one would generally expect the plots of Qn2 through Qn4 to be close together relative to the other quintiles as there is (theoretically) no upper bound to the performance of funds in Qn1 and the lower bound for Qn5 is 0 (in the extreme investors might lose all of their money but no more)[16].

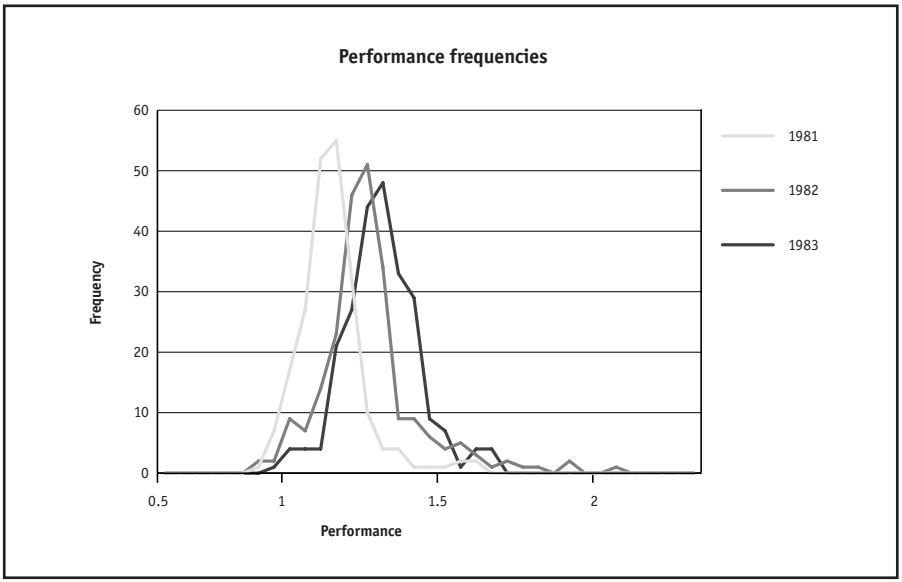
16 There is no evidence of greater volatility amongst funds that perform relatively well. Please see the appendix for further details.

Chart 2



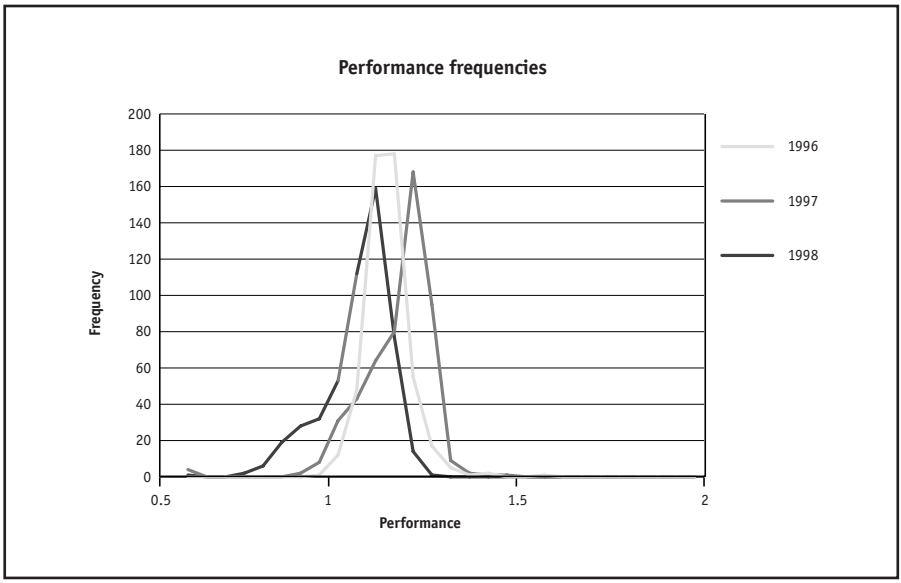
Averaging performance across quintiles may hide differences in how the underlying funds actually perform. For example, the absence of an upper bound on the performance of funds in Qn1 suggests that the average performance could be dragged up by one or two funds that performed spectacularly well. A visual check of this is found by plotting the histogram of performance. Chart 3 plots the distribution (histogram) of actual performance across all funds for the first three years for which we have data. It is apparent that the right hand tails of these distributions are long in comparison with the left hand side. This suggests that the Qn1 results presented earlier are somewhat skewed by the high performance of a small number of funds. The number of funds is not particularly high for these years, which means that these outlying funds have a greater impact on the average for Qn1 than would otherwise be the case.

Chart 3



The distributions for the early 1980s contrast with those for the end of the 1990s that are plotted in Chart 4. In these later years there is not such a long right hand tail to the distribution, leading to a different conclusion about the nature of relative performance. With fewer unusually high performing funds and a larger number of funds to observe in total, the average performance for Qn1 that is recorded for the 1990s appears to be more trustworthy. The earlier conclusion about the narrowing of the difference between quintiles over time also seems reasonable given the evidence from Charts 3 and 4. Note that the distributions appear more narrowly spread for the 1990s, with the possible exception of poorly performing funds.

Chart 4



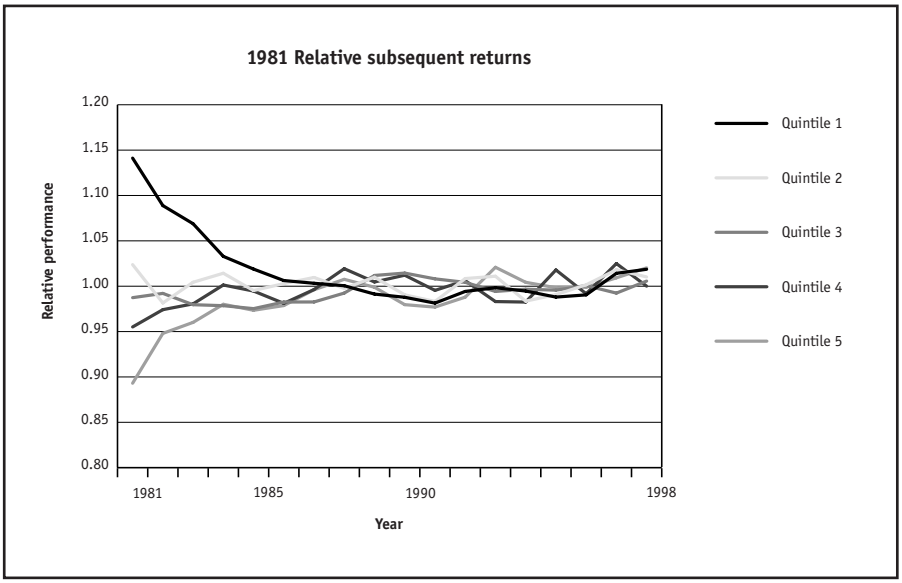
Each of the four sectors from which we take data (five if one counts trackers separately) has different investment objectives. This may have an effect on the measured performance or persistency of funds when analysed by quintile, as in the preceding charts. In fact Charts A2 and A3 in Appendix 1, show little evidence of any difference in investment performance and volatility. The only exception is the UK Smaller Companies sector, which appears to be a little more volatile than the average.

A visual clue as to any persistency in the performance of managed funds is provided in Chart 5. Each fund was ranked by quintile according to its performance in 1981. The performance of each quintile was then divided by the unweighed average across all funds in the same year. The first points in the graph show the difference in the performance of quintiles in 1981 (the same as the corresponding year in chart 2). Each quintile in 1981 then forms a cohort of funds that are followed through the rest of the sample. For example in 1984 the performance of the same funds (so long as they are all still alive) is measured relative to all funds that are in existence throughout 1984. If performance does not persist then we would expect funds that formed part of Qn1

in 1981 or, indeed, any of the 1981 quintiles to have an average subsequent return equal to 1 in 1984.

From Chart 5 we can see that there is a very large difference in the relative performance of funds in the first year (1981). Note also that the performance of each of the portfolios of funds subsequently converges on the mean performance. The order in which middle portfolios (Quintiles 2,3 and 4) are ranked does not even last to the following year. Funds ranked in the top 20% in 1981 do, however, appear to have a consistently better performance than average for a number of following years. But after 5-6 years all of the portfolios, irrespective of their performance in 1981, show no better than an average return. Chart 5 suggests that there may have been some value in using past performance to pick funds. However it would be foolish to draw this conclusion solely on the basis of one year's past performance nearly 20 years ago.

Chart 5



An important question arising from Chart 5 is whether the pattern of persistent performance is repeated for other start dates. Charts A6 and A7 in appendix 1 (start dates 1983 and 1985) continue to show some persistency, although this appears to end around the same year (1987). But Chart 6 fails to show any persistency for funds ranked by quintile in 1987. A similar conclusion may be drawn from Chart 7 (for 1989), although there does appear to be persistency over the first year. Chart 8 (for 1991) shows no persistency at all. Note that even where there is some apparent persistency the size of this, relative to differences between quintiles in the base year, is fairly small (and it declines rapidly).

Chart 6

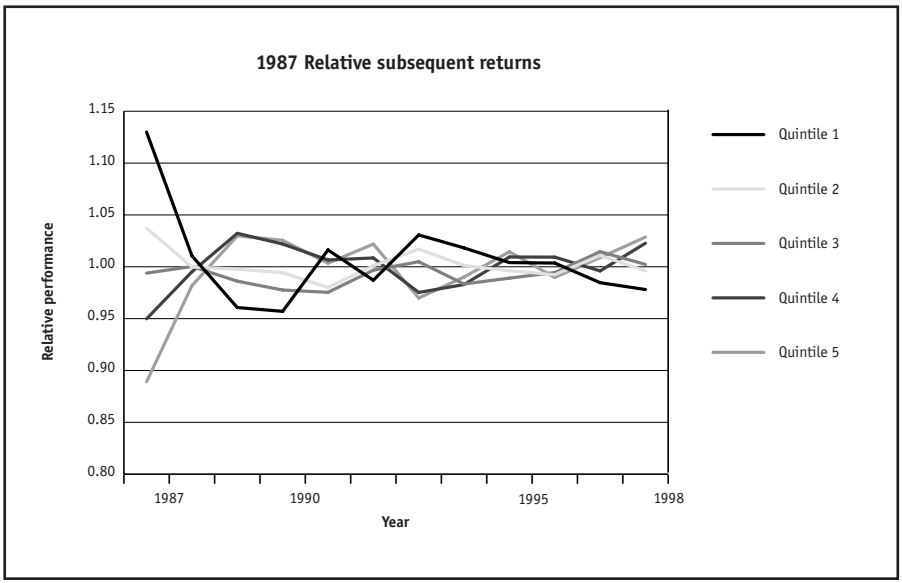


Chart 7

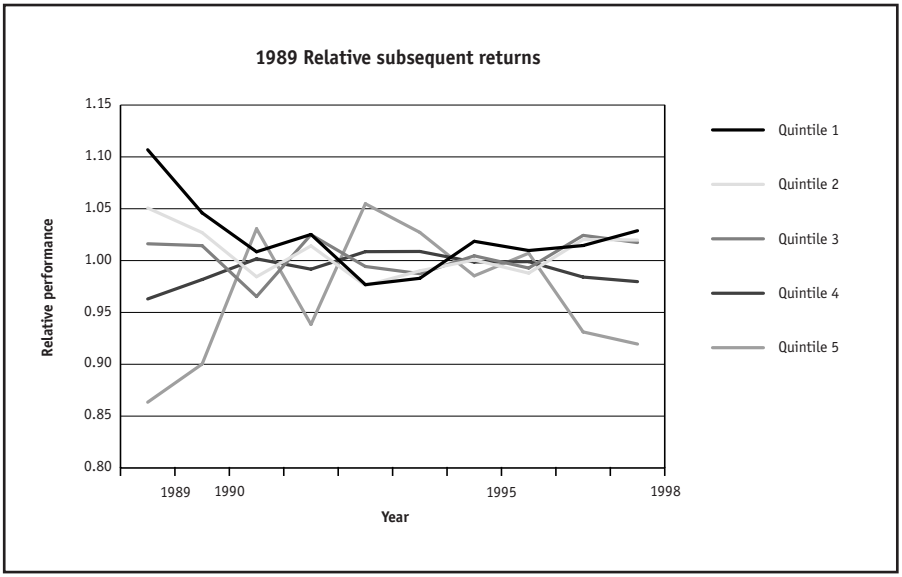
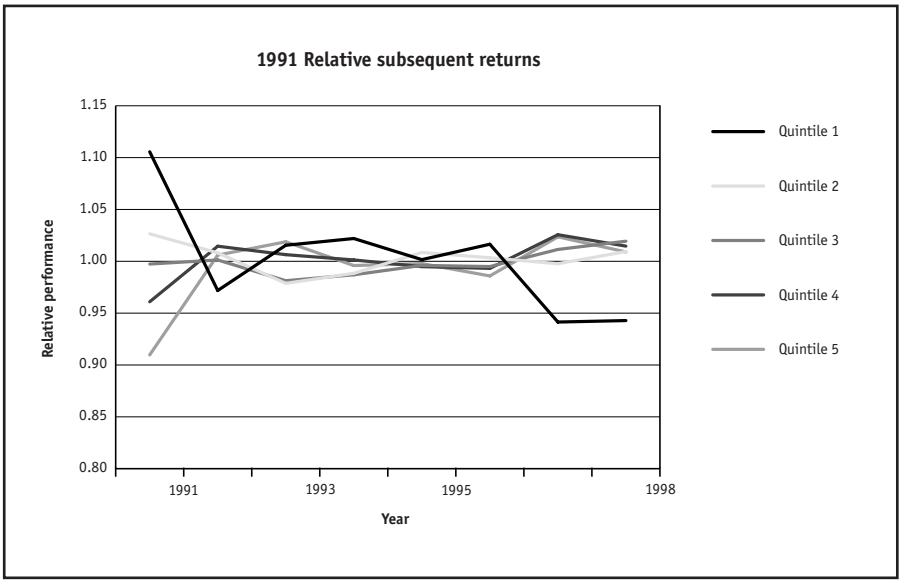


Chart 8



From our visual inspection of the data we might conclude that relative performance persisted during the early 1980s. But from around 1987 the situation appears to have changed. For the last 10-12 years differences in the performance of unit trusts have narrowed (as measured by returns to different quintiles) and there is much less evidence (if any) of persistent performance, good or bad, for this period. On the basis of the simple analysis presented above past performance does not now appear to be a useful or exploitable indicator. However, a better approach to understanding performance and persistency is to examine the market as a whole.

A simple statistical test of whether winners (or losers) repeat may be conducted by examining the number of funds that retain their performance rank from one period to the next. To do this I took all funds which were in existence over an initial period (say 1981 to 82) and a subsequent period (say 1983 to 84). The performance of these funds relative to one another[17] was then ranked for each period. One may then examine, for example, the number of funds that were ranked in the top 20% over 1981 to 82 and which still had this performance ranking over 1983 –84. We can test whether the overall pattern of initial and subsequent performance is different from what would have been expected were there no persistency at all. Three cases are examined in the tables below.

Table 3 shows the results for the period 1981-82 against 1983-84 and, as discussed above, funds were ranked according to quintile. Rows relate to initial fund rankings and Columns to the percentage of funds within the subsequent rankings. So in the first table 27% of funds that were initially ranked in Quintile 2 were subsequently found in Quintile 1. The pattern is different (statistically)[18] from what would have been expected by chance and this is driven by a larger than expected number of funds retaining their ranking (particularly Quintile 5). There is one exception to this retention of rank, with a relatively large number of funds that were initially ranked in Quintile 1 being subsequently ranked in Quintile 5. The statistically significant result for these time periods is made more striking by the relatively small number of funds in existence (186).

17 The results are similar but a little more difficult to interpret if one compares the funds that form the initial group with all of those that are in existence in a subsequent period (new funds are always being brought to the market).

18 A chi squared test shows whether the sum of the difference between the observed and expected number of funds in each cell of the table is significantly different from zero.

Table 3: Initial performance (1981-2) and Subsequent performance (1983-4)

Transition Matrix	Initial 1981-2			Subsequent 1983-4		
		Subsequent				
Initial	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	
Quintile 1	30	14	10	10	35	%
Quintile 2	27	30	10	22	10	%
Quintile 3	19	22	30	24	8	%
Quintile 4	19	24	24	22	10	%
Quintile 5	5	10	27	22	35	%
					Number of Funds	186

Row totals may not sum to 100% because of rounding

A second *transition matrix* was produced for the periods 1991-92 against 1993-94 (see Table 4). Perhaps contrary to expectations, given the graphical analysis above, the pattern for this period is also found to be statistically different from the pattern that would have emerged by chance. However a closer inspection of the table shows that this is the result of reversals in the performance of funds rather than winners or losers repeating.

Table 4: Initial performance (1991-2) and Subsequent performance (1993-4)

Transition Matrix	Initial 1991-2			Subsequent 1993-4		
		Subsequent				
Initial	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	
Quintile 1	20	28	17	7	29	%
Quintile 2	7	12	26	24	32	%
Quintile 3	12	21	22	26	18	%
Quintile 4	22	21	18	26	12	%
Quintile 5	39	18	16	18	8	%
					Number of Funds	380

Table 5: Initial performance (1995-6) and Subsequent performance (1997-8)

Transition Matrix		Initial 1995-6			Subsequent 1997-8		
		Subsequent					
		Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	
Initial	Quintile 1	16	12	15	24	32	%
	Quintile 2	21	22	32	16	10	%
	Quintile 3	22	27	19	18	15	%
	Quintile 4	25	21	18	22	15	%
	Quintile 5	18	18	18	19	28	%
						Number of Funds	368

The analysis was repeated for the period 1995-96 against 1997-98 (see Table 5). There was again some evidence of a reversal in the relative investment performance of funds, with repeat performance only appearing to be in evidence for poor performers (in quintile 5). However the general pattern was not different from a random outcome.

These results raise the possibility that where academic studies have discovered persistency this was driven by earlier periods when capital markets were less efficient. Both market innovations, such as 'Big Bang', and the deepening of the market in general provide an explanation for how market efficiency might have improved. A good overall return achieved by investing in last year's best performing fund was only evident in these early periods. This simple evidence from more recent years therefore supports the view that relative performance does not persist.

4. Testing for Persistency

Measuring Performance

Having discussed the literature and described the pattern of performance using some graphs and basic statistics we now go on to construct our own measure of persistency, to be used with the data already described in the last section. The grounds for developing a different approach are explained next, followed by a discussion of the new methodology. The results are then presented.

Most studies of performance for managed funds rank them according to relative outcomes over an initial period and then examine how this is related to subsequent performance[19]. A potential problem with this approach is that it may fail to pick up persistence where, for example, a fund moves from a high rank to a middle rank and then back again. A solution to this problem is to develop a measure of persistence that tracks a fund over time. However in comparing the investment performance of funds for a number of different time periods one needs to consider investor preferences. It is costly for the typical customer to switch funds or to hedge risk. In this context it is difficult to support an approach that assumes consumers to be neutral to risk. Also we are interested in deviations from the typical or average return and preferences over this are relevant. As a result the customer's preferences over different levels of return form part of the analytical tool.

Ideally therefore a measure of persistency should:

1) Track a fund across all the time periods for which observations on its performance are available

2) Allow for the inclusion of funds that start in the middle of a sample period

3) Allow for the inclusion of funds that are closed during the sample period

19 If the investment performance of funds is ranked rather than used directly the analysis avoids having to provide an explanation for general market movements, an unnecessary task when examining persistency.

4) Be based on a 'reasonable' view of customer preferences and attitude to risk

5) Be flexible (in terms of measuring relative performance and risk preferences)

6) Not be biased by an exceptionally good or bad outcome in a single period

7) Be robust and testable

8) Be easy to understand.

The measure of persistency is therefore grounded in assumptions on consumer attitude to risk and the difference in performance between different fund rankings. To explain how the measure is devised each of these two sets of assumptions is discussed in turn.

Investor preferences

Preceding analyses of persistency have made the assumption that only 'one shot' persistency is of importance[20]. This makes for a rather restrictive view of investor preferences, which this paper seeks to address. Retail investors are likely to be interested in the performance of funds over a longer period of time than a year or two. This is particularly the case when there is an explicit cost (an initial charge or bid / offer spread) when switching funds. Most actively managed funds apply such charges. It is therefore the persistency of returns over a longer period of time that is likely to be of interest, and something which the approach used here seeks to address. However in doing so some further, reasonable, assumptions about investor preferences need to be made. These assumptions are as follows

1) Investors dislike risk.

2) Higher performance is preferred but at a diminishing rate (an individual, call him Bert, would prefer performance that was twice as high but not by twice as much (this is closely related to 1)).

20 For example portfolios of funds may be re-formed annually but the composition of these portfolios is not tracked.

3) Dislike of risk does not change with the performance of the fund that Bert is invested in.

We can derive a structure for the preferences of Bert over different returns (his utility function). The maths for this may be found in Appendix 2. This structure imposes nothing more than the three assumptions explained above. However in order to provide a measure of performance we have to make some choices about the relationship between fund performance and ranking, and also about the extent to which investors dislike risk. The effect of these choices on the measure of persistency is examined carefully, and the sensitivity of the results to these choices is discussed further below.

What about Performance?

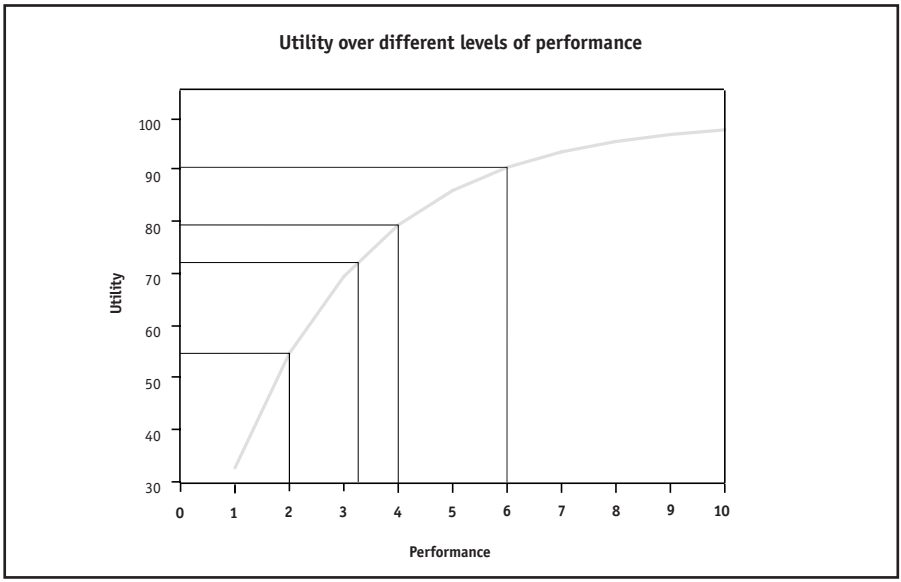
Over time the average performance of funds has changed. Chart 2 above also showed that the average performance of quintiles, and the difference between them, has changed over time. Since some unit trusts will have started during the sample period and others closed, comparing funds on the basis of observed performance figures or their average for any given year is not appropriate. Some funds would appear to have performed relatively well merely because they were around in the early 1980s rather than a later period where average returns were lower. Furthermore, persistency is about the relative performance of funds (whether a fund manager is able to consistently outperform her peers). As a result the following analysis examines the rank (by quintile) of funds and does not introduce a direct measure of performance. An indicative score is given to each rank but the absolute difference between neighbouring ranks is the same (for any single set of scores). In this respect the study is an extension of the methodology that examines the proportion of funds that retain (or change) their rank. The study will examine all years rather than producing a pairwise comparison between initial and subsequent rankings.

Combining Investor Preferences and Fund Performance

Fund performance is ranked by quintile for each calendar year. The difference, in absolute performance, between each quintile each year is assumed to be the same. From Chart 2 this assumption seems more relevant for the later years of the sample than for the early 1980s. Quintile 1 funds seem to have significantly better perfor-

mance for the early period. However Charts 3 and 4 show that there are more funds in the right hand tail of the distribution for the early 1980s. So although the average performance for Qn1 funds appears high to start, the median performance is not very different across the whole sample period. To compare funds with different life spans the performance measure assumes the same investment performance for each quintile for each year. The utility function (set of preferences over the rate of return) of our representative investor, Bert, is then applied to the assumed average performance of each quintile. How we score fund performance and persistency is therefore dependent on the shape of Bert's utility function. The next section of this paper provides results for a number of different assumptions about performance and the shape of the investor's utility function. An example is presented in Chart 9.

Chart 9



The Chart shows how Bert's utility, well being or happiness, changes as the performance of his fund changes. The utility scores will be used in the measure of persistency. Recall that the utility scores will relate to the quintile in which the fund is placed. The difference in performance between neighbouring quintiles is assumed to be the same. So in our chart if funds in quintile 5 have a performance of 2, those in

quintile 4 have a performance of 4 and in quintile 3 a performance of 6. In practice it is the utility scores associated with each quintile that are important. From Chart 9 Bert's utility increases as the performance of the fund increases. So with a performance of 2 we model, in this case, the utility of Bert to be about 55. However if the hypothetical performance of the fund is three times this first level Bert's utility is not three times as high, in fact it is a little less than twice as high. Persistency is addressed by considering Bert's choice of funds when first coming to the market. Assume that a particular fund has a ranking giving a constant performance of 4 across each of two periods (shown on the above graph). This gives Bert utility of 80 in each period. Now consider a fund that has a ranking that gives a performance of 6 in the first period and then 2 in the second. The average performance is again 4 but the average utility is lower (about 73 as shown). So Bert's dislike of risk means that he only values the riskier fund *as if* it had a constant performance of about 3.2. Note that Bert would prefer a fund that always had a higher return, whether this return is more volatile (around a higher mean) or not.

If these outcomes are instead taken as measures of past performance guiding Bert about where to put his savings then the second case is equivalent to Bert taking a bigger bet. This bet has the same expected money payoff (of 4) but has a lower expected utility. In this way we present persistent average performance as better than a more variable performance with the same money outcome. An alternative way to think of this is that Bert does not know when he will want to cash in his investment and so prefers the less volatile fund.

So for each fund the average utility may be calculated, across all of the years for which data on the fund's performance are available. The range of possible average scores becomes rather cumbersome to present in a table for many periods, but an example for two periods is presented below.

In any single period a fund is assigned a utility score according to which performance quintile it falls into. A further example of these scores is provided below.

Table 6: Example utility scores

Quintile	Utility Score
1	67
2	59
3	50
4	38
5	25

If the average utility scores are calculated across two periods then there are 25 possible combinations of scores, as presented in the table below.

Table 7: Average utility over two periods

Average Utility	Subsequent performance rank (period 2)					
		1	2	3	4	5
Rank in Period 1	1	67	63	58	53	46
	2	63	59	54	49	42
	3	58	54	50	44	37
	4	53	49	44	38	32
	5	46	42	37	32	25
Average	48					

There are a number of points to note from the above table. The first is that the average of the 25 values (48) is lower than the score that would be achieved by a fund always maintaining the middle rank (50). The expected (average) score is always 48, irrespective of how many time periods we average the utility scores across. Note also that if a fund were to flip rank between quintile 5 and quintile 1 it would have a lower utility

score (46) than the average. Within this framework a fund could have an average score as low as 25, or as high as 67, suggesting the presence of a persistently bad or good performance ranking.

Testing for Persistence

We would expect some funds to be ranked highly over the entire sample period as a matter of chance, even if relative performance were entirely random. The decision to include past performance in comparative information tables should therefore be taken on the basis of evidence drawn from the entire set of funds. We can then check to see whether more funds have high scores than we would expect by chance (in fact whether the pattern of average utility scores is what would be expected were relative performance random).

Appendix 3 explains that, if relative performance is random, the pattern of average utility scores is known. The average scores will follow a normal distribution. We can only state that the average scores will be normally distributed if we are happy that the distribution of scores, in each year, is statistically independent. They are only independent if relative performance is random. An alternative way to express this is to say that average scores will be normally distributed if there is no persistency in performance. This was verified by running a simulation, giving a large number of hypothetical funds a random rank for each of a number of years and examining the distribution of the average utility scores[21].

As is also explained in appendix 3, a statistical test for normality can be conducted in each case where a distribution of average scores is produced. This forms the test for persistency. If the distribution of average scores is non-normal then this indicates that there is a relationship between past and subsequent performance.

If all funds existed for the entire sample period the process explained above would be sufficient to test for persistency. However a significant proportion of funds were either closed during the sample period or were newly created. Taking the average scores for such funds could seriously bias the results. This is easily explained in an example. If

21 Performance was drawn from a uniform distribution for 500 funds across 18 years in each iteration. A number of utility functions were also later employed to ensure that normality was not driven by choices over this function.

relative performance is random the probability that any given fund will have two top utility scores in a row is 1/25 (recall that funds are ranked by quintiles so this figure is determined as 1/5 times 1/5). The probability that a fund has 18 top utility scores in a row is about 1/3,800,000,000,000. Our measure of persistency should not therefore weight equally the scores of funds for which we have different numbers of observations. Slightly more formally we can say that the variance of the average scores falls with the number of observations on a fund's performance.

The weights that should be applied to the scores can be found exactly. This was done and use of the correct method confirmed by running a simulation. Details may be found in appendix 4[22].

To conclude, this section has discussed a different approach to the measurement of persistency, which allows us to track the relative performance of funds over time, to a greater extent than is the case in most other studies. We have seen how this measure is derived from some 'reasonable' assumptions about investor attitude to risk. This section also states where the implications of these assumptions should be examined, and indicates that results of these tests will be reported further below. This leads us to the results of the formal analysis.

Results

The approach to measuring persistency explained above was adopted for the obtained set of data on UK unit trusts. That is all funds, both live and dead, in existence over the period 1981 to 1998 inclusive in each of four sectors. The process is relatively simple in that for each year all funds were ranked according to performance with each placed into a quintile depending on its rank. In each subsequent year the process was repeated and then the quintile into which each fund fell was compared with the previous year. This comparison allowed a utility score to be assigned to the fund. For example if the fund was ranked in the top quintile both the current and the previous year it would be assigned the highest possible utility score. Table 8 provides an example, but note that the actual numbers do not matter as much as their relative size.

22 In order to test the distribution of average scores for funds with different numbers of observations, each individual score must be weighted by the square root of the ratio of two expected variances. An example of the variances is the expected variance of funds with, for example, nine observations to the expected variance of funds with one more observation (ten).

Table 8: Example average utility scores over two periods

		Past Performance Ranking					
		1	2	3	4	5	
Subsequent	1	9.2	7.8	5.7	2.6	-2.1	
Performance	2	7.8	6.4	4.3	1.1	-3.6	
	3	5.7	4.3	2.2	-1.0	-5.7	
	4	2.6	1.1	-1.0	-4.1	-8.8	
	5	-2.1	-3.6	-5.7	-8.8	-13.5	
	c1=15, c2=-52, k=0.5 x=1.2 -- 4.4 (see Appendix 2)					Average	0.0

More than two years of observations were available for nearly all funds and so the average score could be calculated for each fund. However, as mentioned earlier, a simple average could not be used to compare funds with different numbers of observations. Although the average utility across all cells (as shown equal to 0.0) is, in theory, the same this cannot be said of the probability that any given fund will be close to it. The variance (volatility) of scores falls when averaged over a larger number years. So a fund which had an average score of 9.2 (the highest) over 18 years would be much more indicative of persistency in performance than a fund with the same average across only 2 years. As explained earlier we can set the expected variance of these results equal if we apply a weight (discount) to the scores of the fund for which there are only 2 observations[23]. Since any recorded average score might be the result of chance, a means of testing for persistency is to examine all of the scores. If the pattern displayed is as if the funds' relative performance were random then the test would be unable to demonstrate that there was persistency in performance. Sets of these results are reported below.

Normality tests are based on measures of the Skew (symmetry) and Kurtosis (for want of a better word, pointiness) of the distribution of average scores. These are illustrated in Charts 10 and 11. If a distribution is skewed (value of the statistic is

23 When dealing with quintiles the weight is approx. 0.86^t where t = max observations – observations for the fund concerned. In the example provided this is equal to $0.86^{16} = 0.0895$. This is applied to the difference between the expected value of 0.03 and the score (for example 9.238). This may be applied to the individual or the average score.

different from 0) the implication is that a significant number of funds perform either persistently well or persistently badly. This could indicate that a group of funds performed persistently poorly and that poor performers should therefore be avoided. Alternatively this might indicate that one can pick winning funds on the basis of past performance. The measure of Kurtosis centres on 3. If less than 3 the inference is that relative performance persists across all funds (under such circumstances past performance would be its most useful). If above 3 we should conclude that performance mean reverts. Scores would converge on the mean value more quickly than expected. One scenario which would return such a result is if funds which perform relatively well over a given period are more likely to perform badly over a subsequent period.

Chart 10

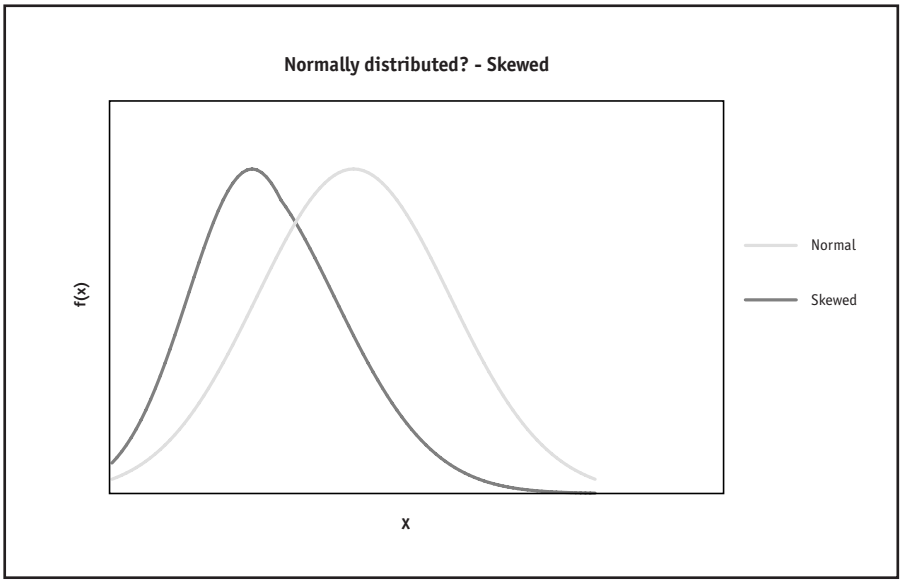
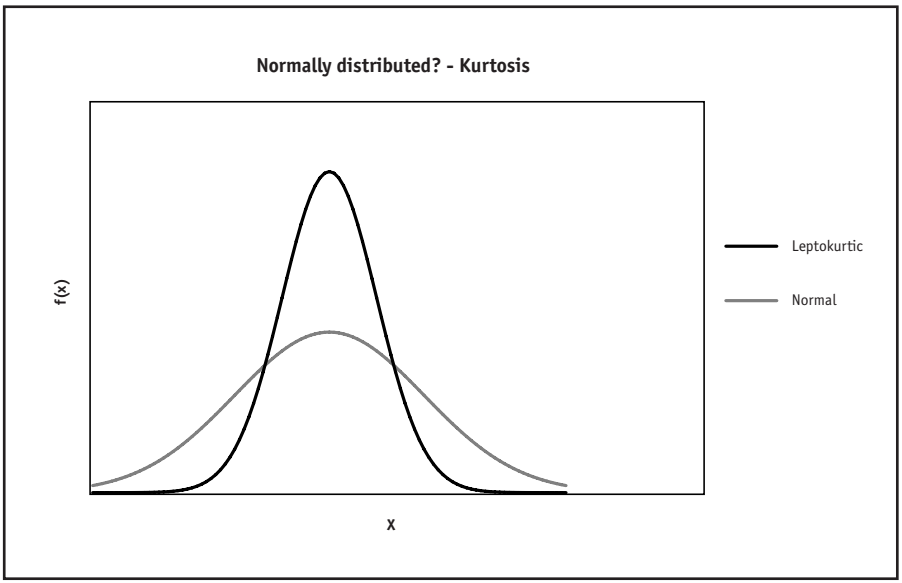


Chart 11



Results for the Period 1987 to 1998[24]

An earlier section examined the basic or raw data. This suggested both the possibility of reversals in the fortunes of some funds (perhaps due to 'style' effects) and a change in the pattern of performance after 1986/7. This coincided with a significant change in the market infrastructure, Big Bang. Since persistency in performance would be a result used to inform current decision making it seems more appropriate to examine the performance of funds after this structural change.

24 Data for 1997 are common to the assumed pre and post big bang periods. There was no clear indication when the effects of the structural change would have been felt. In practice the effects during this year appear to have been neutral given that the nature of the results did not change when this year was dropped.

Data from all funds in existence at some point in time over the period 1987 to 1998 and that had at least two utility scores were therefore examined. Several different sets of assumptions on utility and levels of performance were examined. The results for this period revealed a slightly platykurtic (flat) distribution, with only a few exceptions. However the results on Kurtosis were not statistically different from those of a normal distribution (which is mesokurtic). Further neither tests for Skew, nor the combined Jarque Bera test for normality, found that the distribution of scores was significantly different from what may have been found if the relationship between past and subsequent performance was random. A set of the results relating to this period may be found in the appendix.

The results confirmed that changing assumptions about the performance of funds and the shape of the utility function of the representative consumer affect the test for persistency although not the qualitative nature of the results. Most distributions were platykurtic and the effect of increasing the level of risk aversion (k) and / or the levels of performance (x) was to increase the size of the Jarque Bera test statistic. However for all 'reasonable'[25] values of k and x the test statistics were significantly below the critical value that would have suggested a relationship between the past and the future. A test for normality when risk aversion was removed also failed to find evidence of persistency in performance[26].

There is therefore a lack of evidence to suggest that, from 1987, there is any significant relationship between past and subsequent performance. Therefore for unit trusts in UK sectors, past performance provides no useful information in deciding which fund to invest in. This confirms the conclusions on persistency in investment performance which were drawn earlier, for the period after 'Big Bang'. The general pattern of performance does not suggest that the past provides an indication of future relative performance.

25 For example one could set k and x such that there is virtually no difference in the utility scores of Quintiles 1,2 and 3. Each set of assumptions were made to make sure that this was not the case.

26 The assumption of the same absolute difference between quintiles was retained. Consequently the levels of x would not affect the results on normality.

Results for the Period 1981 to 1998

Before the above results are taken as further evidence that the FSA should not publish figures on past performance in comparative tables the methodology needs to be further validated. A first stage in providing this validation was to run the test for persistency across all years for which data were available. For all sets of assumptions on utility and performance the hypothesis of no relationship between past and subsequent performance was strongly rejected. However, the distributions of scores were strongly leptokurtic (pointy). The conclusion is therefore not that there is meaningful persistence in the returns to managed funds but rather that fund performance converges on the average more quickly than would be expected by chance alone.

There are two explanations which might be put forward for the relatively rapid convergence in scores. The first is that the performance of some funds is more volatile such that there is a greater chance that they will switch between high and low quintile rankings. A preceding section found relatively little evidence for different volatility. However if this explanation is valid then the scores might converge to a common mean. We have modelled investors as disliking risk but the higher expected return to more volatile funds might compensate for this in the calculated scores. A flaw in this argument is that when the degree of risk aversion is increased the results should change. Specifically the scores of more volatile funds would decline relative to the mean with this revealed as a skewed distribution. Despite there being a large change in risk aversion in moving between some utility functions there is no significant measured change in the skewness of the distribution of scores.

A second explanation is that there are stylistic effects acting on performance which, although perhaps lasting for some time, suffer reversals. This corresponds to some of the more recent academic results reported in an earlier section of this paper and could have different implications for the measure of persistency. Some observed short term persistency in scores with reversals would not, necessarily, produce skew in the distribution but would result in the scores being more closely centred on the mean than would be expected by chance. Sets of results relating to the full sample period are reported in the appendix.

From the results the Jarque Bera test statistic is significantly larger than the critical value and we can be more than 99% sure that the distribution of scores did not occur by chance for all sets of choices of k and x (i.e. that past and future performance are related). As k (risk aversion parameter) and / or the level of the x values (performance

measure for the quintiles) increase the value of the Jarque Bera statistic declines. A critique of this might be that the weights given to observations on funds with only a few years observations biases the results. A means of examining such a claim is to consider funds that have more than a specified number of annual observations. All funds with fewer than six observations were therefore dropped and the analysis repeated. The results were, statistically, the same. The conclusion is therefore that the results for the full sample period are not the result of a small weight being given to funds with a few observations.

Results for the Period 1981 to 1987

The more recent investment performance of funds shows little evidence of persistency but this may not be the case for earlier periods. With a smaller number of annual observations and higher associated variance the chances of finding a relationship between past and subsequent performance are lower. However the results for 1981 to 1987 are striking for two reasons. The first is that whereas the results reported for the full sample found strong evidence of convergence on the mean return, in these early years the opposite is true. All of the combinations of utility functions and performance resulted in distributions that were platykurtic (flatter than normal). For a number of utility functions the distribution was also found to be positively skewed. This result indicates the existence of persistency in the returns to funds over the period. Further the positive skew in some of the distributions suggested that persistence was also predominant for well performing funds. This concurs with the graphical analysis presented earlier. Second is that for about half of all the utility functions the Jarque Bera statistic was statistically significant (at least at a 90% level of significance and frequently at a higher level of significance than this). Even where the results did not formally reject the proposition of no persistency the test statistics were marginal, rather than a great distance from the critical value. Sets of results are again provided in the appendix.

In contrast to the results for the full sample period an increase in k and or x resulted in a rise in the Jarque Bera statistic. The difference occurs because of a change in the underlying distribution from being leptokurtic to platykurtic. If the other elements of the utility function are held constant then as values of k and x increase the difference in the utility scores of high versus low ranked funds increases. Consequently a pattern of scores which is intrinsically flatter than normal (platykurtic) will test more strongly

if there is a bigger difference between funds with relatively high average scores and those with low scores. The results on increasing k and x are therefore as expected.

There is evidence that the period 1981 to 1987 was one in which information on the past investment performance of funds could be of use to investors. However in analysing the subsequent period the results do not suggest that past performance is of any use. The general pattern for recent periods is one of no discernible relationship between the past and the future. The results for the entire sample period suggest that there was an adjustment period around Big Bang. Reversals in relative performance may have been the result of more active trading strategies no longer meeting with success. The next paragraph examines turnover more closely. The evidence from this study (the graphs, transition matrices and the tests of the distributions) therefore strengthens findings from the literature that there is little to be gained in using data on past investment performance.

Turnover

The performance of UK unit trusts, as measured in the above analysis, is net of dealing costs (but not other charges such as the bid / offer spread or annual management fees). Conclusions drawn on performance therefore relate both to whether some fund managers have better skills and information and whether they use this advantage effectively, in not over-trading for example. With the exception of small cap stocks one might expect the costs of trading to be broadly similar, irrespective of the stock that is actually traded. If this is the case then the turnover of a fund (the amount that the fund voluntarily[27] bought and sold during the course of the year) will provide an appropriate proxy for the costs of trading.

Since this is both an interesting and important area of fund performance a study of the relationship between fund performance and turnover was conducted. The information on fund turnover is difficult to acquire and time consuming to process so that only

27 There is a debate as to what the most appropriate measure of fund turnover is. The measure that I have adopted is that of unforced sales occurring during the year concerned. This is calculated as total sales minus net outflows (should they occur). This assumes that the fund maintains a sufficiently large 'float' to cover shorter term calls on the fund. The size of fund reserves is, arguably, a choice made by the fund manager who has to balance the potential opportunity cost from not being exposed to equity against the possibility of having to (expensively) sell stock to meet a call for fund withdrawal. In practice the size of net outflows is small relative to the total sales of stock.

a relatively small study has been conducted to date. Data were collected from the returns made by around 100 unit trusts in the UK Growth and Income sector for the years 1997 and 1996. The performance of each of the funds for each year was ranked by quintile as before and the average (unweighed) turnover for each quintile was calculated. The results may be found in the table below.

Table 9: Average turnover

	Avg turnover 97	Avg turnover 96
Quintile 1	26%	41%
Quintile 2	38%	29%
Quintile 3	43%	20%
Quintile 4	46%	21%
Quintile 5	56%	34%

The turnover of poorly performing funds is always higher than the average, suggesting that some of these funds may systematically over-trade. However the evidence in the case of funds that are ranked highly is more mixed. In 1997 high performers had a significantly lower turnover than the average. This is in contrast to 1996 when the turnover of highly ranked funds was the highest found. The correlation of fund turnover between the years is high (0.65) and statistically different from zero, suggesting that some funds may adopt a more active strategy than others. The relationship is a little more confusing when one examines the correlation between fund performance and turnover on a fund by fund basis. For 1996 the correlation coefficient is positive (0.26) and statistically significant. However exactly the opposite result occurs for 1997 when the correlation coefficient is negative (-0.27) and statistically significant. These results may be driven by the turnover rates for funds in Quintile 1, as listed in the table above. Further work is warranted on fund turnover before any firm conclusions may be drawn.

Conclusion

This paper has reviewed a number of academic studies of persistency in the performance of managed funds in both the UK and the US. A closer examination of UK unit trusts has also been undertaken. This provided charts detailing the performance and volatility of funds and some rather more formal analyses. The transition matrices provided a simple statistical test, common in much of the literature. A further set of tests was also conducted using a novel methodological approach. The conclusions of each of these sections were broadly the same, although some conclusions additional to the findings of the literature may be drawn. A summary of the findings follows.

The literature on the performance of UK funds has failed to find evidence that information on past investment performance can be used to good effect by retail investors in choosing funds. The general pattern is one in which investment performance does not persist. Small groups of funds may show some repeat performance over a short period of time, particularly poorly performing funds. However the size of this effect and the fact that it is only very short lived means that there is no investment strategy for retail investors that could usefully be employed. The results from the US literature are similar. Recent debate on the most appropriate methodological approach has suggested that even the weak evidence of investment performance repeating may be exaggerated.

Although the conclusions from the literature seem compelling, further analysis was able to provide a slightly different perspective. The charts detailing the one year performance persistency of quintiles of funds suggested a significant change in the UK around the time of Big Bang. Recent experience appeared to show a complete absence of repeat performance, in contrast to the earlier part of the 1980s. Further confirmation again came from simple statistical tests of transition matrices. There is no strong evidence from these results that performance has repeated in more recent times.

The development of a novel methodological approach allowed for a more consistent examination of performance over the longer term. This also provides an opportunity to examine further the apparent end in the relationship between past and subsequent investment performance. The results concurred with the earlier analyses in finding that there was no persistency in the performance of managed funds after 1987. There was evidence of repeat performance before this point but it would be misleading to suggest that retail investors could use this finding in the present day.

The weight of evidence is that information on past performance cannot be exploited usefully by retail investors.

Appendix 1

Charts showing investment performance may present a misleading picture if funds which perform relatively well are merely rewarding investors for the higher inherent risk that they are adopting. A simple means of examining risk is to calculate the variance of a fund's return over each year[28]. The average variance within each quintile was calculated (quintiles defined by the performance of funds such that the top performing 20% of funds are in Qn1 etc.). Chart A1 plots the average volatility. Most obvious is the spike in volatility in 1987. This coincides both with the switch to electronic trading and also a stock market crash. Also evident is that differences in volatility appear to cease after this 1987 anomaly. Reference to Chart A3, where the re-scaling allows for a closer examination of more normal years, shows that there is no strong pattern to volatility. Examining the volatility of funds in the year after they are ranked in a particular quintile does not show a strong pattern either. Chart A3 suggests that funds which performed well (badly) do not subsequently alter their investment strategy to consolidate (gamble) in the following year for example. A conclusion from this evidence is that there is no substantial difference in the within year volatility of funds according to their relative performance in that year.

28 This approach does not follow the method of most market models but provides a useful first step.

Chart A1

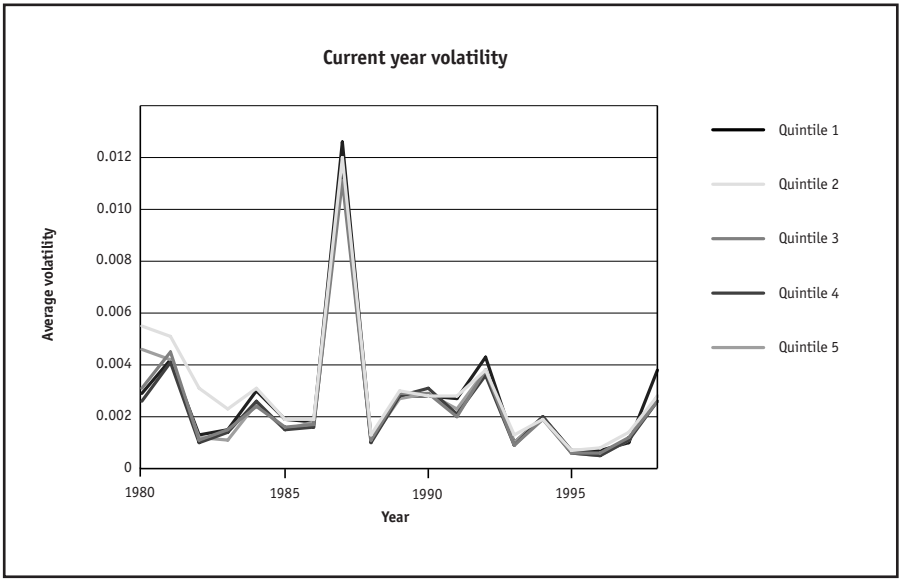


Chart A2

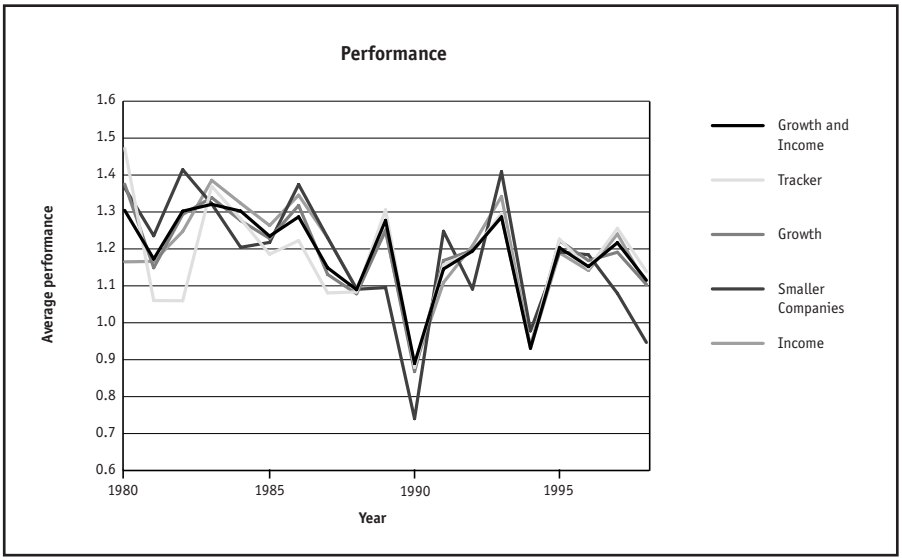


Chart A3

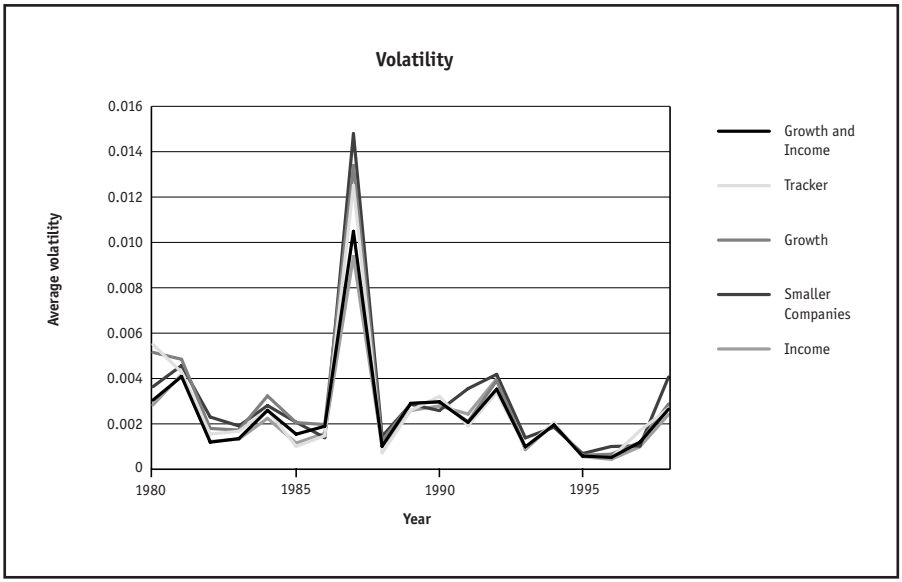


Chart A4

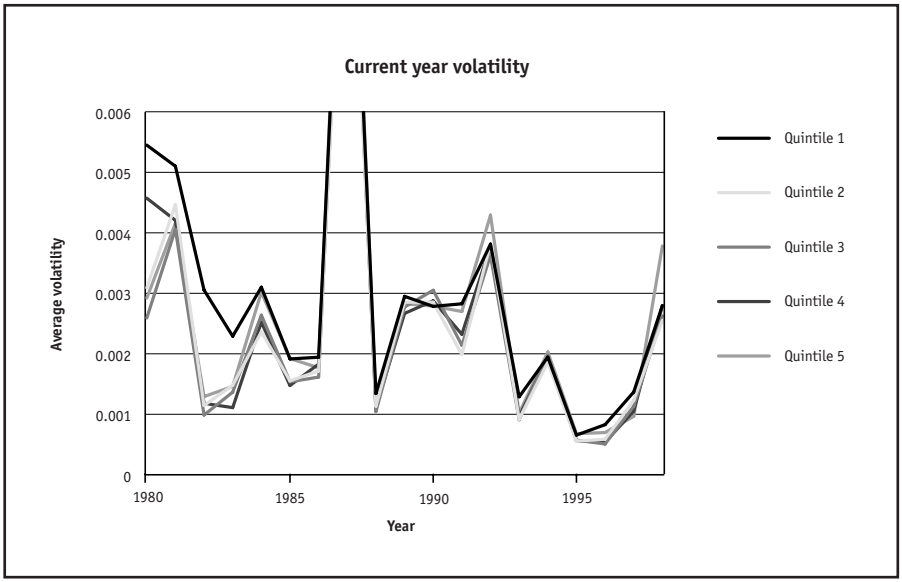


Chart A5

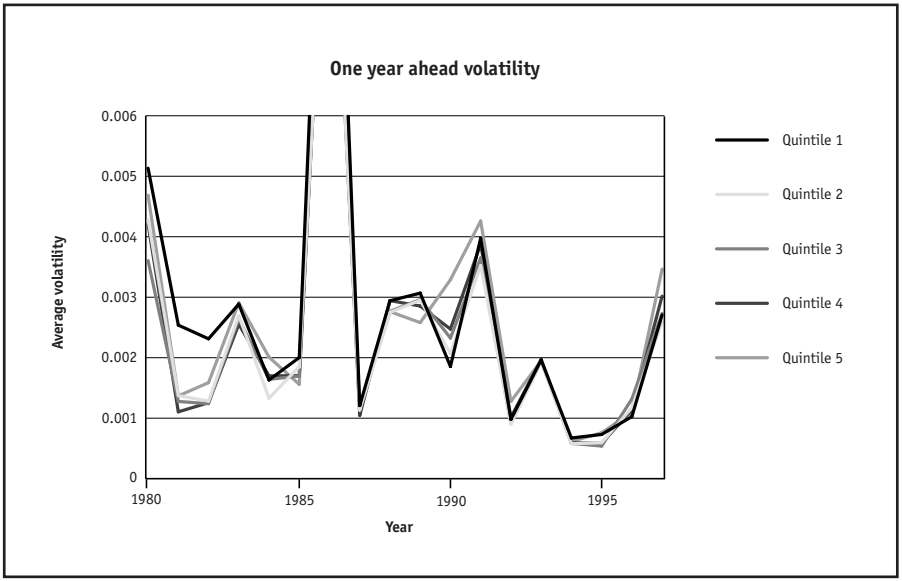


Chart A6

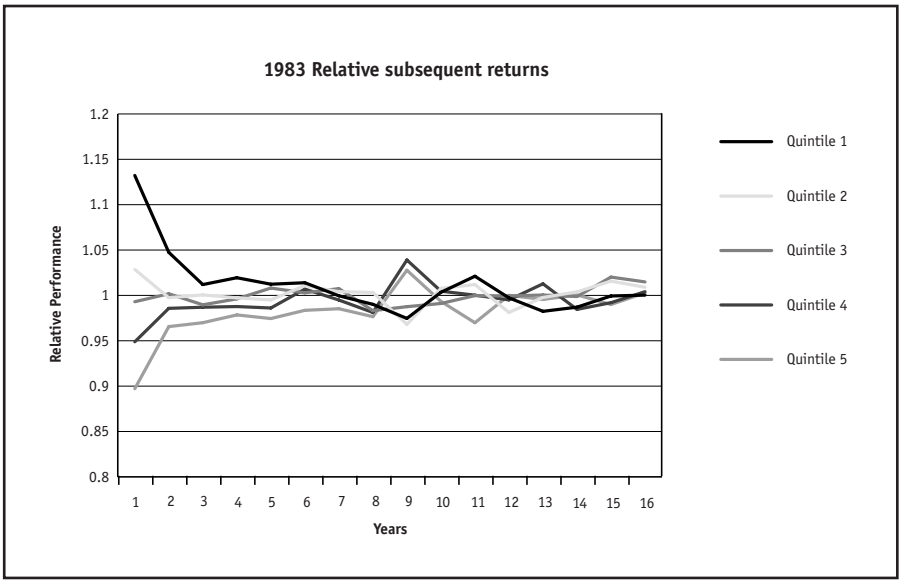
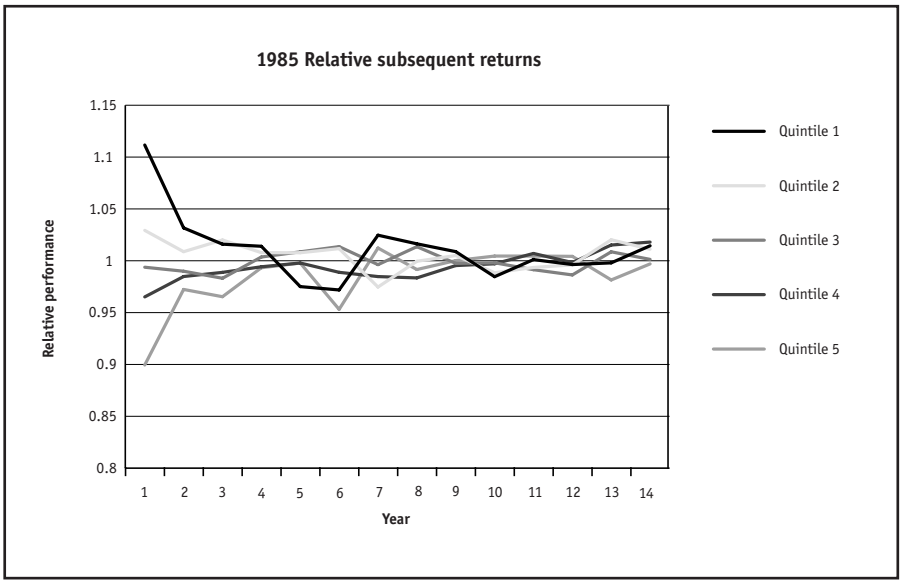


Chart A7



Appendix 2

A risk averse utility function of the form U=U(x) may be defined as having the following properties. From Arrow Pratt, absolute risk aversion is defined as

$$R_A = -\frac{U''(x)}{U'(x)}$$

Where x is the level of fund growth. From assumption 3) in the paper (aversion to risk does not change with the performance of the fund) we set R_A to be constant in x, i.e.

$$R_A = -\frac{U''(x)}{U'(x)} = k$$

Such that re-arranging we get

$$U''(x) + kU'(x) = 0$$

This is a second order differential equation from which we may derive the characteristic polynomial

$$r^2 + kr = 0$$

This has roots r and −k from which we can find the general solution

$$U(x) = C_1 e^{0x} + C_2 e^{-kx}$$

Or, more simply

$$U(x) = C_1 + C_2 e^{-kx}$$

This gives us a utility function with constant absolute risk aversion (k positive), however to satisfy assumption 2) in the paper (higher performance is preferred but at a diminishing rate i.e. that U'>0) we must set $C_2<0$. No further constraints are applied by the assumptions, leaving a relatively free choice of C_2 C_1 x and k. These four parameters determine the shape of the utility function. The level of performance (x) and the difference in performance between quintiles interacts with the level of risk aversion (k) in helping to define the shape of the utility function. As noted above, the assumption of a positive but declining preference for better returns is provided by the assumption that $C_2<0$.

Assuming risk aversion implicitly corrects the average scores of funds that have a higher nominal return but which are more volatile. The extent of this correction naturally depends on the degree of risk aversion that is assumed. If the payoffs are inverted for example[29] the results for normality change. There is then much stronger evidence of persistency in the earlier years of the sample for example, although this weakens following 1987.

Appendix 3

Testing

The base case against which the actual distributions of average funds scores will be measured is random relative performance. If we take the performance in two separate years the probability of being in each cell in the table of past and subsequent performance is the same (1/25). This therefore means that the discrete joint probability distribution function (pdf) can be expressed as the product of the marginal pdfs – i.e. the two variables are statistically independent. This greatly assists in determining the score distribution since, given statistical independence, the asymptotic distribution for the random relative performance case is normal. Further we know the expected score (the average across all scores in the table, since the probability of any single outcome being observed is the same). Appealing to the Central Limit Theorem (that irrespective of the underlying distribution of independent random variables the average must tend to the normal) we can run a standard Chi square test of the calculated scores (relies on the 3rd and 4th moments) to examine the normality of the sample. In essence this tests whether the Joint pdf is truly random.

This is a Jarque Bera test and takes the form;

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]$$

29 Higher utilities correspond to funds that perform consistently badly but that this utility increases at a diminishing rate.

Where Skew (S) is given by

$$S = \frac{\sum(x - \bar{x})^3/n}{s^3}$$

and Kurtosis (K) is given by

$$K = \frac{\sum(x - \bar{x})^4/n}{s^4}$$

With $\bar{x}$ the mean and s the standard deviation.

The JB test statistic approaches the chi squared distribution only slowly, however there are a large number of funds examined within the sample[30] and so the chi squared was appropriate.

As a means of testing this approach I constructed a Monte Carlo simulation with some built in short run persistency. Performance figures were artificially set as random and then some fraction of the preceding period's performance added[31]. Funds were then ranked by quintile as before. The methodology proved not to be as sensitive to these short run effects as a more standard approach (as expected and desired given that the objective of using the methodology is to examine longer term effects). However the test showed that the distribution of scores was platykurtic (flatter than normal) under such circumstances and rejected normality down to relatively small levels of persistency.

30 For ease of exposition the standard chi^2 is listed as the test statistic, critical points of the Bowman Shelton statistic were used where the result appeared marginal.

31 A first order auto-regressive process with the parameter on the lagged value set at less than one. Sensitivity of the method with different underlying utility functions was judged by examining when the Jarque Bera test failed to reject the hypothesis that performance is random. As a benchmark a standard matrix of initial and subsequent performance rankings (by quintiles) was constructed.

Appendix 4

The variance of a discrete distribution, of the form used in this paper is;

$$\sum(x_i - \mu)^2 P_x(x_i) = \varnothing$$

Where x_i is the observed utility score for a given fund in time t and μ is the expected (theoretical) mean.

In order to obtain a common measure or score to evaluate persistency across funds for which we have different numbers of observations we need to examine the variance. If the distribution for funds that have a number of observations (say 9) can be adjusted such that the variance is the same as for funds that have ten annual observations on their score then the average scores of both funds will tend to the same normal distribution. A test for normality on these will then allow an evaluation of persistency. Let x_i be an observation on a fund for which there are 9 observations. The theoretical variance for this distribution is then;

$$\sum(x_i - \mu)^2 P_x(x_i) = \varnothing$$

For funds (j) for which there are ten observations the variance is lower by a factor m

$$\sum(x_j - \mu)^2 P_x(x_j) = m\varnothing$$

To achieve the same variance we multiply both sides of the first equation by m;

$$\sum m(x_i - \mu)^2 P_x(x_i) = m\varnothing$$

And re-arranging get

$$\sum(\sqrt{m}x_i - \sqrt{m}\mu)^2 P_x(x_i) = m\varnothing$$

To give

$$\sum([\sqrt{m}x_i + \mu - \sqrt{m}\mu] - \mu)^2 P_x(x_i) = m\varnothing$$

Or

$$\sum((\sqrt{m}[x_i - \mu] + \mu) - \mu)^2 P_x(x_i) = m\emptyset$$

So the difference between each observed score and the mean μ for a fund with nine observations is multiplied by the square root of m. m is ratio of the theoretical variance for funds with 10 observations to that for funds with 9 observations. The weight is therefore the first term in the above equation;

$$(\sqrt{m}[x_i - \mu] + \mu)$$

The expected variance for any given payoff matrix and number of time series observations per hypothetical fund can be calculated directly (one constructs all possible 5n payoffs for any scenario in which there are n observations on funds and assigns the probability $1/5^n$ for each payoff (the value of μ does not change). This is cumbersome to say the least if n becomes large but a Monte Carlo simulation revealed that the value of m between any distributions for funds with n and n-1 observations is the same. The value of m only changes in relation to the number of ranks (i.e. quartiles or quintiles) and is invariant with respect to the scores or the value of n. The weights are therefore presented as a geometric progression from 0.86^0 where the funds concerned have 18 observations through 0.86^{16} where there are only two observations for the funds.

The shape of the utility function and therefore the relative size of scores is driven by the parameters k and x. Conceivably therefore the possibility of finding (non) normality might be influenced by the choice over these parameters. The expectation is that rejection of the null of normality will not be affected by the choice of k and x except in the extreme. The analysis sets k and x at a number of different levels to see if there is a difference in the results.

Appendix 5

Results for the period 1987 through 1998.

C1	C2	k	x range	Skew	Kurtosis	Jarque Bera	Critical Value 90%	Normal ?
107	-107	1	0.9 - 3.3	0.00494	2.76	1.27	4.61	✔
107	-107	0.3	0.9 - 3.3	0.0491	3.06	0.311	4.61	✔
15	-52	0.5	1.2- 4.4	0.0305	2.86	0.413	4.61	✔
20	-35	0.9	1.35 - 4.95	0.0293	2.71	2.00	4.61	✔

Results for the period 1987 through 1998.

C1	C2	k	x range	Skew	Kurtosis	Jarque Bera	Critical Value 90%	Normal ?
107	-107	1	0.9 – 3.3	0.146	4.52	59.3	9.21	✘
15	-52	0.5	1.2 – 4.4	0.180	4.66	71.7	9.21	✘
15	-52	0.5	2.4 – 8.8	0.112	4.40	49.7	9.21	✘
30	-130	0.9	1.65 – 6.05	0.0837	4.31	43.1	9.21	✘

Results for the period 1987 through 1998.

C1	C2	k	x range	Skew	Kurtosis	Jarque Bera	Critical Value 90%	Normal? ?
107	-107	1	0.9 – 3.3	1.26	2.46	4.00	4.61	✔
15	-52	0.5	2.4 – 8.8	0.220	2.55	4.52	4.61	✔
20	-35	0.9	1.35 – 4.95	0.352	2.69	6.78	4.61	✘
30	-130	0.9	1.65 – 6.05	0.311	2.64	5.87	4.61	✘

References

Bernstein, P.L. Against the Gods: *The remarkable story of risk*. John Wiley & Sons Inc 1998. 2nd Edition.

Blake, D. Lehman,B,N. & Timmerman, A. *Asset Allocation Dynamics and Pension Fund Performance*. Working paper Birkbeck College University of London.

Bogle, J,C. *The Implications of Style Analysis for Mutual Fund Performance Evaluation*. Journal of Portfolio Management. Summer 1998 pp34-42.

Brown, G. Draper, P & Mckenzie, E. *Consistency of UK Pension Fund Investment Performance*. Journal of Business Finance and AccountingVol 24 No 2 March 1997 pp 155- 178.

Brown, S.J. Goetzman, W.N. *Performance Persistence*. Journal of Finance Vol L, No 2, June 1995 pp679-698

Carhart, M.M. *On Persistence in Mutual Fund Performance*. Journal of Finance, Vol LII, No 1, March 1997. Pp57-82.

Elton, E,J. Grunber,M,J & Blake, C,R. The Persistence of Risk Adjusted Mutual Fund Performance. Journal of Business vol 69 no 2 1996, pp133-157.

Fama, E. Efficient Capital Markets: *A review of theory and empirical work*. Journal of Finance Vol. 25, 1970, pp383-416.

Fama, E. & McBeth, J. *Risk Return and Equilibrium: Empirical Test*. Journal of Political Economy Vol. 81 1973 pp 607-636.

Fersun, W, E. & Schadt, R, W. *Measuring Fund Strategy and Performance in Changing Economic Conditions*. Journal of Finance, Vol. LI, No 2June 1996 pp425-456.

Fletcher, J. *An Examination of UK Unit Trust Performance within the Arbitrage Pricing Framework*. Review of Quantitative Finance and Accounting Vol. 8 No 2 March 1992 pp 91-107.

Goetzman, W,N. & Ibbotson, R,G. *Do Winners Repeat? Patterns in Mutual Fund Performance*. Journal of Portfolio Management Vol. 20, Winter 1994, pp9-17.

Grinblatt, M. & Titman, S. *The Persistence of Mutual Fund Performance*. Journal of Finance, Vol. XLVII, No 5, December 1992, pp1977-1984.

Hendricks, D. Patel, J. & Zeckhauser, R. *Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974-1988*. Journal of Finance, Vol. XLVIII, No 1, March 1993, pp93-130.

Lunde, A. Timmerman, A. Blake, D. *The hazards of mutual fund underperformance: A Cox regression analysis*. Working paper, Department of Economics, UC San Diego (1999)

Phelps, S & Detzel, L. *The Nonpersistence of Mutual Fund Performance*. Quarterly Journal of Business and Economics. Vol36, No2 Spring 1997, pp55-69.

A report by Price Waterhouse Coopers for Barclays Global Investors. *25 years of indexing: an analysis of the costs and benefits*.(1998)

Quigley, G. Sinquefield, R.A. *The Performance of UK Equity Unit Trusts*. Report by Dimensional Fund Advisors (1998)

Sauer, D.A. *Information Content of Prior Period Mutual Fund Performance Rankings*. Journal of Economics and Business 1997 pp 549-567.

A report by The WM Company for Virgin Direct Financial Services. *Comparison of Active and Passive Management of Unit Trusts* (1999).

FSA Occasional Papers in Financial Regulation

1	The Economic Rationale for Financial Regulation *David Llewellyn*	April 1999
2	The Rationale for a Single National Financial Services Regulator *Clive Briault*	May 1999
3	Cost-Benefit Analysis in Financial Regulation *Isaac Alfon and Peter Andrews*	September 1999
4	Plumbers and Architects: a supervisory perspective on international financial architecture *Huw Evans*	January 2000
5	Household Sector Saving and Wealth Accumulation: Evidence from balance sheet and flow of funds data *Iftikhar Hussain*	February 2000
6	The Price of Retail Investing in the UK *Kevin R James*	February 2000
7	Some Aspects of Regulatory Capital *Jeremy Richardson and Michael Stephenson*	March 2000
8	Saving for Retirement: How taxes and charges affect choice *Malcolm Cook and Paul Johnson*	May 2000

Available from

FSA Publications
Financial Services Authority
25 The North Colonnade
Canary Wharf
London E14 5HS

and on the FSA website www.fsa.gov.uk