CMPO Working Paper Series No. 00/26

Equity Performance of Segregated Pension Funds in the UK

Alison Thomas
and
Ian Tonks

University of Bristol and CMPO

August 2000

Abstract
We investigate the performance of the UK equity portfolios of 2,175 segregated UK pension funds over the period 1983-97. We find that there is similar pattern in the returns on most of the pension funds and the FT-All Share index, leading us to conclude that most funds in the sample are "closet-trackers". Any measures of outperformance were therefore bound to be small. Over the whole period and across all funds average outperformance was insignificantly different from zero. We investigated the sensitivity of the fund returns to the addition of a size premium, which we found to be significant, and important for the smaller funds in our sample. During three sub-periods we found that there was significant average underperformance during the strong bull market of the mid-eighties, but significant outperformance since 1987. In particular in the period 1987-92 the average outperformance across pension funds was one half of a percentage point per year. Decomposing this abnormal performance we found that most of it could be explained by the ability of both large and small funds to time the size premium. On the whole there were negative returns to both selectivity and to market timing. There was little evidence of any differences in the performance between mature and immature funds.

JEL Classification: G0, G2, N2
Keywords: performance measurement ; pension funds

Acknowledgements
The data used in this study was provided by CAPS Ltd, and we are grateful for the assistance of Alan Wilcock and Ian Ibbotson. This paper has benefited from comments made at seminars at University of Bristol, Lancaster University, London School of Economics and Said Business School, Oxford. We are particularly grateful to comments made by David Ashton, Tim Jenkinson and John O'Hanlon.

Address for Correspondence
CMPO, Department of Economics
University of Bristol
8 Woodland Road
Bristol
BS8 1TN
i.tonks@bristol.ac.uk

CMPO is funded by the Leverhulme Trust.

Non-Technical Summary

This paper investigates the performance of a large sample of UK equity portfolios of segregated UK pension funds over the period 1983-97. In summary we find little cross-sectional variation in the returns on these portfolios leading us to conclude that most funds in the sample are "closet-trackers".

Over the whole period and across all funds average outperformance was insignificantly different from zero. Though during three sub-periods we found that there was significant average underperformance during the strong bull market of the mid-eighties, but significant outperformance since 1987. In particular in the period 1987-92 the average outperformance across pension funds was one half of a percentage point per year.

We investigated the sensitivity of the fund returns to the addition of a size premium, which we found to be significant, and important for the smaller funds in our sample. Decomposing this abnormal performance we found that most of it could be explained by the ability of both large and small funds to time the size premium. On the whole there were negative returns to both selectivity and to market timing.

I Introduction

This paper examines the performance of a sample of UK pension funds' equity investments over the period 1983-97. Trustees of pension funds who are charged with placing their scheme's assets with an external investment manager are faced with two critical decisions. Should they invest their fund's assets in a passive vehicle – a fund that aims to mirror a pre-determined benchmark such as the FT All Share? Or should they seek active management of their assets in the expectation that the additional cost of so doing will be offset by superior returns?

A number of recent UK policy documents have argued that pension contributions in particular should be investing in tracker funds, on the basis that "there is little evidence that active fund management can deliver superior investment returns for the consumer"[1]. The objectives of this project are thus twofold: Firstly, we intend to examine the performance of UK pension funds over the long-term, and to analyse the shift in the distribution of returns relative to an external benchmark as market conditions fluctuate. Thus, we will ask whether fund performance is, on average, better in bull markets or bear markets? Do they add more value in markets that are characterised by a broad spread of activity rather than a narrowly focused market where a handful of major companies dominate benchmark returns? Do they perform better when small or mid-capitalisation stocks lead the majors?

Second we will examine whether the characteristics of the pension fund affect its performance. The two characteristics we focus on are fund size and fund maturity. The pension funds in our sample are funded occupational pension schemes. Occupational pension schemes are usually funded and require contributions throughout the employees working life. In a funded scheme an employee pays into a fund which accumulates over time, and then is allowed to draw on this fund in retirement. These schemes are provided by an employer and may pay on a defined benefit or a defined contribution basis. The fund is administered by trustees, usually nominated by the employer. Defined benefit (or final salary) schemes offer a pension, guaranteed by the employer, usually defined in terms of some proportion of final year earnings, and are related to the number of years of

[1] para. 420, p. 71 Office of Fair Trading (1997). See also Consumers' Association (1997); Department of Social Security (1998); Financial Services Agency (1999)

employment. Defined contribution (or money purchase) schemes are always funded and convert the value of the pension fund at retirement into an annuity.

The trustees of the fund must decide how the funds are managed. There are three methods of managing funded occupational pension schemes. First, under an insured scheme an employer contributes premiums into a scheme, which guarantees to pay a pre-defined benefit at a pre-defined time. The risk of a funding shortfall (ignoring default risk) is borne by the fund manager (typically an insurance company), and not the individual or the corporation. Second the fund management may be outsourced: here the scheme's trustees find one or more external managers who are given a mandate to manage assets against a pre-determined benchmark. If the performance objectives are met, this pension scheme should meet all actuarially defined future liabilities. The risks of a contribution shortfall are thus borne by both the sponsoring company and by the external fund manager (to the extent that a failure to meet the benchmark will result in a loss of assets under management). In this second type of scheme the trustees may opt to join a pooled investment fund, which typically offers a lower fee structure though no mandate flexibility. Alternatively the trustees may request that the fund be managed on a segregated basis, offering greater mandate flexibility, though typically at a higher price.

The data in the current study relates to segregated schemes. The third method is in-house management: a number of large pension schemes are managed by a team of in-house professional. This allows trustees complete flexibility in terms of asset/liability matching. However the risk of a contribution shortfall lies solely with the scheme sponsor

In part the size of the pension fund will depend on the size of the employer, but some large employers may have a number of separate schemes operating for sub-groups of employers. We will examine whether there is an optimal fund size in terms of performance. For market liquidity reasons large funds may be constrained in the portfolio of assets in which they invest, whereas smaller funds may be able to take advantage of investing in a wider range of securities.

In theory the trustees of a pension plan should allocate the scheme's funds into asset classes according to the timing of the future liabilities of the plan. Thus, if the company has a young workforce, a priori, one would expect its pension fund to invest in more risky assets, in expectation that the greater short-term volatility will be rewarded in the long run through superior performance. Similarly, in schemes where there is a high proportion of retirees or where the workforce is old, assets should be skewed towards instruments that have lower risk and more clearly defined cash flows.

Given the wide variation in demographic profile of the members that is likely to exist in a broad sample of pension funds, the empirical investigator would thus expect to find a similarly disparate distribution in the returns earned by the various funds over time. In this paper we investigate whether there is any evidence to suggest that the trustee does indeed actively manage the portfolio with which he is entrusted. We also examine whether the distribution of fund returns suggest that trustees actively allocate assets according to the liability profile of the fund.

The significance of this work for trustees and plan advisors is compelling. At the most fundamental asset allocation level, the conclusions of the analysis of the distribution of returns will aid trustees in their decision as to whether to invest their pension fund monies in an active or in a passive vehicle.

II Previous Evidence on Performance of Managed Funds

Empirical evidence suggests that the performance of the average portfolio manager relative to external benchmarks has been disappointing. The early literature of the performance of mutual funds in the US [Jensen (1968), Crenshaw (1976), Friend, Blume & Crockett (1970), McDonald (1976), Williamson (1972)] found that simple tests of abnormal performance did not yield significant returns.[2] Although on average fund managers do not outperform, in any sample there is a distribution to the performance, and more recently research has investigated whether the outperformers in the sample continue to outperform in the future. Grinblatt and Titman (1992) find that differences in mutual fund performance between funds persist over 5-year time horizons and this persistence is consistent with the ability of fund managers to earn abnormal returns. Hendricks, Patel and Zeckhauser (1993) analysed the short-term relative performance of no-load, growth orientated mutual funds,

[2] The early work of Jensen (1968), and others all established that during bull markets fund managers cannot outperform a market index. However in bear markets, active managers are more likely to outshine passive alternatives

and found the strongest evidence for persistence in a one year evaluation horizon. Malkiel (1995) however argues that survivorship bias is more critical than previous studies have suggested.[3] When an allowance is made for survivorship bias in aggregate, funds have underperformed benchmark portfolios both after management expenses and even gross of expenses. Further he finds that whilst considerable performance persistence existed in the 1970s, there was no consistency in fund returns in the 1980s. Brown and Goetzmann (1995) examine the performance persistence of US mutual funds and claim that the persistence is mostly due to funds that lag the S&P. They demonstrates that relative performance pattern depends on period observed and is correlated across managers, suggesting that that persistence is probably not due to individual managers – it is a group phenomenon, due to a common strategy that is not captured by standard stylistic categories or risk adjustment procedures. This is consistent with herding theories of behaviour (Grinblatt, Titman and Wermers, 1994). They suggest that the market fails to discipline underperformers, and their presence in the sample contributes to the documented persistence. Carhart (1997) demonstrates that common factors in stock returns and investment expenses explain persistence in equity mutual funds' mean and risk-adjusted returns. Only significant persistence not explained is concentrated in strong underperformance by the worst return mutual funds. His results do not support the existence of skilled or informed mutual fund portfolio managers. Daniel, Grinblatt, Titman and Wermers (1997) using normal portfolio analysis shows that mutual fund managers – in particular aggressive-growth funds, exhibit some selectivity ability but that funds exhibit no timing ability. They introduce measure that identifies if a manager can time the market, size, book to market, or momentum strategies. Gruber, (1996) poses the question: why do people buy mutual funds when their performance is so poor? He postulated that it might be because unitised products are bought and sold at NAV so management ability is not priced into the product. If management ability exists then performance should be predictable. Some investors will be aware of this and will invest accordingly. In the UK Blake and Timmerman (1997) examine 2300 UK open ended mutuals over 23 year period (1972-1995), using bid prices and net income- so gross of fees. Over the period the data includes 973 dead and 1402 surviving funds, and by studying the termination of funds, they are able

[3] Malkiel ponts out that only the more successful mutual funds survive. Higher risk funds that fail tend to be merged into other products to hide their poor performance. Also bias from tendency to run incubator funds – run ten different products – see which are best and market those, ignoring the poor record of the rest

to shed light on the extent of survivorship bias. They find economically and statistically very significant underperformance that intensifies as the termination date approaches, and they conclude that survivorship does not alter the results significantly.

Turning to pension funds specifically Ippolito and Turner (1987) examines returns on 1,526 US pension funds and find underperformance relative to the S&P500 Index. Lakonishok, Shleifer and Vishney (1992) provide evidence on the structure and performance of the Money Management Industry in the US in general but focus on the role of pension funds, examining 769 pension funds, with total assets of $129 billion at the end of 1989. They find the equity performance of funds under-performed the S&P 500 by 1.3% per year throughout the eighties. Lakonishok, Shleifer and Vishney emphasise that although there is a long literature on the under-performance of mutual funds, pension funds also under-perform relative to mutual funds on average.

Coggin, Fabozzi and Rahman (1993) investigate the investment performance of a random sample of 71 US equity pension fund managers for the period January 1983 through December 1990, and finds average selectivity measure is positive and average timing ability is negative. Though both selectivity and timing are sensitive to the choice of benchmark when management style is taken into consideration. For example they find that funds that target value strategies yielded outperformance of 2.1 per cent per annum, but funds that adopted growth strategies underperformed by -0.96 per cent.

In the UK Brown, Draper and McKenzie examine the consistency of UK Pension Fund Performance, and finds limited evidence of persistency of performance for a small number of fund managers. Their sample consists of 232 funds 1981-90 and 409 funds 1986-92; all funds retained a single fund manager. Consistency holds over different time horizons, samples and classification schemes. Blake, Lehmann, & Timmermann, (1996) examine a sample of 364 UK pension funds who retained the same fund manager over the period 1986- 1994. They find that the total return is dominated by asset allocation. Average return from stock selection is negative, and average return to market timing very negative. Although UK equity managers comparatively good at selecting equities – although only 16% of sample beat peer group average.

Within market cycles, pension fund trustees typically aim to find portfolio management firms that can provide a consistency of performance regardless of market or economic conditions.

III Measuring Fund Performance

Jensen's technique is to regress the excess returns on the individual fund above the risk free rate $R_{pt} - R_{ft}$ against the excess return on the market $R_{mt} - R_{ft}$, plus any additional factors F_t that *a priori* are expected to determine returns

$$R_{pt} - r_{ft} = a_p + b_p (R_{mt} - r_{ft}) + gF_t + e_{pt} \qquad (1)$$

for each fund p over the t data periods, and save the coefficients a_p , b_p and g. The factors in F_t may include a size premium, book-to-market, and momentum [Carhart (1998)], though in the empirical results reported below we only allow for a size factor.

Under the null hypothesis of no-abnormal performance the a_p coefficient should be equal to zero. For each fund we may test the significance of α_p as a measure of that funds abnormal performance. We may test for overall fund performance, by testing the significance of the mean a when there are N funds in the sample

$$\bar{a} = \frac{1}{N}\sum_{p=1}^{N} a_p \qquad (2)$$

Assuming that the performance of each fund is independent [Cov(ε_p, ε_q) = 0][4], the appropriate t-statistic is

$$t = \frac{1}{\sqrt{N}}\sum_{p=1}^{N}\frac{a_p}{SE(a_p)} \qquad (3)$$

The original Jensen technique made no allowance for market timing abilities of fund managers when fund managers take an aggressive position in a bull market, but a defensive position in a bear market. When portfolio managers expect the market portfolio to rise in value, they may switch from bonds into equities and/or they may invest in more high beta stocks. When they expect the market to fall they will undertake the reverse strategy: sell high beta stocks and move into "defensive" stocks.

If managers successfully engage in market timing then, returns to the fund will be high when the market is high, and also relatively high when the market is low. More generally fund managers may time with respect to any factor. If managers successfully market time, then a quadratic plot will produce better fit (Treynor-Mazuy test).

$$R_{pt} - r_f = a_p + b_p(R_{mt} - r_f) + d_p(R_{mt} - r_f)^2 + e_{pt} \qquad (4)$$

Significance of market timing is measured by d_p. An alternative test of market timing suggested by Merton-Henriksson is

$$R_{pt} - r_f = a_p + b_p(R_{mt} - r_f) + d_p(R_{mt} - r_f)^+ + h_{pt} \qquad (5)$$

where $(R_{mt} - r_f)^+ = Max(0, R_{mt} - r_f)$

Recently Ferson and Schadt (1996) advocate allowing for the benchmark parameters to be conditioned on economic conditions: called conditional performance evaluation, on the basis that some market timing skills may be incorrectly credited to fund managers, when in fact they are using publicly available information to determine future market movements. In which case Ferson and Schadt argue that the predictable component of market movements should be removed in order to assess fund managers private market timing skills. Under a conditional version of the CAPM, the Jensen regression becomes

$$R_{it} - r_{ft} = a_i + b_i(Z_{t-1})(R_{mt} - r_{ft}) + e_{it} \qquad (6)$$

where Z_{t-1} is a vector of instruments for the information available at time t (and is therefore specified as t-1) and $\beta_i(Z_t)$ are time conditional betas, and their functional form is specified as linear

[4] This is a debatable assumption, since separate funds may be managed by the same fund manager. On the other hand one of the characteristics of segregated fund management, is the fund managers design bespoke portfolios for the individual fund

$$b_i(Z_t) = b_0 + B'z_{t-1} \qquad (7)$$

where $z_{t-1} = Z_{t-1} - E(Z)$ is a vector of deviations of the Zs from their unconditional means. Implementing this approach involves creating interaction terms between the market returns and the instruments. Instruments used are: lagged treasury bill rate, dividend yield, default premium (difference between low and high quality corporate bonds), and the slope of the term structure (difference between long and short run government bond yields)

The test for market timing now isolates the effect of public information. The amended Treynor-Mazuy test is

$$R_{pt} - r_f = a_p + b_p(R_{mt} - r_f) + B'z_{t-1}(R_{mt} - r_f) + d_p(R_{mt} - r_f)^2 + e_{pt} \qquad (8)$$

where the sensitivity of the managers beta to the private market timing signal is measured by d_p. The amended Merton-Henriksson test is

$$R_{pt} - r_f = a_p + b_d(R_{mt} - r_f) + B_d' z_{t-1}(R_{mt} - r_f) + d_c(R_{mt} - r_f)^+ + D' z_{t-1}(R_{mt} - r_f)^+ + h_{pt} \qquad (9)$$

where $(R_{mt} - r_f)^+ = (R_{mt} - r_f) * Max[0, R_{mt} - r_f - E(R_{mt} - r_f | Z_{t-1})]$

and $d_c = b_{up} - b_d \qquad D = B_{up} - B_d$

The significance of market timing is represented by the significance of δ_c.

IV Data

The data used in this study was provided by the Combined Actuarial Performance Services Ltd (CAPS). It consists of quarterly returns on UK equity portfolios of 2,175 UK pension funds from March 1983 to December 1997. In addition for each fund-quarter the manager of the fund and the size of the fund is provided. CAPS provide a performance measurement service for about half of all segregated pension fund schemes in the UK. There is one other major provider of pension fund performance: WM Ltd. Chart 1 shows the distribution of pension fund assets across asset categories in the general CAPS database. Typically a UK pension fund invests about 57% of assets in UK equities, and it is the returns on UK equity portfolios which is examined in this study. Our dataset consists of a total of 59,509 observations on quarterly returns and fund size, and the maximum number of Quarters is 56. Table 1, Panel A illustrates the Distribution of fund quarters over the dataset, and shows that 50 per cent funds have 24 or less observations, and the average life of a fund in the data is just less than seven years. This high attrition rate is partly explained by the closure of funds due to the sponsoring companies merging, or becoming insolvent, and also due to the fund switching to alternative performance measurement services.

Table 2 provides some descriptive statistics on the returns to the UK equity portfolios of the pension funds in our dataset. The average discrete quarterly return over all funds over all quarters is 4.32%, compared with an average discrete return of 4.38% for the FT-All Share Index. The overall standard deviation of these returns is 8.67%, and the distribution of returns also emphasises the variability in returns. But these pooled measures disguises an important statistic, which is that the between funds standard deviation is much less than the within fund distribution. This implies that for a particular quarter the distribution of fund returns is tightly packed around the mean, but that over time the variability of returns is much higher. In fact the correlation between the time series values of the FT-All Share index and the average return each quarter across the pension funds is 0.995. The contrast in the within and between standard deviations might be indicative of the herding behaviour of pension funds suggested by Lakonishok et al.

Table 2 also report on the distribution of returns weighted by the value of the fund at the beginning of each quarter. The value weighted average return of 3.80% implies that small funds have a higher return than large funds and this is an issue we will return to later. In the subsequent regression analysis, we require a minimum number of observations to undertake a meaningful statistical analysis, and we imposed the requirement that time series fund parameters are only estimated when there were 12 or more quarterly returns for that fund. This cut-off value of three years accords with the typical fund mandate. Table 2 reports the distribution of returns of the sub-sample of 1724 funds with at least 12 time series observations, and this may be checked with the distribution of returns across the whole sample, to check that the sub-sample is indeed representative. Similarly table 2 also reports the distribution of returns of those 284 funds that remained in existence over all 56 quarters in our dataset.

In Panel B of Table 2 we report statistics of the size of the equity portion of the pension funds in our sample. The size distribution is highly skewed with a large number of very small funds. For example in 1997 the median size fund had an equity portfolio of 28 million pounds. Whereas the largest fund had an equity portfolio of over 9 billion pounds.

In this study we use data on all UK pension funds irrespective of whether they change manager, though normally we think of abnormal returns as being due to fund manager skills, and indeed this is the motivation in the Brown et al (1997), and Blake et al (1999) studies. But survivorship bias is likely to be more of an issue in same manager funds. In addition pension funds may be inherently different for example due to a different mix of contributors/pensioners. Further concentrating on the same fund manager condition ignores movement in personnel, between fund management companies. Pension fund trustees may switch fund managers after movement in personnel.

V Results

In the first row of Table 3 we report the average parameter estimates from regressing equation (1) across 1,714 funds, where the single factor benchmark return is specified as the excess return on the market. It can be seen that the average α is slightly positive but is insignificantly different from zero. We also report the distribution of these parameter values and the t-statistics across funds, and the distribution of the Jensen alphas and the associated t-statistics are plotted in figure 1 and 2. It can be seen that both distributions are symmetrically distributed around the mean. Just over half of the alpha statistics are positive, and about 10 per cent are significantly different from zero. The explanatory power of the individual time series regressions are very high, with the average coefficient of determination being 0.95. In addition the fund betas are typically close to unity: eighty per cent of the funds have betas between 0.95 and 1.08, which is consistent with our earlier finding that the distribution of returns in any quarter is highly correlated with the market index. It would appear that the funds in the sample are "closet-trackers" since they all invest in similar well-diversified portfolios, which mimick the market index.

We then divided the funds into two groups on the basis of fund size. This classification was determined as follows. Over the whole sample we computed the distribution of fund size, over time and across funds. We identified the fourth and eighth deciles of this distribution. Then for each fund we computed the average fund size over the fund's life. Those funds whose average size was less than the pooled distribution's fourth decile were classified as a small funds; those funds whose average size was greater than the pooled distribution's eighth decile were classified as large funds. This classification resulted in 731 small funds and 302 large funds. This classification was clearly arbitrary, but the reason for the asymmetric use of deciles reflected the skewed size distribution in the sample as evidenced in Table 2 Panel B

In Panels B and C in Table 3 we report the results by fund size. Surprisingly, the average alpha coefficient for the 731 funds in the small fund sample is negative, though insignificant. The average alpha coefficient for the 302 large funds is positive, but also insignificant. The interpretation of these results in comparison with the descriptive statistics in Table 2, is that once an adjustment is made for the fund's risk, the outperformance of small funds is less than for large funds. In figures 3 and 4 we plot the cross-section distributions of the fund alphas for large funds and small funds separately.

In table 4 we apply the two tests for market timing, for the single factor CAPM benchmark. The two tests are the Teynor-Mazuy test from equation (4) and the Merton-Henriksson test outlined in equation (5). Both tests produce similar results. The Jensen-alphas reported in Table 3 can be decomposed into a selectivity-alpha, and a market timing delta. The results in Table 4 shows that the selectivity-alphas for both the Teynor-Mazuy and the Merton-Henriksson tests are significantly positive, but that the timing coefficients are significantly negative, meaning that funds appear to be very poor market timers: they increase the betas of their portfolios at the wrong times. These funds appear to increase the beta of their portfolios when the market index is going down, and reduce the portfolio beta when the market index is increasing. These perverse market timing results are consistent with the findings of Coggin et al (1993). The distribution of the selectivity-alphas and the market timing delta are illustrated in figures 5 and 6.

Table 5 reports the results of evaluating fund performance by fund maturity. Fund maturity is proxied by net inflows into the fund. Funds with low and negative inflows will represent relatively mature funds who are running down the size of the fund. Funds with positive inflows will represent more immature funds, perhaps with few existing pensioners. For each observation we compute the net inflows in that quarter as [emv-(1+ret)smv]/smv. For each fund we estimate deciles of the net inflows distribution. Funds for which the fifth decile is positive (mostly positive inflows) are classified as immature funds. Funds for which the sixth decile is negative (mostly negative inflows) are classified as mature funds. This classification results in 619 mature funds and 625 immature funds. Surprisingly this division of the data results in evidence of significant abnormal performance for the sample identified as mature funds, but underperformance for the immature sample.

In table 6 we report the results of extending the single factor model to include an additional size factor. This additional factor allows for the fact that historically, small companies have traditional outperformed their large counterparts. This has been shown to important in the computation of appropriate benchmarks for studies of UK stock returns [Dimson and Marsh (1986)]. The returns on the size factor that we use is the difference between the return on a small firm index (Hoare-Govett Small Companies Index) and the FT-All Share index. In fact in the early 'nineties this premium was negative. The first row of table 6 shows that the coefficient γ_p on the size premium was positive on average and significant, and the majority of funds had a positive exposure to the size premium.

We also tested for timing effects with respect to the market index and the size premium. The δ_p coefficient reports the market timing effect, and the κ_p coefficient the effect of size timing. For the sample overall, from the third and fourth rows of Table 6, Panel A we can see that the both according to the Treynor and Merton tests, the average selectivity alpha is significantly negative, and the average market timing parameter is also significantly negative. These results imply that funds are both poor at selectivity and market timing. However in the case of the Treynor measure the positive exposure to the size premium, is accompanied by a positive average size timing κ_p parameter. This implies that funds are good at timing the size premium. The Merton test is slightly odd because of a negative coefficient on the size factor.

The remaining panels in Table 6 investigate these issues further by examining whether there is a difference in parameter estimates by size of fund, and also over different sub-periods. Panel B and C shows that it was also the case that both large and small funds had a positive exposure to the size premium, and the sensitivity of the small funds was greater: 0.093 rather than 0.051. For both sub-groups of funds, selectivity was significantly negative, market timing was poor, but size timing was significantly positive. The average size timing coefficient of 1.8 for the small fund sample was greater than that for the large funds, and implies that the small funds are more able to time the size premium. This is consistent with the idea that small funds are able to invest in small companies, whereas large funds are unable to take advantage of movements in the size premium, because it is more difficult for them to invest in small companies on account of their larger size.

Chart 2 shows the movement in a number of market indices over the whole period 1984-97. We can identify three distinct periods. The mid-eighties were characterised by a steep bull market, which ended after the stock market crash in the fourth quarter of 1987. There followed a period of slow and not very volatile growth in the indices up to the middle of 1992 when the UK exited the Exchange Rate Mechanism. The third period is identified by a continuation of the steady growth trend but with increased volatility.

The first rows of Panels D, E and F in Table 6, report the results of the two factor model for each of the three sub-periods. In the first bull market phase there is significant underperformance on average, though in the later two sub-periods on average funds outperform the two factor benchmark. The exposure to the size factor is always positive and significant. The inclusion of the timing variables, shows that selectivity is always negative, but that the size timing parameter is positive and significant, in the two earlier sub-periods though negative in the last sub-period, where the outperformance is explained by a positive market timing coefficient. The cross-section distributions of the Jensen-alphas for each of these sub-periods is plotted in figures 7, 8 and 9.

Table 7 expands on the results in Table 6, by examining portfolio performance jointly split by fund size and time period. Over the first sub-period 1984-87 both large and small funds underperformed the two factor benchmark. The decomposition of this underperformance shows that for both small and large funds there is positive and significant size timing. In both cases the addition of the quadratic size premium means that the coefficient on the linear size premium becomes negative. Selectivity for the large funs is positive, but negative for the small funds. Market timing for both groups is negative. In the middle time period 1987-92, both small and large funds display outperformance. Again most of this outperformance is driven by significant size timing, with insignificant selectivity and negative market timing. In the final sub-period 1992-97, both small and large funds outperform the benchmark, with negative size timing over this sub-period, particularly for large funds. The source of the outperformance over this sub-period is market timing. Selectivity of the large funds is positive but insignificant. Small funds exhibit significant negative selectivity.

In Table 8 we re-examine the question of portfolio performance using conditional performance evaluation techniques. The time-series regressions were resticted to those funds having a minimum of 20 quarters, since the parameters in the amended Merton-Henriksson regressions require 11 degrees of freedom. Comparing the results in Table 8 with those in Table 4, it can be seen that in the case of the Treynor test, the conditional estimation does not greatly alter the unconditional results: significant selectivity, but perverse market timing. Though for the Merton test, the conditional tests result in both significant selectivity and market timing.

Finally in Table 9 we report the results of estimating Treynor's market timing test for the 278 long-lived funds, that were in existence over the whole 56 quarters. This sub-set of the data exhibits insignificant selectivity, but positive market timing. Over the three sub-periods, as for the whole sample there is significantly negative selectivity in the first sub-period, but significantly positive selectivity in the two later sub-periods. In all three sub-periods there is positive market timing.

VI Conclusions

We have investigated the performance of the UK equity portfolios of 2,175 segregated UK pension funds over the period 1983-97. This is longest set of UK pension fund data analysed to date, and with such a long dataset we have been able to examine performance over three distinct sub-periods.

We noted at the outset the similarity between pension fund returns and the returns on the FT-All Share index. Most of the pension funds in our sample had an equity beta close to unity. In addition, the coefficient of determination in the regression of fund returns against returns on the market was very high. Both of these findings imply that fund returns were very close to the returns of the FT-All Share Index. It would appear that the funds in the sample are "closet-trackers" since they all invest in similar well-diversified portfolios, which mimick the market index. Any measures of outperformance were therefore bound to be small, and this is what we found.

We investigated the sensitivity of the fund returns to the addition of a size premium, which we found to be significant, and important for the smaller funds in our sample. This is consistent with the idea that larger funds are unable able to take advantage of investing in smaller companies, because of their concerns about the liquidity of their investments.

Over the whole period and across all funds average outperformance was insignificant. However during the sub-periods there was significant average underperformance during the strong bull market of the mid-eighties, but significant outperformance since 1987. In particular in the period 1987-92 the average outperformance across pension funds was one half of a percentage point per year.

Decomposing this abnormal performance we found that most of it could be explained by the ability of both large and small funds to time the size premium. On the whole there were negative returns to both selectivity and to market timing. There was little evidence of any differences in the performance between mature and immature funds.

References

Blake, D. Lehmann, B. and Timmermann, A. "Performance Measurement Using Multi-Asset Portfolio Data; A Study of UK Pension Funds 1986 - 1994", Pensions Institute

Blake, D., B. Lehmann and A. Timmermann (1999) "Asset Allocation Dynamics and Pension Fund Performance", *Journal of Business* forthcoming.

Blake, D. and Timmerman A. (1997) "The Birth and Death Processes of Mutual Funds" *European Finance Review*, 1997

Brown, G. P. Draper and E. McKenzie (1997) "Consistency of UK Pension Fund Performance", *Journal of Business Finance and Accounting* 24, March

Brown, S. and W. Goetzmann (1995) "Performance Persistence", Journal of Finance

Carhart, M. (1997) "On Persistence in Mutual Fund Performance", Journal of Finance,

Coggin, T.D., F.J. Fabozzi, S. Rahman (1993), "The Investment Performance of US Equity Pension Fund Managers: An Empirical Investigation", Journal of Finance,

Consumers' Association, *A Blueprint for Better Pensions* (1997)

Daniel, K., M. Grinblatt, S. Titman and R. Wermers (1997) "Measuring Mutual Fund Performance with Characteristic Based Benchmarks", Journal of Finance,

Department of Social Security (1998), *A New Contract for Welfare: Partnership in Pensions* (HMSO)

Dimson, E. and P. Marsh (1986) "Event study methodologies and the size effect", Journal of Financial Economics, vol. 17, 113-142.

Financial Services Agency (1999), *Comparative Information for Financial Services*

Grinblatt, M. and S. Titman (1992) "Persistence in Mutual Fund Performance", Journal of Finance,

Gruber, M. (1996) "Another Puzzle: The Growth in Actively Managed Mutual Funds", JooURNAL of Finance,

Hendricks, D., J. Patel and R. Zeckhauser (1993) "Hot Hands in Mutual Funds: Short Run Persistence of Relative Performance, 1974 – 1988", Journal of Finance

Ippolito, R.A. and J.A. Turner (1987) "Turnover, fees and pension plan performance", *Financial Analysts Journal*, vol. 43, 16-26.

Lakonishok, J.A., A. Shleifer and R.W. Vishny (1992) "The structure and performance of the money management industry", *Brookings Papers on Economic Activity*, 339-391.

Malkiel, B.G. (1995) "Returns From Investing in Equity Mutual Funds 1971 to 1991", Journal of Finance,

Office of Fair Trading (1997), *Report of the Director General's Inquiry into Pensions*

Table 1: Descriptive Statistics on Pension Fund-Quarters

Fund-Quarters						
No. of Funds		2,175		No. of Quarters		59,509
Distribution of Fund-Quarters						
min	5%	25%	50%	75%	95%	max
1	4	12	24	41	56	56

Table 2: Descriptive Statistics on Fund Returns and Fund Size

Panel A: Returns Across Quarters and Funds

	Returns				FT-All Share Rets
	All	Weighted by smv	>= 12 Quarters	= 56 Quarters	
Mean	0.0432	0.0380	0.0428	0.0444	0.0438
Std. Dev.	0.0867	0.0814	0.0867	0.0858	0.0834
Overall					
Between funds	0.01652		0.0092	0.0030	
Within Funds	0.08628		0.0864	0.0857	
Distribution of returns:					
5%	-0.0725	-0.0689	-0.0727	-0.070	
10%	-0.0543	-0.0537	-0.0543	-0.0536	
25%	0.0016	0.0016	0.0015	0.0021	
50%	0.0463	0.0441	0.0459	0.0469	
75%	0.0896	0.0747	0.0885	0.0926	
90%	0.1525	0.1346	0.1527	0.1511	
95%	0.1825	0.1693	0.1825	0.1823	
Obs.	59,317	59,314	56,403	15,842	56
No. of Funds	2170	2170	1724	278	

Panel B: Fund Size Across Funds

	Size at start of Quarter (£m)		
	March 1983	Dec 1990	Dec 1997
Mean	25.02	50.24	102.27
Std. Dev. Between	85.01	194.45	387.30
Distribution of Fund size:			
Min	0	0.018	0.17
5%	0.307	0.92	3.95
10%	0.441	1.36	6.02
25%	1.06	3.31	12.39
50%	3.20	8.35	28.12
75%	14.25	27.36	70.14
90%	51.64	102.88	221.90
95%	111.30	174.89	356.03
Max	1,113.4	3,823.63	9,108.62
Obs.	833	1131	1004

The table shows discrete returns, and computes arithmetic averages

Table 3 Performance Evaluation, with CAPM benchmark

	No. Funds	a	a t-stat	b	b t-stat	R2
Panel A: All funds						
Average values	1714	0.00017	0.966	1.018	1,280.0	0.953
Distribution of parameters						
10%		-0.0047	-1.4833	0.9525	14.4449	0.9146
25%		-0.0021	-0.7167	0.9911	20.8745	0.9510
50%		0.0002	0.0570	1.0218	30.5056	0.9692
75%		0.0023	0.8101	1.0508	39.8608	0.9796
90%		0.0046	1.4554	1.0802	48.3085	0.9861
No. coeffs >0 (*>1)		898		1173*		
No. of signif coeffs			165		1714	
Panel B: Small Funds (<40% smv)						
Average values	731	-0.00002	-1.5600	1.018	750.26	0.950
Panel C: Large Funds (>80% smv)						
Average Values	302	0.0001	0.2092	1.014	685.9	0.967

For each fund we regress the single factor model (CAPM) $R_{pt} - r_{ft} = a_p + b_p (R_{mt} - r_{ft}) + e_{pt}$. In the first row of the table we report the average parameter estimates from these regressions, and the relevant overall t-statistic for the average value of each parameter, computed as in equation (3) in the case of the α's, and similarly for the other parameters. The cross-fund distribution of the parameter estimates and corresponding t-statistics are displayed in the remaining rows. The final row counts the number of cross-fund parameter estimates that are greater than zero (greater than unity in the case of the β coefficient.

Table 4: Performance Evaluation for CAPM with Market Timing: All observations

	No. Funds	a	a t-stat	b	b t-stat	d(Timing)	dt-stat	Average R2
Panel A: Treynor –Mazuy Method								
Mean Parameter	1714	0.0008	11.055	1.012	1,310.1	-0.0013	-21.152	0.956
10%		-0.0042	-1.1350	0.9431	12.5994	-0.3816	-2.2391	0.9185
25%		-0.0012	-0.4001	0.9847	20.9055	-0.2395	-1.4136	0.9528
50%		0.0009	0.2918	1.0170	30.4643	-0.1068	-0.5497	0.9709
75%		0.0031	0.9912	1.0489	40.1341	0.0859	0.3593	0.9808
90%		0.0055	1.5914	1.0777	50.1130	0.5032	1.1716	0.9872
No. coeffs >0 (*>1)		1054		1099*		93		
Panel B: Merton-Henriksson								
Mean Parameter	1714	0.0018	19.412	1.044	1,107.6	-0.0493	-27.14	0.956

For each fund we regress the single factor model augmented by a market timing term. The Treynor-Mazuy test in (4) is $R_{pt} - r_{ft} = a_p + b_p (R_{mt} - r_{ft}) + d_p(R_{mt} - r_f)^2 + e_{pt}$, and the relevance of market timing is represented by the significance of the δ_p coefficient. The Merton-Henriksson test in (5) is $R_{pt} - r_f = a_p + b_p(R_{mt} - r_f) + d_p(R_{mt} - r_f)^+ + h_{pt}$ where $(R_{mt} - r_f)^+ = Max(0, R_{mt} - r_f)$, and the relevance of market timing again given by the significance of the δ_p coefficient. Panel A reports the results of the Treynor-Mazuy test, including the distribution of the individual fund estimates. Panel B reports only the mean parameter values of the time-series estimates. The relevant overall t-statistic for the average value of each parameter, computed as in equation (3) in the case of the α's, and similarly for the other parameters. All standard errors are robust.

Table 5, Performance Evaluation by fund maturity with Two-Factor benchmark

	No. Funds	a	a **t-stat**	b	b **t-stat**	g	g**t-stat**	**R2**
Panel A: Mature Funds (negative inflows)								
Mean Parameters	619	0.0007	5.7898	1.0207	819.7	0.0684	31.262	0.957
10%		-0.0032	-1.2274	0.9663	14.6397	-0.0398	-0.6931	0.9275
25%		-0.0010	-0.3911	0.9926	22.3514	0.0158	0.2687	0.9575
50%		0.0007	0.2848	1.0214	32.5832	0.0619	1.2663	0.9719
75%		0.0025	0.9450	1.0518	42.6638	0.1166	2.3529	0.9819
90%		0.0044	1.5328	1.0870	51.2683	0.1827	3.1880	0.9870
No. coeffs >0 (*>1)		377		432		499		
Panel B: Immature Funds (positive inflows)								
Mean Parameters	625	-0.00083	-5.6297	1.017465	742.2863	0.088264	31.93461	0.962136
10%		-0.0067	-1.9767	0.9448	13.7904	-0.0528	-0.6877	0.9283
25%		-0.0038	-1.0965	0.9825	19.5388	0.0123	0.1432	0.9578
50%		-0.0005	-0.1905	1.0188	28.7484	0.0708	1.0993	0.9735
75%		0.0018	0.6410	1.0532	38.3126	0.1547	2.2990	0.9830
90%		0.0046	1.4334	1.0865	46.1897	0.2520	3.5261	0.9887
No. coeffs >0 (*>1)		283		394		488		

For each fund we regress the two factor model $R_{pt} - r_{ft} = a_p + b_p (R_{mt} - r_{ft}) + g(R_{HGt} - R_{mt}) + e_{pt}$ for immature and mature funds separately. .For each observation we compute the net inflows in that quarter as [emv-(1+ret)smv]/smv. For each fund we estimate deciles of the net inflows distribution. Funds for which the fifth decile is positive (mostly positive inflows) are classified as immature funds. Funds for which the sixth decile is negative (mostly negative inflows) are classified as mature funds. The relevant overall t-statistic for the average value of each parameter, is computed as in equation (3) in the case of the α's, and similarly for the other parameters.

Table 6: Performance Evaluation by fund size and time sub-samples for Two-Factor benchmark with Market Timing.

	No. funds	a	a **t-stat**	g	g**t-stat**	d	d**t-stat**	k	k **t-stat**	**R2**
Panel A: All observations										
	1714	-0.0001	-0.0635	0.0788	53.924					0.961
Treynor	1714	-0.001	-12.14	0.0388	44.24	-0.0089	-19.67	1.382	51.98	0.966
Merton	1714	-0.002	-13.12	-0.0789	-10.34	-0.0375	-19.51	0.244	41.50	0.965
Panel B: Treynor Test Small Funds (<40% smv)										
	732	-0.00046	-3.150	0.0929	34.593					0.959
	732	-0.0013	-9.513	0.0338	25.051	-0.0486	-13.094	1.835	34.59	0.964
Panel C: Treynor Test Large Funds (>80% smv)										
	302	0.00004	0.667	0.0514	20.003					0.971
	302	-0.0005	-4.458	0.0064	14.73	-0.040	-15.12	1.329	27.56	0.976
Panel D: Treynor Test: 1st Q 1983- 3rd Q 1987										
	845	-0.0033	-26.45	0.107	29.26					0.971
	845	-0.0004	-7.269	-0.2327	-30.922	-0.2083	-49.876	5.026	51.818	0.978
Panel E: Treynor Test: 4th Q 1987 – 2nd Q 1992										
	1019	0.0013	14.262	0.0626	27.968					0.969
	1019	-0.00016	-0.192	0.0757	37.457	-0.088	-11.859	0.933	37.806	0.974
Panel F: Treynor Test: 3rd Q 1992 – 4th Q 1997										
	983	0.0003	2.922	0.0751	35.36					0.950
	983	-0.0003	-3.327	0.0712	35.05	0.67	30.95	-0.251	-11.61	0.959

For each fund we regress the two factor model $R_{pt} - r_{ft} = a_p + b_p (R_{mt} - r_{ft}) + g(R_{HGt} – R_{mt}) + e_{pt}$, with additional quadratic terms for market timing and size premium timing. The Treynot-Mazuy test becomes $R_{pt} - r_{ft} = a_p + b_p (R_{mt} - r_{ft}) + g(R_{HGt} – R_{mt}) + d_p(R_{mt} - r_{ft})^2 + k_p(R_{HGt} – R_{mt})^2 + e_{pt}$. The relevance of market timing is represented by the significance of the δ_p coefficient, and of size timing by the significance of κ_p. The relevant overall t-statistic for the average value of each parameter, is computed as in equation (3) in the case of the α's, and similarly for the other parameters.

Table 7: Performance Evaluation by Time and Size sub-samples with Two-Factor benchmark and Market Timing

No. funds	a	a **t-stat**	b	b **t-stat**	g	g**t-stat**	d	d**t-stat**	k	k **t-stat**	**R2**
Panel A: Treynor Test: 1st Q 1983- 3rd Q 1987, Small funds											
382	-0.0032	-16.193	1.012	507.79	0.1188	20.17					0.968
382	-0.0004	-4.883	0.988	682.04	-0.2281	-18.446	-0.2119	-29.24	5.1675	32.694	0.974
Panel B: Treynor Test: 1st Q 1983- 3rd Q 1987, Large funds											
169	-0.0039	-14.667	0.998	440.82	0.0822	11.776					0.980
169	0.0001	2.446	0.968	615.1	-0.2793	-20.316	-0.2617	-34.15	5.2495	29.124	0.986
Panel C: Treynor Test: 4th Q 1987 – 2nd Q 1992, Small funds											
377	0.0009	5.297	1.033	437.3	0.0687	16.78					0.967
377	-0.0009	-1.961	1.0299	500.4	0.0961	24.091	-0.0281	-4.003	1.1157	22.846	0.973
Panel D: Treynor Test: 4th Q 1987 – 2nd Q 1992, Large funds											
221	0.0017	9.037	1.032	402.13	0.0552	13.342					0.975
221	0.0003	0.4858	1.0321	485.2	0.0583	16.874	-0.1426	-8.577	0.9157	21.365	0.980
Panel E: Treynor Test: 3rd Q 1992 – 4th Q 1997, Small funds											
275	0.0003	1.2976	1.002	274.7	0.0857	18.040					0.940
275	-0.0005	-2.8015	0.9794	321.1	0.0823	16.717	0.5994	14.008	-0.1215	-2.723	0.950
Panel F: Treynor Test: 3rd Q 1992 – 4th Q 1997, Large funds											
213	0.0006	2.4608	1.024	321.95	0.0614	17.301					0.964
213	0.0001	0.4449	0.9902	357.6	0.0556	16.995	0.7901	19.167	-0.4351	-10.596	0.973

For each fund we regress the two factor model $R_{pt} - r_{ft} = a_p + b_p (R_{mt} - r_{ft}) + g_p(R_{HGt} – R_{mt}) + e_{pt}$, with additional quadratic terms for market timing and size premium timing. The Treynot-Mazuy test becomes $R_{pt} - r_{ft} = a_p + b_p (R_{mt} - r_{ft}) + g_p(R_{HGt} – R_{mt}) + d_p(R_{mt} - r_{ft})^2 + k_p(R_{HGt} – R_{mt})^2 + e_{pt}$. The relevance of market timing is represented by the significance of the δ_p coefficient, and of size timing by the significance of κ_p. The relevant overall t-statistic for the average value of each parameter, is computed as in equation (3) in the case of the α's, and similarly for the other parameters.

Table 8: Performance Evaluation with Conditional Estimation for CAPM with market timing

	No. Funds	**Average** a	a **t-stat**	**Average** b	b **t-stat**	**Average** d	d**t-stat**	**R2**
Treynor All (n>12)	1714	0.0018	19.38	1.041	811.7	-0.405	-21.39	0.967
Teynor All (n>20)	1299	0.0020	19.85	1.025	847.0	-0.2583	-20.697	0.964
Teynor Small Funds (n>20)	486	0.0016	8.4993	1.019	430.03	-0.1507	-7.874	0.961
Treynor Large Funds (n>20)	256	0.0021	11.443	1.0176	499.9	-0.2903	-14.638	0.974
Merton-H (n>20)	1299	0.0016	16.466	1.003	517.3	0.1593	2.003	0.970

For each fund we regress the conditional single factor model augmented by a market timing term, where each of the time-series regressions are restricted to those funds having a minimum of 20 quarters, since the parameters in the amended Merton- Henriksson regressions require 11 degrees of freedom. The Treynor-Mazuy test in (8) is $R_{pt} - r_f = a_p + b_p(R_{mt} - r_f) + B'z_{t-1}(R_{mt} - r_f) + d_p(R_{mt} - r_f)^2 + e_{pt}$ where the sensitivity of the managers beta to the private market timing signal is measured by d_p. The amended Merton-Henriksson test is $R_{pt} - r_f = a_p + b_d(R_{mt} - r_f) + B_d'z_{t-1}(R_{mt} - r_f) + d_c(R_{mt} - r_f)^+ + D' z_{t-1} (R_{mt} - r_f)^+ + h_{pt}$ where $(R_{mt} - r_f)^+ = (R_{mt} - r_f)* Max[0, R_{mt} - r_f - E(R_{mt} - r_f|Z_{t-1})]$; and $d_c = b_{up} - b_d$; $D = B_{up} - B_d$. The signifiance of market timing is represented by the significance of δ_c. The reported coefficients are the mean parameter values of the time-series estimates from the individual fund regressions. The relevant overall t-statistic for the average value of each parameter, computed as in equation (3) in the case of the α's, and similarly for the other parameters.

Table 9: Performance Evaluation of Long-lived Funds by time sub-samples for CAPM with Market Timing

	No. Funds	Average a	a t-stat	Average b	b t-stat	Average g	g t-stat
Whole Sample	278	0.00014	0.2485	1.0172	290.99	0.060333	9.279
1st Q 1983-3rd Q 1987	278	-0.0038	-17.755	1.0096	525.4	0.1062	18.624
4th Q 1987–2nd Q 1992	278	0.0026	12.264	1.0346	422.0	0.0683	16.821
3rd Q 1992–4th Q 1997	278	0.0006	3.256	1.0305	325.1	0.0776	20.846

For each fund we regress the single factor model (CAPM) $R_{pt} - r_{ft} = a_p + b_p (R_{mt} - r_{ft}) + e_{pt}$ for the three time sub-periods. We report the average parameter estimates from these regressions, and the relevant overall t-statistic for the average value of each parameter, computed as in equation (3) in the case of the α's, and similarly for the other parameters.