Pensions Institute

DISCUSSION PAPER PI-0106

Performance Benchmarks for Institutional Investors: Measuring, Monitoring and Modifying Investment Behaviour

David Blake and Allan Timmermann

June 2002

ISSN 1367-580X

The Pensions Institute
Cass Business School
City University
106 Bunhill Row London
EC1Y 8TZ
UNITED KINGDOM

http://www.pensions-institute.org/

Chapter 5

Performance benchmarks for institutional investors: measuring, monitoring and modifying investment behaviour

DAVID BLAKE AND ALLAN TIMMERMANN

ABSTRACT

The two main types of benchmarks used in the UK are external asset-class benchmarks and peer-group benchmarks. Peer-group tracking is much more prevalent with pension funds and mutual funds than with life funds. However, the use of customized benchmarks that reflect the specific objectives set by particular funds is increasing. Benchmarks influence the type of assets selected and, equally significantly, the type of assets avoided. Peer-group benchmarks have a tendency to distort behaviour, particularly when combined with a fee structure that does not promote genuine active management. The outcome tends to be herding and closet index matching.

The main alternatives to peer-group benchmarks are: single-index benchmarks with time-varying coefficients, multiple-index benchmarks and fixed benchmarks. The first two alternatives have recently been discussed in the academic literature but have yet to catch on in the practitioner community.

There are also benchmarks based on liabilities. These are generally related to real earnings or consumption growth or to the discount rate on liabilities. Explicit liability-based benchmarking is currently not very common, but is likely to become so in the light of both the increasing maturity of pension funds, various regulatory and financial reporting developments, and the Myners Review of Institutional Investment. Liability-driven performance attribution explicitly takes the liabilities into account.

The US has similar external asset-class and peer-group benchmarks as the UK. Other countries tend to use fixed or bond-based benchmarks.

In conclusion, we find that benchmarks are important, but so are fee structures. They can either provide the right incentives for fund managers or they can seriously distort their investment behaviour.

5.1 INTRODUCTION

The issue of performance benchmarks for institutional investors has generated a great deal of controversy recently. Are they set too low, making them very easy to beat? Are they set too high, making them hard to beat unless fund managers take on excessive risks? Is the frequency of assessment against the benchmark (typically on a quarterly basis) appropriate for long-term investors? Do they introduce unintended (and undesired) incentives, such as the incentive for fund managers to herd together or to avoid holding securities (such as those of micro-cap, small-cap, unquoted or start-up companies) that are not included in the benchmark? How, if at all, should performance against the benchmark influence the fund manager's compensation. How should the fund's liabilities be taken into account when assessing the fund's performance. This chapter examines and assesses the benchmarks that are currently used by institutional investors in the UK. It also looks at possible alternatives to these benchmarks and briefly reviews what happens in other countries.

5.2 WHAT BENCHMARKS ARE CURRENTLY USED BY INSTITUTIONAL INVESTORS?

Performance benchmarks in the UK have been around since the early 1970s. They are an essential part of the investment strategy of any institutional investor and help both to define client/trustee expectations and to set targets for the fund manager. Benchmarks can be set in relation to liabilities and can therefore change if the liabilities change, say, as a result of increasing maturity. Benchmarks might also be influenced by regulations (e.g. a Minimum Funding Requirement[1] (MFR)), accounting standards (e.g. Financial Reporting Standard 17[2] (FRS17)), or client/trustee preferences (e.g. trustees might prefer

[1]Introduced in the UK by the 1995 Pensions Act and operating from 1997, but it was announced in the March 2001 Budget that it would be scrapped.

[2]Issued by the Accounting Standards Board in November 2000 and coming fully into force in June 2003.

to minimize the volatility of employer contributions into a pension plan than minimize the average level of employer contributions, given that, in final salary plans, the pension is funded on a balance of cost basis).

The benchmark, appropriately set, has important implications for how the actions of the fund manager are interpreted. An appropriate benchmark recognizes formally that the strategic asset allocation or SAA (i.e. the *long-run* division of the portfolio between the major categories of investment assets, such as equities, bonds and property) is a risk decision relative to the liabilities, rather than an expected return decision. In other words, the SAA, properly interpreted, is not an investment decision at all: instead it is determined largely by reference to the maturity structure of the anticipated liability cash flows. In contrast, the stock selection and market timing (i.e. tactical asset allocation) decisions are investment decisions and it is the fund manager's performance in these two categories that should be judged against the benchmark provided by the SAA.

5.2.1 Single-index benchmarks and peer-group benchmarks

The two main types of benchmarks used in the UK are external asset-class benchmarks and peer-group benchmarks. These benchmarks are used by both 'gross funds' (i.e. those without explicit liabilities) and 'net funds' (i.e. those, such as pension funds, with explicit liabilities). When external performance measurement began in the early 1970s, most pension funds selected customized benchmarks (which involved tailoring the weights of the external benchmarks to the specific requirements of the fund). Shortly after, curiosity about how other funds were performing led to the introduction of peer-group benchmarks. More recently, following the recognition that the objectives of different pension funds differ widely, there has been a return to customized benchmarks.

The WM Company,[3] for example, uses the following set of external benchmarks to assess the performance of the pension funds in its stable:

- UK equities: FTA All Share Index.
- International equities: FT/Standard & Poor World (excluding UK) Index.
- North American equities: FT/Standard & Poor North America Index.
- European equities: FT/Standard & Poor Europe (excluding UK) Index.
- Japanese equities: FT/Standard & Poor Japan Index.
- Asia – Pacific equities: FT/Standard & Poor Asia – Pacific (excluding Japan) Index.

[3]The WM Company is one of the two key performance measurement services in the UK, the other is CAPS (Combined Actuarial Performance Services).

- UK bonds: British Government Stocks (All Stocks) Index.
- International Bonds: JP Morgan Global (excluding UK) Bonds Index.
- UK index-linked bonds: British Government Stocks Index-linked (All Stocks) Index.
- Cash: LIBID (London Inter-Bank Bid Rate) 7-day deposit rate.
- UK Property: Investment Property Databank (IPD) All-Property Index.
- International Property: Evaluation Associates All Property Index (a US index to reflect the fact that most international property investments are held in the US).
- Total portfolio: WM Pension Fund Index (based on all the funds monitored by WM).

All these indices assume that income is reinvested (gross of tax) and the returns are calculated on a value- and time-weighted basis. These benchmarks have the virtues of being independently calculated and immediately publicly available. However, some of them (most notably cash and international equities and bonds) have weightings that can differ substantially from those of the pension funds. Some indices are subject to measurement problems, particularly the property indices. Further, the external benchmarks include only the securities of relatively large companies.

The WM Company also uses peer-group indices for pension funds:

- WM50 Index for very large funds.
- WM2000 Index for small and medium-sized funds.

These are designed to reflect the fact that UK pension funds have portfolio weights that can differ substantially from those of the external indices. For example, UK pension funds tend to have a higher weight in Europe and a lower weight in the US than a global market-weighted index (ex UK). They also reflect the fact that large (mainly mature) funds have a very different asset allocation from that of smaller (less mature) funds. Both sets of indices are gross of the following costs: transactions costs (dealing spreads and commissions) and running costs (management and custody fees, property security and insurance costs).

Peer-group tracking is less prevalent with life funds than with pension funds. WM has constructed a with-profits universe mainly as a result of the curiosity of life offices to know how competitors are performing, but acknowledges that the product range of life offices is too great to make meaningful peer-group comparisons. Most benchmarks for life funds are based on external indices. In comparison, peer-group comparisons are more common with unit trusts and are used for promotional purposes.

5.2.2 Evaluating the single-index benchmarks

How are they constructed?

The first question that must be asked with any external index-based benchmark is: how was it constructed? Suitable index-based benchmarks have to be constructed on a value- and time-weighted basis. This essentially means that the constituents of the index are weighted according to their market capitalizations and that the timing of reinvested income is not allowed to distort the measured return. Other types of indices such as price-weighted indices (which simply sum up the prices in the index regardless of market capitalization) and geometric indices (which simply multiply together the prices in the index regardless of the market capitalization) would not make suitable benchmarks. This is because it is impossible for any real-world portfolio to mimic the behaviour of either of these two indices. However, while it is impossible for a real-world portfolio to mimic, say, a geometric index, it would not be difficult for the real-world portfolio to beat this index: anyone who knows Jensen's inequality will understand why! (see Blake (2000: 590–591)).

Even with benchmarks constructed on a value- and time-weighted basis, there are practical considerations to take into account before using them to assess performance. First, benchmarks can be constructed without having to incur the kinds of costs that face real-world fund managers, such as brokers' commissions, dealers' spreads and taxes.

Second, the constituents of the benchmark change quite frequently. While this involves no costs for the benchmark, it involves the following costs for any fund manager attempting to match the benchmark. The deleted securities have to be sold and the added securities have to be purchased: this involves both spreads and commissions. In addition, when the announcement of the change is made, the price of the security being deleted tends to fall and the price of the security being added tends to rise and these price changes are likely to occur before any fund manager has the chance to change his portfolio. A bond index-based benchmark is even more expensive to beat: over time, the average maturity of a bond index will decline unless new long-maturing bonds are added to replace those that mature and automatically drop out of the index.

Third, the benchmark assumes that gross income payments are reinvested costlessly back into the benchmark on the day that the relevant stock goes *ex-dividend*. In practice no fund manager would be able to replicate this behaviour: dividends and coupon payments are not made until some time after the ex-dividend date, the payment is generally made net of income or withholding tax, there are commissions and spreads incurred when reinvesting income and the trickle of dividend or coupon payments that are received

at different times are going to be accumulated into a reasonable sum before being reinvested. All these factors cause a tracking error to develop between the benchmark and any real-world portfolio attempting to match the benchmark, and leads to the real-world portfolio invariably underperforming the benchmark. So tracking error has to be recognized as an inevitable part of the process of fund management, both for active and passive strategies.

Why are they difficult to beat?

Apart from these practical considerations, there are other reasons why an institutional investor might find it difficult to beat an external index-based benchmark. First, there may be restrictions placed on fund managers which prohibit them from even attempting to match the index, let alone beat it. We can consider some examples. There can be trustee-imposed prudential limits on the maximum proportion of the fund that can be invested in a single security. For example, most trustees place a limit of 10% on the fund's investment in the shares of a single company. When the market weighting of Vodaphone in the FTSE100 index rose above 10% during 2000, fund managers were obliged to sell Vodaphone shares to bring their portfolios within the 10% limit and the FTSE100 index compilers were asked to introduce a new benchmark in the form of a 'capped' FTSE100 index that limits the weight of any security to 10%. As another example, some countries place regulatory limits on the holdings of certain securities by foreign investors: e.g. for national security reasons there might be limits on the foreign ownership of defence sector stocks.

Second, investors may not wish to be represented in some of the markets covered by the index. For example, a global emerging markets index would cover all continents, but investors might choose to avoid certain regions such as Africa, the Middle East or Russia.

Third, there is the so-called 'home country bias', the preference for securities from the home market. If UK pension funds were fully diversified on a global basis, they would hold less than 10% of their assets in the UK and more than 90% abroad. Yet UK pension funds which are the most diversified internationally of all the world's pension funds hold around 80% of their assets in the UK and only about 20% abroad.

Why should this be the case if the most diversified and hence the least risky portfolio possible is the global index? The only defensible answer to this question is that UK pension fund liabilities are denominated in sterling and, for liability matching purposes, pension fund managers select a high weight for sterling-denominated assets. It cannot really be justified on the grounds of risk. In the last 10 years, UK pension funds would have performed much better had they held the global index: although the Japanese market fell

markedly, the rise in the US market more than compensated for this as well as outperforming the UK by a handsome margin (see, e.g., Timmermann and Blake (2000)).

Finally, even if an index is chosen as a benchmark, no index currently in use contains the shares and bonds of all the companies in the economy, although it should if it is to be an efficient index.

Why is there a bias against small companies and venture capital?

The external indices listed above contain the securities of only relatively large companies. This is a particularly important issue for new companies which find it difficult to obtain equity capital to finance their start-up or to expand in their early years. The gap in the provision of equity finance for small companies in the UK was first identified by the Macmillan Committee on Finance and Industry in 1931 (and is known as the 'Macmillan gap'). The Macmillan gap was still present when the Wilson Committee to Review the Functioning of Financial Institutions reported in 1980 and made these comments about pension funds:

> In law, their first concern must be to safeguard the long-term interests of their members and beneficiaries. It is, however, possible for fiduciary obligations to be interpreted too narrowly. Though the institutions may individually have no obligation to invest any particular quantity of new savings in the creation of future real resources, the prospect that growth in the UK economy over the next two decades might be inadequate to satisfy present expectations should be a cause of considerable concern to them. The exercise of responsibility which is the obverse of the considerable financial power which they now collectively possess may require them to take a more active role than in the past ... in more actively seeking profitable outlets for funds and in otherwise contributing to the solutions of the problems that we have been discussing. (Wilson (1980: 259–260)).

The pension funds' defence against this criticism rested on the argument that the costs of investing in small companies were much higher than those of investing in large companies. The reason for this is as follows. Small companies are difficult, and therefore expensive, to research because they are generally relatively new and so do not have a long track record. Also, their shares can be highly illiquid, and pension funds, despite being long-term investors, regard this as a very serious problem. Further, pension fund trustees place limits on the proportion of a company's equity in which a fund can invest. For example, a pension fund might not be permitted by its trustees to hold more than 5% of any individual company's equity. For a company with equity valued at £1 m, the investment limit is £50,000. A large pension

fund might have £500 m of contributions and investment income to invest per year. This could be invested in 10,000 million-pound companies or it could be invested in 50 large companies. It is not hard to see why the pension fund is going to prefer the latter to the former strategy, even if it could find 10,000 suitable companies in which to invest.

Related to the criticism that pension funds are unwilling to invest in small companies is the criticism that pension funds have been unwilling to supply risk-taking start-up or venture capital to small unquoted companies engaged in new, high-risk ventures. Venture capital usually involves the direct involvement of the investor in the venture. Not only does the investor supply seed-corn finance, he also supplies business skills necessary to support the inventive talent of the company founder. This can help to reduce the risks involved. The reward for the provision of finance and business skills is long-term capital growth. The problem for pension funds is that, while they have substantial resources to invest, they do not generally have the necessary business expertise to provide the required support. Further, while venture capital investments only ever take up a small proportion of the total portfolio, they take up a disproportionate amount of management time. Also the performance in the early years can be poor. As a result, pension funds remain largely portfolio investors rather than direct investors. In other words, they prefer to invest in equity from which they can make a quick exit if necessary, rather than make a long-term commitment to a particular firm.

Not only do pension funds tend to avoid the risks of direct investment, they tend also to be risk-averse when it comes to portfolio investment. They seek the maximum return with the minimum of risk, and the investment managers of pension funds tend to be extremely conservative investors, devoid of entrepreneurial spirit. As G. Helowicz has pointed out, pension funds:

> do not have any expertise in the business of, or a commitment to, the companies in which they invest. Shares will be bought and sold on the basis of the potential financial return. It therefore follows that the potential social and economic implications of an investment decision have little influence on that decision. (Benjamin *et al.* (1987: 98))

The other main factor is the legacy of the great inflation of the 1970s and the stop-go policies of governments at the time. UK investors with highly cyclical venture capital investments experienced substantial losses during every 'stop' phase.

UK pension funds have in recent years responded to the above criticisms. For example, some of the larger funds have established venture capital divisions. But they invest only about one-tenth of what US pension funds invest as a proportion of assets: 0.5% of total assets in 1998 as against 5% in the

US, according to the British Venture Capital Association. The venture capital industry raised three times more funding in 1998 from overseas pension funds and insurance companies than from their UK equivalents: 37% of the total as against 13%. Moreover, most of the venture capital in the UK is used to finance management buy-outs in existing companies, rather than to finance green field site development.

Nevertheless, it appears to be the case that the 'statement of investment principles' and the 'statement on socially responsible investment' required by the 1995 Pensions Act have focused the attention of pension funds on these issues in a way that was absent before the Act. The same is likely to be true of the 'principles of institutional investment' that will be introduced following the Myners Report.[4] It is possible that establishing a suitable venture capital benchmark might help to promote pension fund investment in new start-ups as well. It certainly appears to be the case that behaviour soon follows measurement when a performance benchmark is established: very quickly, the benchmark changes from being a tool of measurement to a driver of behaviour.

5.2.3 Evaluating the peer-group benchmarks

What is the effect of peer-group benchmarks?

This question has recently been addressed by Blake, Lehmann and Timmermann (2000). They find that the answer depends to a large extent on the industrial organization of and practices within the fund management industry.

The UK fund management industry is highly concentrated, with the top five fund management houses accounting for well over 50% of the funds under management (it was as high as 80% in 1998). This contrasts with the US where the top five fund managers account for less than 15% of the market. There is also a much lower turnover of fund management contracts in the UK than in the US, implying that client loyalty can help smooth over periods of poor performance more effectively than in the US. In addition, there is a single dominant investment style in the UK (namely balanced multi-asset management), which contrasts with the much wider range of styles in the

[4]Myners (2001). The principles cover: effectiveness of decision making by well-informed fund trustees, clarity of investment objectives for the fund, adequacy of time devoted to the strategic asset allocation decision, competitive tendering of actuarial and investment advice services to trustees, explicit investment mandates for fund managers, shareholder activism, appropriate benchmarks, performance measurement of fund managers and advisers, transparency in decision making, publication of mandates and fee structures via the statement of investment principles, regularity of reporting the results of monitoring of advisers and fund managers.

US (e.g. value, growth, momentum, reversal, quant and single asset-class management).

Further, the remuneration of the fund manager typically depends solely on the value of assets under management, not on the value added by the fund manager and there is typically no reward for outperforming either the external or peer-group benchmark and no penalty for underperforming these benchmarks. However, the long-term success of any fund management house depends on its relative performance against its peer group. The large fund management houses in the UK have lost business in recent years not because of their poor *absolute* performance, but because of their poor *relative* performance.

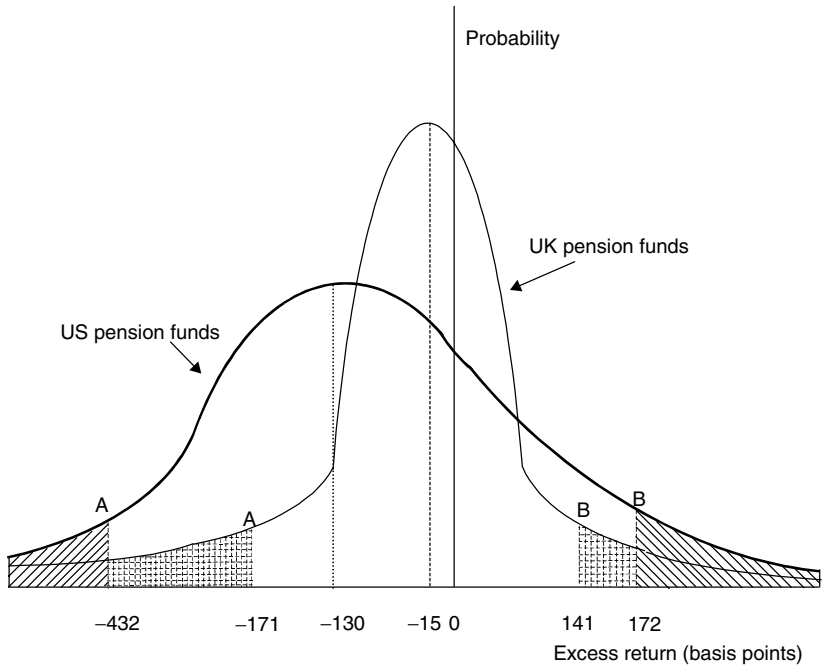
These differences in industrial organization and practice have led to significant differences in investment performance between pension funds in the UK and US. Blake, Lehmann and Timmermann (2000) found that, during the 1980s and 1990s, the median UK pension fund underperformed the market index by a fairly small 15 basis point p.a., whereas the median US pension fund underperformed by a much wider margin of 130 basis points p.a.[5] At the same time, the dispersion of pension fund returns around the median was much greater in the US than in the UK (603 basis points for the 10–90 percentile range, compared with 311 basis points in the UK).[6] These results, illustrated in Figure 5.1, clearly indicate that genuine active fund management is much more prevalent in the US than in the UK: UK pension fund managers display all the signs of herding around the median fund manager who is himself a closet index matcher.

What role do fee structures play?

Fee structures appear to provide a disincentive to undertake active management in the UK, while relative performance evaluation provides a strong incentive not to underperform the median fund manager. While UK pension fund managers are typically set the objective of adding value, their fees are generally related to year-end asset values, not to performance. Genuine *ex ante* ability that translates into superior *ex post* performance increases assets under management and, thus, the base on which the management fee is calculated. However, this incentive is not particularly strong and active management subjects the manager to non-trivial risks.

The incentive is weak because the prospective fee increase is second order, being the product of the *ex post* return from active management and the management fee and thus around two full orders of magnitude smaller than

[5]The US results come from Lakonishok, Shleifer and Vishny (1992: 348).
[6]The US results come from Coggin, Fabozzi and Rahman (1993: 1051).



Source: Blake, Lehmann and Timmermann (2000).
Note: A = 10th percentile of funds; B = 90th percentile of funds

Figure 5.1 The dispersion of returns on UK and US pension funds in excess of the market index

the base fee itself. Moreover, the *ex post* return from active management of a truly superior fund manager will often be negative and occasionally large as well, resulting in poor performance relative to managers who eschewed active management irrespective of their ability. The probability of relative underperformance large enough to lose the mandate is likely to be at least an order of magnitude larger than the proportional management fee. Hence, the risk of underperformance due to poor luck outweighs the prospective benefits from active management for all but the most certain security selection or market timing opportunities.

How successful are active fund managers?

The next result concerns the active management abilities of UK pension fund managers, that is, their skill in outperforming a passive buy-and-hold strategy. There are two principal types of active management: security selection and market timing. Security selection involves the search for undervalued securities (i.e. involves the reallocation of funds within asset categories) and

market timing involves the search for undervalued sectors (i.e. involves the reallocation of funds between sectors or asset categories).

Blake, Lehmann and Timmermann (1999) decomposed the median total return earned by pension fund managers into the following components:

Component	Percentage
Strategic asset allocation	99.47%
Security selection	2.68%
Market timing	−1.64%
Other	−0.51%
Total	100.00%

The most important task of pension fund managers is to establish and maintain the SAA and the decomposition reveals that, of the median total return over the sample period of 12.06% p.a., 12.00% p.a. (or 99.47% of the total) was due to this essentially passive activity. In terms of the active components, the average pension fund was unsuccessful at market timing, generating a negative contribution to the total return of −1.64%. Security selection was more successful, making a positive contribution to the total return of 2.68%. Even so, the overall contribution of active fund management at just over 1% of the total return (or about 12 basis points p.a.) is less than the annual fee that active fund managers typically charge (which range between 20 basis points for a £500m fund to 75 basis points for a £10m fund[7]).

Finally, the study by Blake, Lehmann and Timmermann (2000) found that above-average performance by a particular fund manager (so-called 'hot hands' in investment performance) was very short-lived: it rarely lasted more than a year. Studies of US fund managers have found persistence in performance extending out to two or three years, but no longer (Hendricks, Patel and Zeckhauser, 1993).

Is there a role for performance-related fees?

One way of providing appropriate incentives to those fund managers who believe that they can generate superior investment performance is to use performance-related investment management fees. In one example of this, the fee is determined as some proportion, f_1, of the difference between the fund's realized performance, g_t, and some benchmark or target, g_t^*, plus a base

[7] *Pensions Management* (September 1998).

fee to cover the fund manager's overhead costs, set as a fixed proportion, f_2, of the absolute value of the fund (V_t in period t):

$$\textit{Performance-related fee in period } t = f_1(g_t - g_t^*)V_t + f_2 V_t \tag{5.1}$$

This would reward good *ex post* performance and penalize poor *ex post* performance, whatever promises about superior *ex ante* performance had been made by the fund manager. The fund would have to accept a reduced fee or even pay back the client if g_t was sufficiently below g_t^* (although the latter case generally involves credits against future fees rather than cash refunds).

Another possibility that is less extreme since it does not involve refunds is:

$$\textit{Performance-related fee in period } t = f_i V_t \tag{5.2}$$

where f_i is the fee rate if the fund manager's return is in the ith quartile.

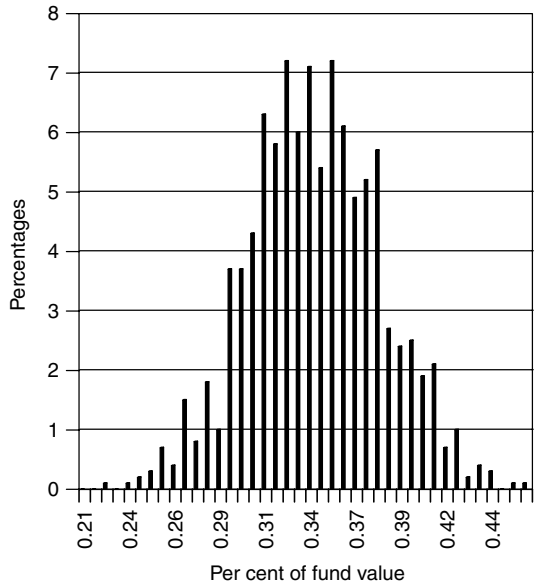
An example of this second type of fee structure is that of the Newton Managed Fund whose particular fee structure is listed in Table 5.1. Figure 5.2 illustrates how this fee structure might work in practice. The chart shows the distribution of fees payable to the manager of a middle-sized fund, based on a Monte Carlo simulation. The 90% confidence interval for the fees lies between 0.22 and 0.45% p.a., while there is a 25% chance that the fee will exceed 0.37% p.a. and a similar chance that it will be less than 0.31% per annum. A mean annual charge of 0.34% implies a total take of approximately 8.9% of the terminal fund value over an investment horizon of 25 years.

The level set for the target g_t^* will have important implications for the outcome. If the target is unrealistic and outside the range of performance expected by a skilled fund manager, the only way the manager can reasonably achieve the stipulated performance is by increasing the volatility of his investment strategy, i.e. by increasing risk. This is highly relevant in practice as some targets are very hard to achieve. Examples of these are: 'beat the median fund by 2 percentage points over a three-year rolling period', or 'be

Table 5.1 Newton Managed Fund

Quartile rank	Fund size		
	Up to £10m	£10–£50m	Above £50m
1st	0.94%	0.59	0.04
2nd	0.79	0.44	0.03
Median	0.69	0.34	0.02
3rd	0.59	0.24	0.01
4th	0.44	0.09	0.01

Source: Newton Fund Managers.



Note: The frequency diagram shows the annual distribution of performance-related fees in a fund with fees calculated according to the following performance scale:

Quartile rank	Fee (%)
1st	0.59
2nd	0.44
Median	0.34
3rd	0.24
4th	0.09

The Monte Carlo simulation assumes the following: a fund with a 25-year investment horizon, a distribution of returns which is normal with a mean of 9% p.a. and a standard deviation of 18%, and 1000 replications. Based on long-run returns reported in Credit Suisse First Boston's *Equity-Gilt Study* (2000), such a portfolio would be invested 35% in equities and 65% in bonds.

Figure 5.2 Frequency distribution of performance-related fees

in the upper quartile of performance'. There is an unconditional probability of 75% of failing to achieve the second target! Clients/trustees are beginning to accept that high targets will most likely be associated with greater volatility in performance, unless the client has *a priori* information that the fund manager is genuinely capable of delivering the target performance.

Clients/trustees are also beginning to accept that targets based on the peer-group median or peer-group distribution are very likely to distort fund

manager behaviour. This is partly because the median performance is really an outcome rather than a target. Whereas a fund manager knows the composition of an external index prior to making his own investments and so knows how much he is overweight or underweight in different securities, he will not know for sure what the asset allocation of the median fund manager is until the end of the performance period. All fund managers will be in the same position and this provides a strong incentive for fund managers not to deviate too far from each other. Hence, we find that there is a tight distribution of fund managers around the median fund manager who, in turn, generates a performance little different from that of a passive index matcher. Those fund managers who beat the median fund by 2 percentage points over a three-year rolling period, or who end up in the upper quartile of performance, are therefore more likely to do so by chance than by skill.

All this suggests that the target g_t^* should be set in relation to an external benchmark rather than to a peer-group benchmark if clients/trustees wish their fund managers to pursue genuinely active fund management strategies. However, this makes quartile-based fee structures virtually impossible to implement, since information on the distribution of returns around the median value of the external index is not collected centrally.

It is particularly important for the fee rate to be symmetric about the target g_t^*, so that underperformance is penalized in exactly the same way that outperformance is rewarded. The worst possible fee structure from the client/trustee's point of view would be one that rewarded outperformance but did not penalize underperformance. An example of this would be:

$$\textit{Performance-related fee in period } t = \max[0,\ f_1(g_t - g_t^*)V_t] + f_2 V_t \tag{5.3}$$

This particular fee structure would simply encourage the fund manager to take risks with the client/trustee's assets. If the fund manager's risk taking paid off, he would receive a large fee. If, on the other hand, performance was disastrous, the fund manager would still get the base fee. All the risk of underperformance (at least in the short term) therefore falls on the client/trustee.

How frequently should fund managers be assessed?

A final issue of importance concerns the frequency with which fund managers are assessed against the benchmark. Despite having very long-term investment objectives, the performance of pension fund managers is typically assessed on a quarterly basis. This is said to provide another disincentive from engaging in active fund management because of the fear of relative underperformance against the peergroup and the consequent risk that an underperforming fund manager will be replaced.

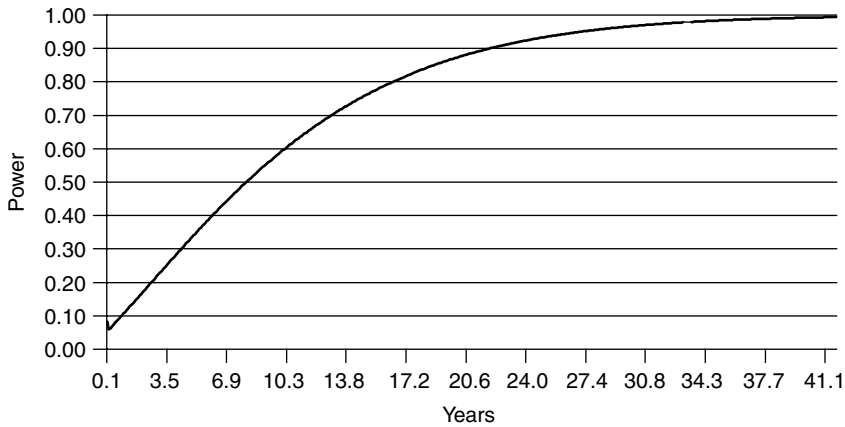
The frequency with which fund managers have their performance assessed ought to be related to the speed with which market anomalies are corrected. Suppose, as argued above, the benchmark has been set in relation to the SAA. Then it is the fund manager's performance in the two active strategies of stock selection and market timing that should be judged against the benchmark provided by the SAA. So the critical question is how long does it take for undervalued stocks to become correctly priced or for market timing bets to succeed? If financial markets are relatively efficient, then pricing anomalies should be corrected relatively quickly. This appears to suggest that a relatively short evaluation horizon is appropriate. To illustrate using a somewhat extreme example, if a market timing bet that involves, say, a significant underweighting of the US stock market, has not paid off after 10 years, then we might be tempted to say that the bet was a bad one.

However, two points speak against the use of relatively short evaluation horizons. The first has to do with time-variations in the investment opportunity set as represented by the relative expected returns and the conditional variances and covariances between the different asset classes. Many studies in the finance literature suggest that the first and second moments of returns on different asset classes vary systematically as a function of the underlying state of the world. Nevertheless, there is considerable uncertainty about how best to model such variations. But it seems reasonable to expect a successful market timing strategy to be linked to the ability to anticipate changes in the underlying economic state. This tends to evolve over fairly long periods of time, as exemplified by the 10-year expansion in the US economy up to 2000. If clients want fund managers to time swings in the business cycle, a long evaluation horizon would seem more appropriate.

The second justification for using a longer investment horizon is that performance is measured with so much noise that it is in effect impossible to assess true fund management skills based on a short performance horizon. Under reasonable assumptions,[8] it is possible to generate the following relationship between the length of the performance record and the power of the test for assessing fund management skills:

Power	Required data record
25%	3.5 years
50%	8 years
90%	22 years

[8]See the Appendix.



Source: See Appendix.

Figure 5.3 Power function – probability of correctly detecting abnormal performance

These figures are derived from Figure 5.3. The power of the test measures the probability of correctly rejecting the null hypothesis that the fund manager generates no abnormal performance. It is clear from the figures that it takes a long time to detect with reasonable confidence that the performance of the fund manager is abnormal. And this result is dependent on an unchanging investment opportunity set which is in itself an unlikely eventuality over a 22-year time horizon.

5.3 WHAT ARE THE ALTERNATIVES?

Recently, the academic literature has begun to investigate alternative benchmarks, based on extensions to the Capital Asset Pricing Model (CAPM). They help to identify the sources of any under- or outperformance by fund managers. There are also fixed benchmarks.

5.3.1 Single-index benchmarks with time-varying coefficients

The external benchmarks considered above are single-index benchmarks that can be justified by the CAPM, invented by Nobel prize winner Bill Sharpe and now one of the cornerstones of modern finance theory.

What is the CAPM?

The CAPM decomposes the expected return on a fund into two parts. The first is the return on a riskless asset such as Treasury bills: all professional investors should be expected to generate a return exceeding that on Treasury

bills! The second is the additional return from taking on 'market risk'. This, in turn, has two components: the 'market risk premium' (otherwise called the 'excess return on the market' or the 'market price of risk'), and the 'quantity' of market risk assumed by a particular fund as measured by that fund's 'beta'.

The market risk premium is measured by the difference between the expected return on the market index and the risk-free rate. The principal market index in the UK is the FTA All Share Index and many equity fund managers have this index as their single-index benchmark. The historical long-run market risk premium for the UK is about 6% p.a.

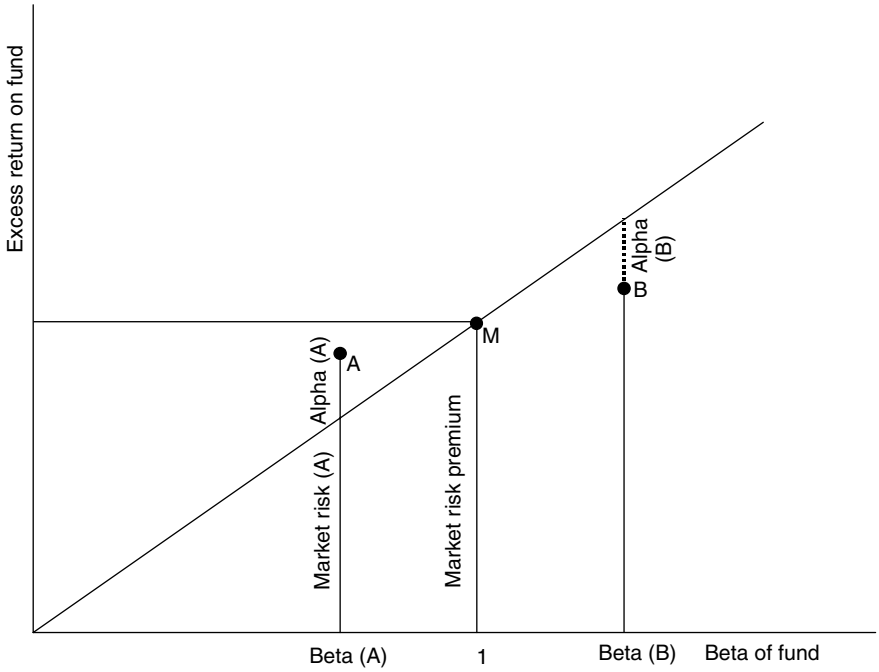
The beta of a fund measures the degree of co-movement between the return on the fund and the return on the market index. Technically the beta is calculated as the ratio of the covariance between the returns on the fund and the market to the variance of the return on the market. It is also equal to the product of the standard deviation of the return on the fund and the correlation between the returns on the fund and the market. These are exactly the same formulae as the slope or beta coefficient in a time-series regression of the excess return on the fund on an intercept and the market risk premium, which explains how a beta coefficient is so named. If the standard deviation of the return on the fund or the correlation between the returns on the fund and the market are high, then the fund's beta will be high. The beta of the market index itself is unity. If the fund beta exceeds unity, the fund is more volatile than the market: a beta of 1.1 implies that the fund is 10% more volatile than the market so that if the market rises or falls by 20%, the fund will rise or fall by 22%.

The CAPM can be expressed as follows:

$$\begin{aligned} \textit{Excess return on fund} &= \textit{Alpha} + \textit{Beta of fund} \times \textit{Market risk premium} \\ &= \textit{Alpha} + \textit{Market risk of fund} \end{aligned} \tag{5.4}$$

where the excess return on the fund is the difference between the realized return on the fund and the risk-free rate. The CAPM is illustrated in Figure 5.4.

If the excess return on the fund exceeds the market risk of the fund, then the fund has generated an above-average performance. The difference between the excess return on the fund and the market risk of the fund is called the fund 'alpha' (sometimes it is called the 'Jensen alpha' after its inventor). A successful fund manager therefore generates a positive alpha. However, it is important to recognize that a fund return exceeding the market index return does not necessarily imply a positive alpha. It is possible for a fund to take on a lot of market (i.e. beta) risk and generate a return higher than the market index return, but nevertheless generate a negative alpha: this would indicate that the market risk assumed by the fund manager was not fully rewarded.



Source: Blake (2000: Fig.14.8); Note: M, excess return on the market index (which has a beta of unity), A – positive excess return of fund A, B – negative excess return of fund B.

Figure 5.4 The Capital Asset Pricing Model

This is the case for fund manager B in Figure 5.4: although B generated a return above that of both the market and fund manager A, A is a more successful fund manager.

How has the CAPM been extended?

This is how a single-index benchmark with constant coefficients for alpha and beta operates within the context of the CAPM. A recent development has been to make the beta coefficient of the CAPM time-varying, that is to allow for predictable time-variation in the beta coefficient on the grounds that fund managers should not be credited with using publicly available information concerning changes in investment opportunity sets when making their investment decisions (see Ferson and Schadt (1996); even more recently, Christopherson, Ferson and Glassman (1998) have extended this procedure to allow for time-varying alpha coefficients).

The beta coefficient is made a linear function of a set of predetermined variables: the lagged values of the short-term yield on T-bills, the long-term yield on government bonds and the dividend yield on an equity index such

as the FTA All Share Index; these are all standard regressors with a long tradition in the literature on the predictability of stock returns. So the beta coefficient in this case is determined as follows:

$$\begin{aligned} \textit{Beta of fund} = {} & B(0) + B(1) \times \textit{T-bill yield lagged} \\ & + B(2) \times \textit{Government bond yield lagged} \\ & + B(3) \times \textit{Dividend yield lagged} \end{aligned} \tag{5.5}$$

When Blake and Timmermann (1998) substituted this beta equation into the CAPM equation above and applied it to UK unit trusts over the period 1972–1995, they found it raised the estimate of alpha for the UK balanced sector from −0.74 to −0.52. In other words it lowered the estimate of under-performance slightly for that sector. It made little difference to other sectors, however.

5.3.2 Multiple-index benchmarks

Another recent innovation has been the use of multiple-index benchmarks. For example, Elton *et al.* (1993) pioneered the use of a 'four-index' benchmark consisting of the excess return on large-cap stocks (i.e. a large-cap risk premium), the excess return on small-cap stocks (i.e. the small-cap risk premium), the difference between the returns on an equity growth index and an equity income index (i.e. a growth *minus* income factor) and the excess return on bonds (i.e. a bond risk premium). The multiple-index CAPM therefore becomes:

$$\begin{aligned} \textit{Excess return on fund} = {} & \textit{Alpha} + \textit{Beta}\ (1) \times \textit{Large-cap risk premium} \\ & + Beta(2) \times \textit{Small-cap risk premium} \\ & + Beta(3) \times \textit{(Growth−Income)} \\ & + Beta(4) \times \textit{Bond risk premium} \end{aligned} \tag{5.6}$$

Again, a successful fund manager will generate a positive alpha after taking into account these four factors. In other words, a successful active fund manager will be one who does more than simply buy a portfolio of large-cap stocks, small-cap stocks, growth stocks and corporate bonds.

A variation on this model has been applied to UK unit trusts by Blake and Timmermann (1998). For the UK Equity General sector, for example, they found the following three-index model for the sample period 1972–1995:

$$\begin{aligned} \textit{Excess return on fund} = {} & -0.16 + 0.86 \times \textit{Market risk premium} \\ & + 0.33 \times \textit{Small-cap risk premium} \\ & - 0.07 \times \textit{Bond risk premium} \end{aligned} \tag{5.7}$$

This indicates that after taking into account market risk, small-cap risk and bond risk, a typical unit trust from the UK Equity General sector generated a negative alpha (i.e. underperformed on a risk-adjusted basis) by 16 basis points p.a. on average.

The wider use of multiple-index benchmarks which include small-cap and micro-cap indices might well help to encourage institutional investors to consider their investments in these sectors more carefully since they would now have a specific reference point in the form of a performance benchmark.

5.3.3 Fixed benchmarks

Another possibility is to use a fixed benchmark. This in a sense is what was implied by the long-term financial assumptions of the MFR[9]:

- Rate of inflation – 4% p.a.
- Effective rate of return on gilts – 8% p.a.
- Effective rate of return on equities – pre-MFR pension age – 9% p.a.
- Effective rate of return on equities – post-MFR pension age – 10% p.a.
- Rate of increase of GMP under Limited Revaluation – 5% p.a.
- Rate of statutory revaluation for deferred benefits – 4% p.a.
- Rate of LPI increase in payment – 3.5% p.a.
- Rate of increase in post-1988 GMPs – 2.75% p.a.
- Rate of increase in S148 Orders – 6% p.a.
- The real rate of return on index – linked stocks is I where $(1 + I) = 1.08/1.04$.

The problem with fixed benchmarks is their arbitrary nature. Even if they are based on historical experience, there is no guarantee that they would provide accurate forecasts for the future. For example, the extraordinary performance of the UK stock market over the last quarter century has generated an equity risk premium approaching 10%. It would be highly inappropriate to use this figure to set a benchmark for equities over the next 25 years.

5.4 BENCHMARKS BASED ON LIABILITIES

5.4.1 Liability benchmarks

What are the key liability benchmarks?

The benchmarks considered so far, appropriately adjusted for the relevant universe, are suitable for any institutional investor without matching liabilities,

[9]See 'Current Factors for Use in MFR Valuation' in Guidance Note 27 of the Faculty and Institute of Actuaries, 1998, B27.11–12.

such as a defined contribution pension fund or a unit or investment trust. They are also used in practice by defined benefit pension funds which do have matching liabilities. However, it is important to consider explicit liability-based benchmarks. For example, the liabilities of a final salary pension plan depend on expected earnings growth; they also depend on other factors such as forecasts of life expectancy and the discount rate used for discounting liabilities.

One natural benchmark would therefore be earnings growth. A related benchmark might be GDP growth. Earnings growth and GDP growth are related in the long run, since the share of wages in national income does not trend significantly over time: in fact in long-run dynamic equilibrium, earnings growth and GDP growth will be the same. However, over the course of any business cycle, the growth rates in these two variables can differ substantially.

Another natural benchmark for pension funds would be the growth rate in consumption expenditure, since a pension plan's purpose is to finance consumption expenditure in retirement. Strictly speaking the weights for the consumption expenditure index should reflect the pattern of expenditure by the elderly, which might have a higher weight in medical expenses and a lower weight in foreign holidays, say, than younger more active cohorts of the population. Again in long-run dynamic equilibrium, the growth rates in GDP and consumption expenditure will be the same (otherwise the savings ratio will tend towards either zero or unity).

Why are they easy to beat?

A benchmark based on the growth rate of liabilities would be a fairly easy one to beat, since the returns on funds with a substantial weighting in equities tend to exceed the growth rate of liabilities whether measured by earnings growth, GDP growth or consumption growth. There is a good technical reason why this should be the case: it has to do with what is known as the 'dynamic efficiency' of the economy.[10]

It is possible for economies to accumulate too much productive capital (that is, the plant equipment and machinery used by workers to produce the goods and services that consumers wish to buy). As more capital is accumulated, its return falls: this is because the additional capital is being applied to increasingly marginal and less productive investment opportunities. When there is too much capital, the return falls below the growth rate of the economy. When this happens, the economy is said to be 'dynamically inefficient': everyone in the economy would be better off if there was less saving and investment and more consumption. With less investment, the capital stock falls (as depreciated

[10]See, e.g., Blanchard and Fischer (1989).

capital is not replaced) and the return on capital rises above the growth rate of the economy as measured by the GDP growth rate. When this happens, the economy is in a state of dynamic efficiency.

Most of the key economies in the world have been assessed as being dynamically efficient.[11] This means that, in such economies, the returns on financial assets such as equities (which represent claims on the capital stock) will on average exceed the growth rate of GDP, even though there will inevitably be some years when this does not happen. So a passive strategy of holding a broadly based equity portfolio will generate a return that is likely to exceed wage growth, GDP growth or consumption expenditure growth in most years.

How should future liabilities be discounted?

The discount rate for discounting future liabilities provides another possible benchmark if it is set independently of the return on the assets in the fund. Some asset-liability models use the weighted-average return on the assets in the fund as the discount rate for liabilities: obviously this could not be used as a benchmark. Others use the yield on long-term government or corporate bonds.

The 1995 Pensions Act's MFR norms, for example, used government bond yields to determine the present value of pensions in payment[12]:

> The current gilt yields to be used for valuing pensioner liabilities should be the gross redemption yield on the FT-Actuaries Fixed Interest 15 year Medium Coupon Index or the FT-Actuaries Index-linked Over 5 years (5% inflation) Index, as appropriate. In the case of LPI pension increases, either fixed-interest gilts with 5% pension increases or index-linked gilts with a 0.5% addition to the gross redemption yield should be used, whichever gives the lower value of liabilities. Similar principles should be applied for other pensions which are index-linked but subject to a cap other than 5%.

The justification for using a bond yield is that pensions-in-payment liabilities are less risky than equities and hence should be discounted at a lower yield. On the other hand, pensions-in-payment liabilities are not risk free, and so the discount rate should be higher than that on Treasury bills. This suggests that a bond yield provides an appropriate discount rate. The Faculty and Institute of Actuaries chose the above government bond yields to calculate pensions-in-payment liabilities under the MFR.[13]

---

[11] See Abel *et al.* (1989).

[12] See 'Current Factors for Use in MFR Valuation' in Guidance Note 27 of the Faculty and Institute of Actuaries, 1998, B27.11–12.

[13] The MFR allowed the accruing liabilities of active workers to be discounting using a weighted average of long-run gilt and equity yields, with the weights reflecting the asset mix in the fund.

However, for financial reporting purposes, the Accounting Standards Board requires, in FRS17, that all pension liabilities (including those relating to the accumulating liabilities of active members as well as pensions in payment) are valued using an AA corporate bond yield.[14]

Whichever particular bond yield is used, a fund with a heavy equity component is likely to beat a benchmark based on either government or corporate bond yields in most years, on account of the sizeable positive equity risk premium in the UK financial markets. On the other hand, since equity values are more volatile than those of bonds, there will also be a greater chance of producing periodic deficits in the fund.

Explicit liability benchmarking, although currently not very common, will soon become so for a number of reasons. First, there is the increasing maturity of pension funds: the crystallization of liabilities in terms of a specific stream of pensions-in-payment will inevitably move pension fund asset holdings towards bonds as the natural matching asset. Second, the financial reporting developments just mentioned will introduce a common liability benchmark for all schemes. Third, the replacement of the MFR with a scheme-specific funding standard, as announced by the government in March 2001 and recommended by the Myners Review (2001), will lead to the introduction of scheme-specific liability benchmarks.

5.4.2 Measuring the performance of pension funds using liability-driven performance attribution

'Liability-driven performance attribution' (LDPA) is the name given to the framework for analysing performance measurement and attribution in the case of asset-liability managed (ALM) portfolios, that is, portfolios whose investment strategy is driven by the nature of the investing client's liabilities.[15]

We can illustrate the LDPA framework using the following balance sheet for an asset-liability managed pension fund[16]:

Assets		Liabilities	
Liability-driven assets	A	Pension liabilities	L
General assets	E	Surplus	S

[14]This was the same yield chosen by the equivalent US accounting standard, FAS87.

[15]See Plantinga and van der Meer (1995).

[16]The components of the balance sheet are measured in present value terms. Also for simplicity of exposition, we assume that L relates to accrued past service; thus future contributions are excluded from the balance sheet: actuaries call this the 'accrued benefits method' of valuing pension liabilities.

Suppose that the 'pension liabilities' (L) generate a predetermined set of future cash outflows. The fund manager can meet these cash outflows by investing in fixed-interest bonds (A) with the same pattern of cash flows; these bonds constitute the 'liability-driven assets' (LDAs) in the balance sheet above.[17] Suppose that the pension fund 'surplus' (S) is invested in 'general assets' (E). These can be any assets matching the risk-return preferences expressed by the pension scheme's sponsor (e.g. equities). The surplus is defined as assets $(A + E)$*minus* liabilities (L).[18] The return on the surplus is defined as:

$$r_S S = r_E E + r_A A - r_L L \tag{5.8}$$

where:
r_S = the rate of return on the surplus
r_E = the rate of return on the general assets
r_A = the rate of return on the liability-driven assets
r_L = the payout rate on the liabilities.

Both the pension liabilities and the liability-driven assets will be sensitive to changes in interest rates. Higher interest rates reduce the present value of pension liabilities. Similarly, higher interest rates reduce the value of fixed-interest bonds, since a given stream of fixed-coupon payments is worth less today when yields on alternative assets are higher.[19]

Assuming that interest rate risk is the only source of risk to this portfolio, we can use equation (5.8) to derive a decomposition of portfolio performance as follows. First, we rewrite the return on the general assets as:

$$r_E E = r_E S + r_E (E - S) \tag{5.9}$$

and the return on the liability-driven assets as:

$$r_A A = r_A L + r_A (A - L) \tag{5.10}$$

[17]If the pension liabilities are indexed to uncertain real wage growth or to future inflation then the liability-driven assets will be the assets most likely to match the growth rate in earnings or in inflation over the long term (e.g. indexed bonds, equities and property). But to keep the analysis simple, we assume that the cash flows on future pension payments are known.

[18]Following the 1986 Finance Act, the surplus in UK pension funds cannot exceed 5% of the value of the liabilities. Following the 1995 Pensions Act, the deficit in pension funds cannot exceed 10% of the value of the liabilities and must be reduced to zero within a maximum of ten years.

[19]It is theoretically possible to structure the liability-driven assets in such a way that the pension fund is immunized against interest rate movements. When this happens, the surplus will not respond to interest rate movements. Immunization is explained in Blake (2000: Chap. 14).

Then we can divide each side of (5.8) by S and substitute (5.9) and (5.10) to get the LDPA[20]:

$$\begin{aligned} r_s &= \frac{r_E S + r_E(E-S)}{S} + \frac{r_A L + r_A(A-L)}{S} - r_L \frac{L}{S} \\ &= r_E + \lambda(r_A - r_L) + \gamma(r_E - r_A) \\ &= r_E + \lambda(r_A - \bar{r}_A) + \lambda(\bar{r}_A - r_L) + \gamma(r_E - r_A) \end{aligned} \tag{5.11}$$

or:

Rate of return on the surplus = *Rate of return on the general assets*
+ *Rate of return on the LDAs due to security selection*
+ *Rate of return on the LDAs due to market timing*
+ *Rate of return from a funding mismatch*

where:

$\lambda = \dfrac{L}{S}$ = financial leverage ratio

$\gamma = \dfrac{L-A}{S} = \dfrac{E-S}{S}$ = funding mismatch ratio

$\bar{r}_A$ = the expected return on bonds when they are correctly priced on the basis of the spot yield curve (i.e. when the future coupon payments are discounted using the appropriate spot yields) (see, e.g., Blake (2000: Chap. 5)).

The four-component LDPA in (5.11) can be explained as follows:

1. The *rate of return on general assets* (r_E). This can be analysed using standard techniques, e.g. comparing performance against a pre-agreed peer-group or external benchmark, as outlined in sections 5.2 and 5.3 above.
2. The *rate of return on the liability-driven assets due to stock selection* in terms of, say, credit quality management or sector management. This follows because r_A is the actual return generated by the bonds chosen by the fund manager, whereas $\bar{r}_A$ is the benchmark return on the bonds if

[20]In the case where the surplus is exactly zero, the decomposition in (5.11) is not defined. The fund manager has just generated a sufficient return to meet the payout rate on liabilities. The LDPA in this case would be based on $r_L = r_E(E/L) + r_A(A/L)$ where (E/L) is the portfolio weight in general assets and (A/L) is the portfolio weight in liability-driven assets (see (5.8)).

they were correctly priced according to the spot yield curve: $(r_A - \bar{r}_A)$ is therefore the excess return arising from the stock selection skills of the fund manager.

3. The *rate of return on the liability-driven assets due to market timing*, that is, from choosing a portfolio of bonds with a maturity structure that differs from that of the underlying liabilities, thereby deliberately leaving the portfolio partially exposed to interest rate risk.
4. The *rate of return from a funding mismatch*, that is, from active management of the liability-driven assets such that part of this category is invested in riskier general assets such as equities.

We can illustrate the LDPA using an example. Suppose that a pension fund has the following balance sheet at the start and end of the year:

Assets	Start year	End year	Liabilities	Start year	End year
Liability-driven assets (A)	900	997	Pension liabilities (L)	1,000	1,107
General assets (E)	150	169	Surplus (S)	50	59
	1,050	1,166		1,050	1,166

We will assume that the liability-driven assets are bonds, while the general assets are equities (and that equities have no yield curve effect). The value of the liabilities is calculated as the present value of the liability cash flows using appropriate spot yields as discount rates. We have the following returns on the components of the balance sheet:

Component	Actual rate of return (%)	Benchmark rate of return (%)
Bonds	$r_A = 10.78$	$\bar{r}_A = 10.66$ (assumption)
Equities	$r_E = 12.67$	$\bar{r}_E = 13.30$ (assumption)
Liabilities	$r_L = 10.70$	

The actual rates of return are found by taking the difference between the end-of-year and start-of-year values as a ratio of the start-of-year values. The benchmark return on bonds is calculated in a similar way but based on start- and end-year present values of coupon payments using appropriate spot yields. The benchmark return on equities is simply the realized return on a relevant index, e.g. the FTA All Share Index.

Using equation (5.11) with $\lambda = L/S = 20$ and $\gamma = (L - A)/S = 2$ (using start-of-year values), the LDPA is determined as follows:

	Component	Return (%)
1.	General assets (r_E)	12.67
2.	Security selection ($\lambda(r_A - \bar{r}_A)$)	+2.40
3.	Market timing ($\lambda(\bar{r}_A - r_L)$)	−0.80
4.	Funding mismatch ($\gamma(r_E - r_A)$)	+3.78
	Total	18.05%

The total rate of return on the surplus of 18.05% is made up of 12.67% from the performance of the general assets, 2.40% from successful stock selection of the bond portfolio, 3.78% from a successful funding mismatch, and a loss of 0.80% from market timing. The security selection and market timing effects are magnified by a high leverage ratio (λ) of 20 (the minimum that is permissible since the surplus may not (in the long term) exceed 5% of liabilities), while the funding mismatch effect is magnified by a smaller funding mismatch ratio (γ) of 2. The positive net return of 1.60% from active fund management (i.e. the sum of the returns from security selection and market timing) and the positive net return from a funding mismatch help to generate a high surplus return. However, this cannot conceal the fact that the fund manager underperformed the benchmark in terms of general assets by 0.63%.

The LDPA therefore tells us a great deal about the investment skills of the pension fund manager when he or she is constrained on the liability side of the balance sheet. The only additional information that is required over the current performance measurement framework is as follows: the present value of the pension liabilities (as determined by the pension scheme's actuary), together with the payout rate on these, and the value of the liability-driven assets, together with a customized benchmark return on these.

5.5 WHAT HAPPENS IN OTHER COUNTRIES?

5.5.1 USA

Benchmarking is usually done on an asset class basis against well-known total return indexes. Thus the performance of domestic equity managers is assessed relative to the S&P 500 total return index, fixed-income managers relative to the Lehman aggregate, etc. The other kind of benchmarking is relative to the

average within a peer group. Thus the average of all equity managers who subscribe to Lipper's performance service becomes the benchmark for all the managers in that 'universe'.

5.5.2 Japan

No definite benchmarks have been established yet in Japan. Tentatively, the annual rate of return from the Treasury bond (with a maturity in excess of 10 years) plus 0.1% is used, which is just equivalent to the investment performance from the Fiscal Investment and Loan Program.

5.5.3 Germany

There are four different pension vehicles in Germany.

(1) Direct commitments (book reserves)
Since there are no separate funds, there is no investment choice. Fifty-seven per cent of total occupational pension liabilities in Germany are financed through direct commitments.

(2) Support funds
There are no portfolio restrictions for support funds whatsoever. Instead, investment decisions are made solely by the employer. Therefore, there is either no communicated benchmark at all, or the employer selects the benchmark on a discretionary basis. There are more than 5,000 support funds in Germany but they account for only 8% of total pension assets.

(3) Direct insurance
Currently, the benchmark is 4% p.a. However, there is a public debate about whether this is too high since interest rates are currently low. Therefore, the government is considering lowering the benchmark to 3.5%. There are numerous direct insurance contracts in Germany and they account for 12% of total pension assets.

(4) Pension funds
Pension funds are the only vehicle where having a proper benchmark would make sense. However, pension funds are not required to make detailed information about their investment returns, etc. publicly available. This kind of information need only be disclosed to the regulator. Currently, there are 180 pension funds in Germany and they account for 22% of total pension assets.

5.5.4 Italy

Mixes of well-known indices like JPM bond and MSCI stocks in varying proportions. The exact benchmark of each pension fund is not made public. While it can be requested from the fund, this is a long process.

5.5.5 Chile

The benchmark is the average of the return of the other pension funds (AFPs). The use of market indices has been rejected because the local market benchmarks are of questionable applicability. Pension funds are subject to a number of investment constraints, not taken into account in the existing benchmark, e.g. the weights in the benchmark are changed every quarter but the pension funds invest with a very long horizon.

5.6 CONCLUSION

Performance benchmarks are important for three key reasons: they help to measure the investment performance of institutional fund managers, they provide clients/trustees with a reference point for monitoring that performance and they can also have the effect of modifying the behaviour of fund managers. But benchmarks are not the only factor of importance: fee structures also have a major impact.

At the same time, there needs to be a much greater understanding by clients/trustees of the nature of active fund management. At its simplest, an active portfolio can be interpreted as a passive portfolio plus a set of active side bets against the market. The passive component of the portfolio is the strategic asset allocation and, if the benchmark is set appropriately, the performance of the SAA should exactly match the benchmark. The active components should beat the benchmark if the fund manager's side bets are successful and it should be possible to assess this fairly quickly if financial markets are relatively efficient.[21]

A good benchmark combined with a suitable fee structure would therefore enable an above-average fund manager to deliver, on a systematic basis, superior investment performance without taking on excessive risks. The fact that the evidence indicates that fund managers cannot systematically deliver superior investment performance over extended periods is more an indication of the efficiency of financial markets than of the ineffectiveness of either the benchmark or the particular fee structure.

[21] Although we also showed that the noise generated by changing investment opportunity sets can make it difficult to assess genuine fund management skill over short horizons.

In addition, a good benchmark would be one that did not have built-in biases either in favour of or against particular asset classes. In particular, a dynamic financial system demands that there is no bias against start-up capital, and so a good benchmark would contain the appropriate market weighting in venture capital securities. A good benchmark might therefore be based on a multiple of indices that covers all the key asset categories as well as liabilities. In turn, a good fee structure has an appropriate performance-related element.

There are, of course, unsuitable benchmarks and fee structures. Peer-group benchmarks provide a strong incentive not to underperform the median fund manager, while fee structures based on the value of assets under management do not provide a particularly strong incentive to engage seriously in active fund management. We should not be surprised to find that the outcome is herding around the median fund manager who, in turn, is doing little more than match the index. In other words, this benchmark and fee structure have the effect of modifying the behaviour of the fund manager from that which was agreed with the client/trustee. This is rational behaviour by the fund manager since his long-term survival in the industry depends on his relative performance against other fund managers. But it is certainly not what the client/trustee intended. Similarly, a fee structure that awarded outperformance of a benchmark without penalizing underperformance would lead to the fund manager taking risks with the client/trustee's assets in a way that the client/trustee did not intend. As a final example, the maturing of net investors such as pension funds suggests that scheme-specific benchmarks that reflect the maturity of a particular scheme's liabilities become increasingly appropriate, while, correspondingly, those based on external or peer-group benchmarks become less so.

Benchmarks are important, but so are fee structures. They can either provide the right incentives for fund managers or they can seriously distort their investment behaviour.

5.7 APPENDIX: DERIVING THE POWER FUNCTION

Suppose a fund's monthly excess returns are generated by the equation:

$$R_t = \alpha + \beta R_{mt} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

where R_t is the excess return on the fund in period t, over and above the risk-free rate of return, β is its beta, R_{mt} is the excess return on the market portfolio in period t, ε_t is the residual in period t and α measures the fund's genuine ability to outperform. How long will it take for the trustees to detect

with reasonable statistical reliability whether the fund produces abnormal performance? To answer this question, suppose that $\alpha = -0.1$ and it is known that $\beta = 1$ and $\sigma = 0.5$. For continuously compounded monthly returns data these parameter values correspond to a fund that underperforms the index by 1.2% per year while the idiosyncratic risk is 6% per year. Assuming that the size of the statistical test for the fund manager's ability to add value, p, is the standard 5%, we can illustrate the difficulty of conducting statistical inference about management skills by calculating the power function for a test of the null hypothesis:

$$H_0(\textit{no abnormal performance}): \alpha = 0$$

against the alternative hypothesis:

$$H_1(\textit{abnormal performance}): \alpha \neq 0$$

We do so by computing how many months of data are needed to ensure a 10, 25 or 50% probability of correctly identifying the fund's abnormal performance. The null hypothesis is rejected if:

$$|Z| \equiv \left| \frac{\bar{\alpha} - \alpha_0}{\sigma/\sqrt{n}} \right| > z_{1-p/2}$$

where $\bar{\alpha} = \sum_{t=1}^{n} (R_t - R_{mt})/n$ is the estimated mean performance and α_0 is the value of α under the null hypothesis of zero abnormal performance. $z_{1-p/2}$ is the $(1 - p/2)$ quantile of the distribution of the performance test statistic. The null is rejected if:

$$\bar{\alpha} < \alpha_0 - z_{1-p/2}\sigma/\sqrt{n}$$

or

$$\bar{\alpha} > \alpha_0 + z_{1-p/2}\sigma/\sqrt{n}$$

Otherwise it is accepted. Suppose that, under the alternative hypothesis, the fund manager's performance is α_1, so that $\bar{\alpha} \sim N(\alpha_1, \sigma/\sqrt{n})$. Then the rejection probability can be computed from:

$$\begin{aligned} P(\bar{\alpha} < \alpha_0 - z_{1-p/2}\sigma/\sqrt{n}) &= P\left(\frac{\bar{\alpha} - \alpha_1}{\sigma/\sqrt{n}} < \frac{\alpha_0 - \alpha_1 - z_{1-p/2}\sigma/\sqrt{n}}{\sigma/\sqrt{n}} \right) \\ &= P\left(Z < \frac{\alpha_0 - \alpha_1}{\sigma/\sqrt{n}} - z_{1-p/2} \right) \\ &= \Phi\left(\frac{\alpha_0 - \alpha_1}{\sigma/\sqrt{n}} - z_{1-p/2} \right) \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative density function for a standard normal variate. Likewise, by symmetry of the normal distribution,

$$P(\bar{\alpha} > \alpha_0 + z_{1-p/2}\sigma/\sqrt{n}) = P\left(Z > \frac{\alpha_0 - \alpha_1}{\sigma/\sqrt{n}} + z_{1-p/2}\right)$$
$$= \Phi\left(\frac{\alpha_1 - \alpha_0}{\sigma/\sqrt{n}} + z_{1-p/2}\right)$$

For example, if $p = 0.05$ so that $z_{1-p/2} = 1.96$ and $\alpha_0 = \alpha_1 = 0$, then $P(Z < -2) = P(Z > 2) = 0.025$, so that the power of the test equals the size of the test at 5%.

However, if $\alpha_0 = 0$, $\alpha_1 = -0.1$, $\sigma = 0.5$, we get the following relation between power (the probability of correctly rejecting the null) and sample size:

$$\begin{aligned} P(Reject\ H_0|\alpha_1, \alpha_0, \sigma, n) &= Power(\alpha_1, \alpha_0, \sigma, n) \\ &= \Phi\left(\frac{\alpha_0 - \alpha_1}{\sigma/\sqrt{n}} - z_{1-p/2}\right) \\ &\quad + \Phi\left(\frac{\alpha_1 - \alpha_0}{\sigma/\sqrt{n}} + z_{1-p/2}\right) \\ &= P(Z < -1.96 + 0.2\sqrt{n}) \\ &\quad + P(Z < -1.96 - 0.2\sqrt{n}) \end{aligned}$$

This relationship is used to calculate the results in the main text.

REFERENCES

Abel, A., Mankiw, N.G., Summers, L. and Richard, J. (1989) Assessing dynamic efficiency, *Review of Economic Studies,* **56**, 1–19.

Benjamin, B., Haberman, S., Helowicz, G., Kaye, G. and Wilkie, D. (1987) *Pensions: The Problems of Today and Tomorrow*, Allen and Unwin, London.

Blake, D. (2000) *Financial Market Analysis*, Wiley, Chichester (especially, Chapters 5, 13 and 14).

Blake, D. and Timmermann, A. (1998) Mutual fund performance: evidence from the UK, *European Finance Review*, **2**, 57–77.

Blake, D., Lehmann, B. and Timmermann, A. (1999) Asset allocation dynamics and pension fund performance, *Journal of Business,* **72**, 429–461.

Blake, D., Lehmann, B. and Timmermann, A. (2000) *Performance Clustering and Incentives in the UK Pension Fund Industry*, Pensions Institute, London, June (www.pensions-institute.org).

Blanchard, O.J. and Fischer, S. (1989) *Lectures on Macroeconomics*, MIT Press, Cambridge, MA.

Christopherson, J., Ferson, W. and Glassman, D. (1998) Conditioning manager alphas on economic information, *Review of Financial Studies,* **11**, 111–142.

Coggin, T.D., Fabozzi, F.J. and Rahman, S. (1993) The investment performance of US equity pension fund managers: an empirical investigation, *Journal of Finance*, **48**, 1039–1055.

Elton, E., Gruber, M., Das, S. and Hlavka, M. (1993) Efficiency with costly information: a reinterpretation of evidence from managed portfolios, *Review of Financial Studies*, **6**, 1–22.

Ferson, W. and Schadt, R. (1996) Measuring fund strategy and performance in changing economic conditions, *Journal of Finance*, **51**, 425–462.

Hendricks, D., Patel, J. and Zeckhauser, R. (1993) Hot hands in mutual funds: short-run persistence of relative performance, *Journal of Finance*, **48**, 93–130.

Lakonishok, J., Shleifer, A. and Vishny, R.W. (1992) The structure and performance of the money management industry, *Brookings Papers: Microeconomics*, 339–391.

Myners, P. (2001) *Institutional Investment in the United Kingdom: A Review*, HM Treasury, London, March.

Plantinga, A. and van der Meer, R. (1995) Liability-driven performance attribution, *Geneva Papers on Risk and Insurance*, **20**, 16–29.

Timmermann, A. and Blake, D. (2000) *Determinants of International Portfolio Performance*, Pensions Institute, London, September (www.pensions-institute.org).

Wilson, Rt Hon. Sir H. (chairman) (1980) *Report of the Committee to Review the Functioning of Financial Institutions*, Cmnd. 7937, HMSO, London, June.