Towards evaluating consumer outcomes in the retail investment products market:

A methodology

Prepared for
Financial Services Authority

September 2008

Executive summary

One of the FSA's strategic aims is to 'help retail consumers to achieve a fair deal'.[1] This has particular importance in the retail investment products market, where many products can be both complex and opaque. This, coupled with the inherent stochastic nature of the performance of investment products, poses some significant challenges for retail consumers, as it can be difficult to ascertain the quality of the purchase—not only at the point of the sale, but even after the event. To mitigate potential market failures and consumer detriments, various rules and principles have been put in place to help retail consumers in this market to achieve a fair deal. With a view to understanding further the effect of these measures, the Financial Services Authority (FSA) commissioned Oxera to develop a methodology to assess the extent to which the outcomes for consumers in this market have been *improving over time*.

The methodology has been developed in parallel to a separate project, commissioned by the FSA, to put together a dataset that will enable the empirical analysis to be conducted. Since the data project was ongoing at the time of writing this report, it has not been possible to test the methodology using actual data.

The following summarises the main elements of the methodology developed based on the FSA's project specifications. The methodology defines the relevant consumer outcomes for measurement and sets out the types of analysis to be undertaken using the dataset that will become available, as well as additional data (not currently collected), to enable conclusions to be drawn about improvements in these outcomes over time.

Definition and measurement of relevant consumer outcomes (section 2)

The aim of purchasing investment products is to increase the financial resources available to the consumer at some time in the future. The primary dimensions of a product are therefore the net return it delivers over the holding period, measured after all charges, and the risk associated with that return.

Several FSA regulatory interventions aim to prevent unsuitable sales. However, as understanding suitability requires an examination of individual purchases and circumstances, it was agreed that an examination of suitability lay beyond the scope of this study. Oxera was asked to assume that the products purchased (or advised upon) are suitable in terms of the consumer's risk profile and individual circumstances. The question of interest is whether, among the products that are suitable in the regulatory sense, consumer outcomes have improved over time—ie, the focus is on evaluating improvements in the performance of products that are similar in their risk and other relevant characteristics.

Given the suitability assumption, the main mechanism through which regulation may improve outcomes is via its impact on the product charges that consumers pay. Regulation may reduce charges through specifying requirements on the type of product sold and their charging structures (eg, the introduction of stakeholder products in 2001), enhancing disclosure (eg, the launch of the comparative tables in 2001 or the 'menu' in 2005), or facilitating competition in general.

---

[1] FSA's three strategic aims are: promoting efficient, orderly and fair markets; helping retail consumers achieve a fair deal; improving business capability and effectiveness.

Product charges drive a wedge between the gross returns on the underlying investments and the net returns enjoyed by consumers. However, while higher charges imply a larger wedge between gross and net returns, they do not necessarily imply a lower level of net returns. One of the activities paid for by these charges is the management of the investments. If more or better investment management raises the gross returns by more than the (additional) charges needed to pay for it, the net returns to consumers will be higher, notwithstanding the higher charges.

It is this potential relationship that governs whether higher charges provide a better or worse product (along the return dimension). Investment management can also reduce the risk rather than increase the average gross return. Under these circumstances, higher charges may increase the wedge and lower net returns, but this would be compensated for by a reduction in the risk of the investment.

Combining these two dimensions of value to consumers, the relevant questions to be addressed are therefore:

–   how have the charges paid by consumers evolved over time?
–   what is the relationship between charges and the risk–return performance of products?

In addition to return and risk as the primary dimensions, consumers may value a range of other dimensions or product attributes, including brand, flexibility (eg, with respect to holding period), transparency (on matters such as how the product actually works), consistency (products deliver what they promise to, on a consistent basis), and quality of service.

These dimensions are relevant if consumers value them and are prepared to pay a premium price them—ie, they are relevant if consumers are willing to trade off these attributes against a product's risk–return performance.

While it is possible, at least in principle, to measure whether these additional dimensions have a *price (*eg, whether brand is related to charges or the risk–return performance of products), it is neither practical, nor indeed feasible, to derive the *value* that consumers attach to these dimension(s) within the framework of this analysis. The valuation is likely to be highly idiosyncratic—eg, some consumers may prefer branded products, while others may not. Also, there is no readily available information or metric to describe and classify products according to all of these dimensions.

Net return and risk are the primary dimensions for which a (near) universal value, in terms of 'better' or 'worse' outcomes for consumers, can be established. Other things being equal, consumers would prefer average net returns to be higher rather than lower, and they would prefer lower, rather than higher, risk associated with those returns. When defined along these dimensions, the required analysis of consumer outcomes involves three main steps.

–   An assessment of the development of product charges over time, including analysis of whether average charges have declined or consumer purchases of lower-charge products have increased.
–   An evaluation of the factors that determine product charges and any observed trend in charges, including, in particular, regulation and changes in the regulatory environment.
–   An evaluation of the factors that determine the risk–return performance of retail investment products, including, in particular, analysis of the relationship between charges and performance and whether this relationship is consistent with that established in the literature.

If product charges paid by consumers are found to have fallen over time (due to regulation), then—as long as it can also be established that there is no positive link between charges and gross performance—this can be taken as evidence of improvements in consumer outcomes in the market.

Analysis of trends in charges and their determinants (section 3)

When consumers purchase a retail investment product, they will incur explicit charges, which commonly include:

–   **initial charges**—reflecting the costs of developing and marketing the product, and administrative costs associated with its purchase. They usually also reflect the initial commission paid by the product provider to the intermediary distributing the product. There may also be a one-off charge in terms of an exit fee upon redemption of the product;

–   **annual charges**—the ongoing charges reflect the annual management charges (AMC) as well as other expenses. They may also reflect a trail commission paid to the distributor. The actual total ongoing costs are commonly measured by the total expense ratio (TER).

However, these charges, which are levied and disclosed at provider level, do not necessarily capture what consumers actually pay for the product, for two main reasons:

–   **rebates and discounts**—the distributor of the product may rebate to the consumer (part of) the initial or trail commissions received from the product provider. In addition, the stated charges (initial or AMC) may be discounted by the product provider such that the actual charges are lower than those stated;

–   **additional distribution fees**—these may be levied at the distribution level, adding an extra layer of charges to those levied at the provider level. For example, the consumer may pay a direct fee to the adviser or other intermediary, but may be rebated all or part of the commission; in this case, the rebated commission is not a charge that the consumer pays, but the fee is.

It is the combined impact of rebates/discounts and additional distribution fees that will determine whether consumers end up paying more or less than the charges stated by product providers. However, while the incidence of rebating and discounting has increased in recent years, the significance of this for analysing trends in charges is unclear because a significant proportion of retail transactions is still not affected by rebates or additional fees.

The dataset that is being put together will provide the FSA with historical data on stated charges (AMC and initial charges) and the TER for unit trusts, open-ended investment companies and unit-linked life funds, covering a period of up to ten years (see Appendix 1). The data available also includes the relevant metrics to calculate the risk–return performance of individual funds and other characteristics. However, it does not contain information on any rebates or discounts that may apply to the stated charges or additional fees levied at the distribution level.

The methodology first specifies the hypotheses and empirical tests that can be conducted with the available data—eg, one can test whether stated charges at the provider level have fallen over time. Depending on the findings, it then sets out what additional evidence is available or needs to be collected in order to draw conclusions about the total charges actually paid by consumers.

Evaluating trends in charges over time (section 3.3.1)
The starting point is the empirical analysis of trends in charges using the relevant metrics available from the dataset being created (the stated AMC, initial charges and TER). This involves tracking over time both the average (unweighted and weighted mean, median) and the dispersion around the average (minimum, maximum, inter-quartile range, etc).

While the analysis can be conducted across all funds to give a picture for the market as a whole, it should also be undertaken separately for funds in the same class (eg, using the IMA sector classification) and for those that share similar risk characteristics (see below).

Initial charges (as well as any exit fees) can be annualised using standard holding periods (eg, five years) and combined with the annual charges. Different scenarios for holding periods should be considered, especially if the aim is to evaluate the relationship between the holding period (or flexibility in the choice of holding period) and the total one-off and ongoing charges.

If it is observed that there has been a reduction in charges at the provider level over time, it may be possible to draw the conclusion that the total charges paid by consumers have also fallen over time—provided that charging structures at the distribution level have not changed to offset the reduction in provider charges.

Detailed data gathering and empirical analysis of trends in the distribution of retail investment products were outside the scope of the study. Nonetheless, the report presents some evidence based on publicly available data as well as interviews conducted with providers and intermediaries (section 3.2.2):

–   There have been shifts in distribution channels over time. Over the past ten years, sales via intermediaries (including platforms) have grown overall. Platforms have newly emerged in the market, including, for example, fund supermarkets, which give direct access to consumers and now constitute a significant distribution channel for retail investment products.

–   Rebates and discounts—in particular, of the initial charges—are observed in parts of the market; for example, when the product is bought by the consumer via a financial adviser which uses a fee-based form of remuneration, or where the sale is channelled through a platform (wraps or fund supermarkets). There may, however, be an additional fee to remunerate the intermediary.

While some evidence is available, there is a lack of comprehensive data to measure the evolution of charging structures both within and across distribution channels. If the FSA wishes to quantify the total charges actually paid by consumers and the trends over time, new data would need to be analysed. The report presents suggestions on methods to achieve this (section 3.4).

Evaluating shifts in consumer purchases towards low-charge products (section 3.3.2)
Even if charges for the purchase of specific products or categories of products have not fallen overall, consumers may have become better off (ie, pay less) over time if they have shifted their purchases from high-charge products to otherwise similar products with lower charges.

The dataset that is being created contains the data required to test the relationship between fund inflows and charges (at the provider level) and whether low-charge funds have, over time, experienced a greater inflow than high-charge funds.

In the first instance, the analysis can be based on a simple comparison of fund inflows between low- and high-charge groups of funds, both within classes of similar funds and across classes. It can then be extended to an econometric approach involving a regression of fund inflows as the dependent variable and initial and annual charges as the main independent variables. The regression setting allows the analysis to control for fund size, fund age, investment style, past performance, and other factors that may independently affect inflows. In addition to testing the relationship between charges and fund inflows at any

point in time, the regression can be specified to test whether there has been an increasing trend (if any) in the purchase of lower-charge funds over time.

Similar tests—conducted by academic researchers in the US mutual fund market—have found a negative relationship between fund inflows and initial charges, but no relationship (or, indeed, a positive relationship) between inflows and annual charges (or the TER).

To test shifts in consumer purchases due to differences in total charges, including those at the distribution level, new data would need to be gathered and analysis conducted on:

– trends in rebating and discounting practices, as well as additional distribution fees over time, per distribution channel;
– the shift in consumer purchases across distribution channels, from the more expensive to the cheaper ones.

Evaluating the role of regulation and other determinants of charges (section 3.3.3)
If reductions in charges (or shifts towards lower-charge products) are identified, the relevant hypothesis to be tested is to what extent these trends can be attributed to regulation or changes in regulation.

In principle, the impact of regulation can be assessed using event-study analysis: if a notable reduction in charges is observed at a specific point in time which coincides with the timing of a specific and clearly identifiable regulatory intervention, this may be taken as evidence that the break is due to regulation.

However, the impact of regulatory intervention may be more gradual. In this case, if a downward trend in charges is observed, analysis is required to assess whether it is likely or plausible that any observed reduction in charges is attributable to regulation as opposed to other explanatory factors. While it is not possible to control for the many other factors that might have affected charging structures in the market, methods are available—over and above simply asking the relevant market participants—to test for the importance of the alternative explanations. As described in the report, the main potential alternatives include competition, economies of scale, and technological change.

Analysis of the link between charges and performance (section 4)

Even if there has been a reduction in charges over time (for regulatory or other reasons), outcomes may not necessarily have improved for consumers, given that higher charges may be associated with better-performing products (eg, due to more or better management). What ultimately matter are the net return and the risk associated with that return. Hence, the final part of the empirical analysis is an assessment of the link between performance and charges.

There is a considerable body of literature on the determinants of fund performance, including charges. These studies, many of which are US-focused and apply to equity funds only, tend to conclude that there is no evidence of a positive relationship between charges and fund performance. To the contrary, many studies find a statistically significant negative effect on net performance.

The relevant charges for the empirical tests are those that apply at the provider level since any better (gross) performance is likely to be related to more or better investment management (rather than a particular form of distribution). That is, the analysis involves testing for the link between performance and the stated AMC, or, as is more commonly done in the literature, the TER.

Drawing from the methods proposed in the existing literature, the dataset that is being put together can be used to test the relationship between performance and charges for retail investment products in the UK, and whether this relationship is consistent with that found in the literature. In addition to charges, the dataset contains the data required to calculate monthly returns, as well as data on other characteristics that may affect performance.

Methodologies for risk adjustment (section 4.3)
To test hypotheses regarding the link between charges and performance, performance must be measured so that it reflects both the return and the risk of the different products. This can be done in one of two ways:

–   **risk classification**—products are classified according to their risk. The relationship between returns and charges can then be estimated within each risk class, without further adjustment of the raw returns. The classification can, for example, be based on investment-style classifications (eg, IMA sectors) that may form an ex ante proxy for risk (section 4.3.1). Alternatively, historical returns can be used to estimate the risk and classify products according to the ex post estimated risk (section 4.3.2);

–   **risk adjustment**—the returns for each product are adjusted for its risk based on historical returns data (section 4.3.3). The relationship between risk-adjusted returns and charges can then be estimated across products with different risk characteristics.

Each approach has its attractions, as set out in the report, and there are benefits in using both to cross-check the results and robustness of the conclusions drawn.

The risk of a product is typically measured by the volatility of returns, with the standard deviation being the relevant metric. Risk-adjusted returns can then be calculated as the Sharpe ratio or the variance coefficient. Alternatively, the capital asset pricing model or other factor models can be used to estimate a fund's 'alpha', as the relevant measure of risk-adjusted performance.

However, these measures may not appropriately capture risk if returns are not normally distributed, or if, from a consumer's perspective, downside risk is of particular concern. In this case, an asymmetric measure of risk that captures the likelihood of 'unusually bad' returns may be preferred, such as the semi-deviation, which measures deviations below a certain benchmark rate of return (eg, the risk-free rate). The Sortino ratio can be used as the corresponding measure of risk-adjusted returns. The report discusses several measures of risk and risk-adjusted returns in more detail.

Evaluating the link between charges and performance (section 4.4)
There are two separate empirical questions on the link between charges and performance:

–   Do products with higher stated charges at the time of purchase deliver better performance in the future? The relevant test relates the AMC at the time of purchase to future performance (eg, returns generated over five years following the purchase).

–   Is there a positive contemporaneous relationship between the actual charges incurred and performance? The relevant test relates the TER over a period to the returns generated during the same period.

The relationship estimation is either performed within risk classes (using raw returns) or across risk classes (using risk-adjusted returns). Irrespective of the chosen approach for dealing with risk, the two main empirical methods to estimate the relationship between charges and performance include:

– **ranking**—this entails ranking products according to their (risk-adjusted) returns and, separately, according to their charges. The two rankings can then be compared, also using standard statistical tests, to assess the correlation between the two variables. The hypothesis that higher charges are associated with better-performing products can be rejected if there is no (or a significantly negative) correlation in the rankings;

– **regression**—the econometric approach involves a regression of (risk-adjusted) returns as the dependent variable on charges, while controlling for other factors that may affect performance (eg, fund size). If the coefficient on charges is estimated to be insignificantly different from zero (or significantly negative), this indicates that higher-charge products do not show better performance. While, in principle, more robust than the ranking approach, it can be more complex and requires a range of assumptions to hold for the results to be valid.

The report discusses these empirical methods in more detail. It also sets out additional hypotheses that can be tested in relation to performance, using the dataset that is being created—eg, whether funds with higher turnover show better performance; whether brand is positively related to performance; and whether purchases flow to the better-performing funds.

Finally, the methodology set out here is not the final step needed to enable the FSA to carry out the important task of measuring the full impact of its regime of regulation of retail investment business. The FSA may wish to consider the benefits of collecting relevant data that is not currently available, including to facilitate assessment of changes in transaction quality.

Contents

List of tables

List of figures

1 Introduction

1.1 Content and scope of study

Retail investment products can be both complex and opaque. This, combined with the inherent stochastic nature of investment product performance, makes it difficult for consumers to assess potential outcomes at the point of sale and to ascertain, after the sale, whether they have been able to achieve their desired outcomes. To mitigate potential market failures and consumer detriments associated with these information problems, the Financial Services Authority (FSA) has introduced a series of regulatory initiatives, including various conduct of business rules (such as disclosure requirements) for both product providers and advisers, suitability requirements for advisers, and high-level principles (such as Treating Customers Fairly).

With a view to further understanding how regulations affect consumer outcomes, the FSA commissioned Oxera to develop a methodology to assess whether the outcomes for consumers in the retail investment products market have been improving over time.

An analysis of consumer outcomes in this market would usually include an assessment of the suitability of the products purchased and, if purchased on the basis of advice given, whether this advice was appropriate. However, as understanding suitability would require examining individual purchases and circumstances, it was agreed that an examination of suitability lay beyond the scope of this particular study. In developing the methodology, Oxera was asked by the FSA to assume that the products sold (or advised upon) are suitable for the consumers who purchase them, in terms of the risk profile (or certain other product characteristics) and taking account of their individual circumstances. The question of interest is whether, among the suitable products, consumer outcomes have improved over time—ie, the analysis essentially focuses on evaluating improvements in the performance of products that are similar in their suitability characteristics, particularly with respect to risk or other characteristics.

The aim of purchasing investment products is to increase the financial resources available to the consumer at some future time period. As such, the critical dimensions of a product are the return it delivers after all charges, and the risk associated with that return. Although other dimensions of a product (or how it is sold) will be valued by consumers and are considered in the methodology, the methodology is mainly built around product performance in terms of returns (net of charges) and risk as the primary dimensions of relevant consumer outcomes in the retail investment products market.

When defined in these terms, the required analysis of consumer outcomes involves three main steps.

– An assessment of the development of product charges over time, including an analysis of whether average charges have declined or consumer purchases of lower-charge products have increased.

– An evaluation of the factors that determine product charges and any observed trend in charges, including, in particular, regulation and changes in the regulatory environment.

– An evaluation of the factors that determine the risk–return performance of retail investment products, including, in particular, analysis of whether higher charges are associated with better-performing products.

If it is found that product charges paid by consumers have fallen over time then, as long as it can also be established that there is no positive link between charges and gross performance, this can be taken as evidence of improvements in consumer outcomes in the market.

This methodology project was commissioned by the FSA in parallel to a separate project, conducted by another external consultant, which is devoted to generating the required data on product charges, performance and other product characteristics to which the methodology could be applied. The dataset that is being put together covers unit trusts, open-ended investment companies (OEICS) and unit-linked life funds, and these products define the scope for this study—ie, other products that constitute the retail investment products market, such as retail investment trusts or bank structured funds, are excluded from the scope, largely for data reasons. At the time of completing the methodology, the dataset was still in the process of being created, and it has therefore not been possible to fully test the methodology against actual data.

The dataset captures explicit charges and performance (as well as other product characteristics) at the product provider level, and the initial aim of this study was to develop a methodology that can be applied to this dataset. However, retail investment products can be sold through a variety of distribution channels and in a number of packaged forms.

Each of these distribution channels and product packages carries with it the possibility of different charging levels for what is fundamentally the same underlying product. As a result, the product charges, as stated by providers, are not necessarily indicative of the actual total charges paid by consumers—ie, they may not be a good indicator of the difference that exists between the gross returns on a fund (for which data is directly available) and the net returns actually earned by the consumer after all charges. Hence, basing any analysis on stated charges at the provider level may result in erroneous conclusions being drawn about trends in charges and their impact on net returns.

Therefore, although the initial aim was to develop a methodology that can be applied to the new dataset on provider charges and product performance, the approach taken was a wider one to reflect charges at the distribution level and their interplay with provider charges (eg, through rebates and discounts)—even if data on the latter has not been historically collected to allow a full empirical analysis.

1.2 Approach adopted in developing the methodology

A large body of literature has emerged over the past 20 years on the determinants of investment product performance, including charges. Much of this has related to mutual funds sold in the USA, although there is also evidence of more research being undertaken in the UK. A comprehensive review of this literature was undertaken to draw from the existing findings and to identify the empirical methods adopted by academic researchers to measure product performance and the link between performance, charges and other characteristics.

There may be relevant dimensions, other than performance in terms of net return and risk, which determine the quality of consumer outcomes in the retail investment products market, such as product transparency, flexibility, or whether the product has a recognisable brand name. To identify these other dimensions, a series of meetings were held with representatives of the relevant trade associations, individual firms and FSA experts. These meetings also served to discuss any other issues of relevance to the methodology.

One issue that was commonly highlighted was the difference between the stated product charges and the actual total charges paid by the consumer. To gain a better understanding of the nature and extent of this difference and the potential implications for the methodology, a number of interviews were conducted with product providers, financial advisers and operators of fund platforms (wraps and fund supermarkets).

While it was outside the scope of this study to gather data and build a complete picture on the difference between stated product charges and the actual total charges paid, the interviews confirmed that there could be significant limitations in any empirical analysis using data on charges at the provider level only without taking into account the charging structures for different distribution routes. The methodology was adapted to reflect these considerations: it sets out what analysis can be undertaken with the data available and makes suggestions on what additional data may need to be collected to allow further analysis and to obtain a more robust measure of net returns after all charges.

1.3 Structure of the report

This report summarises the analysis undertaken and methodology developed, and is structured as follows:

–   section 2 provides the conceptual background by defining 'quality' of consumer outcomes in the retail investment products market and how these are measured;

–   section 3 sets out the methodology for the analysis of charges, both of how charges have developed over time and the determinants of charges (including regulation);

–   section 4 describes the methodology for the analysis of the link between product charges and performance;

–   the appendices contain a review of the relevant literature and a description of the dataset that is being put together in the separate project.

2 Consumer outcomes: definition and measurement

When making an investment, consumers are faced with a wide variety of products. These products will have both an expected outcome (or at least an expected probability of outcomes) and, with the passage of time, an actual outcome, which may be within or outside the expected range. The same underlying product may be available through different distribution channels, which may cost the consumer different amounts—hence, the same product may be available at different prices. In addition, the consumer may have chosen the product (and distribution channel) on their own, or they may have advice or help with their choice (which may have a cost to it, which will again change the overall price). Finally, the consumer may, or may not, have a good understanding of the probabilities of different outcomes and the other characteristics of the product they (are about to) buy.

This degree of the complexity, combined with the possibility that the consumer does not have a good understanding of the characteristics of the products potentially available to them, means that defining 'better' and 'worse' outcomes for the consumer is not straightforward.

As with many products, consumers will also value the different characteristics of products differently. Thus, even if the dimensions of better or worse could be fully defined and measured, there would be no guarantee that all consumers would attach the same ranking, let alone apply the same value to, the range of products on offer.

However, there are two key dimensions of retail investment products for which a (near) universal value (in the form of better or worse) can be applied:

– the net return on the investment over a given holding period;
– the risk associated with that return.

Other things being equal, consumers would prefer returns to be higher rather than lower, and they would prefer lower, rather than higher, risk associated with those returns.

This section further explains these two primary dimensions which determine the quality of consumer outcomes in the retail investment products market. It also sets out a range of secondary dimensions—such as brand, flexibility and transparency—and how these can be incorporated in the analysis.

2.1 Defining relevant consumer outcomes

2.1.1 Net return and variability of return (risk) as the primary dimensions
The retail investment products that form the focus of this study are vehicles for saving and investment. The main purpose of purchasing such a product is therefore to increase the financial resources available to the consumer at some future time period. As such, a critical dimension of the product is the level of the future financial resource that will actually be delivered as a result of purchasing the product now.

In purchasing the product, the consumer will have to pay a price now, in cash (or equivalents), and will receive a return in the future, in cash (or equivalents). Other things being equal, a consumer will be unambiguously better off if the total cash returned to the consumer at the future date is higher rather than lower.

This relationship between the total outlay paid by the consumer upon purchase and the total returned to the consumer at the future date forms one main dimension of the relevant

consumer outcomes. The relevant metric can be calculated as the net return to the consumer, expressed as a percentage. In the framework of analysis, it is assumed that, other things being equal, a higher net return is better than a lower net return.

However, the actual return that a consumer will earn on a retail investment product is unknown at the time the purchase is made. The actual return is unknown both absolutely (ie, what return will actually be delivered by product A?) and relatively (ie, which of product A or product B will actually deliver a higher return?). All that can be known at the time of purchase is an average expected return and an expected probability that the actual return is higher or lower than the expected average—this is captured by the notion of risk. It is assumed that the consumers of these products unambiguously prefer low-risk products to high-risk products, other things being equal.

Overall, therefore, the main measures of relevant consumer outcomes in the market for retail investment products are taken to be the net returns and the risk associated with those returns.

Different products will have different risk profiles, rendering a comparison between products difficult. In particular, if different consumers value the probability of different outcomes differently—eg, some consumers may be more wiling than others to accept a greater downside risk for a higher possible upside return—the relative riskiness of products will depend on the preferences of the individual consumer. As a result, even when consumers all prefer higher returns over lower returns, and lower risk over higher risk, they may trade off risk and return differently. As a result, there is no universal method of ranking products with different risk profiles. The most that can be universally agreed is that if two products have exactly the same risk profile, but one has a higher average return than the other, the one with the higher return is better.

This study focuses on net returns, and, in particular, the charges that drive a wedge between gross and net returns of products with similar risk profiles. As further discussed in section 4.3, several different measures of risk, and ways of adjusting returns for risk, are proposed for the empirical analysis.

2.1.2 Other dimensions of consumer outcomes

Although net return and risk are the primary dimensions along which products deliver better or worse outcomes for consumers, there are other dimensions that consumers may value. As identified also in discussions with the industry, these other dimensions include:

–   brand—consumers may value products or providers which have well-recognised household names;

–   flexibility—consumers may value products which are flexible in terms of redemption or holding periods (eg, holding the product for a shorter period of time than expected at purchase);

–   transparency—consumers may value products with better information or disclosure of relevant aspects, such as how the product actually works or the associated risks;

–   consistency—consumers may value products which deliver what they promise to deliver on a consistent basis;

–   quality of service—consumers may value products or providers which have a high standard of service (eg, in terms of timing of redemption payments, the responsiveness of the provider to queries, or, indeed, the quality or amount of information provided).

In the framework of analysis, these dimensions are considered of secondary importance to net return and risk—after all, the aim of purchasing an investment product is to secure a

return that is consistent with the risk assumed. Nonetheless, the dimensions are relevant if consumers value them and are prepared to pay a premium price (or accept a lower net return) for products with particular attributes—ie, consumers may be willing to trade these attributes off against the product's risk–return performance.

The valuation of these dimensions is likely to be even more idiosyncratic than that of return and risk. For example, some consumers may prefer to buy brand names,[2] whereas, for others, brand is irrelevant; some consumers may be willing to pay a premium price to receive product documentation of a particular form or with detailed content, whereas others derive no value from such documentation; and so on. Therefore, it is not practicable to seek to *value* these dimensions so as to be able to derive an overall ranking of the products, taking into account differences in these dimensions.

Nonetheless, within the methodology proposed, it is possible, at least in principle, to measure whether dimensions such as brand have a *price*. Whether this price (if there is one) is above or below the value that consumers place on this dimension would be idiosyncratic to each individual consumer.

For example, in order to test for the price of brand, one can rank products or their providers according to their brand perception. Products from providers with a well-known brand can then be compared with products from those with more limited brand recognition. A significant difference in the level of charges (or net returns) between the two groups would indicate, other things being equal, the premium price of brand (if any). The empirical tests in relation to brand are further discussed in sections 3 and 4.

Testing for the importance of the other dimensions of consumer outcome is also possible, at least in principle. However, compared with a classification of products or providers according to brand, it would seem more difficult, if not impossible, to derive a classification according to the transparency, consistency or quality-of-service dimensions unless extensive primary research is undertaken.[3]

For these reasons, the methodology is built around net return and risk as the primary dimensions for which a (near) universal value, in terms of better or worse outcomes for consumers, can be established.

Moreover, the main question of interest is whether consumer outcomes have been improving over time and the role of regulation in this regard. Even if the other dimensions have a price or value that could be measured, this would not alter the conclusions that can be drawn from the analysis set out in sections 3 and 4 of this report—unless there is evidence that the price or consumer valuation of these other dimensions has changed over time in a way that offsets any trends that are (or are not) identified along the net return and risk dimensions.

2.2   The suitability assumption

The methodology adopted does not seek to evaluate the quality of advice or suitability of consumer purchase decisions. The differing consumer valuations of different characteristics of retail investment products means that many, if not most, products on the market cannot be said to be universally suitable, or universally unsuitable, for all consumers.

The question of whether a product is 'suitable' (in the regulatory sense) is therefore not taken into consideration in this study. Rather, the analysis assumes that all products purchased are

---

[2] For example, some consumers may be willing to pay a premium for branded products because they believe (correctly or otherwise) that such products provide higher net returns or a lower risk profile (including the possibility that, in the case of a very negative outcome, the potential reputational damage to the brand may result in some kind of compensation).

[3] As regards the flexibility dimension, the empirical analysis can be conducted to establish the extent to which total charges vary with respect to the holding period. This is briefly discussed in section 3.3.1.

suitable—ie, the analysis essentially focuses on evaluating performance (and improvements in the performance) of products that are similar in their risk or other characteristics, irrespective of whether or not the particular purchases that underlie the data are suitable (in the regulatory sense).

There might, however, be a relationship between the suitability of the product for a particular consumer and its risk-adjusted net returns. Indeed, if a class of product showed a consistent pattern of delivering significantly lower average returns and significantly higher riskiness, there may be few, if any, consumers for whom these products would be suitable. The analysis set out here, when applied to the data, may reveal such a class of products. However, in general, because product suitability is idiosyncratic to a consumer's valuation of the different characteristics of the product, it is not possible to draw any conclusions about the suitability of a particular product without access to that consumer's valuations. Therefore, for this analysis, an assumption is made that the products are suitable for at least some consumers.

The question of interest is whether, among the suitable products available on the market, the outcomes of consumer purchases of these products have improved over time, and, if so, whether any such improvements can be attributed to regulation as opposed to other factors.

Given the suitability assumption, the main mechanism through which consumer outcomes in the market are likely to improve as a result of regulation is via product charges—ie, via reductions in charges or increased consumer purchases of lower-charge products. If it is found that the charges paid by consumers for similar products have fallen over time, or that consumers are increasingly purchasing those products with lower charges, then—as long as it can also be established that there is no link between charges and the gross risk–return performance of the product—this can be taken as evidence of improvements in consumer outcomes in the market (as a result of regulation or otherwise). This, in short, defines the structure of the methodology and empirical tests set out in sections 3 and 4.

2.3 Defining the main variables for measurement

The following sets out the definitions of net returns and risk for measurement purposes, focusing on aspects that are relevant at the conceptual level; sections 3 and 4 discuss these in more detail for empirical measurement purposes.

2.3.1 Net returns

Unlike most other products, investment products have both a price and output that are measured in the same units—money. There is therefore a direct relationship between the price paid by consumers for the product and its net performance, in terms of the returns it delivers if the net return is measured from the point at which the consumer first decides to make a purchase with cash (or its equivalents) and the point at which the product is turned back into cash (or its equivalents).

The operational definition of net returns adopted in this study captures the following:

– the initial outlay, which includes the price of both the product itself and obtaining the product, is the total net amount of money handed over by the consumer to the retailer of the product, including any additional monies handed over to other intermediaries who are a necessary part of that purchase process;

– the total returned to the consumer is defined as the total amount of money returned to the consumer that can be turned into cash (or its equivalents);

– the net return is calculated as the percentage difference between these flows (ie, money out and money in), over the relevant time period.

Figure 2.1 illustrates one type of typical flow which, if it were over a year, would result in a net return of 5%. Details on the relevant charges in the process are discussed below.

Figure 2.1   Illustration of net returns to consumers
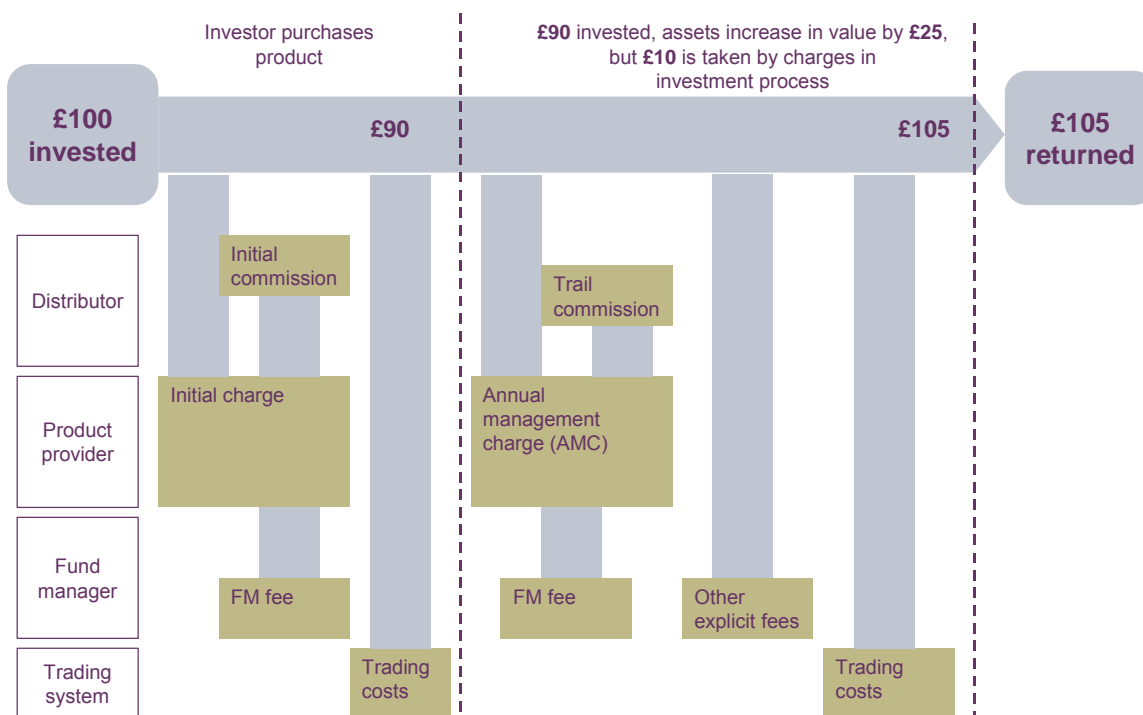
Source: Oxera.

As is illustrated in the figure, although the concept of the net return is reasonably simple, calculating it may not be straightforward. In particular, the presence of different layers of charges, discounts, rebates and other forms of activity mean that the apparent outlay required by the consumer may not be the same as the actual outlay, and the apparent return may not be the same as the actual amount of money the consumer sees being returned.

2.3.2   Risk, or variability, of (net) returns

The returns to a retail investment product are uncertain, and will vary through time. In addition, different products will deliver different returns over the same time period, and the relative performance of different products will not be consistent through time. Some measure(s) of the risk of individual products is needed to enable comparisons to be made, and for the construction of risk-adjusted returns to compare products with different risk profiles. Two general approaches that have been adopted include:

–   a priori classification of products with the same/similar risks;

–   measurement of the variability of the returns of a specific product through time in order to establish its specific risk.

Either approach creates a risk category or measure for any particular investment product. Section 4 provides more detail on how these two approaches can be put into practice, and a number of variants that might be used to meet different analytical purposes.

2.3.3   Charges paid for by consumers

The definition of the net return to consumers can be calculated purely on the basis of the money the consumer pays out upon purchase and the money the consumer gets back when the product is cashed in. How that return is generated, and the uses to which the consumer's

money is put between when they hand it over and when they get it back, does not necessarily alter the net return. However, there is a clear relationship (even if complex) between what the money is used for and the net return enjoyed by the consumer. In particular, the more money that is used up in charges or fees to those providing or distributing the product, the higher the return on the part of the money actually used to purchase the underlying, return-generating investments must be in order to deliver a given rate of net return.

Product providers, advisers and other intermediaries need to be remunerated for the costs they incur in producing, administering and distributing the product. This drives a wedge between the return on the underlying investments and the net return enjoyed by the consumer. In value terms, this wedge corresponds to the sum of the charges (explicit or implicit) that are incurred for the various services that lie between the consumer and the gross return on the underlying investment.

There are many different charges that could be incurred during the process and are paid for directly by the consumer. Some of the explicit charges may be rebated or discounted such that the level of charges, as stated, may not be the level of charges actually paid. In this framework of analysis, charges that are rebated to the consumer do not form part of the wedge between gross and net returns, and so should not be included in any calculation of the 'price' paid by consumers.

While higher charges imply a larger wedge between the gross returns of the investment and the net returns earned by consumers, they do not necessarily imply a lower level of net returns. One of the activities paid for by these charges is the management of the investments. It is therefore possible that there is a positive link between the size of the wedge and the level of gross returns. If more or better investment management raises the gross returns by more than the charges needed to pay for it, the net returns to consumers will be higher, notwithstanding a larger wedge.

It is this potential relationship that governs whether or not higher charges (ie, a larger wedge) provide a better product (along this dimension). It is also possible that more investment management could reduce the risk rather than increase the average gross return. Under these circumstances, a higher level of charges would increase the wedge, and lower the net return, but this would be compensated for by a reduction in the risk of the investment.

Combining these two dimensions of value to the consumer, the relevant questions of interest are: how have charges evolved over time, and what is the relationship between charges and the risk–return performance of products. What these charges actually are, and how to measure them, is considered in more detail in section 3, which also sets out the approach that can be adopted to examine how charges have evolved over time and the determinants of charges, including regulation in particular. Section 4 describes in more detail the approaches to examine the relationship between charges and the risk–return performance.

3 Analysis of product charges

This section presents the methodology to address two main questions of interest: have the product charges paid by consumers fallen over time, and, if so, can any such reductions be attributed to regulation as opposed to other factors? It begins by setting out the relevant charges paid by consumers when purchasing retail investment products. It then explains the required and available data as well as the empirical tests to be conducted in order to examine the main questions of interest.

3.1 Relevant charges for consumers

When consumers purchase a retail investment product, they will incur explicit charges. As disclosed by product providers, these commonly include an up-front charge paid at the time of purchase in addition to ongoing charges that are incurred over the period that the product is held.

– **Initial charges**—these reflect the costs of developing and marketing the product, as well as administrative costs associated with its purchase.[4] They may also reflect the initial commission paid by the product provider to the intermediary distributing the product. Although less frequent, there may also be a one-off charge in terms of an exit fee upon redemption of the product.

– **Annual charges**—the ongoing charges reflect the annual management charges (AMC), as well as other administrative expenses and explicit trading costs[5] associated with the assets underlying the retail investment product. They may also reflect a trail commission paid to the distributor. The actual total ongoing costs are commonly measured by the total expense ratio (TER).

However, these charges, which are levied and disclosed at provider level, do not necessarily capture what consumers actually pay for the product, for two main reasons:

– **rebates and discounts**—the distributor of the product may rebate to the consumer (part of) the initial or trail commissions received from the product provider. In addition, the stated charges (initial or AMC) may be discounted by the product provider such that the actual charges are lower than those stated;

– **additional distribution fees**—these may be levied at the distribution level, adding an extra layer of charges to those levied at the provider level. For example, the consumer may pay a direct fee to the adviser or other intermediary, but gets rebated all or part of the commission; in this case, the rebated commission is not a charge the consumer pays, but the fee is.

It is therefore the combination of actual charges at both provider and distribution level that determine the total charges paid by consumers. Figure 3.1 below presents a stylised illustration of the layers of charges that reduce yields and effectively drive a wedge between the gross returns on the investment and the net returns actually earned by consumers.

---

[4] The initial charge may take the form of a bid–offer spread.

[5] Implicit trading costs that depend on the liquidity of the underlying assets and are commonly measured by effective bid–ask spreads of the underlying assets are not further discussed. The focus is on explicit charges that drive a wedge between gross and net returns—in particular, when the gross return is already measured after any implicit trading costs have been taken into account.

Figure 3.1   Illustration of the layers of relevant explicit charges (or the wedge between gross and net returns)
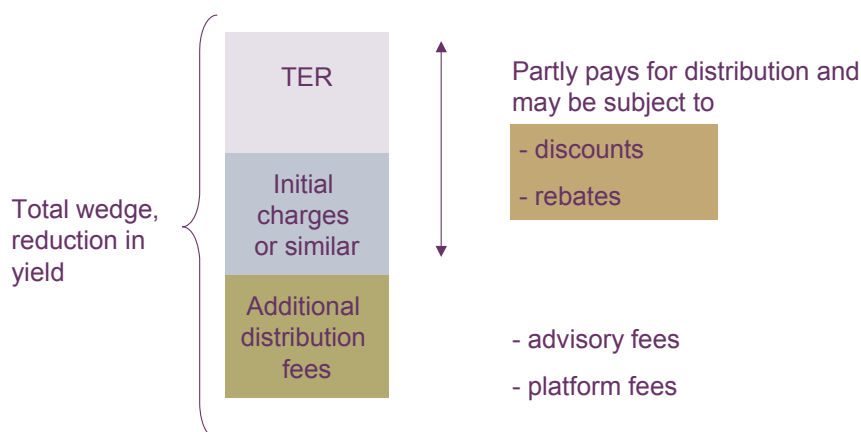
Source: Oxera.

As the total charges can vary by distribution channel, consumers may incur different charges on the same product, depending on how they purchase the product.

As regards intermediated sales, there are various types of intermediaries with different charging structures. Table 3.1 summarises how these intermediaries may be remunerated for their services and how they may employ rebates or discounts for the initial charge or AMC levied at the provider level. The table is a stylised illustration only, drawing from interviews with product providers and intermediaries, as well as other sources.

Table 3.1   Illustration of charging structures by distribution channel

Distributed via	Main remuneration method	Rebates/discounts offered on provider charges
Independent financial adviser (IFA)	Two possible methods: 1. product provider pays commission to IFA (commission-based advice) 2. consumer pays a fee to IFA (fee-based advice)	Initial commission may be rebated to consumer—in particular, where the IFA is remunerated on a fee basis. Rebating may also apply to the trail commission IFA may negotiate a discount on the initial charge or AMC
IFA with wrap	In addition to the remuneration for the IFA (as above), there may be a commission payment from the product provider to the wrap platform provider or an additional fee paid by the consumer for the wrap	Rebating and discounts are common
Fund supermarket	Commission paid by product provider. Possible additional fee paid by consumer	Rebates and discounts are common
Stockbroker, execution-only broker, etc	Fee paid by consumer	Rebates and discounts are common
Direct by product provider	Paid via initial charge and AMC	Often no rebates or discounts. Provider keeps commission

Source: Oxera, based on various sources, including interviews with providers and distributors.

For example, where the consumer purchases a product via an IFA, the IFA may rebate a proportion of the commission it has received from the product provider, in particular where the IFA is remunerated on a fee basis directly from the consumer. Instead of cash being

returned to the consumer, the rebate may take the form of a purchase of additional units of the product at no additional cost to the consumer. It may also take the form of a direct reduction or discount in the charge the consumer has to pay on purchasing the product. In addition to a rebate on the initial charge, it may be that rebates or discounts are offered on ongoing management charges. These rebates and discounts may be (partly or fully) offset if the consumer reimburses the IFA by paying a direct fee on the services provided.

While distribution via IFAs or directly by the provider used to be the main route for consumers to purchase products, there have been changes over the past few years, in particular with the emergence of provider platforms and discount brokers (or execution-only brokers) which serve the retail markets.

Provider platforms are services used by intermediaries (and sometimes consumers directly) to view and administer investment portfolios; they include wraps and fund supermarkets.[6] Wraps allow advisers and consumers to manage portfolios of investments online; they also allow product providers and advisers to outsource certain administrative duties. Unlike wraps, fund supermarkets are online services that sell products directly to consumers.

There appears to be considerable variation in the way platforms are remunerated. Two main methods are used:

- some platforms charge an explicit fee to the consumer, and any product discounts or rebates received are then passed on;
- other platforms receive a share of product charges negotiated with the product provider.

Without the presence of a platform, resources would be spent by the product provider and adviser to undertake the administrative tasks. With a platform, these resources are spent at the platform level and reimbursed accordingly, which can mean that there is no change to the actual price of the investment. For example, without the platform, the AMC of, say, 1.5% may be shared equally by product provider and adviser, with the provider keeping 75 basis points (bp) and paying 75bp in commission to the adviser. Where a platform is used, the product provider may pay 25bp to the platform as remuneration for the outsourced administrative services and keep 50bp. In this case, there is merely a redistribution within the investment value chain, but the total price to the consumer is unchanged. However, there may be an additional fee for the use of the platform, which in turn would increase total charges. Alternatively, efficiency gains from employing a platform may free up resources that could be returned to the investor in the form of further rebates or discounts, thereby reducing total charges.

Like wraps, fund supermarkets and discount brokers selling directly to consumers in the retail market may be able to buy products on a wholesale basis from the product provider, reducing the sales costs of the provider and thus gaining a discount on the purchase. They may then sell these products on to consumers at a higher rate than that charged by the product provider, but pass some of the saving on to consumers in the form of rebates or discounts.

Fund supermarkets or discount brokers may purchase products directly from the product provider, often at the creation price without initial charge. The purchase is treated as an aggregate product which is then separated and sold on an individual basis to the consumers of the supermarket or discount broker. The account records of the product provider may therefore show a single sale of a product rather than multiple sales to the end-consumers. This will also have an impact on the manner in which the AMC is charged. In this case, the wrap pays a single AMC to the product provider, while at the same time it may charge a different AMC to each of its customers.

---

[6] For a definition and further explanations, see FSA (2007), 'Platforms: the role of wraps and fund supermarkets', Discussion Paper 07/2, June.

It is not only the distribution channel that can affect the charging structure and level of total charges paid by a consumer. The way in which a product is packaged may also have an effect.

– Tax wrappers—consumers can purchase retail investment products as part of tax-efficient wrappers, such as ISAs or stakeholder pensions. While charges arise for these wrappers, the stated charges of the underlying funds may be heavily discounted or may not actually be paid (eg, no initial charge may be payable).

– Multi-manager funds—in addition to tax wrappers, there has been an increase in sales, directly to retail investors, of multi-manager funds. Multi-manager strategies enable a retail investor to invest in a range of funds and asset classes while delegating the administration and choice of fund selection to another manager. There are two principal types of multi-manager funds:

– fund of funds (FoF), in which the investor invests in a fund that itself holds a portfolio of units in mutual funds (unit trusts or OEICs), private equity funds, hedge funds, or investment trusts;
– manager of managers (MoM), in which the investor invests in a fund that uses the capital to create separate mandates for other fund managers to manage.

In terms of charges, an investor in a multi-manager fund faces a similar charging structure to investing in a mutual fund, but these charges go towards remunerating the multi-manager fund operator as well as paying for the charges of the underlying funds (in the case of an FoF) or the management fees (in the case of an MoM fund). However, at least in the case of an FoF, the price at which the fund can buy underlying fund units may well be lower than the price an investor would face if buying units in the underlying funds directly.

Overall, therefore, product charges levied at the provider level are a key component of the price consumers pay for investing in retail investment products. However, the charges that are explicitly stated by product providers and others may not actually reflect what consumers pay—total charges, or the total wedge between gross and net returns, may be lower (due to rebates and discounts) or higher (due to additional fees levied at the distribution level). To the extent that charging structures at the distribution level differ, consumers may pay less or more for the same product, depending on the chosen distribution channel. Total charges for the same underlying product may also differ, depending on how the products are packaged (eg, wrapped or not).

3.2   Available data on charges

The above suggests that an analysis of total charges paid by consumers would need to capture charges at both the provider level and the distribution level, distinguishing between different distribution channels for different types of investment product. However, as discussed next, there are limitations in the data available, which has implications for the actual empirical analysis that can be undertaken.

3.2.1   Lack of data on charging structures at the distribution level
In parallel to this study, the FSA has commissioned a separate project to create a dataset using historical data on product charges, performance and other characteristics available from third-party data vendors. The available data refers to unit trusts, OEICs and unit-linked life funds. The data on charges contains one-off and ongoing charges stated at the provider level—ie, initial charges,[7] AMC, as well as the TER as a measure of total expenses charged

---

[7] The initial charge may take the form of a bid–offer spread.

against the fund and, where applicable, exit charges. Appendix 1 provides further details on the dataset.

As summarised in Figure 3.2, the dataset that is being put together does not contain information on any rebates or discounts that may apply to the stated product charges; nor does it contain information about any possible additional fees levied at the distribution level.

Figure 3.2   Summary of data limitations
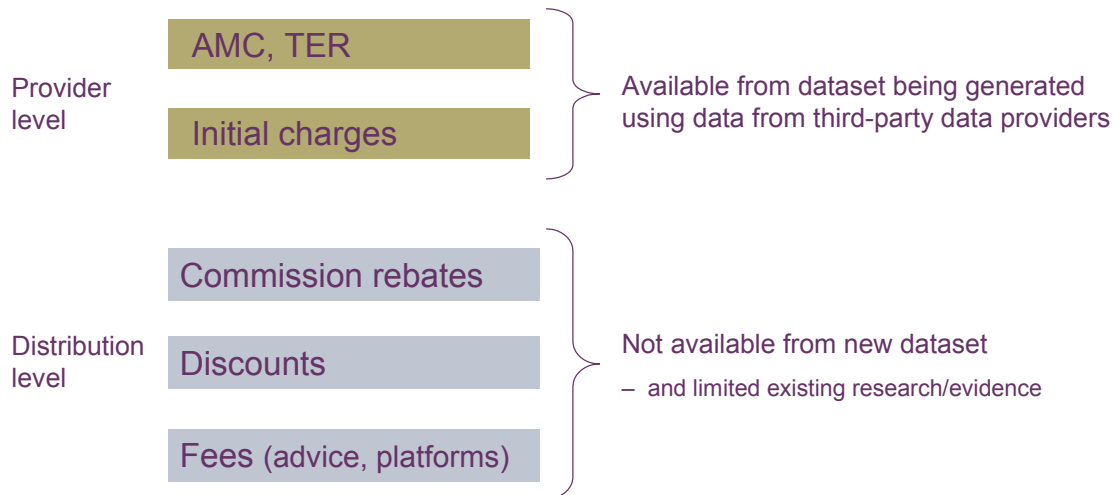
Source: Oxera.

Based on the new dataset, the empirical analysis is thus restricted to stated charges at the provider level, and the methodology developed in this study has to take account of these data limitations. Data gathering and empirical analysis was beyond the scope of this study. Nonetheless, to develop the methodology, it was important to gain a better understanding of the significance of the lack of data at the distribution level, and, if required, to advise the FSA on any additional data that may need to be gathered to allow empirical testing and conclusions to be drawn about the main questions of interest.

Therefore, a number of interviews with product providers and intermediaries were conducted to obtain at least some evidence on charging structures at the distribution level—in particular, in relation to the importance of rebating and discounting. Also, any available evidence on distribution in the retail investment products market was reviewed to assess whether any data is available on charging structures for different distribution channels.

A summary of the evidence obtained is presented below, followed by a discussion of the implications for the methodology and its empirical application.

3.2.2   Evidence from interviews and other sources

The following summarises some of the evidence available to assess both the shifts in distribution channels and the charging structures for different distribution channels, as gathered from public sources as well as from the discussions with industry representatives. While more is available from public sources, the evidence presented has been selected to illustrate the main points only:

–   there have been significant shifts in product purchases between distribution channels over time, which is relevant given the difference in charging structures between channels;
–   rebating or discounting is observed and is common in at least some parts of the market, in which case stated product charges do not reflect what consumers actually paid to purchase the product.

Evidence of shifts in sales between distribution channels

Publicly available data shows that the distribution channels for retail investment products have changed significantly over the last ten years. For example, Figure 3.3 shows the distribution channels for retail sales of UK domiciled unit trusts and OEICs between 1998 and 2007. Sales via intermediaries (including platforms), as classified in the data, have grown significantly over time to capture 85% of the market in 2007.

Figure 3.3   Growth and predominance of intermediated sales
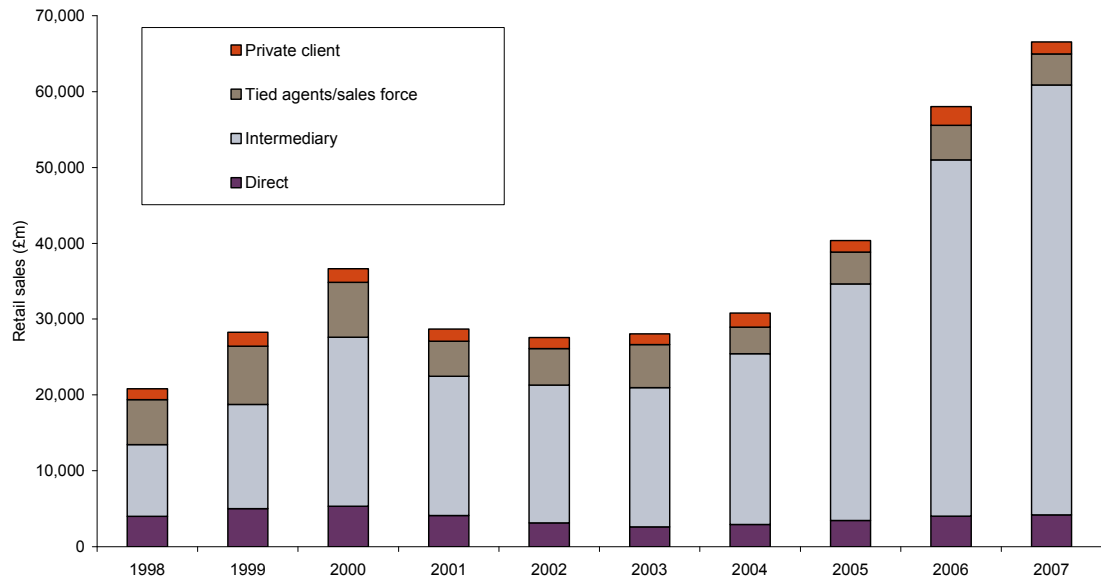
Source: Investment Management Association (IMA).

In particular, fund supermarkets have substantially increased their market share, and one of the reasons for the growth of the intermediary channel at the expense of other channels is the growth of fund supermarkets, which are counted within intermediary sales in Figure 3.3.

Figure 3.4 below can illustrate more clearly the growing importance of fund supermarkets in the gross sales of ISAs. The proportion of sales of UK domiciled unit trusts and OEICs wrapped in an ISA which cannot be allocated to any of the traditional sales routes had grown from 0% in 2001 to 39% by 2007—this mainly includes sales via fund supermarkets.

Figure 3.4   Growth of ISA sales via fund supermarkets

Sales of ISA UK unit trusts and OEICs

16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0

Legend:
- Private client
- Tied agents/sales force
- Intermediary
- Direct
- Unallocated (includes fund supermarkets)

1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
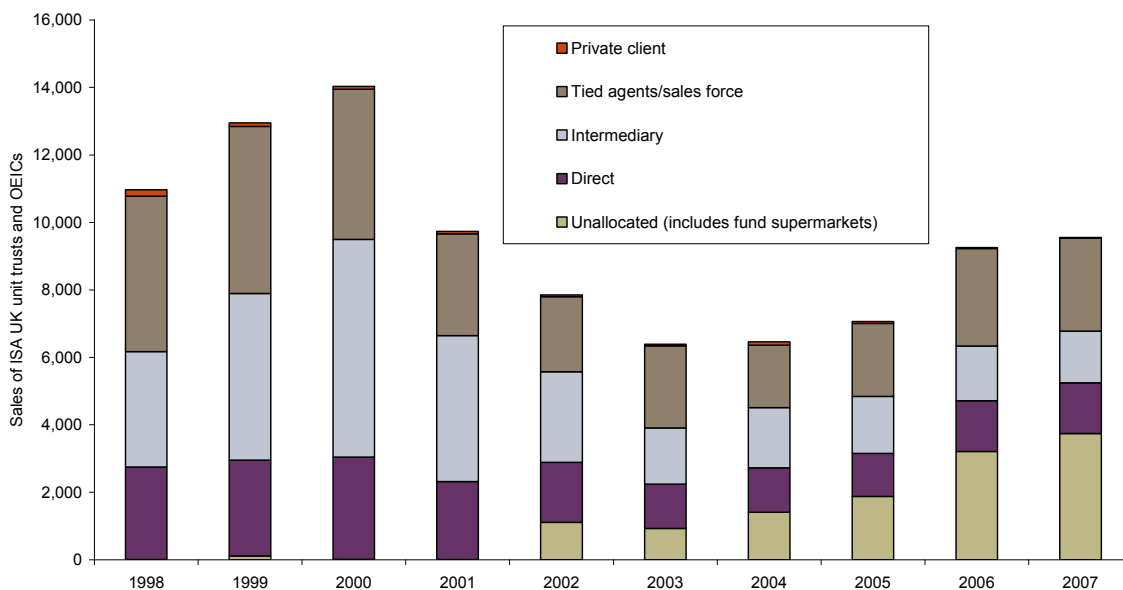
Source: IMA.

Although these results focus on particular segments of the overall retail investment market only, they provide an indication of the changing importance of different distribution channels over time, with the emergence of new distribution channels, such as fund supermarkets.

Discussions with industry representatives confirmed these findings. For example, a product provider indicated that now less than 4% of its gross sales are made directly to consumers. As regards provider platforms (fund supermarkets as well as wraps), interviewees noted the considerable growth. For example, IFAs indicated that they now commonly use wrap platforms, and some firms noted that platform services now constitute a significant part of their overall business.

To the extent that stated product charges (initial and AMC) are more likely to be paid (without rebate or discount) when the product is purchased directly from the provider, as opposed to an intermediary, the shift away from direct sales implies that the proportion of consumer purchases subject to the full stated charges may have fallen over time. Put differently, the growth of intermediated sales makes it less likely that stated product charges at the provider level reflect the total charges actually paid by consumers. Instead, these shifts—in particular, the growth of platforms—give increased scope for rebating or discounting of these charges by either the provider or the intermediary (but with the possibility of additional fees being levied at the distribution level).

Evidence on rebating and discounting practices

Although it was not possible to gather comprehensive data on charging structures at the distribution level, publicly available evidence, as well as views gathered from interviews with industry representatives, provide an indication of rebating and discounting of stated product charges. There appears to be considerable variation not only across, but also within distribution channels.

Interviewees noted that for a typical product the stated initial charge would be about 4–5% of the purchase price, and the AMC about 1.5%. However, all interviewees agreed that, in particular, the initial charges are commonly rebated or discounted—either by the intermediary in the form of a rebate of the initial commission received, or by the provider in the form of a discount to the stated initial charges.

Based on the interviews, the initial charges are fully (or at least partly) rebated or discounted when the consumer buys the product:

– via a financial adviser, which uses a fee-based form of remuneration;
– through a funds supermarket or other form of discount broker.

The main instances when consumers will be likely to pay the initial charges as stated include:

– if they go directly to the product provider; or
– if their adviser is remunerated on a commission basis—however, one of the IFAs interviewed noted that it applies a partial rebate even to commission-based business (with a full rebate applying to its fee-based business).

Rebating or discounting of the AMC was seen to be less common, and several interviewees noted that, unlike with initial charges, the AMC gives an indication of the actual charges paid by consumers. However, there are instances when the AMC is also subject to a rebate or discount, or where the discount broker or platform is able to buy fund units from providers at the institutional price for its retail customers.

As regards additional fees at the distribution level, the platform operators consulted (wrap providers and fund supermarkets) do not levy an up-front or annual fee for their services, but are remunerated from part of the AMC; for example, the platform may receive 25bp from the provider as remuneration for the administrative services outsourced to the platform. Two of the IFAs interviewed operate mainly or exclusively on a fee basis (with full rebating of the initial charge); the fees levied can vary by customer, complexity of the transaction and advice given, etc.

One network of financial advisers commented that the use of fee-based remuneration, although increasing, had been slow to develop and that most of the advisers in its network were still commission-based. Although detailed data was not available, it was estimated that around 70% of business through the network advisers would be on a commission basis, without any rebating, and hence that, in the majority of cases, consumers would be paying the stated product charges. However, the network noted that this is changing; there has been a growing trend towards rebating—five years previously, hardly any rebating was observed and the full initial charge was payable. No information is collected at the network level about the size of rebates applied by individual advisers.

The results of surveys of IFAs give a further indication of the use of fee-based remuneration as well as commission rebating. Consistent with the information provided by the network firm interviewed, Figure 3.5 below shows that the majority of IFA sales continue to be remunerated on a commission basis, in particular with commissions paid at the point of sale.

The extent of fee-based remuneration among IFAs gives an indication of the level of rebating, as a fee-based charging structure would be likely to involve a greater degree of rebating of the commission provided by the product provider to the IFA. Figure 3.5 shows that the majority of sales are remunerated by commissions, in particular commissions paid at the point of sale rather than those paid repeatedly after the initial sale. Nonetheless, fee-based advice, with commission rebating, is observed, even if still only in a small part of the market.[8]

[8] On a historical basis, the results of this survey suggest that there has not been significant growth in fee-based advice since 2005. For example, the percentage of advisers deriving only 0–24% of their business on a fee basis remained between 80% and 90% during 2005 to 2008; the percentage deriving all or a significant part of their business from fees (ie, more than 75%) remained at less than 5% among the surveyed advisers.

Figure 3.5   Commission- versus fee-based remuneration of IFAs
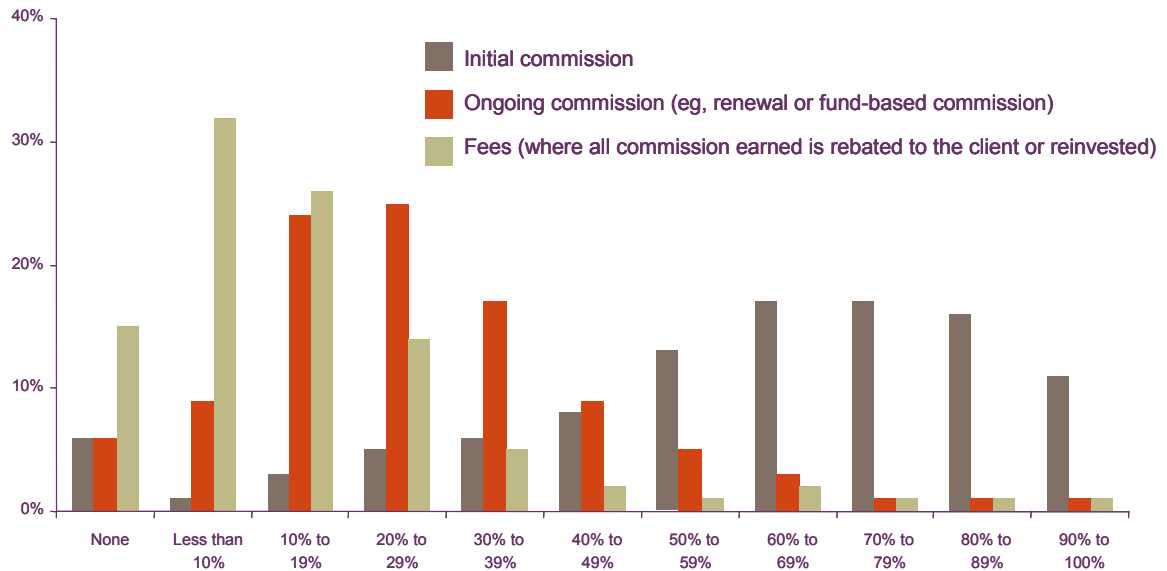
Note: The figure shows the distribution of responses to the question: 'What percentage of your income comes from each of …?'
Source: IFA Census (2007).

Figure 3.6 presents survey results directly aimed at establishing the frequency of rebating by IFAs. Although dating back to 2004, the majority of IFAs indicated that they rebate (at least occasionally) some or all of the commission they receive. However, there is no indication of the value of the rebates offered, and no data was available to examine changes in rebating practices over time.

Figure 3.6   Frequency of commission rebating by IFAs
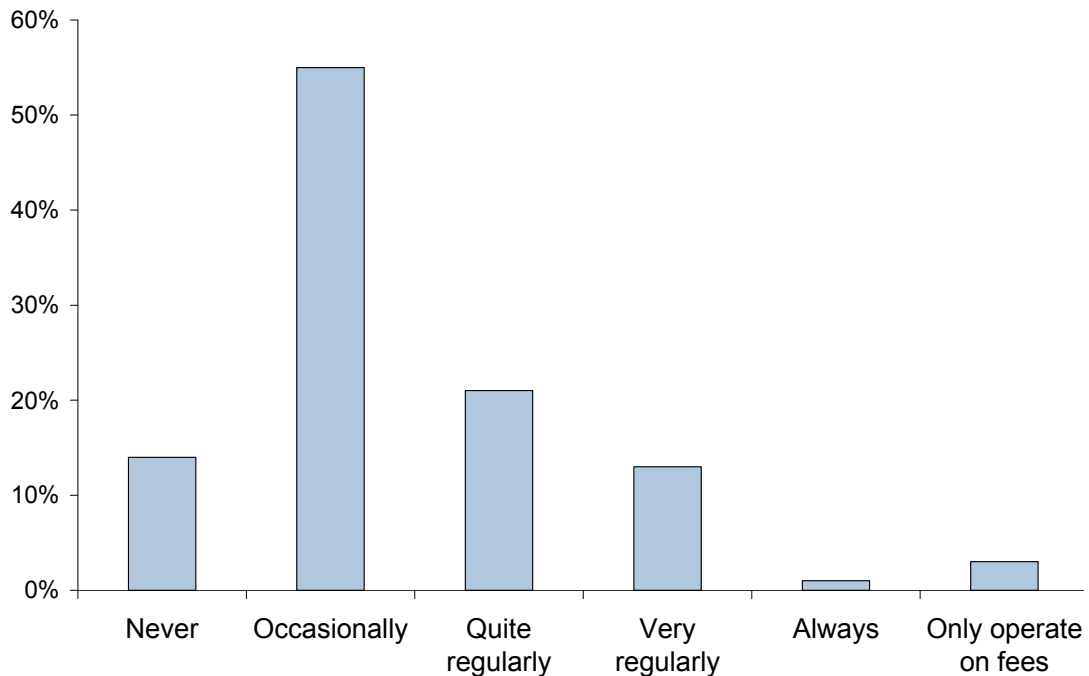
Notes: The figure shows the distribution of responses to the question: 'How often do you rebate commission to your clients?'
Source: IFA Census (2004).

Similar survey data for other distribution channels was not available. Hence, the evidence of increased rebating or discounting—be it due to the shift to fee-based advice, the emergence of platforms, or other reasons—remains largely anecdotal, based on interviews.

3.2.3 Implications for methodology and empirical analysis

While existing research is limited and it was beyond the scope of this study to gather data on a comprehensive basis, the available evidence suggests that stated charges at the provider level (in particular, the stated initial charge) do not give an accurate indication of the total charges paid by consumers when purchasing retail investment products.

The availability of comprehensive data at present on charges at the provider level only has implications for the methodology and empirical analysis that can be conducted. From a methodological perspective, the data limitations mean that the 'ideal' methodology and analysis differ from what can be feasibly achieved given the data available.

A useful starting point for the methodology and actual analysis is therefore to specify the hypotheses and conduct the empirical tests that can be examined with the available data. For example, data is available to test whether provider charges have fallen over time— ie, one can test the hypothesis of whether there has been a reduction in that part of the total wedge between gross and net returns that is due to provider charges.

If this is found to be the case, one can then take the analysis a step further and specify the conditions under which the reduction in provider charges translates into a reduction in total charges paid by consumers. For example, one such condition is that rebating/discounting practices have been stable over time, or indeed have become more widespread (with possibly larger rebates and discounts that are not offset by fees at the distribution level). Anecdotal evidence may support that this condition, or other conditions are indeed met; however, new data may need to be systematically gathered to test robustly whether this is the case.

The following details the empirical approaches that can be adopted to analyse charges in the retail investment products market, setting out which hypotheses can be tested (and how) using the available data, and what inferences can be drawn from the results. It also discusses which additional hypotheses can be examined, or which additional tests would be required to strengthen the inferences, if new data were available. Suggestions on how to gather additional data are discussed separately in section 3.4.

3.3 Methodology for empirical testing

The key empirical questions are how charges in the retail investment products market have evolved over time and what the impact of regulation has been on any trends observed. The following sets out the main hypotheses to be tested in order to address these questions and discusses the methods to conduct empirical tests, taking into account the data actually available.

3.3.1 Evaluating trends in charges over time

The first main hypothesis to be tested is whether charges of retail investment products have fallen over time. Given the data available, the pragmatic approach is first to examine trends in stated provider charges over time. The next step is to explore what data is available or would need to be gathered to test trends in total charges, including the actual charges paid at both the production and distribution level (ie, taking account of the level of discounting/rebating as well as the amount of any additional fees paid).

Trend in provider charges

The dataset that is being put together contains data on (see Appendix 1):

– the AMC at share class or fund level (annual, up to ten years);
– the TER at fund level (annual, up to ten years);
– the initial charge at fund level (monthly, up to ten years).[9]

Although the data gathering was ongoing at the time of writing this report, it appears that data is available for the majority of funds (unit trusts, OEICs and life funds). While data on the AMC and TER is available for all funds sold in the UK irrespective of domicile, the initial charges data is restricted to UK-domiciled funds only.

This allows an analysis of how different charges have evolved on an annual basis over time (up to ten years, given the data available). The analysis can in principle be conducted across all funds to give a picture for the market as a whole.

However, since charges are likely to differ by type of fund, its investment style and other characteristics, an analysis within categories of funds is likely to be more informative. In particular, as discussed in section 2.2, a key assumption in this study is that the chosen product is suitable. As such, the analysis should be undertaken separately for similar types of fund that share similar risk and other characteristics.

The new dataset will contain, for each fund, detailed data on the type of fund, its investment style, monthly fund performance, and so on (see Appendix 1). Using this information, funds can be classified within product and risk categories. As further explained in section 4, the risk classification can be performed on the basis of readily available classifications (eg, those set by the IMA); alternatively, funds can be classified after estimating risk using the monthly performance data. The former, and more simple, approach may be sufficient to gain an understanding of how charges have evolved over time, especially if it is found that there is limited variation in charges between funds with the same (IMA or other) classification.

For each product/risk category, the comparison of charges will involve a tracking over time of:

– the average annual charges—unweighted mean, weighted average to take account of differences in the size of funds (measured by total net assets), and median to control for possible outliers;

– the dispersion of charges around the average—minimum, maximum, inter-quartile range and other percentiles where required.

Changes over time in the average and dispersion of charges of funds (within or across product/risk categories) may be notable from inspection of the calculated descriptive statistics. However, where required, standard statistical tests can in principle be conducted to assess the significance of any changes observed.

The analysis can be performed separately for the AMC (to measure changes in stated annual charges), the TER (to measure changes in the actual total expenses charged against the fund) and the initial charge (to measure changes in up-front costs). The initial charge (as well as any exit fees) can also be annualised and combined with annual ongoing charges, using assumptions about standard holding periods. A five-year holding period is a good starting point, although different scenarios of shorter and longer holding periods should be considered,[10] especially if the aim is to evaluate the relationship between holding period (or

---

[9] The initial charge may instead take the form of a bid–offer spread. Data on exit charges is also available where applicable.

[10] A five-year holding period is used, for example, by Khorana, Servaes, and Tufano (2007) to annualise initial and exit charges. Barber, Odean and Zheng (2005) use a seven-year holding period, in line with IMA (2006), which states that a seven-year holding period for funds may be considered typical.

flexibility in the choice of holding period) and the total one-off and ongoing charges incurred.[11]

The results of this analysis will allow conclusions to be drawn about the evolution of stated charges at the provider level. However, the conclusions that can be drawn about the evolution of the total charges paid by consumers will be more limited.

If it is found that stated charges at the provider level have:

– *fallen over time*, it can be concluded that total charges will also have fallen *if* charging structures at the distribution level have remained largely unchanged over the same time period or if distribution charges have also fallen overall (eg, because of greater use of rebates and discounts that have not been offset by increases in additional distribution fees);

– *not changed (or indeed have increased)*, this does not necessarily imply that there has not been a reduction in total charges over time. There may have been changes in charging structures within distribution channels that result in lower total charges, or there may have been a shift in purchases of the same funds towards cheaper distribution channels.

That is, an analysis of trends in total charges requires charges to be examined at both the provider and distribution level, along with consideration of the interplay between the two (via rebates and discounts).

Trend in total charges
To quantify trends in total charges, data would be required on all charges paid by consumers, after taking into account rebates, discounts or any additional distribution fees levied on product purchases through different distribution channels. If data were available on total charges, the empirical tests to be conducted would be similar to those discussed above for provider charges—ie, test for changes in the average and dispersion of total charges, by product category, and, in addition, by distribution channel.

Anecdotal evidence is available to suggest that rebates and discounts are observed in the market—in particular for the initial charge—and that such practices have become more prominent in recent years. However, as discussed above, no systematic data was available on trends in charging structures within distribution channels; nor was there any data on the actual extent of rebating and discounting (and additional distribution fees) across the different distribution channels and over time.

Section 3.4 provides some suggestions on methods that could be employed to generate the new data required to allow analysis of trends in total charges.

While the dataset does not contain data on charging structures at distribution level, it may allow a partial (if imperfect) test of trends in the cost of distribution, as follows: it has been increasingly common for providers to divide the same fund into two or more share classes designed to segment their market between institutional and retail investors. Different annual charges can apply to the different share classes. To the extent that the difference in the AMC between institutional and retail share classes of the same fund is a reflection of the cost of distribution, and if a sufficiently large sample is available over time, it may be possible to track the difference between institutional and retail AMC to draw inferences about whether

---

[11] Evaluating total ongoing and initial charges over different holding periods would help in establishing whether product flexibility has a price (see section 2). For example, the analysis can show whether there is a general pattern such that products with lower total charges for normal/longer holding periods have higher total charges when held for a truncated/shorter time. For example, this will be the case if, over normal/long holding periods, products with high initial (or exit) charges and low annual charges deliver lower total charges, while, over the truncated/short holding periods, products with low initial (or exit) charges but high annual charges produce lower total charges.

the cost of distribution has fallen over time—ie, for funds with institutional and retail share classes, the difference between AMCs could be calculated and the (average) difference compared over time.

However, there are limitations to this analysis, and inferences need to be qualified accordingly. In particular, the split between retail and institutional share classes may not be perfect—for example, as confirmed in interviews with providers and intermediaries, some consumers may be able to purchase institutional share classes, with their lower charging structures. Nonetheless, if the share class data is readily available in the dataset, this empirical analysis may be worth pursuing to provide a proxy indication of distribution costs.

3.3.2 Evaluating shifts in consumer purchases towards lower-charging products

Even if charges for the purchase of specific products or categories of products have not fallen overall, consumers may have become better off (ie, pay less) over time if they have shifted their purchases from high-charge products to otherwise similar products with lower charges.

The separate hypothesis to be tested is therefore whether low-charge funds have, over time, experienced a greater inflow than high-charge funds. As with the measurement of trends in charges, the trends in relative inflows should be measured for funds within the same product/risk category (or after controlling for the product type and risk).

The dataset that is being put together allows a testing of whether relative inflows are driven by differences in stated charges at the provider level. Broadly speaking, within each broad product/risk category, funds could be grouped according to their level of charges (eg, the stated AMC, although the grouping could also be conducted on the basis of the TER or stated initial charge, depending on the specific question to be addressed). Within each group of fund, the average inflow could be calculated so as to compare the changes in those inflows over time between the high- and low-charge groups.

The inflow or rate of growth of the different groups of funds requires data on monthly or annual gross and net inflows to the funds. Total net assets can also be used to construct a flow measure, by calculating the monthly change in total net assets, adjusted for by the return earned to isolate the inflow from the return generated on existing assets. As done in Barber et al. (2005) and others, the net inflow (F) of new money (in percentage terms) can be calculated as:

$$F_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1}(1 + r_{i,t})}{TNA_{i,t-1}}$$

where $TNA_{i,t}$ is total net assets of fund i in time period t, and $r_{i,t}$ denotes the fund returns over the period.

The analysis could in the first instance be based on a simple comparison of inflows between high- and low-charge groups of funds (ie, comparison of mean, median or other descriptive statistics).

Alternatively, an econometric approach could be adopted. This could involve a (panel data or cross-section) regression of the inflows to individual funds in a month (or year if the data is annualised) on the annual and/or initial charges for that fund, as well as a number of control variables. For example:

$$F_{i,t} = \alpha + \beta_1 \cdot AMC_{i,t} + \beta_2 \cdot InitialC_{i,t-1} + \gamma_1 \cdot Class_i + \gamma_2 \cdot TNA_{i,t} + \gamma_3 \cdot r_{i,\,t-1} + \varepsilon_{i,t}$$

where F denotes the net inflow to fund i in period t, AMC and InitialC are the ongoing and initial charges, Class is the chosen indicator variable to classify the fund (eg, IMA

classification), TNA is total net assets (to control for size), and $r_{i,t-1}$ denotes the return earned in the previous period (to control for purchases that may be driven by past performance).[12]

A significantly negative coefficient β indicates a negative relationship between fund inflows and the level of charges, which in turn indicates that consumers tend to purchase more lower-charge products. With a panel regression, the model can also be specified to test whether this trend in the purchases of lower-charge products has increased over time.[13]

Conducting these tests on the basis of provider charges is informative to assess whether differences in the stated AMC, TER or initial charge have an impact on consumer purchases, and whether over time there has been a shift towards cheaper but otherwise similar products. Similar tests have been performed in the literature, such as by Barber et al. (2005), who examine flows into US equity mutual funds and find a negative relationship between fund flows and initial charges, but no relationship (or indeed a positive relationship) between fund flows and TER.[14]

In order to identify shifts in consumer purchases due to differences in total charges, including those at the distribution level, additional tests would be required, and new data would need to be collected. In particular, in addition to the shift between products, what needs to be tested is whether consumers have, over time, shifted their purchases of the same types of product to cheaper distribution channels. This requires analysis of:

– rebating and discounting practices as well as additional distribution fees over time, per distribution channel;
– the shift of purchases across distribution channels, from the more expensive to the cheaper ones.

As discussed above, such data has not been systematically collected over time, which means that a historical analysis based on the available data is not possible. New data would need to be gathered to allow such analysis in the future.

3.3.3 Evaluating the impact of regulation and other determinants of charges
If reductions in charges (or shifts towards lower-charge products) are identified, the relevant question from the FSA's perspective is to what extent these trends can be attributed to regulation or changes in regulation.

Regulation could have an impact on charges (and related purchase decisions) through several different channels. First, there may be direct requirements that relate to the types of product sold or their charging structures. An example of this would be the introduction of stakeholder products in 2001, which included a cap on the AMC. Second, regulation could

---

[12] The detailed specification and variables to be included in the regression can be defined only once the new dataset has been put together. Barber et al. (2005) conducts cross-section regressions in which the dependent variable is the quarterly fund flow (scaled by TNA). The main independent variables include: total expenses (annual expenses plus one-seventh of a fund's front load (ie, initial charge)); annual expenses; the front load; a dummy variable which takes the value of 1 if the fund appeared in the highest expense decile in year t-1; the annual (adjusted) fund returns for the previous 12 months; the monthly standard deviation of the fund's return over the previous 24 months; the beginning-of-quarter TNA (in log form); and the fund age (in log form).

[13] Data on some variables (including rebates and discounts of charges) may not be available, and excluding these variables could result in the regression model suffering from omitted variable bias. This problem can be mitigated by extending the above models. If the omitted variables are positively related to fund inflow, the missing part of the inflow equation is likely to be positive and can therefore be captured by a one-sided error term. Thus, the model is augmented by $\varepsilon_{i,t} = u_{i,t} + v_{i,t}$ (where $u_{i,t}$ is a zero mean error random error, and $v_{i,t} \geq 0$). This model can then be estimated using stochastic frontier analysis (SFA), and estimates of the one-sided error term can be viewed as the impact of the unobserved variables on fund inflow. For an application of SFA techniques in the mutual fund market, see Annaert, den Broeck and Vennet (2001), who use SFA techniques to evaluate the determinants of fund performance.

[14] Similar findings are reported in Wilcox (2003) and Sirri and Tufano (1998).

affect charges through requirements to disclose information, including information on charges.[15]

In principle, the impact of regulation can be assessed using event study analysis, whereby the changes in charges (or changes in the differential growth rate between high- and low-charge products) are related to specific events of regulatory changes. A simple before-and-after comparison around a regulatory event could then establish whether any observed changes can be attributed to that event.

The difficulty with this approach in practice is that specific regulatory events (and their timing) cannot always be clearly identified and/or the effects of those interventions may take time to materialise. Over the period for which data on product charges is available (ten years—ie, since 1998/99), there are a number of regulatory changes that could be considered as candidate regulatory events.

– **Financial Services and Markets Act 2000**—this Act increased the powers of the FSA and ensured that only authorised firms could operate within the retail investment area.

– **Introduction of stakeholder products, 2001**—legislation in 1999 and 2000 introduced a suite of products targeted at increasing the participation of lower-income individuals in retail investment. A number of criteria must be followed to enable a product to be treated as a stakeholder product, the most relevant being a cap on the AMC. The first products became available in 2001.

– **Comparative tables, 2001**—launched in a consultation paper in 1999, the tables went live in 2001. Their aim is to allow investors to compare initial and annual management charges on products offered by product providers. An independent comparison tool was not previously available to investors.

– **Depolarisation and the menu, 2005**—depolarisation of financial advice reversed the policy of designating firms to be either IFAs or tied agents, which had been in existence since 1987. Two requirements came into force alongside depolarisation: advisers have to produce an initial disclosure document for the investor detailing the type of advice that their client would be receiving; and the adviser's charges and remuneration have to be disclosed (the 'menu'). The intention is to increase the transparency of charges and quality and type of advice given by an adviser, and to show how the adviser is remunerated.

– **Treating Customers Fairly, 2006**—as part of the move to more principles-based regulation, Treating Customers Fairly (TCF) requires firms to put in place the relevant measures to test and demonstrate that they are treating customers fairly.

It is outside the scope of this study to determine which are the relevant regulatory events to evaluate, conceptually and empirically, or whether these have had an impact on charges (or related purchase decisions). However, it appears that the types of event set out above would be expected to show more gradual changes, if any, and potentially overlap in their timing. The typical process of industry dialogue, discussion paper, feedback, consultation paper, feedback, and policy statement also suggests that changes in regulation are likely to have a more gradual impact.

As such, rather than employing a strict event-study approach based on before-and-after comparison around the event date, the alternative (and 'softer') approach may be:

---

[15] In the context of this study, the focus is on improvements in consumer outcomes, but the impact of regulation on charges may also be negative—eg, the costs of complying with information requirements may increase total costs for providers and intermediaries and put pressure on margins; firms may then attempt to pass on the increased costs in the form of higher charges unless efficiency gains can be achieved or competitive pressure prevents them from doing so.

– to establish first whether there has indeed been a reduction in charges over time (or a shift in purchases towards cheaper products);

and, if so, to:

– assess how far it is likely or plausible that any observed reduction (or shift) is attributable to regulation as opposed to other explanatory factors.

As regards the latter step, if a notable reduction in charges (or related shifts in purchases) is found (either around a specific point in time or more gradually), different methods can be adopted to inform the analysis of the likely impact of regulation. If the reduction (or shift) indeed appears to be a structural break at a specific point in time that coincides with the timing of a specific and clearly identifiable regulatory intervention, this may be taken as sufficient evidence that it is regulation, rather than other factors, that explain the break, especially if no other changes that could have the same impact can be identified at the same point in time.[16]

If, on the other hand, the reduction (or shift) is more gradual, as may be expected, the first and most simple approach would be to interview or survey product providers or intermediaries to seek their opinion on the drivers behind changes in charging structures. While not the focus in the interviews with industry representatives conducted as part of this study, the issue was raised by some interviewees. The main regulatory event that was identified as having had an impact on charges was the introduction of stakeholder pensions, although market developments, such as the emergence of provider platforms, were seen as important drivers of changes in charging structures overall. Further interviews or a more comprehensive survey could be conducted to provide information about the importance of regulation compared with other factors.

Non-regulatory factors that may explain the trend in charges

To establish further whether **provider platforms** have had an impact on charging structures and whether these platforms might have contributed to an overall reduction in the total charges paid by consumers, more comprehensive data would need to be gathered—at provider, platform and adviser level—to estimate total charges and how these compare with the charges of products not distributed via platforms. This data would need to comprise information in particular on the remuneration of the platforms, the availability and size of discounts and rebates, and the level of additional platform fees—data that is not available for analysis in the dataset being put together as part of this project.

Based on this dataset, it may be possible to conduct some additional tests on the importance of other factors that may have driven provider charges over time. For example, an alternative explanation for a reduction in provider charges (if observed) may be increased **competition**. One, albeit imperfect, way of testing this would be to examine charges for different categories of fund (eg, funds that share the same IMA classification) and to compare how charges have developed, on average, in those fund categories where competition is likely to be more intense than in other categories. Relevant measures for the intensity of competition would be the number of funds (or the increase in the number of funds) and the concentration of fund assets within the category.[17] Similarly, in order to test for the importance of **economies of scale**, it may be possible to explore the differences in charges and their evolution over time for funds of different size.

That is, if the analysis of charges—either stated charges at provider level, or, if possible, total charges—reveals a downward trend (in discrete steps or gradually), further analysis can be

---

[16] Testing for a break in the time series of charges can also be done using an econometric approach rather than mere inspection of changes in the data around the time of the regulatory event of interest. This approach can also be extended to test for more gradual changes using a dynamic econometric model.

[17] For example, Kheswani and Stolin (2005) use the Herfindahl index of concentration of assets under management to measure competition within IMA sectors.

undertaken to explore whether the observed decline in charges is indeed due to regulation or instead may be attributable to non-regulatory factors. The analysis involves checking the plausibility of alternative explanations, in line with the methods outlined above.

Brand and other determinants of charges

The dataset that is being put together can also be used to test for other determinants of product charges, if not over time then at least in the cross-sectional setting. As examined in other studies, possible determinants include factors such as the size of a fund, the type of fund, its investment style, the fund's age, its domicile (onshore or offshore), portfolio holdings and turnover.[18]

In addition, it may be that the brand of a product or its provider may explain charges—eg, among the funds in a given product/risk category, well-known household brands may be the more expensive ones. Testing this hypothesis is relevant, as explained in section 2.1, because it could be argued that, despite the higher charges (if identified), buying a well-known brand may be valued by consumers in its own right. (In addition, it may be that higher charges produce superior overall net returns, so that brand and/or higher charges are a signal of better performance—see section 4.)

Assuming that funds and their providers can be categorised according to recognised household brands and no brands, the data available can be used to test for the 'price' of brands for consumers (but not the value consumers derive from brands).[19] As regards empirical methods, simple comparison tests can be conducted whereby, within a product/risk category, average charges (AMC, TER or initial charges) are compared between funds with and without a brand name.

Alternatively, an econometric approach can be pursued, using regressions of the level of charges of individual funds on a brand indicator variable as well as other control variables, for example:

$$\text{charges}_{t,i} = \alpha + \beta \cdot \text{brand}_i + \gamma_1 \cdot \text{class}_i + \gamma_2 \cdot \text{TNA}_{t,i} + \gamma_3 \cdot \text{age}_{t,i} + \gamma_4 \cdot \text{domicile}_{t,i} + \varepsilon_{t,i}$$

where charges can be the AMC, TER or initial charges,[20] brand is an indicator variable that takes the value 1 if the fund (or provider) is a well-recognised brand, class is an indicator of the product/risk category of the fund (eg, IMA sector classification) and TNA is total net assets (to control for size). The fund's age and its domicile (eg, UK or non-UK) can also be included among the other control variables.[21]

For the purposes of this analysis, it is not necessary to create a single precise definition of what constitutes a well-known brand. Rather, if a relationship with charges does indeed exist, this should manifest itself in at least one of the classifications adopted (assuming that at least one is approximately right). If this analysis indicates that the independent influence of brand is significant (in terms of explaining charges), it may then be worthwhile increasing the sophistication of the classification for a more precise measurement of the relationship.

If a positive relationship between charges and brand is identified, this would in itself not imply anything about whether consumers are worse or better off buying a premium brand. This depends on whether premium brands deliver superior risk–return performance (this is

---

[18] See, for example, Khorana et al. (2007).

[19] As with brand, it would in principle be possible to price (not value) other aspects that may be argued to represent measures of the quality of consumer outcome, including the transparency, consistency, and quality-of-service aspects discussed in section 2, as long as products can be classified according to these aspects. However, classification of individual funds into the 'poor' and the 'good' is difficult based on readily available data, as already discussed in section 2.

[20] Where the dependent variable is the AMC or TER, the initial charge could be included as an explanatory variable in the regression to test whether there is a relationship between annual and initial charges—eg, funds with higher initial charges may have lower operating expenses.

[21] The detailed specification and measurement of the variables can be determined only once the new dataset has been put together.

discussed in section 4) and whether consumers attach a value to brands over and above their 'price'. If brand does matter, the analysis of trends in charges and product purchases, as set out above, needs to control for brand—either in a regression setting or by conducting a separate analysis of products with or without brand name. The same holds for other attributes of a product.

3.4 Possible methods for gathering new data

While some evidence is available on charging structures that apply at the distribution level, new data would be required to allow a systematic assessment of the total charges paid by consumers, after taking account of rebates, discounts and any additional fees that may be levied at the distribution level.

Gathering data is costly, and the costs need to be weighed against the benefits. If the analysis of stated charges at provider level suggests a significant decline over time, it may not be necessary to collect detailed data at the distribution level in order to enable conclusions that total charges paid for by consumers are likely to have fallen (given that it may not be plausible to argue that additional distribution charges will have offset the decline). If, on the other hand, stated charges have remained unchanged (or indeed have increased) over time, further analysis may be required; hence further data may need to be collected to show that there may nonetheless have been a reduction in total charges, but due to changes at distribution level. As such, the need for more data and the extent of the data-gathering exercise is partly dependent on the outcome of the analysis of the trend in stated charges at the provider level.

With these considerations in mind, the following first sets out the evidence that is already available to allow at least a partial analysis of charges at distribution level, without requiring separate data to be gathered. It then summarises different approaches on how to gather new data if the objective is to conduct a more comprehensive assessment—if not historically, then at least going forward.

3.4.1 Evidence from new dataset and existing FSA data sources
In addition to the evidence summarised in section 3.2.2, further analysis is possible using the data that will become available in the new dataset that is being put together for this project, as well as the data already available to the FSA. This includes the following.

– **AMC of retail and institutional share classes**—the new dataset will contain information on the stated initial and annual charges at provider level. As discussed in section 3.3.1, this dataset may allow at least a partial analysis of the cost of distribution, by examining the difference in the annual charges that applies to retail share classes of a particular fund and the institutional share classes. Subject to the caveats (see section 3.3.1), this analysis gives an indication of how the cost of distribution has evolved over time if the difference in charges between retail and institutional share classes reflects the cost of retail distribution (and assuming that a sufficiently large sample of funds with multiple share classes is available in the dataset over time).

– **Retail Mediation Activities Returns (RMAR)**—some of the data collected by the FSA as part of the RMAR can be used as evidence on charging structures at the distribution level. In particular, retail intermediaries are required to provide the FSA with income statements which contain data on gross and net commissions. These could be used to provide a proxy indicator for the level of commission rebating that occurs, since the difference between the two numbers should reflect the difference between the gross commission that the intermediary receives from a provider and any commission that it rebates to the consumer. However, from interviews with industry representatives, it became clear that this is an imprecise indicator to use. Although it may give an indication of the extent of rebating of cash payments that are made from the provider to

the intermediary and then passed on to the consumer, the data does not reflect any discounting of charges at the provider level—for example, in the form of fund units being purchased at a lower (discounted) price from the provider. Further analysis of the RMAR data would be required to understand whether it could be used to proxy rebating practices in the market.

The RMAR data can also be used to further examine the prevalence of fee-based advice in the market, since intermediaries are required to separately disclose income from fees as opposed to commissions. This data allows an analysis of the number of firms deriving income from fees instead of, or in addition to, commissions and the proportion of fees out of total income.

3.4.2 Potential new sources of data

The available data is incomplete and will allow a partial analysis only. New data will need to be gathered if the objective is to obtain detailed information on charging structures at the distribution level (and changes within and across distribution channels over time).

The new data can be gathered by means of surveys, but it may also be possible to extend or adjust the data returns that firms are already required to complete and submit to the FSA on a regular basis, including the following:

– **RMAR**—while the RMAR already contains some data to allow a partial analysis of rebating and the frequency of fee-based advice (see above), the data requests to intermediaries filing the RMAR would need to be extended to allow analysis of other indicators. For example, if the RMAR required the reporting of the number or value of products sold by intermediaries, it would be possible to derive the net total income or the cost per unit or value sold, which could then be tracked over time.

– **Product Sales Database (PSD)**—product providers submit data on their product sales to the FSA, and the PSD represents a useful source of information on the sales by product providers. However, it gives only a high-level indication of the distribution channel through which a particular product has been sold. To enable an individual product to be tracked more closely through the value chain, it would be necessary to extend the PSD to include more comprehensive information on the distribution channel used for a product and its sales values, as well as volumes at each individual stage. In particular, it would be useful to track the degree to which products are being sold through newly emerging distribution channels such as discount brokers or other forms of platform provider.

Such additional reporting requirements impose costs on regulated firms, and the feasibility and costs would have to be assessed and weighed against the potential benefits.

The alternative approach to gathering data would be through specifically designed surveys. There are three potential sources for gathering the data: providers, intermediaries or consumers directly.

– **Providers** could be surveyed about the distribution channels for their products and the frequency and level of rebating or discounting of the stated charges. However, discussions with intermediaries and providers have highlighted that much of the rebating that does occur is implemented by the intermediary without necessarily the provider being aware of the extent to which it is taking place; neither would the provider know about the level of any additional fee negotiated by the intermediary. Thus, provider information may not give the complete picture. Nonetheless, provider information may serve to inform about the different charging structures and to identify any discounting practices, and the changes in those practices, at the provider level.

– A survey of **intermediaries** would need to be conducted in order to obtain a clearer picture of the extent of discounting, commission rebating and any additional fees levied at the distribution level. The survey would need to cover different types of distribution channel, including IFAs (as per the survey evidence shown in section 3.2.2), discount brokers or fund supermarkets, as well as the products channelled through wrap platforms. At the same time, information from intermediaries could be collected on unit costs, net income and the value of products sold in order to assess possible trends in the income received and expenses incurred at this stage in the value chain for retail investment products. Such information could be used, in conjunction with data from providers, to identify the extent to which redistribution was occurring in favour of the consumer.

– As an alternative to approaching providers or intermediaries, it may be possible to undertake a survey of **consumers** themselves in order to identify the *actual* charges they paid on products they had purchased. Such an approach would require a cohort of consumers to be identified who had recently purchased retail investment products. The FSA has already identified relevant groups of consumers for its financial capability baseline study, as well as more recent research that has been undertaken on how consumers use information when purchasing retail investment products.[22] However, any such survey would need to ensure that consumers are fully aware of, and can identify, the charges that they had incurred up front and/or on an ongoing basis. It would also be necessary to identify consumers who have used a variety of distribution channels for their purchases, which may not be the case with existing cohorts. The establishment of a new cohort that enabled this and fulfils the other required criteria could be more time-consuming and problematic than a survey of intermediaries.

Any type of survey would need to be conducted in regular intervals in order to identify trends, and the costs of the exercise would need to be evaluated against the benefits. Overall, a survey at the consumer level appears to be the least practical even if it is at this level that total charges should in principle be measured. A survey of providers may be considered more practical than a survey of intermediaries, not least given the relative number of firms that would need to be included to form a representative sample; however, the information available at the provider level may not be sufficient to build the complete picture.

Whatever method is used to gather the data, it is unlikely that detailed information would be available on a historical basis. However, if systematic analysis of total charges and the impact of regulation is to be pursued in future, it will be necessary to start gathering data now in order to achieve this going forward. With further regulatory changes planned in the years to come, following the Retail Distribution Review, this could become particularly relevant.

---

[22] Financial Services Authority (2008), 'Services and Costs Disclosure: Qualitative Research with Potential and Recent Purchasers of Financial Products', Consumer Research Paper 65a, Stage 1, February.

4 Analysis of the link between charges and performance

While charges are important, what ultimately matters is net returns and the risk associated with those returns. Even if charges are found to be falling over time (due to regulation or other reasons), outcomes may not necessarily have improved for consumers, given that higher charges may be associated with better risk–return performance (eg, due to more or better management).

Section 4.1 summarises the existing academic literature on the relationship between investment product performance and charges. Drawing from the academic literature, section 4.2 provides an overview of the framework for measuring and testing the relationships of interest. Section 4.3 sets out the methodologies used for empirically measuring the risk-adjusted performance of different products. Section 4.4 describes the approaches for testing the specific hypotheses on the relationship between investment product performance and charges (or other product characteristics).

4.1 Overview of findings in the academic literature

There is a considerable body of academic research examining investment product performance and how it is related to charges and other product characteristics, such as the underlying portfolio turnover and fund size. Typically, performance is measured by risk-adjusted net returns (after expenses at fund level), and charges are measured by the TER.

However, most studies focus on mutual funds investing primarily in equities, and in the US market. This may limit the extent to which the methods and conclusions from the literature can be readily applied to the hypotheses being tested regarding a wide range of retail investment products in the UK. Nevertheless, the existing academic research provides a basis for formulating a robust methodology for measuring risk-adjusted performance, and for modelling the relationships between performance and charges, and other characteristics.

Table A2.1 in Appendix 2 provides an overview of the key academic studies examining the relationship between investment product returns and charges. The studies generally conclude that there is no evidence that higher fund charges are associated with superior net (risk-adjusted) returns. To the contrary, charges are typically found to have a statistically significant negative effect on net performance.

For example, Quigley and Sinquefield (2000) find that worst-performing (in terms of risk-adjusted net return) UK equity trusts investing in equities have higher expense ratios. Gruber (1996) and Elton, Gruber and Blake (1996) conclude that there is a negative relationship between expense ratios and risk-adjusted net returns of US common stock funds. Malkiel (1995) and Carhart (1997) quantify the link using regression analysis, and conclude that a 1% increase in expense ratios leads to decreases of 1.92% or 1.54% respectively in net returns. Malkiel (1995) hypothesises that increasing expenses have a 'multiplier' effect on net returns (ie, the coefficient is greater than one, in absolute terms) because other relevant factors, such as turnover, may be positively correlated with the expense ratios.

Some academic studies do not find a statistically significant relationship between investment product charges and net performance. Fletcher (1999), for example, does not find that higher expense ratios are linked with either lower or higher risk-adjusted net returns on UK unit trusts investing in US equities.

There have been a very limited number of studies looking at funds investing in assets other than equity. One such paper is by Blake, Elton and Gruber (1993), who examine the

relationship between the risk-adjusted net returns on US bond funds and expense ratios, and find a negative association.

Some of the papers summarised in Table A2.1 have also looked at the relationship between performance and other characteristics, such as size or turnover of the fund. For example, Carhart (1997) and Droms and Walker (1994) find that there is no relationship between net returns and the turnover of the underlying portfolio. Fletcher (1999), Carhart (1997), Droms and Walker (1994) do not find a statistically significant relationship between performance and size of the fund. Elton, Gruber and Blake (1996) find a significant negative relationship between expense ratios and the size of US common stock funds, implying that larger funds have lower charges in their sample.

4.2   Overview of the analysis of the link between performance and charges

In line with the approaches adopted in the literature, the process of getting from the raw data to conclusions about the relationship between performance and charges involves two main stages for the empirical analysis:

–   the risk adjustment stage, where the returns and risk of investment products are estimated, and the returns are adjusted for risk (or risk is otherwise controlled for in the analysis);

–   the relationship estimation stage, where the hypotheses of interest are tested by estimating the relevant relationship between performance and charges (or other characteristics).

Figure 4.1 illustrates the two stages, and the methods available to implement each stage. Sections 4.3 and 4.4 explain each stage in turn.

Figure 4.1   Framework for testing hypotheses about performance
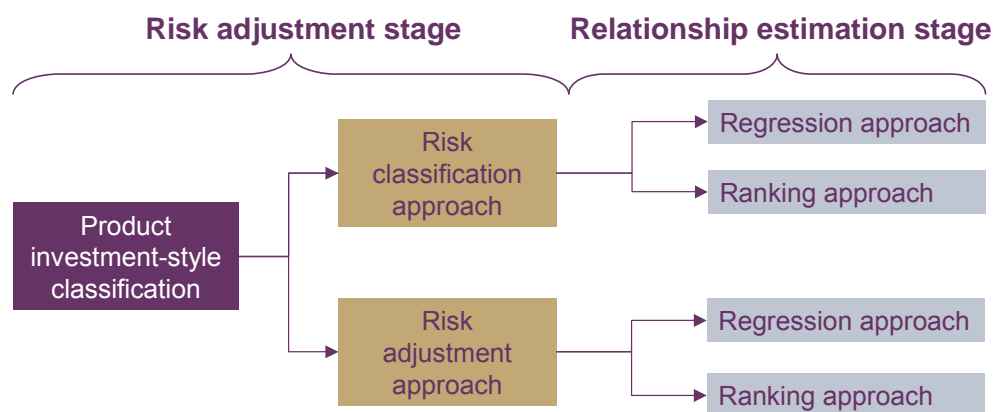
Source: Oxera.

Risk adjustment stage (see section 4.3)—different approaches are available to measure net returns on the investment product while controlling for risk.

1)   Products are classified according to their investment style (eg, IMA classification), where the classification is used as a proxy for risk. Here, raw net returns can be used in the relationship estimation stage, and the relationship estimation is performed for each investment-style class separately.

2)   If existing investment-style classifications do not provide a reliable proxy for risk, risk can be estimated directly, based on the returns data, and products can be classified according to this estimated measure of risk (eg, standard deviation of returns). As with

1), raw net returns can then be used for the relationship estimation stage, and the relationship estimation is performed separately for each risk class.

3) Instead of a risk classification based on the estimated risk or investment style, risk-adjusted returns can be calculated (eg, Sharpe ratios). Here, risk-adjusted net returns are the relevant performance measure used for the relationship estimation stage, and the estimation can be performed across products with different risk or investment style.

4) Given that product investment style may capture other important performance or consumer choice aspects, approaches 2) and 3) may be preceded by the classification of products according to their investment style.

Relationship estimation stage (see section 4.4)—there are two approaches to measuring the relationship between performance and charges, or other investment product characteristics of interest: the regression approach and the ranking approach. The former involves regressing the (net risk-adjusted) returns on charges, while controlling for other investment product characteristics, such as size and turnover. The latter entails first ranking (and grouping) products according to their performance, and then comparing (average) charges across these groups.

Within the framework, the choice of the final approach or approaches adopted for testing the hypotheses will largely be driven by three main considerations: data, the relationship(s) tested, and robustness checks.

– Sample size—for example, investment-style classification of products followed by risk classification, which is based on the estimated return risk, may lead to a small number of observations (ie, products) in each risk class, especially for less common product types. If this is the case, regression analysis *may* not be an appropriate approach as it may suffer from small-sample problems, and the ranking approach may be less problematic in this case (since, technically, only a few observations are needed to perform the ranking test).[23]

– The relationship(s) being tested or measured—if the relationship between only two variables is being tested (eg, performance and charges), then both the regression and ranking approaches may be appropriate. However, only the regression approach can be used if it is essential to control for other variables that may affect the relationship (eg, fund size or turnover). Moreover, the rank correlation coefficients cannot be aggregated (ie, averaged) in a meaningful way across different groups in the way that regression coefficients can be averaged. This is further discussed in section 4.4.1.

– Robustness checks—testing the hypotheses of interest using different approaches enables the robustness of the conclusions drawn to be checked. For example, conclusions drawn from an analysis that is based on investment-style or risk classifications can be used to check the conclusions drawn from analysis using risk-adjusted returns across risk classes. Thus, it can be of benefit to use alternative methods as a way to cross-check results and ensure the robustness of the conclusions reached on the hypothesised relationships.

---

[23] For example, if there are 100 unit trusts with investment objective 'global growth', and risk classification based on historical returns results in ten risk groups (deciles) with ten observations in each, the regression approach on the ten cross-sections of ten investment products is not appropriate for measuring the relationship between performance and charges, and the ranking approach may be more appropriate (although the aggregation of the correlation coefficient across the different risk classes may not be meaningful in this case).

4.3   Methodology for empirical measurement of risk-adjusted performance

To be able to test the hypotheses regarding the link between performance and charges (or other characteristics), performance must be measured so that it reflects both the return and the risk of the products. In essence, this can be done in two ways, as shown in Figure 4.1 above—the first is simply to classify the products according to risk, and then examine the relationship between net returns and charges within each risk class. The second is to calculate risk-adjusted returns directly, based on historical returns data.

4.3.1   Investment-style classification as a control for risk
The investment style or objectives of a product can be an important determinant of the product's risk–return profile. For example, products investing in growth equities can be expected to have higher return and risk than products investing in government bonds.

Therefore, classification of products according to their investment style or objectives may provide an approximate classification of products according to their risk profiles. If such a classification provides a proxy for risk then the raw returns can be compared within each investment-style class; similarly, the relationship between raw returns and charges may be examined within each class.

However, investment-style classification may be a poor proxy for investment risk if the expected return variability within each class is, in fact, very different across individual products. To an extent, this can be tested by examining the variability of risk within each investment-style class, measured using historical returns data. If the estimated risk varies significantly across individual products then investment-style classification may not be sufficient to control for risk, and risk adjustment using ex post risk measures (estimated using historical data) may be necessary.

Even if the analysis is ultimately performed using ex post risk measures, classifying products by investment style and conducting tests within each class has merits for various reasons. First, investment style or objectives may be a dimension of consumer choice in itself. Also, if consumers follow 'fashion', products using a certain investment style may see (temporary) boosts in fund inflows and size, which in turn may affect both performance and charges.

Second, the theoretically relevant risk measure is the *expected* variability of returns, and investment style may at least proxy for this ex ante risk. The analysis within investment-style classes may strengthen the analysis—in particular, the specific behaviour of different types of fund (eg, property compared with small cap equity) may mask both positive and negative relationships between charges, and both the average returns and volatility of returns, which may not be uncovered when all funds are examined together.[24]

Third, an investment-style classification can be simple to implement, given readily available classifications of funds. For example, the dataset that is being put together contains a variable that indicates the IMA classification for UK-domiciled funds. This classification has 31 categories, as listed in Table 4.1.

---

[24] For example, if, on average, property funds had higher charges than small cap funds, and on average over the period of analysis property funds had a higher return, the analysis of the two types of fund taken together might indicate that higher charges lead to higher returns. However, a more robust test would be needed to see whether there was a relationship between charges and returns within property funds and within small cap funds. Only if the relationship held *within* each class of investment style would the conditions for a positive relationship between charges and returns *overall* be satisfied.

Table 4.1   IMA classification (UK sector classification)

All funds				
Income funds		**Growth funds**		**Specialist funds**
Immediate income	**Growing income**	**Capital protection**	**Capital growth**	
UK gilts	UK equity income	Money market	UK all companies	Specialist
UK index		Protected/guaranteed funds	UK smaller companies	Technology and telecommunications
Linked gilts			Japan	Personal pensions
UK corporate bond			Japan smaller companies	Absolute return
UK other bond			Asia Pacific including Japan	
Global bonds			Asia Pacific excluding Japan	
UK equity and bond income			North America	
			North America smaller companies	
			Europe including UK	
			Europe excluding UK	
			Europe smaller companies	
			Cautious managed	
			Balanced managed	
			Active managed	
			Global growth	
			Global emerging markets	
			UK zeros	

Source: IMA (http://www.investmentuk.org/statistics/sector_definitions/fund_class.asp).

4.3.2   Measurement of risk based on historical returns

Rather than, or in addition to, using investment-style classification as an indication of risk, historical returns can be used to estimate directly the risk for each product. The estimate can then be used to control for risk, when examining and comparing investment product returns. In modern financial theory, the 'riskiness' of a product is typically measured by the volatility of the returns, with the standard deviation being the relevant metric.

However, as discussed below, the standard deviation has a number of drawbacks as a measure of risk, and alternative measures may be preferred. For example, downside risk may capture the risk of the product better under various circumstances—the semi-deviation (or semi-variance) is a common metric for the downside risk. The different measures are illustrated using actual returns data in Box 4.1.

Standard deviation

Standard deviation is the a statistical measure of variability which, when applied to investment returns, measures the volatility of returns. Standard deviation of investment product returns over the time period t=1…T is defined as:

$$\sigma = \sqrt{\sum_{t=1}^{T} (r_t - r_{mean})^2 \Big/ (T-1)}$$

where r_t is return on the investment product over month t, and r_{mean} is the average return over the time period from t=1…T. The monthly standard deviation can be annualised by multiplying it with a square root of 12.

Thus, the standard deviation simply measures the magnitude of deviations from the mean return, and, as such, captures the risk of not achieving the mean return. While it is a standard

measure for return volatility and hence investment risk, the measure has a number of drawbacks.[25]

– Positive and negative deviations from the mean affect the standard deviation in the same way—ie, positive deviations are 'as bad' as negative deviations. This is not what consumers may intuitively understand by risk, since they may perceive positive deviations are 'good'.[26]

– When comparing investment products in terms of risk, the standard deviation leads to intuitively unreasonable results when the average returns are very different. For example, suppose product A has a mean return of 3% and product B has a mean return of 10%, and both have a standard deviation of 6%. Based on the standard deviation, both products are equally risky. However, because product A has a much lower mean return, a higher proportion of returns are likely to be negative, whereas only a small percentage of product B returns are likely to be negative—hence, consumers may assess product B as being less risky.

– Standard deviation is an appropriate measure of variability when the returns are approximately normally distributed. However, investment fund returns are more often than not skewed and/or exhibit kurtosis (see also Box 4.1). One reason for this is the presence of non-linearities in the returns, which may result from portfolio rebalancing strategies, capital protection and other factors. Hence, the standard deviation may not capture the variability of returns, or the product risk, appropriately.

When assessing the risk of investment products, it is the presence of skewness and/or kurtosis in the return distribution that is most likely to undermine the use of the standard deviation. Hence, in order to assess the appropriateness of standard deviation as a measure of risk, tests should be performed on the returns data—if the return distribution significantly deviates from the normal distribution for a large proportion of products then asymmetric measures of risk, such as the semi-deviation, discussed below, may be more appropriate.

Semi-deviation

The likelihood or frequency of 'unusually bad' returns may capture the notion of risk for retail investors better than the standard deviation. Hence, a downside risk measure may be more appropriate, and it may overcome some of the problems associated with non-symmetric return distributions. One of the most recognised measures of downside risk is the semi-deviation, or semi-variance,[27] which measures deviations below a certain benchmark, B:

$$\text{semi} - \text{deviation} = \sqrt{\sum_{t=1}^{T} (r_t - B_t)^2 \Big/ (T - 1)} \text{ for } r_t < B_t.$$

Thus, only the returns that are below the benchmark B are included in the calculations. What would be a suitable benchmark B? The average return on the given investment product may serve as a benchmark (in this case, B may be constant). An alternative is the risk-free rate—the advantage of using a single benchmark for all the investment products is that semi-deviations are easily comparable across products.

In the above definition, T in the denominator may be either the total number of time periods or the number of time periods when the returns were below the benchmark—the choice may depend on the type of benchmark used and on the question examined more generally. For

---

[25] For an overview of investment risk measurement, see Balzer (1995).

[26] This does not pose a problem for the given application if the returns are symmetrically distributed around the mean—ie, if there is no skewness in the distribution of returns or if what is being compared has the same skewness.

[27] For example, Markowitz (1959) commented: 'analyses based on [semi-variance] tend to produce better portfolios than those based on [variance].' Semi-deviation can be shown to encompass various other measures of downside risk, such as the probability of shortfall and the expected shortfall—see Balzer (1995).

example, if the risk-free rate is used as the benchmark, and semi-deviation is to take account of the fraction of return observations below the benchmark, then the total number of time periods may be used in the denominator.

Box 4.1   Illustration of return distribution and estimates of risk and risk-adjusted returns for a sample of four different funds

Distribution of returns

The chart below illustrates the monthly return distribution (covering six years of data) of four funds with different investment objectives: a UK corporate bond fund, a protected/guaranteed fund, a global emerging markets fund, and a UK income equity fund. Two important observations emerge. First, as can be expected, the estimated distributions (probability density functions) and hence risk–return profiles are very different across the investment-style classes. Second, the return distributions are not symmetrical, but exhibit skewness (and kurtosis). This means that the standard deviation may not capture the risk of the funds appropriately.
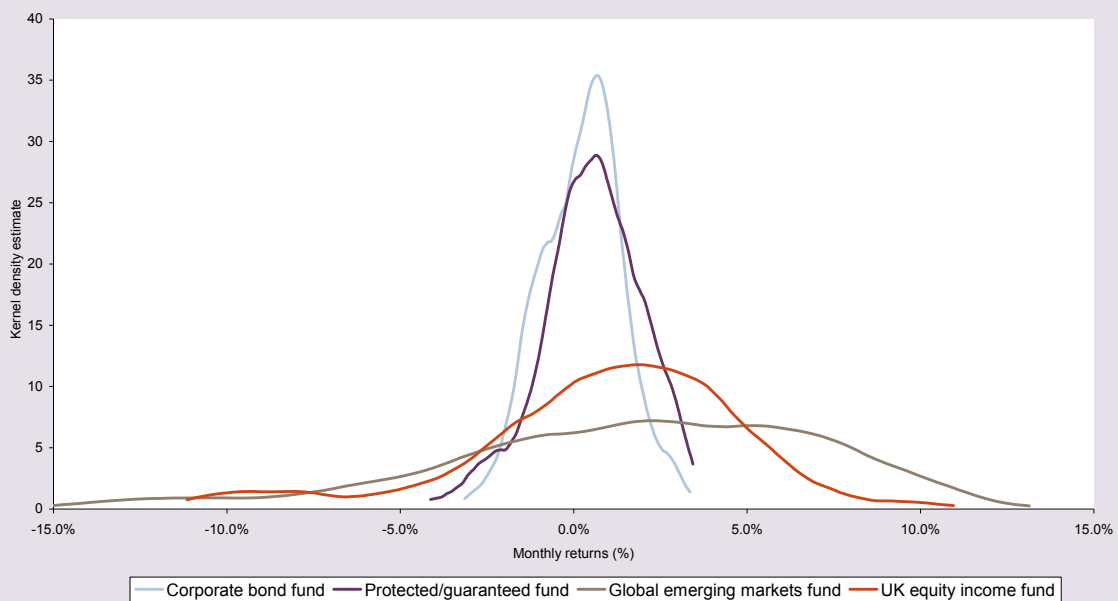
Source: Datastream and Oxera calculations.

Estimates of risk and risk-adjusted returns

The table below shows different estimates of risk (standard deviation and semi-deviations) and risk-adjusted returns, calculated based on the risk estimates (ie, Sharpe ratio and different Sortino ratios—see section 4.3.3 below for a definition). Semi-deviations, and hence Sortino ratios, are calculated with respect to two 'benchmarks' (an average return and a return of–1%) both for the whole sample and for the returns below the benchmark. (The difference between the whole sample semi-deviation and below-the-benchmark semi-deviation is that the former takes into account how many returns in the sample were below the benchmark. The –1% benchmark here was chosen for illustrative purposes only—in practice, a benchmark of 0% or the risk-free rate could be used.)

	Corporate bond fund	Protected/ guaranteed fund	UK equity income fund	Global emerging markets fund
Average return	0.2%	0.5%	0.9%	1.7%
Standard deviation	1.1%	1.4%	3.8%	5.3%
Semi-deviation (below average)	1.3%	1.5%	4.6%	6.0%
Semi-deviation (below −1%)	0.7%	1.6%	4.5%	5.9%
Semi-deviation (below average, whole sample)	0.8%	1.1%	3.0%	4.1%
Semi-deviation (below −1%, whole sample)	0.3%	0.5%	2.2%	2.9%
Risk-adjusted returns				
Sharpe ratio	0.20	0.37	0.23	0.31
Sortino ratio (below average deviation)	0.17	0.33	0.20	0.28
Sortino ratio (below −1% deviation)	0.31	0.32	0.20	0.28
Sortino ratio (below average deviation, whole sample)	0.27	0.49	0.30	0.41
Sortino ratio (below −1% deviation, whole sample)	0.79	1.10	0.40	0.58

Source: Datastream and Oxera calculation.

In this example, the ranking of performance would differ depending on the risk-adjusted return measure chosen. For example, according to the Sharpe ratio and the Sortino ratio based on below-average returns semi-deviation, the ranking of the funds would be (starting with the best-performing): protected/guaranteed fund, global emerging markets fund, UK equity income fund and corporate bond fund. However, according to the Sortino ratio based on below −1% return semi-deviation, the ranking would be: protected/guaranteed fund, corporate bond fund, global emerging markets fund and UK equity income fund.

Thus, in principle, the choice of risk-adjusted return measure may have an impact on the ranking of performance and hence on any conclusions regarding the relationship between performance and charges. Sharpe ratio and Sortino ratio based on below-average returns semi-deviation is likely to produce similar rankings of performance.

4.3.3 Methodologies for risk adjustment

An estimate of investment product risk based on historical returns enables the returns to be adjusted by the risk so that the product performance and the relationship between performance and charges can be assessed. The academic literature has focused on quantifying the risk-adjusted return using a single measure, either via the Sharpe ratio or other more complex models. An alternative way to control for risk when measuring product performance is to classify or group products according to their risk, such that the returns of products within each risk class are directly comparable.

Similar to what was set out in Figure 4.1 above, Figure 4.2 illustrates the two methodologies that can be used to measure risk-adjusted performance. The first step in both cases is investment product risk estimation, as explained in sections 4.3.1 and 4.3.2 above. The following describes the methodologies in more detail.

Figure 4.2   Methodologies for risk adjustment

I: Classification approach

II: Risk-adjustment approach
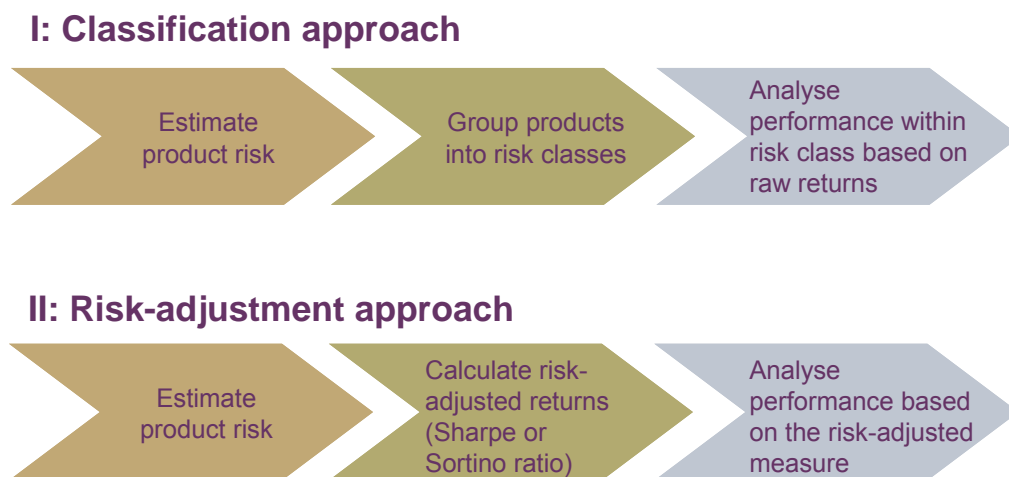
Source: Oxera.

Risk-adjustment approach

Academic research has largely focused on the risk-adjustment approach. The two key methods used have been the Sharpe ratio and, in the majority of the papers reviewed, measurement based on the capital asset pricing model (CAPM). For various reasons, discussed in Appendix 2, the CAPM-based approach or other more complex modelling of risk-adjusted returns is not pursued here.[28] Rather, the focus is on the Sharpe and Sortino ratios.

The **Sharpe ratio** is often used as a measure of risk-adjusted return. Sharpe (1966) proposed to divide the fund's average excess return by the standard deviation of its returns, thus adjusting the return for total risk.[29] A higher Sharpe ratio means either higher excess return or lower risk, or both; a lower Sharpe ratio implies either lower excess return or higher risk, or both.

The Sharpe ratio is measured as the ratio of the difference between product i's expected return and the risk-free rate, and the volatility of the returns.[30] The ratio can be estimated using historical returns—average returns estimate the expected returns, and the standard deviation estimates the risk (ex post volatility).

$$\text{Sharpe}_i = \frac{E(r_i) - r_f}{\sigma_i}$$

The Sharpe ratio is relatively simple to compute and understand. Moreover, it may be considered an attractive way of measuring consumer outcomes since modern financial theory postulates that consumers select financial assets with the greatest Sharpe ratio.[31]

---

[28] Using the CAPM framework for the risk adjustment requires that a number of assumptions regarding the returns are satisfied. Many of these may not hold for the retail investment products, such as normally distributed returns, existence of an appropriate 'market' or 'benchmark' return, etc. Moreover, the application of the CAPM is likely to be impractical given the number and diversity of the investment products examined.

[29] The Sharpe ratio adjusts for the 'total' risk of the product, not just the non-diversifiable risk. If investors hold a diversified portfolio of financial assets, adjustment for the non-diversifiable risk only (rather than for total risk) may be preferable in most applications. This is why academic studies tend to measure risk-adjusted fund performance using CAPM-based models, which take into account that some of the individual fund's risk is diversifiable.

[30] A related measure is the variance coefficient, which is calculated as the ratio of the standard deviation to the expected return. The variance coefficient measures the risk per unit of return, unlike the Sharpe ratio, which focuses on 'excess' returns (ie, returns over and above the risk-free rate).

[31] Modern financial theory is based on an assumption of 'quadratic preferences' of investors, and hence assumes that investors trade off risk and return in the manner described by the Sharpe ratio.

Although a comparison between products using the Sharpe ratio is unambiguous when only one of the variables is different, when both actual returns and volatility differ there is an implied trade-off between the two that is not necessarily grounded in actual consumer/ investor preferences (ie, consumers may not necessarily 'trade off' risk and return in this manner, even though this assumption underlies finance theory). As a result, the Sharpe ratio may not produce an unambiguous consumer ranking when products differ significantly along both the risk and return dimensions. For example, the Sharpe ratio is the same (0.5) for product A with average excess return of 2% and volatility of 4% and product B with average excess return of 4% and volatility of 8%, but investors may not necessarily rank the products A and B in an unambiguous manner (eg, some may prefer A, while others may prefer B).

The Sharpe ratio is also subject to the same criticisms as the standard deviation as a measure of risk—ie, if the standard deviation, or volatility, of returns does not capture the risk well then the Sharpe ratio is flawed as a measure of risk-adjusted returns. This may be addressed by using other ratios, such as the Sortino ratio, which is based on the semi-deviation as an alternative measure of the risk.

The **Sortino ratio** has a similar rationale and can be calculated in a similar way to the Sharpe ratio, except that the risk is measured by the downside rather than the overall volatility of returns. Thus, the semi-deviation of returns, rather than standard deviation, is used as an estimate of downside risk:

$$\text{Sortino}_i = \frac{E(r_i) - r_f}{\text{semi} - \text{deviation}_i}$$

The ways that the semi-deviation can be estimated were set out in section 4.3.2 and illustrated in Box 4.1 (eg, as a downside deviation with respect to the average return, or with respect to a different benchmark).

It is possible, in principle, that different conclusions are drawn regarding the relationship between performance and charges, depending on whether risk-adjusted returns are measured by the Sharpe or the Sortino ratio(s).[32]

Using Sharpe and/or Sortino ratios for measuring the risk-adjusted returns imposes an important restriction on the way consumers are assumed to trade off return and risk (ie, as a ratio), and the conclusions drawn about the relationship between risk-adjusted returns and charges are necessarily constrained by this assumed 'valuation' by the consumers. This is one of the limitations of using either Sharpe or Sortino ratios as a measure of risk-adjusted returns. Because this valuation may be idiosyncratic to individual consumers, it may be necessary to report the analysis along the two dimensions of risk and return separately. The risk classification approach enables this to be done while still controlling for risk.

Classification approach

The risk classification approach entails classifying products according to their risk, as estimated by either standard deviation or semi-deviation. Raw (net) returns can then be used to test the different hypotheses within each 'risk class'. In this way, risk is controlled for when analysing the returns and their relationship with charges—at the same time, no structure is imposed on how consumers trade off risk and return.

Another attractive feature of the classification approach is that hypotheses can be tested with respect to the two elements of product performance separately—that is, a relationship between returns and charges can be analysed, while controlling for risk, but also the

---

[32] In fact, different conclusions regarding the relationship between charges and risk-adjusted returns depending on the metric used may provide an insight into *what* kind of risk-adjusted performance (if any) is affected by charges. For example, if no statistically significant relationship is found between the Sharpe ratio and charges, but a positive relationship is found between the Sortino ratio and charges, it may be concluded that higher charges do not 'fund' reductions in overall volatility of returns, but that they deliver lower downside variability, which may improve consumer outcomes.

relationship between risk and charges can be examined, while controlling for returns. The latter would require a classification according to returns, rather than risk (ie, classification into average returns classes), and then analysing the relationship between charges and the risk metric.

Once the product risk is estimated either by the standard deviation or the semi-deviation of returns, the products can be ranked and divided into quartiles or deciles (depending on the size of the sample)—this ensures that products with similar risk measures are within the same risk class. An alternative to simple classification based on quartiles or deciles is to perform a standard cluster analysis or apply latent class models to group 'similar' products together—in this case, several measures of risk can be used simultaneously to establish similarity in terms of estimated risk.

Thus, the risk classification approach is appealing in that it is intuitive and simple to apply, and enables the hypotheses regarding the two elements of product performance—risk and return—to be tested separately. However, there are two main difficulties with using this approach to control for risk. First, if there are not many products within each risk class, the analysis of the relationship between performance and charges may not be reliable from a statistical point of view. Second, it can be difficult to aggregate the estimated relationships across the risk classes, which in turn makes it difficult to obtain an overview of the results and to draw conclusions. [33] One way around this is to present a 'count' for how many risk classes (or products) the estimated relationship was positive, negative, or zero.

4.3.4 Practical issues in performance measurement
Net return measurement
In the literature, net returns are measured by the changes in NAV, or bid prices, and assuming that all distributions (eg, dividends) are reinvested. For example, Cuthbertson et al. (2006) demonstrated that the net returns, r^{net}, are calculated as follows:

$$1 + r^{net}_t = \frac{NAV_t}{NAV_{t-1}} \prod_{j=1}^{J} (1 + \frac{DIST_j}{RENAV_j})$$

where NAV_t is the net asset value of the fund at the end of period t, J is the number of dividend or capital gains distributions during the period, $DIST_j$ is the jth distribution (in money terms) and $RENAV_j$ is the NAV at which the jth distribution was reinvested. Net returns exclude all fund expenses and asset transactions costs. Gross returns can be obtained by adding TER to the net return over the corresponding period.

Fletcher (1999) has an alternative presentation of the formula for the (net) monthly returns of UK unit trusts:

$$r^{net}_t = \ln(\frac{p_{t+1} + d_t}{p_t})$$

where p_t is the offer price of the unit trust at the start of month t, d_t is the gross dividend paid (in the ex-dividend month) in period t, and it is assumed that dividends are reinvested at the end of the month (ie, at p_{t+1}).

The dataset that is being put together allows the calculation of these return measures on a monthly basis, given that it contains monthly time-series data of NAV, bid prices and distributions for UK unit trusts, OEICs and unit-linked funds.

---

[33] Even if a cross-section regression is run for each risk class, the magnitude of the coefficients is meaningful only within each risk class, since the coefficients are likely to differ in magnitude across risk classes due to higher risk classes having a higher level of return. The rank correlation coefficient may be aggregated across the risk classes, but there is no meaningful interpretation of the aggregate measure.

Once monthly raw returns are calculated from the price and distributions data, **average returns and risk** can be measured based on the time series. Typically, no less than three years of monthly returns are used (ie, 36 observations) for the estimation of average returns and risk over time. Depending on the hypothesis that is examined, average returns and risk can be calculated over 3–5 years or over a decade or more. For example, if the question concerns a performance comparison over time, it may be preferable to use average return and risk over 3–5 years than over a longer time period, and then compare the averages over time.

When calculating the raw returns for further analysis and hypothesis testing, it is important to check whether there are any **outliers** that may require some adjustments to the data—there may, for example, be outlier returns in specific periods for a given investment product, or an investment product may be an outlier itself in terms of average return and risk. Outliers may need to be removed because they are likely to represent realisations of highly unlikely events, and statistical techniques can be adopted to identify those outliers. Here it is useful to recall that it is the *expected* performance in terms of returns and risk that captures consumer outcomes, and historical data is merely used to estimate these variables. As a result, it is possible that historical data contains observations which are outliers in that they are unlikely realisations of the underlying process.

Survivorship-bias-free measurement

Academic research documents widely that the exclusion of 'dead' (eg, merged or liquidated) investment funds from studies of returns over time can induce a serious survivor bias in the reported performance measures. That is, the exclusion of funds that did not survive until the end of the sample period tends to bias upwards the performance estimates. This, in turn, may lead to an incorrect finding of a positive relationship between the performance and charges where there is none.[34] For example, Blake and Timmermann (1998) estimate that survivor bias is around 0.8% per annum for UK unit trusts.[35] Quigley and Sinquefield (2000) report a survivor premium of 2.3% per annum and a survivor bias of 0.7% per annum for UK unit trusts investing in UK equity.

In order to avoid survivorship bias when measuring performance, the sample should include all funds in existence at some time during the sample period, even if they were subsequently liquidated or merged into other funds, and did not survive until the end of the sample period.

The key practical issue associated with survivorship-bias-free measurement is how to measure performance and risk of a fund if the fund was either liquidated or merged during the period of measurement. There are a few principles that may be followed when constructing a survivorship-bias-free dataset. For the purpose of this analysis, the ideal approach would be to track the assets directly, thus mirroring what would have been likely to happen to actual investors in these funds. Where this is not possible, some reasonable assumptions need to made about how investors would be likely to react to the merging or closure of a fund.

– If a fund merged into another fund during the measurement period, two approaches can be followed (Gruber 1996).

– It can be assumed that investors place their money in the fund that continued to exist after the merger.[36] Thus, the returns of the target fund post-merger would be the same as those of the acquiring fund.

---

[34] See Fletcher (1999), p. 463.

[35] Blake and Timmermann (1998) also find that survivorship bias tends to be more acute for unit trusts investing in international equity than for unit trusts investing in UK equity.

[36] There is some limited evidence that the target fund investors' money tends to stay within the acquiring fund following the merger. See Gruber (1996), Allen and Parwada (2006), Jayaraman, Khorana and Nelling (2002).

– It can be assumed that investors place their money in the average surviving fund within the same product and investment-style class. Thus, the returns of the target fund post-merger would be equal to the average returns in the given investment-style class.[37]

– Fund performance tends to be 'abnormal' around the time of termination (ie, merging into another fund or liquidation). For example, Blake and Timmermann (1998) find that non-surviving funds tend to pay negative risk-adjusted returns during the months before their termination. Although it would be possible to remove these observations from the analysis, given that these 'abnormal' performances will have been experienced by investors, this does not seem appropriate for analysing the general relationship between charges and performance.

– Funds that were in existence for only part of a year tend to be dropped from the analyses altogether, since the data is unlikely to be reliable (eg, Malkiel 1995).[38]

Methods similar to those adopted in the academic research can be applied when conducting empirical analysis on the dataset that is being put together for this study. The dataset does contain information on all funds (in existence at some point over the past ten years and possibly longer), and where funds have been liquidated or merged, this is indicated in the data.

4.4 Methodology for empirical testing of the relationship between performance and charges

In line with the framework outlined in Figure 4.1, once returns and risk have been measured and the risk-adjustment undertaken, the hypotheses about the relationships of interest can be tested. In particular, it can be tested whether higher charges deliver better investment product performance.

Given that any better performance is likely to be related to better or more extensive investment management, the main tests focus on establishing the link between performance and charges at the provider level. As done in the literature, this means examining the relationship between returns and the stated AMC or the TER. As discussed below, it is, however, possible (data issues aside) to extend the analysis to examine links with distribution and the total charges paid for by consumers. (In this case, better performance for consumers can arise if, by spending more on the distribution process, consumers buy a better-performing fund given any level of management charge at the provider level.)

The following sets out the empirical analysis to be undertaken and hypotheses to be tested to provide information on the link between performance and charges.

4.4.1 Methods for measuring the relationships
As already outlined in section 4.2 and Figure 4.1, there are two main types of approach to empirically examining the link between product performance and charges (or other characteristics): regression and ranking. Both approaches are valid, and both have their advantages and disadvantages.

Ranking approach to testing relationships
In essence, the ranking approach entails ranking products according to their performance, and, separately, according to the level of charges (or other characteristics of interest), and

---

[37] For example, if the investment product were a unit trust investing in UK income equities, the returns on the product after the merger into another fund would be equal to the average returns of the surviving unit trusts investing in UK income equities.

[38] If the fund was liquidated during the measurement period and was in existence for less than three years during the measurement period, dropping this fund should also be considered, as the number of return observations may be insufficient for reliable measurement.

then comparing the two rankings. The more similar is the ranking according to performance to the ranking according to charges, the higher the correlation between the two variables, and vice versa.

Although the ranking approach can be implemented by simply observing whether there is similarity in the rankings according to two variables, a more robust way is to perform standard statistical tests on the level of correlation between the rankings. One of the most common statistical tests for this purpose is the Spearman rank correlation test, which produces a correlation coefficient between the rankings of two variables, and takes a value between –1 and +1. Moreover, a significance test is also usually reported by standard statistical packages, indicating whether the measured correlation is statistically significant from zero.

The main advantage of the ranking approach to testing the relationship between performance and charges is that the rank correlation test is non-parametric and does not require any assumptions about the distribution of the variables to be satisfied. Hence, rank correlation tests may be preferred to the regression approach when the regression model is likely to be misspecified.

However, there are a few possible drawbacks to using the ranking approach. For example, other potentially relevant factors, such as fund size, cannot be controlled for when measuring the relationship between performance and charges. Another possible drawback is that, where the ranking approach is adopted separately to products within different risk classes, the estimated rank correlation coefficients cannot be aggregated (averaged) across the risk classes in a meaningful way.

Regression approach to testing relationships

The regression approach involves regressing the (risk-adjusted) returns on the level of average charges and/or other characteristics, such as size of the fund. The following univariate specification can be estimated based on cross-section data (ie, data for each investment product in a single given time period):

$$r_i = \alpha + \beta x_i + \varepsilon_i$$

where r_i is the risk-adjusted return of product i (or the raw return, if the risk classification method has been used to control for risk and the regression is run for each risk class), and x_i is the characteristic of interest, such as charges. β, the coefficient estimated from the regression, then measures the relationship between risk-adjusted returns and the characteristic of interest for the given sample of investment products. x_i can also be a vector of investment product characteristics (eg, charges, fund size, investment style, and turnover), in which case, a multivariate regression is estimated. In this way, other factors or characteristics that are potentially related to product performance can be controlled for.[39]

The attractiveness of the regression approach is that it is a robust method for measuring the association between two or more variables. Even if the relationship between returns and charges is of key interest, the regression framework enables other factors that may potentially influence the returns to be controlled for. Other advantages are that the regression coefficients, β, can be tested for statistical significance, as well as 'aggregated' across

---

[39] Instead of standard regression techniques, SFA could be used. This approach is an extension of the econometric approaches discussed in this report, but involves replacing the 'standard' error term with a composite error term comprising two components: a noise component and an inefficiency component. This has the advantage of removing noise, as far as possible; such noise might be due to data measurement issues, luck by the fund manager, atypical performance at a given point, etc. See Annaert, den Broeck and Vennet (2001) for a further discussion and application of SFA techniques to evaluate the performance determinants in a sample of European equity mutual funds. Non-parametric approaches, such as DEA, may also be applied to understand what drives fund performance. The advantage of these non-parametric approaches is that they do not require specific assumptions about the functional relationships. It is also possible to use semi-parametric approaches that attempt to increase the flexibility of the parametric approaches.

different classes, if the regression is done separately for different classes or groups of products (eg, according to risk classification).

However, the regression approach requires a number of assumptions to be satisfied for the results to be unbiased and meaningful. For example, the error terms in the regression need be homoscedastic and normally distributed; and no relevant variables should be omitted from the regression. Hence, as with all regression analysis, it is crucial to perform diagnostic tests in order to validate the use of the specified regression model and, if these fail, to re-specify the model as appropriate. Moreover, a cross-section regression should be implemented on a sample of reasonable size (in contrast, fewer observations are needed for the Spearman rank correlation test).

The following sets out how the main questions of interest can be examined by formulating a hypothesis and then testing it using either the ranking or regression approach.

4.4.2 Testing the link between management charges and performance
There are two separate empirical questions on the link between charges and performance.

–   Do products with higher stated charges at the time of purchase deliver better performance in the future? Put differently, do stated charges at the time of purchase have any predictive power about performance? The relevant test relates disclosed charges at the time of purchase (eg, the AMC) to future performance (eg, returns generated over five years following the purchase).

–   Does more expensive management deliver higher returns? Put differently, is there a contemporaneous relationship between actual charges incurred and performance? The relevant test relates the TER over a period to the returns generated during the same period.

The following sets out both types of test, although it is of note that the academic literature has focused on the second type.

Hypothesis: Products with higher stated charges at the time of purchase do not deliver better performance in the future
The key question captured by this hypothesis is whether product charges disclosed or prevailing at the time of the purchase decision have predictive power in terms of future performance. If the evidence shows that higher charges are associated with superior future performance (ie, there is a positive relationship between the two), then it could be concluded that the disclosed or prevailing charges are a signal of future performance (ie, the hypothesis can be rejected). However, if the evidence shows no relationship (ie, it is statistically insignificant) or a (statistically significant) negative relationship between the two, then it can be concluded that higher disclosed or prevailing charges do not lead to better outcomes in terms of future performance.

The hypothesis can be tested by measuring the relationship between charges at time t and average (risk-adjusted) returns over the next n periods, t+1 … t+n. Since the charge that is typically disclosed (or readily available to the consumer) at the time of purchase is the AMC, it is the AMC at time t for each product that can be compared against the performance in the subsequent periods. However, the TER at time t may also appropriate.

Therefore, in order to measure and test the relationship of interest, the following analyses can be undertaken.

–   The universe of products may be classified according to their type (eg, unit trusts) and investment style (eg, IMA classification), and the analysis may then proceed for each class separately.

- Product risk is estimated, based on the returns data—standard deviation or semi-deviation is calculated based on monthly returns in period t+1 to t+n. There are then two ways to proceed for risk adjustment:

  - products are either sorted according to the risk measure and grouped into risk classes (eg, deciles); or
  - the Sharpe or Sortino ratio is calculated for each product, based on the estimated risk measure and average monthly returns in period t+1 to t+n.

- The relationship between AMC for each product at time t and the average (risk-adjusted) performance over period t+1 to t+n is quantified and tested. One way to measure the relationship is to regress the net (risk-adjusted) returns on the AMC—the estimated coefficient is the measure of the relationship for each investment style and/or risk class. The regression coefficients can be tested for statistical significance, and they can be averaged across investment-style and/or risk classes to obtain an estimate of the 'aggregate' relationship. Another way to measure the relationship is to perform a rank correlation test between the average product performance from t+1 to t+n and the AMC at time t.

Box 4.2 sets out the steps for testing the hypothesis using the risk-adjusted returns (Sharpe or Sortino ratios) and regression analysis, as an example of one of the possible approaches.

Box 4.2   Example: Testing the hypothesis that more expensive products do not deliver better performance in the future

Hypothesis testing using Sharpe or Sortino ratios and regression analysis
Suppose the following data is available, corrected for survivorship bias: monthly returns of each investment product for at least four years, from t to t+n (n≥3), and annual data on the AMC over the same time period.

- Calculate the average return and standard deviation (or semi-deviation) over the time period t+1 to t+n. The average returns ($r_{t+1...t+n}$) and standard deviations ($sd_{t+1...t+n}$) can be monthly or annualised.

- Calculate Sharpe or Sortino ratios for each product over the time period t+1 to t+n:
  S-ratio$_{t+1...t+n}$=$r_{t+1...t+n}$ / $sd_{t+1...t+n}$, where sd is either standard deviation or semi-deviation of returns, and S-ratio is either Sharpe ratio or Sortino ratio.

- Classify products according to investment style (eg, using IMA classification). The analysis proceeds within each class—ie, the relationship between prevailing charges and subsequent performance will be measured for each class separately.

- Estimate the following regression (for each IMA class), where i denotes an investment product:[40] S-ratio$_{t+1...t+n, i}$ = α + β·AMC$_{t, i}$ + ε_i

- β, the estimated slope coefficient, measures the relationship between AMC and future (risk-adjusted) performance of the investment products within each IMA class. A statistical significance test on the estimated β confirms whether the hypothesised relationship exists: if β is not significantly different from zero, product charges, as disclosed or prevailing at the time of the investment decision, have no predictive power for future performance (ie, hypothesis can be accepted). The hypothesis also can be accepted if β is statistically significant but negative, whereas the hypothesis can be rejected if β is statistically significant and positive.

[40] Diagnostic tests on the econometric model should be performed, to ensure that the standard OLS assumptions, such as homoscedastic and normally distributed error terms, are satisfied. If the regression model is misspecified, alternative formulations need to be considered, such as the inclusion of additional explanatory variables in the regression, or performing a rank correlation test instead of the regression.

Possible extensions

– It may be possible to track over time the relationship between the disclosed AMCs and future performance of the fund by estimating a set of cross-section regressions:

$$\text{S-ratio}_{t+1...t+n,\, i} = \alpha_1 + \beta_1 \cdot \text{AMC}_{t,\, i} + \varepsilon_{1,\, i,}$$
$$\text{S-ratio}_{t+2...t+n+1,\, i} = \alpha_2 + \beta_2 \cdot \text{AMC}_{t+1,\, i} + \varepsilon_{2,\, i,} \text{ etc}$$

– The relationship between disclosed AMCs and future performance may also be tracked over time by using panel data analysis: $\text{S-ratio}_{t+1...t+n,\, i} = \alpha + \beta \cdot \text{AMC}_{t,\, i} + \varepsilon_{t,\, i,}$ where S-ratio$_{t+1...t+n,\, i}$ would be a moving average of the Sharpe ratio or Sortino ratio.[41] Further extension could also be considered within this model. For example, coefficients can be allowed to vary over time in either a parametric or non-parametric fashion.

– In order to obtain an 'aggregate' relationship between the disclosed AMCs and future performance, the β estimated for each investment-style class (eg, IMA class) can be averaged, on either an equal- or value-weighted basis. An alternative is to estimate a single cross-section regression including all investment products, where the explanatory variables also include dummy variables for each investment class. In this manner, a single β estimate is obtained while potential differences between products are controlled for by inclusion of the dummy variables.

– The regression framework[42] allows other variables that may affect performance to be controlled for, such as size and turnover. Hence, additional variables may be included as the explanatory variables for the risk-adjusted performance, such as average total net assets (TNA) or average turnover during the same time period over which the performance is measured:

$$\text{S-ratio}_{t+1...t+n,\, i} = \alpha + \beta \cdot \text{AMC}_{t,\, i} + \gamma_1 \cdot \text{TNA}_{t+1...t+n,\, i} + \gamma_1 \cdot \text{turnover}_{t+1...t+n,\, i} + \varepsilon_{t,i}$$

Hypothesis: Higher charges are not associated with higher net returns

This hypothesis tests whether higher charges of a product reflect better management of the underlying assets—ie, do more expensive products deliver higher risk-adjusted net returns? For the hypothesised relationship to exist (ie, no link between charges and performance), one would need to observe either a negative or no relationship between charges and contemporaneous returns. If the data analysis showed a statistically strong positive relationship, then the hypothesis may be rejected and the conclusion reached that higher-priced investment products do deliver higher risk-adjusted net returns.

The hypothesis can be tested by measuring the relationship between average charges over the period from t to t+n and average (risk-adjusted net) returns over the same time periods. n would typically be no less than two (ie, the average returns and charges should be measured over at least a three-year period). Since the TER captures the actual expenses charged

---

[41] These approaches can also be applied using unadjusted returns and controlling for risk in the regression. In the panel data set-up, the simple specification could, for example, take the form: $r_{t+1...t+n,\, i} = \alpha + \beta \cdot \text{AMC}_{t,\, i} + \gamma \cdot \text{semi-deviation}_{t+1...t+n,\, i} + \varepsilon_{t,\, i,}$ where $r_{t+1...t+n,\, i}$ would be a moving average of returns, and semi-deviation$_{t+1...t+n,\, i}$ would be a moving average of risk. Again, within this approach, the coefficients can be allowed to vary over time. Instead of standard regression approaches, SFA techniques can be applied by extending the model to decompose the error term $\varepsilon_{t,\, i} = u_{i,t} + v_{i,t}$, where $u_{i,t}$ is a zero mean error random error and $v_{i,t} \geq 0$, capturing the effect of omitted variables (given that data on some variables may not be available) as well as the level of inefficiency of the product compared with the benchmark portfolio. For an application of SFA techniques in the fund market, see, for example, Annaert, den Broeck and Vennet (2001). Further extensions to this model could also be considered, including parameterisation of $u_{i,t}$ with the aim of establishing, for example, whether charges affect the inefficiency.

[42] Rather than using a regression approach, it may be possible to use non-parametric approaches such as DEA. For example, if products are grouped within bands of charges (eg, high, medium and low), DEA may be used with the two dimensions of risk and return as the key outputs, with controls for various facts that are likely to determine those outputs. No significant or systematic differences in the estimated DEA frontiers between the different charges groups would indicate that higher charges do not result in better performance.

against a fund—and hence the actual 'cost' of delivering the net returns to consumer—the TER are the relevant measure for testing this relationship (as is also done in the literature).

Therefore, in order to measure and test the relationship of interest, the proposed framework and the methodologies can be applied to the available data, similar to what was described above, but the main difference being that the analysis focuses on the *contemporaneous* link between performance and charges (here the TER).

–   The universe of products may be classified according to their type (eg, unit trusts) and investment style (eg, IMA classification), and the analysis would then proceed for each class separately.

–   Product risk is estimated, based on the returns data—standard deviation or semi-deviation is calculated based on monthly returns in years t to t+n. Then there are two ways to proceed for risk adjustment:

    –   either products are sorted according to the risk measure and grouped into risk classes (eg, deciles); or
    –   the Sharpe or Sortino ratio is calculated for each product, based on the estimated risk measure and average monthly returns in years t to t+n.

–   The relationship between the average TER over period t to t+n for each product and average (risk-adjusted) performance over the same period t to t+n is quantified and tested. One way to measure the relationship is to regress the net (risk-adjusted) returns on the average TER, possibly including other relevant factors, such as fund size—the estimated coefficient is the measure of the relationship for each investment style and/or risk class. The regression coefficients can be tested for statistical significance, and can be averaged across investment style and/or risk classes to obtain an estimate of the 'aggregate' relationship. Another way to measure the relationship is to perform a rank correlation test between the average product performance and average TER for each product.

Box 4.3 sets out the steps for testing the hypothesis using the risk classification approach for controlling for risk and the rank correlation for measuring and testing the significance of the hypothesised relationship, as an example of one of the possible approaches (unlike Box 4.2, which illustrated how to test relationships using the risk-adjustment and regression approach).

Box 4.3   Example: Testing the hypothesis that higher charges are not associated with higher net returns

Hypothesis testing using classification approach and rank correlation test
Suppose that the following data is available, corrected for survivorship bias: monthly returns of each investment product for at least three years, from t to t+n (n≥2), and annual data on TER over the same time period.

–   Calculate average returns, standard deviation (or semi-deviation) and average TER for each investment product over the time period *t* to *t+n*. The average return and standard deviation can be monthly or annualised.

–   Classify products according to investment style (eg, IMA classification). The analysis proceeds within each class—ie, the relationship between current charges and future performance will be measured for each investment-style class separately.

–   Sort the products according to their measure of risk (standard deviation or semi-deviation of returns), and group them, for example, into deciles or quartiles, depending on the number of products within each class.

–   For each risk class, perform a rank correlation test (eg, Spearman rank correlation test)

on the (raw) returns and TER of each product.

–   The estimated rank correlation coefficient measures the relationship between the TER and the performance of the investment products within each risk class. A positive correlation coefficient shows that there may be a positive association between net returns and the TER (controlling for risk), and a negative correlation coefficient shows that there may be a negative association. A statistical significance test on the estimated rank correlation coefficient confirms whether the estimated relationship is statistically significant.

–   Given that this will produce several rank correlation coefficients within each investment-style class (eg, ten, if the products were classified into risk classes by splitting them by deciles), it may be desirable to summarise these results in order to draw overall conclusions about the direction and strength of the estimated relationship between performance and charges. This can be done by counting the number of statistically significant positive and negative rank correlation coefficients, and statistically insignificant coefficients. This summary can then be analysed and it can be concluded whether there is sufficient evidence to accept or reject the hypothesised relationship.

Possible extensions
–   Rather than measuring and testing the relationship between returns and charges, the relationship between risk and charges can be examined. To do so, products can be classified into 'return classes' and the rank correlation test can be performed on the risk and charges of each product, within each 'return class'.

4.4.3 Testing the link between portfolio turnover and performance

In addition to testing the link between performance and charges, other determinants of performance can be evaluated, such as portfolio turnover. In the academic literature, it is recognised that turnover of the underlying assets is a proxy for the trading costs incurred in managing the underlying assets. Therefore, testing the link between portfolio turnover and performance can be seen as a way of determining whether more actively managed funds deliver superior performance—ie, whether the additional trading costs incurred are more than compensated by either higher investment returns or lower volatility of those returns, or both.

Since the new dataset will contain data on portfolio turnover, it will be possible to conduct tests on the relationship between portfolio turnover and product performance.

Hypothesis: Products with higher turnover do not deliver higher net returns
The hypothesis can be tested by measuring the relationship between average turnover over the period from t to t+n and average (risk-adjusted) returns over the same time periods. n would typically be no less than three (ie, the average returns and charges should be measured using monthly data over at least a three-year period).

As with the hypothesised relationship between charges and performance described above, the analysis can be implemented using either risk-adjusted returns or a risk classification to control for risk when comparing returns, and then using either regression or ranking approach to estimate the direction and strength of the relationship.

–   The universe of products may be classified according to their type (eg, unit trusts) and investment style (eg, IMA classification), and the analysis would then proceed for each class separately.

–   Product risk is estimated, based on the returns data—standard deviation or semi-deviation is calculated based on monthly returns in period t to t+n. Then there are two ways to proceed for risk-adjustment:

    –   either products are sorted according to the risk measure and grouped into risk classes (eg, deciles); or

– the Sharpe or Sortino ratio is calculated for each product, based on the estimated risk measure and average monthly returns in years t to t+n.

– The relationship between average turnover over period t to t+n for each product and average (risk-adjusted) performance over the same period t to t+n is quantified and tested. As before, one way to measure the relationship is to regress the net (risk-adjusted) returns on the average turnover, possibly including other relevant factors such as fund size. Another way to measure the relationship is to perform a rank correlation test between the average product performance and average turnover for each product within each class.

A statistically insignificant or a statistically negative relationship between turnover and performance would indicate that funds with higher turnover (ie, funds that are more actively managed) do not deliver increased value to consumers in terms of superior performance.

4.4.4 Testing additional hypotheses regarding product performance
The following sets out additional hypotheses that could in principle be tested in relation to product performance, in line with the issues discussed in sections 2 and 3.

Hypothesis: Distribution channel affects investment product performance or charges
The potential link between performance and charges discussed so far has focused on charges at the fund level (AMC, TER), with returns measured net of these charges rather than the total charges that consumers may actually pay. While charges at the distribution level will clearly affect the net returns actually earned by consumers, it would seem less obvious how distribution (unlike fund management) could affect the gross returns of a product.

Nonetheless, data permitting (and in line with the discussion in section 3), there is a case for examining net (risk-adjusted) returns earned by consumers using different distribution channels. If the distribution channel does affect total charges—and, hence, net performance—then systematic differences would be expected to be observed in average returns and average total charges across distribution channels.

In principle, a regression approach could be used to estimate the impact of the distribution channel on performance or total charges. The more straightforward method would be to compare the average performance or total charges across the distribution channels, although this implies that other relevant factors cannot be controlled for when examining the hypothesised relationship.

As discussed in section 3, the empirical analysis is made difficult by data limitations on charging structures within distribution channels and on shifts in consumer purchases across distribution channels over time. Products will not generally have one 'main' distribution channel, but can be distributed via various channels. Hence, hypotheses similar to those on the relationship between net risk-adjusted returns (net of fund-level charges) and charges at the fund level may not be testable.

However, if the distribution channel systematically affects performance and fund-level charges (stated AMC and TER) and if some products have one main distribution channel, this would have implications for the testing of the other hypotheses. Like other relevant characteristics, the distribution channel may therefore need to be controlled for to examine the relationship between product performance and TER (or stated AMC).

Hypothesis: Brand does not affect product performance
As set out in section 2.1, there may be aspects other than risk and return that may determine the quality of consumer outcomes in the retail investment products market, such as brand, flexibility or transparency. Within the framework of analysis proposed, it is possible to test for the cost or 'price' of these characteristics in terms of the risk–return performance of a

product. However, the framework does not allow the value of the characteristic to the consumer to be derived. In principle, if the consumer's valuation of the characteristic is higher than the price (as measured by the characteristic's association with the risk-adjusted net returns), this characteristic would be worthwhile.

While the importance of quality aspects, such as 'transparency', may be more difficult to test given the lack of data to classify products according to these aspects, testing the importance of brand should, in principle, be possible.

As discussed in section 3.3.3, the empirical analysis would involve classifying products according to the strength of the brand of a product (or its provider), and then testing for the relationship between brand strength and performance. Similar to the tests on the link between performance and charges (or turnover) described above, the approach could involve either a comparison based on ranking or a regression of performance on an indicator variable of brand strength. The regression would allow control for other variables that may affect performance, such as charges, fund size and turnover:

$$r_{t,\,i} = \alpha + \beta \cdot brand_i + \gamma_1 \cdot TER_{t,\,i} + \gamma_2 \cdot TNA_{t,\,i} + \gamma_3 \cdot turnover_{t,\,i} + \varepsilon_{t,i}$$

where r denotes (risk-adjusted) returns, brand is the indicator variable of interest, and TER, turnover and TNA control for charges, portfolio turnover and fund size. The returns can be measured by risk-adjusted returns (eg, using Sharpe or Sortino ratios) or by raw net returns if the risk is controlled for by undertaking the analysis separately for each risk class.

4.4.5 Testing the link between performance and fund inflows

In addition to testing the hypotheses regarding the link between product performance and charges, it may be of interest to test whether 'money is smart'—ie, whether consumers purchase products that perform better (or have performed better) in terms of delivering higher net (risk-adjusted) returns.

A similar test was discussed in section 3.3.2 regarding whether consumers tend to purchase products that have lower charges (and whether there is evidence that this trend has increased over time).

Hypothesis: Consumers do not tend to purchase the better-performing products

In order to test the relationship between current fund inflows (F) and current or past performance, as measured by returns, a regression of the following simplified form can be estimated:

$$F_{t,\,i} = \alpha + \beta \cdot r_{t-1,\,i} + \gamma \cdot F_{t-1,\,i} + \varepsilon_i$$

where F denotes fund inflows and r is the (risk-adjusted) return, measured in the current period t or in previous periods (t-1) for each product i.[43] Fund flows and returns in earlier periods (t-2, t-3, etc) can be included in the regression analysis in addition to other control variables that may influence fund inflows (see section 3.3.2).

In the dataset that is being put together, no historical data is available on fund inflows (gross or net). However, as a proxy measure, total net assets can be used to construct a flow measure, by calculating the change in total net assets, adjusted for by the return earned over the period (see section 3.3.2).

A positive and statistically significant estimate of the β coefficient(s) would result in the rejection of the hypothesis and instead indicate that fund flows are positively related to performance—ie, consumers tend to purchase funds that perform (or have performed) better.

---

[43] For example, Gruber (1996) adopts a similar regression-based approach to test whether fund flows are predictable.

A1 Summary of dataset

In a separate project commissioned by the FSA, a dataset is being put together drawing from the data available from third-party vendors (see Table A1.1). At the time of drafting this study, the dataset had not yet been completed, so the list of available data is indicative only.

The dataset, when complete, will cover data on charges, fund performance and other fund characteristics for unit trusts, OEICS and unit-linked funds. Initial charges (and exit charges) are available for funds domiciled in the UK only, but data on the other variables is expected to be available for all funds sold in the UK, irrespective of domicile. The time period covered includes the past ten years (possibly longer), and the data on performance (and other characteristics) is available on a monthly basis. Separate performance data at share class level is available for funds with multiple share classes.

Table A1.1 Summary of dataset

Data item	Comments
Fund identifier	SEDOL, ISIN, etc
	Identifiers available to associate share classes with fund
Status	Active, merged, liquidated, with date of liquidation or merger
Launch date	
Other fund characteristics	Product type, investment style, IMA sector, etc
Management company	Name and contact information
Successor name for merged funds	
Fund manager benchmark name	
Asset class codes	
Annual management charge (AMC)	Annual, 10 years
Total expense ratio (TER)	Annual, 10 years
Initial charge	Quarterly, 10 years. Exit charge where applicable
Prices	Monthly NAV (or bid-offer prices), 10 years
Dividends/distributions	Dividend per share and date of distribution, 10 years
Portfolio turnover	Annual, 10 years
Total net assets	Monthly, 10 years

Source: Based on information provided by consultant commissioned by FSA to put together the dataset.

A2   Review of relevant literature

The following sets out further details on the measurement of risk-adjusted returns, as proposed in the literature. It also summarises the studies on the relationship between charges and performance.

A2.1   Measurement of risk-adjusted returns

The need to adjust return by the risk is emphasised in academic empirical studies, which tend to use risk-adjusted performance measures to examine questions on performance and fund characteristics, performance persistence and survivorship bias. Various approaches to the risk adjustment of returns have been used in the academic literature. The two key ones are the Sharpe ratio as a measure of risk-adjusted return, and regressions based on the CAPM, often called Jensen regressions, which is a single-factor model. There are also other methods for estimating risk-adjusted returns, such as Fama–French regressions (a multi-factor model). These methods are briefly reviewed and discussed below.

A2.1.1   Sharpe ratio
The Sharpe ratio is often used as a measure of risk-adjusted return. Sharpe (1966) proposed dividing the fund's excess return by the standard deviation of its returns, thus adjusting for total risk.[44] A higher Sharpe ratio means either higher excess return of lower risk, or both; whereas a lower Sharpe ratio implies either lower excess return or higher risk, or both.

The Sharpe ratio is relatively simple to compute and understand. Moreover, it may be considered an attractive measure of consumer outcome, since modern financial theory postulates that consumers select financial assets with the highest Sharpe ratio. One of the disadvantages of using this ratio as a risk-adjusted measure of return in some applications is that investors, optimally, hold a diversified portfolio of financial assets, and therefore adjustment only for non-diversifiable risk (rather than for total risk) may be preferable. This is why academic studies, most of which examine investment fund performance and performance persistence, tend to measure risk-adjusted fund performance using CAPM-based models (or 'factor models'), which take into account that some of the individual fund's risk is diversifiable.

A2.1.2   Factor models
Factor models link portfolio excess return to various factors, such as market return, 'size' and 'book-to-market' risk factors, via 'factor loadings' or betas. The simplest factor model is CAPM, implemented via Jensen regression, with only market return as the factor. Fama–French's three-factor model is another widely used model, as is Carhart's four-factor model, an extension of the three-factor model. Some researchers have proposed other, more complicated, parametric and non-parametric factor models, such as conditional models and models to incorporate market-timing parameters (see Cuthbertson, Nitzsche and O'Sullivan 2006 for an overview).

Factor models measure a fund's performance by 'alpha', which represents a risk-adjusted return. It is often interpreted as showing whether the fund outperforms its benchmark (hence it is called 'abnormal' return). Thus, this performance measure takes into account the

---

[44] Excess return is measured as the raw return less the risk-free rate.

performance of the 'market'. Although factor models have mostly been applied to analyse the performance of equity funds, there is also some limited literature on other types of funds.[45]

CAPM-based measurement (Jensen regressions)

A financial product's return is linked in some way to the market conditions and fluctuations. Hence, an investor cannot diversify some of the financial product's risk (referred to as the 'non-diversifiable' or 'market' risk). The CAPM establishes the following relationship between expected excess returns, R, and excess market returns, R^m: $R = \beta \cdot R^m$, where β measures the asset's return sensitivity to the market returns.

The difference between actual and expected returns, as predicted by the CAPM, represents the risk-adjusted return on the product, α: $\alpha = R - \beta \cdot R^m$. Importantly, the risk adjustment is for the non-diversifiable risk only, which is captured by the β parameter. Often α is called the 'abnormal return', as it measures by how much the product outperforms the market (or its benchmark)—a positive α means outperformance, whereas a negative α means underperformance.

In practice, α—the risk-adjusted return on a product—is estimated using a Jensen regression. For any fund p, the Jensen alpha α_p is given by:

$$\alpha_p = R_p - E(R_p) = R_p - \beta \cdot E(R_m);$$

where R_p is the actual average excess return and R_m is the excess return on the market portfolio or 'benchmark'.

The excess return on the market portfolio, R_m, should be the return on the market index corresponding to the fund's investment category—ie, the return on the relevant 'benchmark'.

Alphas are usually estimated in a time-series context, by running the following regression for each fund p:

$$R_{pt} = \alpha_p + \beta_p R_{mt} + u_{pt}$$

where u_p is a zero-mean error term. The regression constant α_p is then an estimate for the fund's alpha, and is a measure of 'abnormal' or risk-adjusted return. Regarding the choice of frequency and sample size, the academic literature tends to use monthly (Carhart 1997) or quarterly (Malkiel 1995) return data, and anything from three (Elton, Gruber and Blake 1996) up to, for instance, 20 years (Blake and Timmermann 1998) of monthly return data is used.[46]

Fama–French three-factor model

The Fama-French three-factor model 'extends' the simple CAPM model by adding 'factors' that have been found to explain performance. The following regression can be estimated:

$$R_{pt} = \alpha_p + \beta_p^M R_{mt} + \beta_p^{SMB} R_{SMBt} + \beta_p^{HML} R_{HMLt} + u_{pt}$$

R_{SMB} is the return on the 'Small Minus Big' portfolio, which is a zero-investment portfolio that is long in small stocks and short in large stocks. R_{HML} is a 'High Minus Low' portfolio, which is a zero-investment portfolio being long in high book-to-market stocks and short in low book-to-market stocks.[47]

---

[45] See, for example, Blake, Elton and Gruber (1993) for analysis of bond funds, and Lee (1997) for analysis of property funds.

[46] Alphas can also be estimated for each fund for each time period, in which case, the Fama–MacBeth approach should be followed (see description in Carhart, 1997).

[47] A widely used extension of the three-factor model is Carhart's four-factor model (Cahart 1997), where the fourth factor captures one-year momentum in stock returns.

Drawbacks of factor models

CAPM-based measurement of risk-adjusted returns via Jensen regressions and Fama–French and other factor models has been widely used to measure the risk-adjusted returns on retail investment products, such as mutual funds, unit trusts and others, but mainly for products investing in equity. However, there are a number of drawbacks to using these approaches in this study:

–   the assumptions of CAPM may not necessarily hold for all types of investment product (eg, normal returns);

–   the choice of 'benchmark' return may be somewhat arbitrary and not necessarily a good proxy for the market return that the theory underlying the CAPM considers.

CAPM-based measures of the risk-adjusted returns are likely to be particularly problematic in the case of retail investment products investing in non-equity (financial) assets, both because CAPM assumptions may not hold and because 'benchmark' returns may not proxy well enough the market return (ie, it may be difficult to pick the right benchmark).

A2.2   Evidence on the relationship between performance and charges

The academic literature tends to examine the link between product performance and charges using two different approaches: regression and ranking.

–   **Regression approach**—this involves regressing the (net risk-adjusted) returns on TER. In some cases, multivariate regressions are estimated, whereby other investment product characteristics, such as size and turnover, are controlled for. The literature does not present arguments for or against using univariate regressions as opposed to multivariate regressions, although, in the papers summarised in the table below, there is a tendency to use the former. Typically, cross-section regressions are run for each time period (eg, month), and coefficients are then time-averaged. However, in some cases the variables are first time-averaged and then a cross-section regression is estimated.

–   **Ranking approach**—this typically involves three steps. First, investment products are ranked and possibly grouped according to their performance (eg, risk-adjusted net returns). Second, average charges for each 'group' are estimated. Lastly, average charges are compared across groups to establish whether there is a relationship between the ranking of the products and the ranking of the charges (eg, rank correlation tests can be performed).

A summary of some of the main academic studies in this area is provided in Table A2.1.

Table A2.1 Evidence on the relationship between performance and charges

Paper	Finding	Details
Carhart (1997)	Negative relationship between risk-adjusted net returns and expense ratios (a coefficient of –1.54)	Data: monthly returns of US equity mutual funds
		Risk-adjusted net returns each month are estimated using the Carhart four-factor model (based on the Fama–MacBeth approach)
		To estimate the relationship between performance and charges, univariate cross-section regressions are estimated for each month and the coefficients averaged over time
Malkiel (1995)	Negative relationship between average annual net returns and expense ratios (a coefficient of –1.92)	Data: annual returns of US equity mutual funds; no adjustment for survivorship bias
		Returns are not adjusted for risk, and are time-averaged in order to perform a univariate cross-section regression
		No statistically significant relationship is found between average annual gross returns and expense ratios
Droms and Walker (1994)	No significant relationship between raw or risk-adjusted net returns and expense ratios	Data: annual returns of international equity funds
		Raw returns are measured as annual returns net of expenses, and adjusted returns are measured by the Sharpe ratio (calculated using net returns)
		Single-year, cross-sectional multiple regressions are estimated, including also factors such as total assets and turnover
Elton, Gruber and Blake (1996)	Worst-performing funds in terms of risk-adjusted return have higher expense ratios	Data: monthly returns of US common stock funds
		Risk-adjusted returns each month are estimated using an augmented Fama–French model
		Other findings: worse performance is also associated with worse stock-picking skills; expenses decline as the size of the fund increases; and successful funds do not increase fees compared with unsuccessful funds
Gruber (1996)	Negative relationship between risk-adjusted returns and expense ratios	Data: monthly returns of US common stock funds
		Risk-adjusted returns estimated using a four-factor model (the fourth factor is a bond index)
		Average expense ratios are compared across deciles of funds, where the ranking is according to the risk-adjusted performance. The Spearman rank correlation test is used to test the difference between rankings
Blake, Elton and Gruber (1993)	Negative relationship between risk-adjusted returns and expense ratios (a coefficient of around –1)	Data: monthly returns of US bond funds
		Risk-adjusted returns each month are estimated using CAPM-type regression with bond indices (based on the Fama–MacBeth approach)
		Risk-adjusted returns are regressed on expense ratios to measure the relationship between performance and charges
Quigley and Sinquefield (2000)	Worst-performing funds in terms of risk-adjusted net return have higher expense ratios	Data: monthly returns on UK unit trusts investing in equities
		Risk-adjusted returns are estimated using the Fama–French model
		'Portfolios' of best-, medium- and low-performing unit trusts are formed, and average expenses calculated for each 'portfolio', and then compared
Fletcher (1999)	No significant relationship between risk-adjusted net returns and expense ratios	Data: monthly returns on UK unit trusts investing in American equities
		Risk-adjusted net returns are estimated against different benchmarks using unconditional and conditional CAPM regression
		The relationship between returns and charges is measured using monthly cross-section regressions; coefficients are then averaged. No relationship between average annual charges and performance is found
Droms and Walker (1995)	Risk (as measured by the standard deviation of returns) is positively related to expense ratios (a coefficient of around 5, statistically significant)	Data: annual returns of equity mutual funds; the analysis is affected by survivorship bias
		Risk is measured by standard deviation of total returns over 20 years. (Coefficient of variation—standard deviation divided by average returns—is also used as a measure of risk)
		Mutlivariate regression is used to assess the relationship between different characteristics and risk

A3   References

Allen, D.E. and Parwada, J.T. (2006), 'Investors' response to mutual fund company mergers', *International Journal of Managerial Finance*, **2**:2.

Annaert, J., Van den Broeck, J. and Vennet, R.V. (2001), 'Determinants of mutual fund performance: a Bayesian stochastic frontier approach', Ghent University Working Paper.

Balzer, L.A. (1995), 'Measuring investment risk: A review', *Journal of Investing*, **3**, 47–58.

Barber, B.M., Odean, T. and Zheng, L. (2005), 'Out of sight, out of mind: The effects of expenses on mutual fund flows', *Journal of Business*, **78**, 2095–120.

Blake, C., Elton, E. and Gruber, M. (1993), 'The performance of bond mutual funds', *Journal of Business*, **66**:3.

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

Carhart, M. (1997), 'On persistence in mutual fund performance', *Journal of Finance*, **52**, 57–82.

Cuthbertson, K., Nitzsche, D. and O'Sullivan, N. (2006), 'Mutual fund performance', Cass Business School Working Paper.

Droms, W. and Walker, D. (1994), 'Investment performance of international mutual funds', *Journal of Financial Research*, **17**, 1–14.

Droms, W. and Walker, D. (1995), 'Determinants of variation in mutual fund returns', *Applied Financial Economics*, **5**, 383–9.

Elton, E., Gruber, M. and Blake, C. (1996), 'The persistence of risk-adjusted mutual fund performance', *Journal of Business*, **69**:2.

Fletcher, J. (1999), 'The evaluation of the performance of UK American unit trusts', *International Review of Economics and Finance*, **8**, 455–66.

Gruber, M. (1996), 'Another puzzle: The growth in actively managed mutual funds', *Journal of Finance*, **51**, July, 783–810.

IMA (2006), 'Asset Management Survey', Investment Management Association, London.

Jayaraman, N., Khorana, A. and Nelling, E. (2002), 'An analysis of the determinants and shareholder wealth effects of mutual fund mergers', *Journal of Finance*, **57**:3.

Kheswani, A. and Stolin, D. (2006), 'Mutual fund performance persistence and competition: A cross-sector analysis', *Journal of Financial Research*, **29**, 349–366.

Khorana, A., Servaes, H. and Tufano, P. (2007), 'Mutual Funds Fees Around the World', HBS Finance Working Paper No. 901023.

Lee, S. (1997), 'The components of property fund performance', *Journal of Real Estate Portfolio Management*, **3**:2.

Malkiel, B.G. (1995), 'Returns from investing in equity mutual funds 1971 to 1991', *Journal of Finance*, **50**, 549–72.

Markowitz, H.M. (1959), *Portfolio Selection: Efficient Diversification of Investments*, John Wiley & Sons, New Jersey.

Quigley, G. and Sinquefield, R.A. (2000), 'Performance of UK equity unit trusts', *Journal of Asset Management*, **1**, 72–92.

Sharpe, W.F. (1966), 'Mutual Fund Performance', *Journal of Business*, **39**, 119–38.

Sirri, E. and Tufano, P. (1998), 'Costly search and mutual fund flows', *Journal of Finance*, **53**, 1589–622.

Wermers, R. (2000), 'Mutual fund performance: an empirical decomposition into stock-picking talent, style, transaction costs, and expenses', *Journal of Finance*, **55**, 1655–95.

Wilcox, R.T. (2003), 'Bargain hunting or star gazing? Investors' preferences for stock mutual funds', *Journal of Business*, **76**, 645–63.

www.oxera.com