CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Mutual fund performance: measurement and evidence
Authors	Cuthbertson, Keith;Nitzsche, Dirk;O'Sullivan, Niall
Publication date	2010-05-01
Original Citation	CUTHBERTSON, K., NITZSCHE, D. & O'SULLIVAN, N. 2010. Mutual Fund Performance: Measurement and Evidence. Financial Markets, Institutions & Instruments, 19 (2), 95-187. doi: 10.1111/j.1468-0416.2010.00156.x
Type of publication	Article (peer-reviewed)
Link to publisher's version	http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0416.2010.00156.x/abstract - 10.1111/j.1468-0416.2010.00156.x
Rights	© 2010 New York University Salomon Center and Wiley Periodicals, Inc. This is the pre-peer reviewed version of the following article: CUTHBERTSON, K., NITZSCHE, D. & O'SULLIVAN, N. 2010. Mutual Fund Performance: Measurement and Evidence. Financial Markets, Institutions & Instruments, 19 (2), 95-187, which has been published in final form at http://dx.doi.org/10.1111/j.1468-0416.2010.00156.x
Download date	2026-11-17 11:01:02
Item downloaded from	https://hdl.handle.net/10468/1269

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

MUTUAL FUND PERFORMANCE

Keith Cuthbertson*, Dirk Nitzsche* and Niall O'Sullivan**

First version 10th November 2005
This version 12th December 2006

Abstract:

We evaluate the academic research on mutual fund performance in the US and UK concentrating particularly on the literature published over the last 20 years where innovation and data advances have been most marked. The evidence suggests that ex-post, there are around 2-5% of top performing UK and US equity mutual funds which genuinely outperform their benchmarks whereas around 20-40% of funds have genuinely poor. Key drivers of relative performance are, load fees, expenses and turnover. There is little evidence of successful market timing. Evidence on picking winners suggests past winner funds persist, particularly when rebalancing is frequent (i.e. less than one year) - but transactions costs and fund fees imply that economic gains to investors from actively switching into winner funds may be marginal. However, recent research using more sophisticated sorting rules (e.g. Bayesian approaches) indicate possible large gains from picking winners, when rebalancing monthly. The evidence also clearly supports the view that past loser funds remain losers. Broadly speaking results for bond mutual funds are similar to those for equity mutual funds but hedge funds show better ex-post and ex-ante risk adjusted performance than do mutual funds. Sensible advice for most investors would be to hold low cost index funds and avoid holding past 'active' loser funds. Only very sophisticated investors should pursue an active investment strategy of trying to pick winners - and then with much caution.

Keywords : Mutual fund performance, persistence, smart money.

JEL Classification: C15, G11, C14.

* Cass Business School, City University, London, UK
** Department of Economics, University College Cork, Ireland

Corresponding Author:
Professor Keith Cuthbertson, Cass Business School, 106 Bunhill Row, London, EC1Y 8TZ.
Tel. : +44-(0)-20-7040-5070, Fax : +44-(0)-20-7040-8881, E-mail : k.cuthbertson@city.ac.uk

Electronic copy of this paper is available at: http://ssrn.com/abstract=955807

CONTENTS

1. INTRODUCTION

2. INDUSTRY STRUCTURE

3**. PERFORMANCE**

- **3.1 Equilibrium Models**
- **3.2 Performance Measures**
  - **A. Factor Models**
  - **B. Characteristic Based Measures**

4. EX POST PERFORMANCE: EVIDENCE

- **4.1 Measuring Returns**
- **4.2 Size and Power**
- **4.3 Survivorship Bias**
- **4.4 US and UK Results**
- **4.5 Luck and Performance: A Tail of Two Tails**
- **4.6 Market Timing**

5. PICKING WINNERS: PERSISTENCE

6. PERSISTENCE: EVIDENCE

- **6.1 Predictability: Statistical Measures**
- **6.2 Recursive Portfolio Method**
- **6.3 Stock Holdings And Trade Data**

7. FUND FLOWS AND PERFORMANCE

- **7.1 Performance and Fund Characteristics**
- **7.2 Investment Flows And Performance: Is Money Smart?**
- **7.3 Mutual Fund Managers**

8. CONCLUSIONS

Electronic copy of this paper is available at: http://ssrn.com/abstract=955807

1. INTRODUCTION

The mutual fund industry in the USA and UK has increased dramatically over the last 30 years and now accounts for a substantial amount of private sector saving and substantial new net inflows of saving into risky financial assets. For example in January 2005 about 50% of US households held *equity* mutual funds, comprising an aggregate investment of around $4.5 trillion in 4,500 funds. Most funds are 'active' in that they either try to pick 'winner stocks' or they engage in market timing (i.e. predicting relative returns of broad asset classes) and active funds generally charge higher fees than 'index' or 'tracker' funds (which mimic movements in broad market indexes). In the US and UK about 70% of institutional funds are actively managed and this rises to over 90% for retail funds. In this article we summarize and evaluate the academic research on mutual fund performance in the US and UK, in the context of policy debates on long term saving, concentrating particularly on the literature published over the last 20 years where innovation and data advances have been most marked[1].

Two key issues on fund performance have been central to recent academic and policy debates. The first is whether active funds have an (ex-post) abnormal fund performance in terms of gross returns (but after adjustments for risk) which is positive and whether any outperformance accrues to fund managers (as a whole) or to investors[2]. A second major issue is whether abnormal fund performance can be identified ex-ante and for how long it persists in the future. If fund returns persist then it may be possible for investors to re-allocate their savings towards 'winner funds' and enhance their long term abnormal returns (relative to a passive index strategy) – in short, "money may be smart".

Both of the above approaches (particularly the second) are usually interpreted as direct tests of the EMH in a market where entry barriers are relatively low, there are many professional traders who operate in a competitive environment and information is available at relatively low cost – precisely the conditions under which the EMH is expected to be valid. So, mutual funds provide a way of testing the behavior of investors against the classic paradigm of finance theory where individuals are assumed to make rational decisions in relatively frictionless and low information cost markets, which leads to the elimination of inferior financial products and the growth in successful ones.

---

[1] Space constraints imply we cannot survey the empirical literature from other countries – indeed many countries have little reliable MF data or the MF industry is relatively undeveloped. While there are some useful academic surveys on the performance of the mutual fund industry these are mainly in book form and now require updating – see for example, Friend, Blume and Crockett (1970), Sirri and Tufano (1993), Grinblatt and Titman (1995), Pozen 1998, Bogle (1999). For an excellent survey of the issues in funded versus pay-as-you-go savings see Lindbeck and Persson (2003).

[2] Taxes on capital gains and dividend disbursements also influence the return to investors, although lack of data on individuals' tax liabilities makes any adjustments difficult - so most studies use pre-tax returns.

With increasing longevity, a 'savings gap' is predicted for many countries in 20 years time as forecasts of state pensions suggest inadequate basic provision for future pensioners (e.g. see Presidential Commission on Social Security Reform 2001, OECD 2003, Turner 2004, 2005, 2006). Will voluntary saving in mutual and pension funds over the next 20 years be sufficient to fill this gap? A key element here is the attractiveness of savings products in general and also the choice between actively managed and passive (or index/tracker) funds. If voluntary saving is not sufficient to fill the savings gap, then the issue of compulsory saving may have to be considered together with the degree of investor choice allowed within this framework. Over long horizons small differences in net returns can lead to large differences in terminal wealth and hence in resources available for retirement. For example for {10,20,30,40} year horizons, a 1% p.a. lower gross return or higher charges leads to differences in terminal wealth of {10.5%, 22.1%, 35.0%, 49.2%} respectively. The performance of active versus passive mutual funds and their associated charges, is therefore of key concern, particularly to long term investors.

The behavioral finance literature (see Barberis and Thaler 2003 for a survey) has provided theoretical models and empirical evidence for the US and UK which suggests that active stock picking 'styles' such as value-growth (LaPorta et al 1997, Chan et al 1996, Chan and Lakonishok 2004) and momentum (Jegadeesh and Titman 1993, 2001, Chan et al 2000, Hon and Tonks 2003), as well as market timing strategies (Pesaran and Timmermann 1994, 1995, 2000, Ang and Bekaert 2006) can earn abnormal returns after correcting for risk and transactions costs. Large sections of the mutual fund industry follow active strategies and more recently there is an ongoing debate on whether mutual (and pension) funds should invest in hedge funds and private equity, which also follow a wide variety of active strategies. The question is therefore whether ex-ante, one can find actively managed funds (after correcting for risk and transactions costs) which outperform index funds and whether investors switch funds out of 'loser fund' into 'winners'. To the extent that any 'savings gap' is to be filled by investment in mutual funds, the need to evaluate risk adjusted performance in a tractable way, while taking account of the inherent uncertainty in performance measures, is of key importance both for individual savers and for overall policy on long-term savings.

The rest of this article is organized as follows. In section 2 we discuss the organization of the mutual fund industry concentrating on the US and UK (although other countries have similar set ups). In section 3 we examine theoretical approaches to evaluating the ex-post performance of mutual funds and in section 4 we evaluate the evidence on US and UK ex-post performance. Section 5 analyses different approaches to measuring persistence in performance while in section 6 we discuss empirical evidence on ex-ante performance and performance persistence.

In section 7 we analyze the relationship between fund characteristics, fund flows and performance and the question of whether 'money is smart'. In section 8 we present our conclusions[3] .

2. INDUSTRY STRUCTURE

Mutual funds are pooled investments which enable investors to enjoy economies of scale in gaining access to well diversified portfolios of securities[4]. They also provide liquidity to the investor as funds can be traded between the investor and the trust manager, although investors cannot directly short-sell funds. Not surprisingly, the mutual fund industry is larger in countries with strong rules, laws and regulation, where the population is more educated and wealthier and where defined contribution pension plans are more prevalent (Khorana, Servaes and Tufano 2006).

Mutual funds are created and managed by a mutual fund management company which is registered with the SEC in the USA and the FSA in the UK. Investors buy shares (units) in the fund but the number of shares in issue varies according to demand, hence the term 'open ended'. This implies that the share price always reflects the underlying net asset value and unlike investment trusts, is not affected by market sentiment towards the mutual fund itself. The management company often manages a 'complex' or 'family' of funds with various investment objectives and often they elect the same individuals to the boards of each of their funds within a fund family. New funds are continually introduced, while less successful funds may merge with other funds (often within the same family) or may cease to exist.

In the US and UK funds are sold by two main methods. The first involves sales by broker-dealers in the US or independent financial advisers IFA's in the UK. Funds sold in this way usually involve a front-end or back-end load fee, with the fund determining the load but it is retained by the selling broker (or IFA) as compensation to the investment advice provided. A

[3] Because of space limitations we largely ignore the empirical literature on the performance of closed-end funds, hedge funds and pension funds, although our analysis of alternative methodologies can also be applied to these asset classes. A survey of recent work on the performance of closed-end funds can be found in Barberis and Thaler (2003), while for hedge fund performance see Fung and Hsieh (1997), Brown, Goetzmann and Ibbotson (1999), Brown, Goetzmann and Park (2001), Agarwal and Naik (2002), Capocci and Hubner (2004) and Kosowski, Naik and Teo (2006). For the performance of pension funds see Thomas and Tonks (2001), Tonks (2005), Blake, Lehmann and Timmermann (1999), Del Guercio and Tkac (2002) and Goyal and Wahal (2004). We also largely ignore the literature on the corporate governance and agency problems of mutual funds – see for example, Brown, Harlow and Starks (1996), Nanda, Narayan and Warther (2000), Nanda, Wang and Zheng (2004), Zhao (2004), Elton et al (2004), Meier and Schaumberg (2004), Gaspar, Massimo and Matos (2006), Bris, Gulen, Kediyala and Rau (2005), Davis and Kim (2005).

[4] In the UK mutual funds are often referred to as Unit Trusts although their correct designation is Open Ended Investment Companies, OEIC. In the US 'unit trusts' purchase assets but do not subsequently trade them.

fund's management company receives a management fee based on a percentage of net assets (and paid out of the fund's assets) – for investment advice and administrative services. Part of these annual expenses are used to pay for direct marketing and (usually if there is no load fee) to compensate selling brokers (or IFA's) – in the US the latter are known as "12(b)-1" fees and are limited to 1% p.a. of the fund's assets by the National Association of Securities Dealers, NASD.

The second sales method is direct marketing of funds via newspapers etc and 'fund supermarkets' such as *Charles Schwab's One Source* and here there is no load fee. 'Fund supermarkets' offer a menu of 'participating funds' which can be held in a single consolidated account. In the US, directly marketed funds may use the 12b-1 fee to pay for advertising or shelf space at a fund supermarket. To the extent that the 12b-1 fee is used for advertising this implies that current shareholders bear the cost of attracting new shareholders[5].

In 2003 the average mutual fund had assets of $941m but the large standard deviation of around $3,400m indicates a few large funds - the largest having assets of over $82.2bn, with an interquartile range of $49-565m. There were about 6,600 funds and 16,000 share classes of which 56% charged a load fee, 66% charged a 12b-1 fee and 32% had neither[6]. The SEC requires that all US mutual funds report an annual total expense ratio, (as a percentage of total net assets) TER which comprises operating expenses (i.e. management, legal, accounting and custodial fees) plus the 12b-1 fee – but excluding trading costs associated with purchases/sales. In the US the average (NAV weighted) TER in 2003 was 76bp, with operating expenses accounting for 80% (61bp) and 12b-1 fees for 20% (15bp) of the TER. A few funds (87) are index funds with average operating expenses of 42bp with a standard deviation of 19bp and ranging from 8 to 85bp[7]. The impact of explicit costs on performance for both active and index funds is a major area of investigation in the mutual fund literature. In the US and UK, trading costs include brokers commissions and bid-ask spreads and are not included in the management fee and the TER[8].

In the US mutual funds are pass-through entities for tax purposes and the fund does not pay any taxes on its holdings – dividend and capital gains *realizations* are passed on equally to

---

[5] Mahoney (2004) states that for the US, the Investment Company Institute reports that 63% of 12b-1 fees are paid to brokers.

[6] Figures are from the CRSP US fund database

[7] The Vanguard 500 Index Fund had 7bp per year trading costs and an expense ratio of 28bp per year (on average) over 1975-1994 (Wermers 2000). Trading costs fell to 3 bp per annum over 1990-94.

[8] In the US, total commissions are reported to the SEC but not spread costs on the grounds that the latter are difficult to measure in a meaningful way. Funds do not publish commissions as the SEC believes this could be misleading if commissions and spread costs are inversely related.

all the fund's shareholders (regardless of when the shares were created).[9] In the US, investors who purchase shares in a fund which have accrued but unrealized gains face the prospect of distributions of realized gains even if they have only held the fund for a short time – this is the 'capital gain overhang' (Bergstresser and Poterba 2002).

Restrictions or covenants to mutual fund activity vary across funds. But in 2000 about 90% of funds could not buy on margin, around 65% could not short securities, about 25-30% could not trade either index futures or individual options. In addition around 20% were not allowed to borrow money and about 20% of funds could not hold restricted securities (e.g. those obtained by private sale). But of those funds that are *not* restricted in the above fashion a maximum of only about 12% of funds actually used any one of the above techniques over the 1996-2000 period - and there are no appreciable trends in this figure over this period (Almazan et al 2004).

Funds may be market 'tracker' or 'index' funds - which are not actively managed. With the recent appearance of Exchange Traded Funds ETFs, investors may also 'track' a diversified position in a given style category (e.g. small stocks, telecom stocks).[10] Most *active* funds can be allocated to different categories/styles, often by combining a returns database with other independent databases on styles – although the allocation to different styles may sometimes be problematic. Broadly speaking funds may be 'domestic' or 'international' and for each of these categories funds may be predominantly or exclusively equity, bond, property or money market funds. Studies of actively managed funds often concentrate on 'equity funds' although it is worth noting that these funds do not hold all of their assets in equities. The performance of particular subsets of 'equity funds' such as aggressive growth, growth and income, growth, equity-income and small companies are often undertaken[11]. These equity funds may be limited to holdings of domestic equities (particularly for large developed economies like the US) but other studies

---

[9] Prior to the US Taxpayer Relief Act 1997, funds which realized more than 30% of their capital gains from positions held for less than three months did pay taxes. This act also removed restrictions against short sales (the 'short-short' rule) and derivatives trades. Mutual funds held in tax deferred form include those in IRA and 401K accounts.

[10] ETFs are also redeemable at market value at any time of the trading day (and for example, not just at 4pm New York time as for US mutual funds) and ETFs often have special tax privileges.

[11] In the US 'self declared' fund styles are overseen by the SEC but it is not always the case that the style name accurately represents the underlying assets in the portfolio. The SEC rules mandate that a fund name must imply that it has at least 80% of their assets in securities of this type/name but there is much leeway in interpretation of the rule. For example, there are no restrictions in using 'value' and 'growth' and although you must abide by the rule as regards 'small cap', 'midcap' and 'largecap', this only applies if these are precisely defined in the prospectus for the fund. Also, the SEC only requires full disclosure of portfolio holdings twice per year, so often investors cannot accurately track any style drift between these reporting periods. Morningstar (see *Morningstar Mutual Funds OnDisc Operations Manual*) classifies funds into the five 'categories' noted in the text but reserves the term 'style' for 'growth' and 'value' or a blend of the two based on P/E and BMV (relative to similar ratios for S&P500 index). Morningstar also classifies funds into 'small', 'medium' and 'large' styles, based on market cap. Thompson/CDA- Spectrum files and the CRSP mutual fund files have somewhat different investment categories from each other and from Morningstar – so allocation to a particular category requires some judgement – Wermers (2003a). We use the terms 'category' and 'style' interchangeably.

consider funds with domestic and foreign equities. Academic studies of bond mutual funds are less prevalent and are only briefly mentioned in this study[12].

3. PERFORMANCE

In the first part of this section we discuss theoretical models of the mutual fund industry which focus on fund performance before examining risk adjusted performance measures based on both factor models and the characteristics of stock holdings.

3.1 EQUILIBRIUM MODELS

Is it reasonable to expect some funds to under or over-perform their benchmarks over very long horizons or that investors can use ex-ante trading rules which result in positive and persistent abnormal returns? In short, are there models which provide a rationale for 'smart money' behavior in which investors can 'beat the market'. In a seminal article Grossman and Stiglitz (1980) argue that in equilibrium, expected abnormal returns should not be zero, otherwise there would be no incentive to gather and process costly information. Taking up the idea that information processing is costly, Berk and Green (2004) use a general equilibrium competitive model (with no moral hazard or asymmetric information) to analyze fund flows, ex-post returns and performance persistence. The model is very similar to the standard perfectly competitive model of the firm where decreasing returns to scale coupled with low barriers to entry ensure that any short-term abnormal profits (e.g. due to lower costs or skill in stockpicking) are quickly competed away. Hence managers are skillful, but in equilibrium firms earn zero abnormal profits and past performance cannot be used to predict future performance or to infer the average skill level of managers.

In the model, fund managers do have differential skill (stock picking ability) at a gross return level, which investors learn about based on past returns - investors then chase recent 'past winners'. This performance-flow relationship is endogenous in the model and is nonlinear (convex) because the cost function is non-linear – past winner funds attract disproportionate cash inflows. But net cash inflows are subject to diminishing returns, so successful funds find it ever more difficult to earn abnormal net returns - any profits are therefore short-lived and in equilibrium funds earn zero abnormal returns.

Essentially, skilled managers pursue an *active* strategy until the expected net return on the marginal dollar inflow equals the return on a passive index fund (with similar risk

[12] See for example, Blake, Elton and Gruber (1993), Elton, Gruber and Blake (1995), Huij and Derwall (2006).

characteristics) – after which any additional cash inflow is placed in index funds. Managers are not allowed to borrow or short-sell. The performance-flow relationship means that skilled managers can extract rents by charging a management fee proportional to funds under management – managers aim at all times to maximize fees but given the performance-flow relationship this involves maximizing expected returns[13]. The manager faces increasing marginal costs on actively managed funds which establishes an optimal size for managed funds with any excess invested in the passive fund (at zero cost). As additional fees are paid on index funds, this is how the active managers extract their rents. Indeed, it is the response of fund flows to past returns which facilitate the competitive process, but as it is the fund managers who are skilled, then they (and not the investors) expropriate the economic rents. However, in equilibrium neither the size of the fund nor the manager's ability help predict *investor* returns (which are net of management fees). When investors do not know who the skilled managers are, they infer this from past returns (in a Bayesian framework), so fund inflows are generated by rational investors but it is this (plus diminishing returns), which makes future returns to investors unpredictable.

A key prediction of the model is that extremely good performance should result in large fund inflows, while poorly performing funds experience relatively smaller outflows (i.e. the performance-flow relationship is non-linear). However, funds with very large losses should close down. It is these competitive forces that ensure that both 'good' and 'bad' performance does not persist in equilibrium.

The model of Lynch and Musto (2003) shares some common assumptions with Berk and Green (2004), since investors learn about manager's abilities based on past returns. But in the Lynch and Musto model exogenous differences in ability lead to differences in performance persistence, because there are no diminishing returns to funds which experience cash inflows. The model does not seek to explain persistence per se but it is the effect of *differential* persistence between past winner and loser funds which leads to a convex performance-flow relationship. The decision about *changing strategy* is the key feature of the model and is assumed to be more likely for past losers than past winners. If past winners persist (because their successful strategy does not change), then we would expect past performance to lead to large inflows into winner funds. Similarly if past losers persist then we expect to see cash outflows. But why do *severe* past losers barely experience lower future outflows than moderate past losers? It is the latter that the model focuses on.

[13] This rules out dynamic strategies such as managers with good (bad) performance early in the year, reducing (increasing) risk later in the year – Brown, Harlow and Starks (1996), Chevalier and Ellison (1999a).

Lynch and Musto use a two period model and when time-1 fund return is low the manager is likely to switch strategies and this decreases the predictive power of past returns for future returns - hence the model predicts poor performance has only weak persistence. Conversely, when time-1 fund return is high, the manager retains the same successful strategy and hence past 'winners' persist. The convex flow-perfomance relationship follows from this prediction of differential persistence in returns. Since past winners continue with their winning strategy, they experience strong cash inflows. The worse the performance of 'bad' funds the greater the probability that investors believe a successful change of strategy will take place and hence they tend not to move funds disproportionately, from very poorly performing funds. Hence, the model provides a rational explanation for continued investment in what are very poorly performing funds – although one would expect that with no change in strategy and persistent poor performance, such funds would eventually experience substantial outflows.

These two key models of the flow-performance relationship give clear but different predictions. In both models managers have differential skill, which investors infer from past performance. But in the Berk and Green model, the average ex-post abnormal fund performance (net of costs) is zero and performance persistence (of both past winners and losers) is weak or non existent *in equilibrium.* In the Lynch and Musto model the performance of past winners persists (because they do not change their successful strategy) but past losers experience less persistence. For both models two key questions remain. First, how long does any disequilibrium and hence (good or bad) abnormal performance last? Second, is abnormal performance economically significant and exploitable by either fund managers or investors - these are empirical questions to which much of this survey is devoted.

The equilibrium model of Nanda, Narayan and Warther (2000) is somewhat different to the above models since it is more concerned with explaining the co-existence of load and no-load funds, based on the idea that some (but not all) investors create systematic liquidity costs for funds. The latter can be reduced by funds charging loads. Expected fund returns are increasing in manager ability and decreasing in (dollar) funds under management – this relationship is exogenous to the model. The managers set fund fees to maximize expected profits (in a risk neutral world) and a positive cash inflow from investors ensues as long as expected returns are positive. The managers earn rents from their innate skill and each fund has an optimal size where the marginal return from ability equals the marginal cost of increased size. Managers whose ability is revealed to be high (i.e. those with skill) form load funds - but in order to attract investors with low liquidity needs they share some of the rents by charging lower management fees, so that investors in load fees earn positive expected returns in equilibrium. However, managers whose ability is revealed to be low, form no-load funds that attract investors with high

liquidity needs. The model therefore has two key predictions for the performance-cost relationship. First, investor returns in load funds exceed those in no-load funds and second, investors in funds with high-load fees will earn a higher return than those in low-load funds – we investigate the empirical evidence for these predictions in section 7.1.

The above theoretical models seek to explain some of the stylized facts of the performance-flow relationship and the strength of short-run and long-run abnormal performance of past winner and loser funds. We now turn to methods used to measure such abnormal performance.

3.2 PERFORMANCE MEASURES

Risk adjusted mutual fund performance is usually measured using either factor models or characteristic based performance measures. Factor models are parametric, while the latter are semi-parametric and match fund stock holdings with specific stocks with similar characteristics. Both approaches have their strengths and weaknesses which are discussed below.

A. FACTOR MODELS

An m-factor model of a stock-j's excess return $r_{j,t} = R_{j,t} - R_{f,t}$ is:

[1] $$r_{j,t} = \beta_j^{'} F_t + \varepsilon_{j,t}$$

where $R_{f,t}$ = risk free rate, $F_t = \{F_{1,t}, F_{2,t}, \ldots\ldots, F_{m,t}\}'$ are the m-factors and $\beta_j^{'} = \{\beta_{j,1}, \beta_{j,2}, \ldots\ldots, \beta_{j,m}\}$ are the m-factor loadings. A robust factor model should explain the cross-section of *average* returns on any set of assets or portfolios and in practice portfolios are often chosen to maximize the spread in average returns and hence increase the power of the tests. For example, the cross section of average returns on stock portfolios based on a double sort, on book-to-market value BMV and size (into 25 quintiles) cannot be explained by CAPM betas but are adequately explained by the Fama-French (1992,1993) three factor (3F) model, which includes the market return, 'size' and 'book-to-market' risk factors[14]. The 'size' risk factor may be due to the greater sensitivity of earnings prospects to economic conditions with a resulting higher probability of distress during economic downturns. Also, small firms may embody

[14] The Fama-French size factor SMB ('small minus big'), is the difference between the returns on a portfolio of small stocks (e.g. smallest third) and large stocks. The book-to-market value factor, HML ('high minus low'), is a measure of the difference between the returns on high versus low BMV stocks. To construct these series the portfolios are usually rebalanced every year. (For the US, data on these variables can be found on Kenneth French's website, http://mba.tuck.dartmouth.edu/pages/faculty/Ken.French)

greater informational asymmetry for investors than do large firms. Both effects imply a risk loading for size and a higher required return. Fama and French (1992) also report a strong positive relationship between the cross-section of average stock returns and BMV which may be due to high BMV firms being 'fallen angels' and hence tend to have a higher risk of bankruptcy and financial distress.

The cross-section of average returns on stocks sorted into portfolios on the basis of their prior one-year raw returns are not well explained by the Fama-French 3F model but a one-year 'momentum' factor MOM captures this 'anomaly' (Jegadeesh and Titman 1993, 2001).[15] The interpretation of MOM as a risk factor is more tenuous than the other factors mentioned above but since momentum investing is a mechanical strategy (requiring little or no skill), then it can be argued that it should not be counted as part of a fund's abnormal return.

A mutual fund-i at time t, with asset proportions $w_{k,t}$ (k = 1, 2, …, N) has a required return (to compensate for factor risk) equal to $\beta_{i,t}^{'} EF_t$ and the fund abnormal performance is given by $\alpha_{i,t}$ in the regression (Admati et al 1986, Lehmann and Modest 1987):

[2] $$r_{i,t} = \sum_{j=1}^{N} (w_{j,t}\alpha_k + w_{j,t}\beta_j^{'})F_t \; + \sum_{j=1}^{N} w_{j,t}\varepsilon_{j,t} = \alpha_{i,t} + \beta_{i,t}^{'} F_t + \varepsilon_{i,t}$$

where $\varepsilon_{i,t} = \sum_{j=1}^{N} w_{j,t}\varepsilon_{j,t}$, $\alpha_{i,t} = \sum_{j=1}^{N} (w_{j,t}\alpha_j)$ and $\beta_{i,t} = \sum_{j=1}^{N} (w_{j,t}\beta_{j,i})$. Note that the fund's parameters will be time varying if either the factor betas or the fund weights are time varying, but in practical applications these parameters are often assumed to be constant over the investment horizon considered.

If one believes that the factors represent risk, then alpha represents the stock picking skill of fund managers. For those who are agnostic about the factors acting as proxy variables for 'risk', these regressions are usually referred to as *performance attribution models.* The logic is that after considering all systematic influences on fund returns over time, a positive estimated alpha ($\hat{\alpha}_{i,t} = \overline{r}_{i,t} - \hat{\beta}_i^{'} \overline{F}_t$) indicates that the fund has an average return which exceeds that from following 'mechanical' investment strategies - in addition, a positive alpha implies that an investor can combine the fund and the factors to obtain a Sharpe ratio higher than that which can be obtained using the benchmarks alone.

UNCONDITIONAL MODELS

These have factor loadings that are assumed to be time invariant. Carhart's (1997) four factor (4F) performance measure is the alpha estimate from:

[3] $$r_{i,t} = \alpha_i + \beta_{1i} r_{m,t} + \beta_{2i} SMB_t + \beta_{3i} HML_t + \beta_{4i} MOM_t + \varepsilon_{i,t}$$

where $r_{m,t}$ is the excess return on the market portfolio, SMB_t, HML_t and MOM_t are zero investment factor mimicking portfolios for size, book-to-market value and momentum effects, respectively. If $\beta_{4i} = 0$ the model is the Fama-French (1992, 1993) 3F model while Jensen's (1968) alpha is the intercept from the CAPM one-factor (or market) model. Subject to having the correct factor model of equilibrium returns, a positive (negative) and statistically significant value of alpha indicates superior (inferior) risk adjusted performance and stock picking skills.

CONDITIONAL MODELS

Conditional models (Ferson and Schadt 1996) allow for the possibility that a fund's factor betas depend on lagged public information variables. This may arise because of under and over-pricing (Chan 1988 and Ball and Kothari 1989), or changing financial characteristics of companies such as gearing, earnings variability and dividend policy (Mandelker and Rhee 1984, Hochman 1983, Bildersee 1975). Also, an active fund manager may alter portfolio weights and consequently portfolio betas depending on public information. Thus there may be time variation in the portfolio betas depending on the information set Z_t so that $\beta_{i,t} = b_{0i} + B_2^{'}(z_t)$, where z_t is the vector of deviations of Z_t from its unconditional mean. For the CAPM this gives:

[4] $$r_{i,t+1} = \alpha_i + b_{0i}(r_{b,t+1}) + B_i^{'}(z_t * r_{b,t+1}) + \varepsilon_{i,t+1}$$

where $r_{b,t+1}$ = the excess return on a benchmark portfolio (i.e. market portfolio in this case). Christopherson, Ferson and Glassman (1998) assume that alpha (as well as the beta's) may depend linearly on z_t so that $\alpha_{i,t} = \alpha_{0i} + A_i{}'(z_t)$ and the performance model is:

[5] $$r_{i,t+1} = \alpha_{0i} + A_i^{'}(z_t) + b_{0i}(r_{b,t+1}) + B_i^{'}(z_t * r_{b,t+1}) + \varepsilon_{i,t+1}$$

---

[15] The MOM variable is the difference in returns between a portfolio of previously high return stocks (e.g. top 1/3rd

Here, α_{0i} measures abnormal performance after controlling for (i) publicly available information, z_t which may influence abnormal return and (ii) adjustment of the factor loadings based on publicly available information. The above conditional models are easily generalized in a multifactor framework. Most studies follow Ferson and Schadt (1996) and Christopherson, Ferson and Glassman (1998) where the Z_t variables are often taken to include the one-month T-Bill yield, the dividend yield of the market factor and the term spread. An alternative (semi-parametric) approach to mimic conditional models is to assume the unconditional model has time-varying parameters (e.g. Kalman filter random coefficients model – Mamayski et al 2004).

UNCERTAINTY AND EXTRANEOUS INFORMATION

There are obvious uncertainties in using a parametric approach and evaluating performance is likely to involve model error, so sensitivity tests using alternative parametric models - including allowance for time varying parameters – are often employed. Also, a standard Bayesian approach would utilize priors on the fund's parameters and use the posterior alpha as the performance measure. Recent work has used information *extraneous* to the fund, to increase the precision of that fund's alpha estimate (e.g. data on non-benchmark returns, or on the performance of the average fund). The extraneous information may be informative for a fund's alpha per se, but in addition we can apply a Bayesian approach by also incorporating priors about the extraneous information itself (e.g. priors about the average performance of *all* funds, might influence the posterior estimate of a particular fund's alpha). We discuss these approaches below.

MARKET TIMING

In addition to stock selection skills, models of portfolio performance also attempt to identify whether fund managers have the ability to market-time. Can fund managers successfully forecast the future direction of the market in aggregate and alter the market beta accordingly? (see Admati et al 1986). Treynor and Mazuy (1966) TM assume a successful market timer adjusts the market factor loading using $\beta_{it} = \theta_i + \gamma_{im}[r_{m,t}]$ while Henriksson and Merton (1981) HM, view market timing as the payoff to a call option on the excess market return $\beta_{it} = \theta_i + \gamma_{im}[r_{m,t}]^+$ where $\left[r_{m,t}\right]^+ = \max\{0, r_{m,t}\}$ - hence the estimation equation for the market model is:

[6] $$r_{i,t} = \alpha_i + \theta_i(r_{m,t}) + \gamma_{im} f[r_{m,t}] + \varepsilon_{i,t}$$

) and previously poor return stocks (e.g. bottom 1/3rd), usually with rebalancing every year.

where $f[r_{m,t}]$ is the appropriate non-linear function and $\gamma_{im} > 0$ indicates successful market timing. These two market timing models can be easily generalized to a multifactor conditional-beta model, where β_i also depends on the public information set, z_t (Ferson and Schadt 1996).

Most studies of market timing use the above parametric approach but linear parameterisations of market timing are highly restrictive and do not separate out the fund managers quality of information about market timing and the aggressiveness with which she reacts to this information. Also, heteroscedasticity and skewness can bias the tests (Breen et al 1986) while Jagannathan and Korajczyk (1986) show how "artificial" timing bias may arise when funds invest in securities with option-like characteristics while Goetzmann et al (2000) show that if funds undertake daily market timing but tests use monthly data then the HM measure has low power and is biased (i.e. there is "interim trading bias" – see Ferson and Khang 2002). Finally, Coles e1t al (2006) show that using the wrong model (e.g. HM when TM is true) or the wrong benchmark timing index can lead to substantial biases in measuring selectivity and timing ability.

Two approaches to mitigate the above problems are the use of holdings data in a parametric framework and the use of non-parametric tests. Holdings data allows one to directly calculate the beta of fund-i β_i, from the individual (estimated) betas of its constituent stocks. A direct parametric test of market timing (Jiang et al 2005) is to run the regression $\beta_{i,t} = \alpha_i + \gamma_{im} g[r_{m,t}] + \varepsilon_{i,t}$ where $g[r_{m,t}]$ is either $r_{m,t}$ or $[r_{m,t}]^+$. Using a novel non-parametric approach Jiang (2003) provides a test of market timing which can identify the quality of the manager's market timing ability and does not rely on linear factor models.[16] The intuition behind the approach is that the manager should increase beta if she thinks the market will rise in the future. For any two non-overlapping time periods t_1 and t_2 the statistic $\theta = 2\ x\ prob(\beta_{t_1} > \beta_{t_2} \mid r_{m,t_1+1} > r_{m,t_2+1}) - 1$ is greater than zero for successful market timing. With no market timing ability β has no correlation with the market return and hence *prob(.)* = 1/2 and θ = 0 (θ < 0 implies adverse market timing). Now consider the triplet sampled from *any* three periods for mutual fund-i's excess returns, $\left\{ r_{i,t_1}, r_{i,t_2}, r_{i,t_3} \right\}$ where $r_{m,t_3} > r_{m,t_2} > r_{m,t_1}$. A fund manager who has superior market timing ability (regardless of her degree of aggressiveness) should have a higher average beta in the t_2 to t_3 period, than in the t_1 to t_2 period. The measured

[16] The non-parametric test has good size properties and is robust (to outliers, non-normality and heteroscedasticity, differences in timing frequency and data frequency) – although it does require serially uncorrelated returns.

value of beta in these two periods is $\beta_{12} = (r_{i,t_2} - r_{i,t_1})/(r_{m,t_2} - r_{m,t_1})$ and $\beta_{23} = (r_{i,t_3} - r_{i,t_2})/\ (r_{m,t_3} - r_{m,t_2})$. The sample analogue to θ is therefore:

[7] $$\hat{\theta}_n = \binom{n}{3}^{-1} \sum_{w} sign(\beta_{23} > \beta_{12})$$

where w represents the triplets in the data where $w \equiv R_{m,t_1} < R_{m,t_2} < R_{m,t_3}$ and $sign(.)$ assumes a value of {1, -1, 0} if the argument is {positive, negative, zero}. Under certain relatively weak assumptions $\hat{\theta}_n$ is asymptotically $N(0, \sigma^2_{\hat{\theta}_n})$.[17]

B. CHARACTERISTIC BASED MEASURES

It was noted above that in measuring abnormal fund performance with conditional factor models we assume that any time variation in portfolio weights are related to publicly observable variables. But data on a fund's asset holdings or trades (i.e. buys/sells) allows one to construct a series for fund returns which accurately reflects the changing weights on the characteristics of the stocks held by the fund (e.g. small/large cap stocks, high/low BMV stocks). Below we outline 'characteristic based' performance measures which use stock holdings and trade data of funds and then discuss some strengths and weaknesses of the approach.

Grinblatt and Titman (1989) pioneered this approach which was then extended by Daniel, Grinblatt, Titman and Wermers DGTW (1997), whereby each stock held (or traded) by the mutual fund is matched with 'benchmark stocks' that have similar characteristics in terms of size, BMV and momentum. For all stocks listed on the NYSE, AMEX and NASDAQ for which data on the characteristics are available, a triple sort (based on quintiles) according to market capitalization, BMV and past 12 month returns, is undertaken. This results in 125 benchmark portfolios with distinct size, book-to-market and momentum attributes. Benchmark portfolios are rebalanced recursively each year and value weighted *quarterly returns* are computed for each of the 125 portfolios. The variable CS ("Characteristic Selectivity") measures a fund's stock selection ability ("stock picking skills") :

[8] $$CS_t = \sum_{j=1}^{N} w_{j,t-1}[R_{j,t} - BR_t(j, t-1)]$$

[17] see Serfling (1980), Abrevaya and Jiang (2005) on the properties of U-statistics.

where $R_{j,t}$ is the return on stock-*j* in quarter-*t* and $BR_t(j,t-k)$ is the return on the benchmark portfolio in quarter *t* to which stock-j was allocated during period t-k according to its size, BMV and momentum characteristics. The weight $w_{j,t-k}$ is the fund's portfolio weight in stock-*j* at the end of quarter *t-k*[18]. If CS_t is positive then the fund held stocks at the end of quarter t-1, which have on average outperformed their characteristic benchmark returns in period t. The CS measure can also be applied to stocks sold/purchased during period t, which might be more informative of genuine skill, than stocks passively held. If the 'characteristics' are thought to mimic true underlying risks then CS is a genuine risk adjusted abnormal return, otherwise CS is a performance attribution measure and shows whether the fund has outperformed stocks with similar characteristics.

The variable CT ("characteristic timing") measures whether the fund generates additional return by exploiting the predictability in the returns of the (size, BMV, momentum) benchmark portfolio:

[9] $$CT_t = \sum_{j=1}^{N} [w_{j,t-1} BR_t(j,t-1) - w_{j,t-5} BR_t(j,t-5)]$$

A positive value for CT implies the fund increased its holding in stocks whose benchmark returns have risen (on average) over the past year. CT provides an alternative measure of 'market timing' to the parametric HM and TM measures and the non-parametric approach of Jiang (2003), but here 'timing' refers to prediction of benchmark returns rather than the (excess) market return. Finally, "average selectivity":

[10] $$AS_t = \sum_{j=1}^{N} [w_{j,t-5} BR_t(j,t-5)]$$

measures the benchmark return at t, due to asset allocation at t-5 into stocks with specific characteristics. A positive AS indicates that the fund was holding stocks 5 quarters previously with characteristics that are *currently* experiencing high returns - this would not normally be considered a measure of skill but instead can be interpreted as a combination of inertia and good luck, rather like the return to passively holding momentum stocks.

---

[18] When we use a trade data measure of CS, then $w_{j,t-k}$ is the dollar purchases/sales of stock-j as a proportion of total purchases/sales of all stocks. Note that when using stock holdings the characteristics based measures assume these stocks are not sold within the quarter – an issue we take up in section 5.1.

The ex-post gross return on the fund's stock holdings (or trades) is then the sum of these elements : $R_t = CS_t + CT_t + AS_t$. For each mutual fund, the variables CS, CT and AS can be averaged over different time horizons (e.g. h = 1, 2, 3, … quarters) after portfolio formation, to give an estimate of persistence at each horizon, using an event study framework. In addition the cross-section of average CS measures for each fund can be correlated with other fund characteristics (e.g. size, fees, turnover) to determine which fund attributes contribute to stock picking skills.

It is argued that the CS index gives a better measure of risk adjusted abnormal return than do conditional factor models, where time-varying weights are assumed to be related to a subset of observable publicly available data and therefore do not *directly measure* changes in stock holdings[19]. In contrast, the CS index explicitly takes account of the changing proportions held in different stocks. Also, in the presence of non-linear factor return premia the characteristics approach may provide better benchmarks than linear factor models. Note however, that the CS-holdings measure ignores trades that take place between reporting dates, an issue we address in section 5.

Stock *trades (*buys/sells) are 'active' decisions and are only likely to take place when the trader has superior information which outweighs any additional trading costs - whereas continuing to hold a stock may be a relatively passive decision reflecting no strongly held views. Thus with the characteristics approach it is argued that using stock *trades* (i.e. buys/sells) may be more powerful in uncovering good (or bad) stock picking or timing skills, than using holdings data (or factor models). However note that the above argument may carry less force for funds which experience very large cash inflows - if these funds feel they need to move quickly into stocks to avoid diluting returns by holding cash.

The drawback in using holdings and trade data is that it is usually only available for stocks (and not other assets held by the fund) and we still have to define the appropriate 'characteristics' of the stocks, which is bound to be somewhat arbitrary (although no less arbitrary than choosing the factors in factor models). Because of a lack of comprehensive data on all asset holdings (and trades) by funds, the characteristics based approach has only been applied to *stock* holdings and the sales/purchases of stocks by US equity mutual funds. Also, the CS index is often measured gross of (stock) transactions costs and management fees so 'returns' to investors are not directly measured. As equity funds also hold other assets (e.g. cash, bills,

[19] A limitation in using quarterly holdings noted by Elton, Gruber, Krasny and Ozelge (2006) is that compared with monthly data you may miss upwards of 20% of a typical fund's trades.

bonds) then a positive value for any of the above characteristic based measures (CS,CT,AS) need not translate into a positive (gross or net) return on the fund *as a whole.*

Models of abnormal return based on CS or alpha can be applied to individual funds or portfolios of funds (e.g. an average across all fund styles or just all growth funds). However, averages across a large portfolio of funds may be of little practical use if individual investors feel constrained in the number of funds they are willing to hold (perhaps due to load fees, rebalancing or information costs)[20] – although this would be mitigated if the chosen portfolio of funds comprised a limited number of fund families or in the case of institutional investors if they are subject to low transactions costs. So, performance measures applied to individual funds or 'small' portfolios may be of more practical use. Also, it is clearly of interest to see if performance measures (for either individual funds or portfolios of funds) are stable over sub-periods – if they are, then this suggests persistence (i.e. good (bad) funds stay good (bad) funds) and the possibility of profitable ex-ante trading rules – an issue we take up in section 6.

4. EX POST PERFORMANCE: EVIDENCE

We begin this section by outlining some data issues in measuring fund returns, then in section 4.2 we discuss size and power of our performance measures and in section 4.3, estimates of survivorship bias. In section 4.4 we examine the ex-post *average* risk adjusted performance of US and UK equity mutual funds while in section 4.5 we discuss the performance of *individual* funds, particularly those in the tails of the performance distribution. Finally in section 4.6 we examine the evidence for successful market timing by funds.

4.1 MEASURING RETURNS

Fund returns can be gross or net of various charges depending on whether we wish to measure returns to the fund (or fund managers) or returns to the investor/customer. *Net returns* are returns to investors (before deduction of any load fees or payment of personal taxes) and are calculated as:

[11] $$1+R_t^{net} = (NAV_t / NAV_{t-1}) \left(\prod_{j=1}^{J} (1+\frac{DISTN_j}{RENAV_j})\right)$$

where NAV_t is the net asset value of the fund at end of period t, J is the number of dividend or capital gains distributions during the period, $DISTN_j$ is the jth distribution in dollars and

[20] Rebalancing costs are incurred if the portfolio is equally or value weighted. Rebalancing costs for equally weighted portfolios occurs when funds die or are merged and also to accommodate the long term rise in the number of funds. In addition a value weighted fund requires rebalancing as market prices fluctuate.

$RENAV_j$ is the NAV at which the jth distribution was reinvested. Net returns are therefore after deduction of all fund expenses and all security level transactions costs.[21] *Gross fund* returns *(i.e. pre-expenses but post transactions costs)* R_t^g are usually defined as $R_t^g = R_t^{net} + TER_t$ where TER is the total expense ratio[22] – the latter includes management fees and other administrative costs of running the fund (e.g. custodial fees, commissions to third parties selling the funds and advertising '12b-1' fees). Load fees are not included in the above 'returns' and are generally paid to brokers as a sales commission. In some studies 'fund return' is the return on the largest shareclass, while others use the value-weighted return of all individual shareclasses (see Wermers 2003a).

A further issue is the use of the term 'gross' returns when applied to asset *holdings data* – these returns are usually pre-expenses *and pre-transactions costs* of asset purchases/sales and we designate these returns as $R^g_{assets,t} = \sum_{i=1}^{N} w_{i,t-1} R_{i,t}$, where $w_{i,t-1}$ = weight of asset-i in the fund portfolio at end of period t-1, $R_{i,t}$ = return on asset-i during the holding period and N = number of assets in the fund. If all of the N-assets of the fund are incorporated then $R^g_{assets,t}$ = $R_t^{net} + TER_t + TrC_t$, where TrC_t = security level transactions costs. However, often the asset holdings considered only comprise a fund's *stock* holdings and we designate this gross return as R^g_{stk} [23]. Of course, unlike *fund* returns (R_t^g and R_t^{net}), returns based solely on stock holdings do not represent returns to investors on the fund's assets as a whole. Comprehensive data on mutual fund holdings and trades (buys/sells) is not usually available but by merging databases a reasonably long time series for funds' *stock* holdings (and trades) is available quarterly but only for the US[24].

Since data on the identity of fund managers over time is somewhat sparse, studies which measure returns to specific fund managers are much less prevalent than those which measure

---

[21] This is the case for CRSP and Morningstar data bases. Other US data bases include Lipper and some researchers compile their own survivorship-bias-free data bases (Elton et al 1996a) or daily returns data bases (Busse 1999). For UK returns data the main source is Standard and Poors Micropal - in calculating NAV, prices are measured bid-to-bid.

[22] TERs are annual figures so for monthly returns 1/12th of the annual TER is added to net returns.

[23] To muddy the waters further, returns on stock holdings after deduction of *estimated* transactions costs are often referred to as 'net returns' (to stocks held/traded) - but note that this is before deduction of TERs and hence such figures do not represent returns to the *investor* from the fund's stock holdings.

[24] Mutual funds are required by the SEC to report holdings twice a year but most funds publicly disclose portfolio holdings quarterly and these are available from private vendors such as Thompson Financial (formally CDA/Spectrum).

returns to the fund itself. This is not a major drawback if a fund's style is largely a group decision. However, it is clearly of interest for investors to assess whether performance is at the fund or fund manager level, since this may determine relative fund flows and the question of whether 'money is smart' (see section 7.3).

For the US, the CRSP (Centre for Research in Security Prices) and Morningstar monthly data bases are most frequently used in academic studies, with the latter providing more detail on fund composition. Some fund databases for the US and UK have substantial coverage of both 'alive' and 'dead' funds over a fairly long time horizon, so survivorship bias can be measured and taken account of in empirical work. In principle, nonsurviving funds may have good or bad performance. They may cease to exist because they were successful and then merged with other funds or they may have been forced to close due to bad performance. The CRSP database is 'free' of survivorship bias but the Morningstar data base is not. However, all databases contain errors and very often have to be 'cleaned up'. For example, Elton et al (2001) report that the CRSP data base suffers from *omission bias,* since some funds have monthly data, some annual and some have no returns data at all - and this effects the measurement of average alpha by around 40 bp per annum.

4.2 SIZE AND POWER

Whatever metrics we choose to measure fund performance, model error and the size and power of the test statistics need to be assessed. This applies *a fortiori* if we have small samples or test statistics that do not follow standard distributions (e.g. because of non-normality due to excess kurtosis or skew in extreme winner or loser portfolios) - however, most tests of ex-post performance using $\hat{\alpha}$ and CS are usually based on standard t-tests and critical values[25]. One way to address these issues is to create random portfolios of *stocks* (rebalanced say each month) whose broad characteristics mimic those of actual mutual fund holdings (in terms of stock turnover, fund size etc). The presumption is that these random stratified portfolios of stocks which form our set of simulated mutual funds, should earn a zero abnormal return, if our chosen performance model is well specified. In addition, to assess the power of standard tests we can now add y% p.a. to the return of each of the (randomly selected) stocks in the simulated mutual fund portfolios and test the null of CS or $\hat{\alpha}$ equal to zero. The proportion of rejections of the null (e.g. determined by the number of $t_{\hat{\alpha}} > t_c$) gives an indication of the power of the test, at a particular significance level.

[25] t-statistics usually incorporate a Newey-West (1987) correction to standard errors for any serial correlation or heteroscedasticity.

Kothari and Warner (2001) show that the factor based regression approach using 36 observations to estimate α_{3F}, α_{4F} or the CS(holdings) measure have misspecification around 0.5-1% p.a. in absolute terms and the size of the test for the two alpha measures are as high as 12% (at a 5% nominal significance level)[26]. Using the alphas with 36 observations gives reasonable power of around 80% for quite moderate abnormal returns of 3% p.a. and power increases for α_{3F}, α_{4F} measures when longer data periods are used[27]. The CS measure based on *trades* has high power but only if abnormal returns accrue over the first 6 months after the trade. These simulation results provide baseline information when interpreting the statistical significance of ex-post abnormal returns using actual data. The results suggest the need to investigate performance in the tails of the cross-section performance distribution, as high abnormal returns substantially increases the power of the tests - and also to only include funds with $T_i \geq 36$ observations to avoid bias and size distortions. Note that the above results on size and power are based on standard test statistics but these are invalid if residuals are non-normal – the latter points to possible improvements using bootstrap techniques – an issue which we discuss in the next section.

There is one further statistical issue to be aware of when summarizing the performance of mutual funds. It is true that when testing alpha for a single fund (or a single portfolio, such as the average fund) then 'luck' is correctly measured by the significance level (γ) chosen. However, when we use a given significance level say γ = 5% for the alphas of each of M-funds, the probability of finding *at least one* lucky fund from a sample of M-funds is much higher than 5% (even if all funds have true alphas of zero)[28]. For example, if we find 20 out of 200 funds (i.e. 10% of funds) with positive estimated alphas at a 5% significance level, then some of these will merely be lucky. The false discovery rate (FDR) measures the proportion of lucky funds among a group of funds which have been found to have significant (individual) alphas and hence 'corrects' for luck amongst the pool of 'significant funds'. For example, suppose the FDR among our 20 winner funds is 80% then this implies that only 4 funds (out of the 20) have truly significant alphas. As we see below this adjustment is frequently ignored although recently Barras et al (2005) apply the FDR to adjust for luck.

---

[26] Huij and Verbeek (2006) also find that multifactor performance estimates suffer from systematic biases because the style proxies do not take account of transactions costs and short-sales restriction – they argue that such biases are reduced by using composite style proxies based on funds' net returns.

[27] Power increases quite rapidly as T increases and is very high for 5 or 10 years of monthly data.

[28] This probability is the compound type-I error.

4.3 SURVIVORSHIP BIAS

Malkiel (1995) on US data (1982-90) finds survivorship bias of 1.4% p.a. (value weighted fund returns) and a survivor premium of 6.5% p.a.[29] Elton et al (1996a) find survivorship bias of 90 bp and this inflates average alpha performance measures by between 40 and 100 bp over the 1977-93 period (depending on the model and sample length used) – they also find that survivorship bias is more concentrated in small funds and growth funds. With average alphas of US funds measured at around minus 70 bp p.a. then survivorship bias is quantitatively important and particularly so for data prior to 1983 (Elton et al 2001). It is also perhaps worth noting that for a *matched* sub-sample of funds in the CRSP and Morningstar databases, differences in the average alpha of either the 25 largest funds or 25 smallest funds (estimated using 60 months of data) are around 15 and 25 bp, respectively - so differences exist between databases even for the *same* funds. Tracking funds over time because of name changes, mergers, change in investment objectives and treatment of restricted funds (e.g. those closed to new investors) also create difficulties when compiling and comparing results from different data bases (Wermers 1997). Results on survivorship bias for UK funds is sparse but Quigley and Sinquefield (2000) using UK data over 1978-1997 (751 funds) report a survivor premium of 2.31% p.a. and a survivor bias of 0.7% p.a., the latter being close to that of Blake and Timmermann (1998) of 0.8% p.a.

An associated type of survivorship bias arises even when the sample of funds includes all nonsurviving funds, since some measures of performance (e.g. alpha) require funds to have existed for a minimum length of time. If this minimum is set at a 'high' level, this may reduce estimation error but may bias performance findings upwards. Using recent US data (January 1975 - December 2002) Kosowski et al (2005) find that only including (domestic) equity funds with greater than 60 monthly observations (rather than all funds) indicates that this type of survivorship bias is only around 20 bp per annum for net returns, and Wermers (1999) for the 1975-99 period based on gross returns from *stock holdings* data reports a similar figure. Simulation of mutual fund returns shows that the influence of survivorship bias on tests of persistence can bias results in either direction (Brown et al 1992, Carpenter and Lynch 1999, Carhart et al 2002b), depending on whether a single period or multiperiod survival rule is imposed.

[29] Survivor premium is the difference between the annual compound raw returns of portfolios of surviving and nonsurviving funds. Survivor bias is the difference between the annual compound returns of the surviving funds and the full set of both surviving and nonsurviving funds. In addition, 'incubation bias' arises if funds only begin to report to a database after a period of successful performance – this problem is more prevalent for hedge funds than mutual funds.

The above results demonstrate the importance of including all funds that have existed over the data period under study and more recent US and UK studies should not suffer from acute survivorship bias.

4.4 US AND UK RESULTS

We now turn to the ex-post performance of US and UK mutual funds, both active and index funds. In our survey of past work we frequently indicate the data period used, how many funds were included in the study, as well as some representative results, so the reader can better evaluate the relative contribution of past work to the key issues addressed. Statistical significance at the 10%, 5% and 1% significance levels are denoted by superscripts '*', '**', and '***' respectively and p-values are also reported. Unless stated otherwise 'statistically significant' refers to a 5% significance level (or better).

US STUDIES

Adjustment for risk is always a contentious issue for actively managed funds, so Elton, Gruber and Busse (2004) focus their attention on 52 US, S&P500 *index funds* (January 1996-December 2001) - they use the CAPM-alpha or simply the fund's differential (net) return over the market index as measures of abnormal performance[30]. The (equally weighted) average index fund's differential return is minus 0.485% p.a. and the CAPM-alpha performance is minus 0.410% p.a.[31] – these figures provide a useful yardstick with which to assess the relative performance of index funds versus the average active fund. These average underperformance figures for index funds closely match the average annual total expense ratio (TER) of 0.444%, implying that underperformance arises because they incur advertising, rebalancing, cash flow and management fees in order to closely track the index. But given the wide spread in TERs for index funds of 0.06% to 1.35% p.a. (with 25th and 75th percentiles of 0.25% and 0.59%), it is also worth noting that it may be possible to find an index fund at near zero cost (Elton, Gruber and Busse 2004)[32].

For active funds, early studies used Jensen's alpha as a measure of risk adjusted performance. For example, Ippolito (1989) using a sample of 143 mutual funds finds that most earn abnormal returns sufficient to cover their expenses over the period 1965–1984 - results which were in contrast to some earlier studies (Friend, Blume and Crockett 1970, Jensen 1968,

---

[30] They also consider tracking error and tax efficiency of index funds versus the index itself – but we concentrate on performance results.

[31] The maximum spread in alphas across index funds is quite considerable at -1.53% to 0.228% p.a.

[32] The tracking errors across funds as measured either by the absolute value of $\beta_i - 1$ or the R-squared of the CAPM regressions, range from 0.005 to 0.021 and 0.9991-1.0000, respectively and hence are very small.

Sharpe 1966). However, Elton, Gruber, Das and Hlavka (1993) show that after correcting for non-S&P500 stocks in the benchmark market index, positive pre-expense alphas disappear - a result supported by Malkiel (1995) who finds that only a *small* number of mutual fund managers (out of a total of 322 over the period 1971-91) have statistically significant Jensen's α_i^g - but there is no evidence that such ability exists at the net return level.

Chen, Jegadeesh and Wermers (2000) is a key study which uses stock trades data to investigate stock picking skill. They find using raw returns that stocks purchased by funds outperform stocks sold by them (over the next year) – this provides prima facie evidence for active funds as a whole, having some skill in picking stocks (before expenses and trading costs). Wermers (2000) examines the reasons for differences between returns based on *stock holdings* data and the overall returns on the fund. Over the 1974-1994 period he examines the relationship between buy-and-hold (value weighted) average gross returns on funds' *stock holdings* $\overline{R}_{stk}^{g}$ and CRSP net returns. The difference $\overline{R}_{stk}^{g} - \overline{R}_{CRSP}^{net}$ = 2.3% p.a. of which 0.7% is due to the lower returns of non-stock holdings with expenses and trading costs also contributing roughly 0.7% p.a. each to the reduction in funds' gross returns from their stock holdings[33]. So, although gross returns on the average funds' *stock holdings* cover expenses and trading costs – other fund assets further reduce the overall return on the fund. The latter is taken up by Shukla (2004) who is concerned that measuring fund returns using asset holdings data ignores the *interim* trading costs between the two portfolio composition dates (which are usually quarterly) – this could bias results on performance using the CS index, given the high average turnover of fund assets (of around 100% in 2003). Regardless of management fees charged, any revisions to a portfolio would only add value for investors if the return on a revised portfolio (net of added trading costs) is higher than the return on a (buy-and-hold) passive portfolio. He shows that in the US, such a return differential is zero on average across funds (for up to 6 month horizons) – so investors do not obtain any additional return due to frequent portfolio revisions by active traders, as a whole[34]. However, there is some (weak) evidence that portfolio revisions by *growth funds* add value and there is also a wide dispersion in returns to trading across different funds – the return differential to trading is largest for small funds and those with more concentrated portfolios. So there may be some individual funds that are successfully trading at the margin.

---

[33] Transactions costs are estimated from the trade data using Keim and Madhavan's (1997) estimates of institutional execution costs for stocks (commissions plus price impact), based on cross-section regressions.

[34] Data used is from Morningstar Pricipia CD-ROM, for 458 funds, August 1995-December 2002. The return differential to trading (net of transactions costs) over the return on the passive portfolio is $R_t^{diff} = (R_t^g - TrC_t) - (R_t^{g,psv} - TrC_t^{psv})$, where R_t^g = gross return from active trading, $R_t^{g,psv}$ = gross return on passive portfolio, with transactions cost on the passive portfolio of TrC_t^{psv} - which we are assumed to be zero.

In a novel approach Baker, Litov, Wachter and Wurgler (2005) examine the returns of fund holdings and trades around earnings announcement dates, arguing that 'announcement returns' are likely to be 'abnormal'. In an event study framework they classify funds into those with weight increases (decreases) and examine subsequent returns for 3 days around the earnings announcement dates. Stocks in which funds have experienced increasing (decreasing) weights have 20 bp per annum higher (21 bp lower) returns in the subsequent earnings announcement periods (relative to matched CS-returns). This *difference* in future CS returns of around 40 pb per annum is similar to that for the raw returns difference, indicating stock picking skill (before costs) *within* characteristic groupings.

How do US funds perform on average when we consider net returns to investors rather than gross returns to the fund itself. On a CRSP net return (value weighted) basis Wermers (2000) find α_{4F}^{net} = -1.16%***p.a. indicating underperformance by the average fund over 1974-94. Bringing the latter figure up to date (January 1975-December 2002) using around 1,700 mutual funds, Kosowski et al (2005) find that the (equally weighted unconditional 4F) net return alpha is about minus 0.5% p.a. – so the 'average active fund' underperforms its benchmarks[35]. The cross-section standard deviation of these α_i across the 1,700 funds is high, indicating the possibility that some funds are performing very well (and others very badly)[36]. In an interesting study Kosowski (2006) finds that the average alpha (depending on the model used) is 3-5% p.a. higher in recessions than in expansions, demonstrating that investment in mutual funds (and particularly growth orientated funds) may provide value added in times when marginal utility of wealth is high[37].

UK STUDIES

Much less empirical work on performance has been done on UK funds. Leger (1997) estimates the CAPM-alpha on 72 UK investment trusts in four non-overlapping five-year samples between 1974 and 1993 and finds little evidence of statistically significant ex-post performance. Quigley and Sinquefield (2000) examine the performance of all UK equity mutual funds (including

---

[35] Barras et al (2005) over the same time period find α_{4F}^{av} = -0.44%p.a. (p=0.16) but use 1472 funds whereas Kosowski et al (2005) have 1,788 funds (both studies use funds with $T_i \geq 60$ observations).

[36] For example the top and bottom ranked funds have 4F net return alphas of 4.2% and minus 3.6% *per month*, respectively while the 95th and 5th percentile funds have alphas of 0.4% p.m. and minus 0.5% p.m. respectively (Kosowski et al 2005).

[37] Over 1965-2002, Kosowski (2006) uses both a split sample technique (based on NBER recession/expansion periods) and a conditional regime switching model.

non-survivors) between 1978 and 1997 (752 funds) using gross returns (i.e. before deduction of management fees). An equally weighted portfolio funds gives α_{3F}^{g} of about minus 1% p.a. (t=2.3) and poor average performance is found across all four investment styles (i.e. growth, income, general equity and smaller companies). Similarly, Fletcher (1997) over 1980-1989 using an APT model for growth, income and general equity categories finds no statistically significant abnormal performance[38].

The above studies strongly suggest that the average US or UK fund (even within specific sectors) does not earn abnormal positive net returns. Of course, this does not preclude funds in the tails of the distribution having statistically significant alphas - an issue we take up in section 5.4.

4.5 LUCK AND FUND PERFORMANCE: A TALE OF TWO TAILS

As we have seen there is considerable evidence for the US and the UK that the *average* active fund under-performs its benchmark returns. However, it is also found that some subgroups of funds do seem to out-perform their benchmarks (e.g. US growth oriented funds, Chen et al 2000, Wermers 2000). Wermers (2003 - 'big bets') provides prima facie evidence that there may be a small number of funds taking 'big bets' which on average outperform funds taking small bets and the former also have positive risk adjusted returns[39] - but it is clear that isolating such funds is a difficult task and there are some funds taking 'big bets' that do not outperform. So, can we be sure that such 'group' out-performance is not due solely to 'good luck'? Kosowski, Timmermann, White and Wermers (2005) is the first paper to explicitly control for luck when measuring *individual* fund performance using a cross-section bootstrap methodology with US monthly net returns, while Barras et al (2005) examine luck for *a group* of top performing funds using the false discovery rate. We discuss these in turn.

Funds in the extreme tails of the performance distribution are likely to exhibit non-normality in their idiosyncratic risk - they may also be 'short lived' funds so standard asymptotic results do not apply. Hence, test statistics based on standard critical values may give misleading inferences. Under these circumstances bootstrapping procedures are required. The simplest approach is to apply the bootstrap on a fund-by-fund basis - but this excludes information in the cross-section of luck across *all* funds.

---

[38] Benchmark portfolios are estimated by an asymptotic principal components technique as outlined in Connor and Korajczyk (1986).

[39] Funds undertaking 'big bets' are defined are those with a large *standard deviation* of market index adjusted returns, while abnormal returns are measured as either average market adjusted returns or Jensen's alpha or the unconditional 4F-alpha. These variables are measured over 3-year non-overlapping periods and bi-variate *contemporaneous* regressions are run for each 3-year period over 1975-2000.

To highlight the importance of a cross-section bootstrap consider whether the stellar performance by some funds even over a run of many years, is due to luck or skill. Are these 'stars' just the lucky ones amongst the cohort of all fund managers?[40] For example, if we are told that a particular mutual fund has an abnormal average return of say 10% pa then we might well be impressed. But if we are told that this return of 10% p.a. was achieved by the best performing fund out of say 1,000 funds, we should be less impressed. This is because in a large universe of 1,000 funds there will always be some funds that perform well (badly), simply due to chance. The issue then arises as to how we can separate 'skill' from 'luck' for *individual* funds, particularly when idiosyncratic risks are highly non-normal – as is the case for funds in the extreme tails, in which investors are particularly interested.

Suppose we are interested in the performance of the best fund (in the ex-post data) and whether this is due to skill or luck. We could 'replay history' by only considering the idiosyncratic risk of the ex-post best performing fund. But when we replay history for the second or third etc. ranked fund in the *ex-post* data, it is quite conceivable that one of these funds now has the 'best' performance. This would hold *a fortiori* if the distribution of idiosyncratic risk for the second, or third, etc. ranked funds have relatively large variance. Clearly, re-running history for just the ex-post best fund ignores the other possible distributions of luck encountered by all other funds – these other 'luck distributions' provide highly valuable and relevant information. Put more technically, in picking out the best fund (ex-post) we have 'ordered' it as 'number one' and because of that fact we need to compare its performance with all other funds which have the potential to be 'number one', if we are to separate skill from luck for the best fund – this is the theory of order statistics[41].

Briefly, the simplest ('residual only resampling') cross-section bootstrap procedure is as follows. Suppose the estimated factor model is $r_{i,t} = \hat{\alpha}_i + \hat{\beta}_i' F_t + e_{i,t}$, for i = {1, 2, …, M) funds, where T_i = number of observations on fund-i, F_t = vector of risk factors and $e_{i,t}$ = residuals of fund-i. For each fund-*i*, we draw a random sample (with replacement) of length T_i from the residuals $e_{i,t}$ and use these *re-sampled* bootstrap residuals $\tilde{e}_{i,t}$ to generate a simulated excess

---

[40] Peter Lynch of the Megellan fund is cited as a star manager (Marcus 1990) and the US based Schroder Ultra Fund earned 107% over 3 years ending in 2001, and was closed to new investors as early as 1998.

[41] Unfortunately, analytic results from order statistics are only available for well defined distributions but since idiosyncratic risk across each fund is different and does not follow known distributions, we have to resort to bootstrap procedures (see Efron and Tibshirani 1993, Politis and Romano 1994).

return series $\tilde{r}_{i,t}$ for fund-i, under the null hypothesis of no abnormal performance (i.e. setting α_i = 0 so that $\tilde{r}_{i,t} = \hat{\beta}_i' F_t + \tilde{e}_{i,t}$). This is then repeated for all funds. Next, using the simulated returns $\tilde{r}_{i,t}$, the performance model is estimated and the first bootstrap estimate of alpha $\tilde{\alpha}_i^{(1)}$ for each fund is obtained. The $\tilde{\alpha}_i^{(1)}$ estimates for each of the M-funds represent sampling variation around a true value of zero (by construction) and are entirely due to 'luck'. The $\tilde{\alpha}_i^{(1)}$ {i = 1, 2, …, M} are then ordered from highest to lowest, with the maximum bootstrap alpha denoted $\tilde{\alpha}_{\max}$. The above process is then repeated B times. Suppose we are only interested in the performance of the best ('max') fund. The B-values of $\tilde{\alpha}_{\max}$ provide the distribution of $f(\tilde{\alpha}_{\max})$ under the null α_i=0 for all funds. We now compare the *ex-post* $\hat{\alpha}_{\max}$, with its appropriate 'luck distribution' $f(\tilde{\alpha}_{\max})$ - if $\hat{\alpha}_{\max}$ is greater than the 5% upper tail cut off point from the empirical distribution $f(\tilde{\alpha}_{\max})$ then we reject the null that its performance is due to luck (at 95% confidence) and we infer that the fund has genuine skill. Note that the distribution $f(\tilde{\alpha}_{\max})$ uses the information about 'luck' represented by all the funds and not just the 'luck' encountered by the 'best fund' in the ex-post ranking. This is a key difference between Kosowski et al (2005) and many earlier studies that use the bootstrap methodology on a fund-by-fund approach.

The above methodology can also be applied to other performance statistics such as t_{α_i} which has superior statistical properties (Hall, 1986, 1992)[42]. For example, after bootstrapping on t_{α_i} (using a maximum of 1,704 US funds, January 1975-December 2002), Kosowski et al (2005) find that funds ranked above the top 5th percentile (i.e. a maximum of about 90 funds) have genuine skill with α_{4F}^{net} in excess of 4.8% p.a.[43] The proportion of funds with positive alpha-skill

---

[42] t_{α_i} has better sampling properties than α_i - the obvious reason being that the former 'corrects for' high risk-taking funds (i.e. σ_{ε_i} large), which are likely to be in the tails. Put another way, if the distribution of alpha for each fund is *niid* (under the null) but each fund has a different σ_{ε_i} then the cross-section distribution of the alphas $f(\alpha_i)$, will *not be normal*, but the distribution of $f(t_{\alpha_i})$ remains normal – this is the basis for t_{α_i} (but not α_i) being a 'pivotal statistic' and hence it is preferred as the performance metric in the bootstrap. However, it is important to note that even when using t_{α_i} as the performance metric, we cannot generally infer what the tails of the cross-section distribution $f(t_{\alpha_i})$ will look like (e.g. if ε_i are drawn from a mixture-normal distribution) – this is why we need the cross-section bootstrap.

[43] This result is largely invariant to use of an conditional/unconditional 4F model, to the minimum number of monthly observations used ($18 < T_{\min} < 120$), sorting on alpha rather than t-alpha and to the type of bootstrap (e.g. block bootstrap to account for serial correlation, or a bootstrap taking account of contemporaneous cross-fund correlations in idiosyncratic risk).

is as high as 30-40% in the 1975-89 period, but falls to around 5% in the 1990-2002 period, when there are far more funds under management[44]. These skilled funds are all found to be in either the aggressive growth or growth styles, rather than in 'growth or income', and 'balanced-income' sectors.

Bootstrapping using *gross* returns and the CS(holdings) measure, indicates stock picking skill for funds at or above the 20th percentile. Coupled with the above result using net returns, this implies that funds ranked between the top 5% and top 20% earn positive *gross* returns from their stock picking skills, but they do not earn enough to cover management fees and transactions costs. It is only the top 5% of funds which genuinely earn positive abnormal returns *for investors*. As with results based on t-alpha, genuine 'good skill' when using CS is found mainly for growth and aggressive growth styles.

Barras et al (2005) take a somewhat different approach to 'luck' than Kosowski et al (2005) focusing on the 'false discovery rate' FDR - that is, the proportion of lucky funds among funds with significant (individual) alphas. This methodology unlike Kosowski et al does not require the multivariate distribution of all funds' performance but adjusts the count of the number of significant funds, for those funds that are merely lucky. The Barras et al (2005) procedure therefore deals with the 'true' statistical significance within a *portfolio* of funds, whereas Kosowski et al (2005) approach applies to the statistical significance of *individual* (ordered) funds. For 'all' US equity funds (1975-2002), Barras et al find a FDR of 55% amongst the 52 'winner' funds (at a 5% significance level), so only 23 of these (which constitutes 2% of all funds) have genuine skill - and such skill is found to reside in growth styles rather than in the 'growth and income' style (for the latter the FDR=100%, so all significant positive alphas are due to luck). Also, although there are only a small number of genuinely skilled funds, they all lie in the extreme right tail of the alpha-distribution, so a portfolio of these funds has a reasonable high expected alpha of 2.35% p.a. (at γ = 5% significance level, p. 23)[45].

Using a similar bootstrap approach to Kosowski et al (2005) on UK data (842 funds, 1975-2002), Cuthbertson, Nitzsche and O'Sullivan (2006) find only 12 funds out of the top 20 funds have genuine skill (at a 10% significance level). As one moves further towards the centre of the performance distribution (i.e. at or below the top 3% of funds) there is no evidence of stock

[44] Similarly, risk adjusted returns to hedge funds are higher in the early 1990's than in the later 1990's (Agarwal and Naik 2002).

[45] Note that the Kosowski et al(2005) and Barras et al(2005) definitions of luck are very different. The former is based on the significance of an ordered fund's *individual* alpha at a specific quantile of the multivariate performance distribution, for a given significance level, while the latter calculate the *number* of genuinely significant funds from a portfolio of funds which have been found to be individually statistically significant (at a given significance level).

picking ability – the bootstrap indicates that any positive $t_{\hat{\alpha}}$'s are due to luck rather than skill. They also show that 'genuine' top performers are not necessarily those with an ex-post ranking right at the 'top'. This makes it extremely difficult for the 'average investor' to pinpoint *individual* active funds which demonstrate genuine skill, based on their track records. For the UK, in contrast to the US results, skill appears to reside with equity income funds rather than 'all company' or small company funds[46].

Note that the above results on top performing funds are broadly consistent with a competitive equilibrium since we expect to observe very few funds with positive risk adjusted returns over long horizons. This is because funds with genuine skill and high past returns have large inflows (see section 7.2) and with increasing marginal costs to active management, this should lead to zero long run abnormal returns to investors (Berk and Green 2004).

At the negative end of the performance scale using net return 4F-alphas, UK and US results strongly reject the hypothesis that most poor performing funds are merely unlucky.[47] Between 20% and 40% of funds have negative abnormal performance which is due to 'bad skill' rather than 'bad luck' (Kosowski et al 2005, Cuthbertson et al 2006). Similarly, Barras et al (2005) using the FDR find around 20% of all funds have genuinely 'bad skill' and these funds are spread throughout much of the left tail (and across all investment styles). The Barras et al approach applies to 'counts' of funds within a portfolio of 'significant funds', rather than to the statistical significance of *specific* funds - but they give a broadly similar picture to Kosowski et al (2005).

The results for the left tail of the performance distribution are *not* consistent with the competitive model of Berk and Green(2004), since 'bad skill' should lead to large outflows from these funds and the return on such funds who subsequently survive should, in equilibrium, equal that on a passive (index) fund. The continued existence over long time periods, of a large number of funds which have a truly inferior performance (which cannot be attributed to bad luck), may indicate that many investors either cannot correctly evaluate fund performance or find it 'costly' to switch between funds, hence any approach to a competitive equilibrium appears to be very slow.

---

[46] They also find that positive performance amongst onshore funds is due to genuine skill, whereas for offshore funds, positive performance is attributable to luck.

[47] When recursive OLS (or the Kalman filter-random coefficients model) is applied to a *portfolio* of UK funds based on bootstrap rankings, Cuthbertson et al (2006) find considerable stability in the estimated portfolio alphas (as well as the market return and SMB factor loadings). This suggests that over several years, there is genuine constant outperformance amongst a few top funds and genuine underperformance amongst many poorly performing funds. Also, at the bottom of the performance distribution for US funds, the CS measure is not statistically different from zero, so these funds earn zero abnormal returns gross of fees and transactions costs (but as we have seen, negative net abnormal returns to investors).

EXTRANEOUS AND PRIOR INFORMATION

Recent work has emphasized the use of information extraneous to a specific fund to help pick superior funds – this information can be purely data based or involve priors in a Bayesian framework (or both). The idea that badly measured (standard) fund alphas might be improved with the use of prior information on that fund and the formation of Bayesian alphas is well known. However, Jones and Shanken (2005) suggest that the alphas of *all other funds* may be used to improve estimates of an individual fund's alpha. The intuition for this can be demonstrated as follows. Assume *all* funds' alphas are distributed as $\alpha_i \sim niid(\mu_\alpha, \sigma_\alpha)$ and you are concerned with the estimate of alpha-XYZ. If residuals are independent across funds and the number of funds M is very large then we would have very accurate estimates of $(\mu_\alpha, \sigma_\alpha)$. In the absence of any information about α_{XYZ} it would seem reasonable to take $(\mu_\alpha, \sigma_\alpha)$ as an estimate of its alpha and precision – even though we do not know where XYZ lies in the cross-section distribution. If we now have a (standard) estimate of α_{XYZ} it would seem sensible that our best estimate of XYZ's alpha is a weighted average of α_{XYZ} and μ_α (with relative weights depending on the precision of the two estimates). This they refer to as *'learning across funds'* since all other fund returns influence the estimate of any one fund's alpha.

One further twist can be added to this approach. It follows from the above that *priors* about $(\mu_\alpha, \sigma_\alpha)$ should influence priors about an individual fund XYZ, which in turn influence that fund's posterior alpha – so the priors for each fund are not independent. This avoids the problem that with independence across residuals, then as the number of funds increases the maximum-alpha fund increases without bound even when all true alphas are zero – a standard result from the theory of order statistics. Hence with prior independence, as the number of funds increases, it will always be possible to find an active fund with a positive alpha - thus supporting active investment even when investors are very skeptical about managers' skill (as found by Baks, Metrick and Wachter 2005). But with prior *dependence* across funds this nonsensical outcome is precluded. This is because the maximum alpha-XYZ is shrunk towards μ_α and if there is no skill across funds the latter will tend to zero – in addition, as the number of funds increases σ_α approaches zero and the relative weight given to μ_α in the pooled estimate for alpha-XYZ also increases. The maximum posterior estimate for alpha-XYZ is therefore bounded[48]. An example

[48] There is a useful parallel here with the cross-section bootstrap procedure of Kosowski et al (2005) which also compares the maximum ex-post alpha with the cross-section of risk across all funds (which may not be independent), thus avoiding misleading inferences by just bootstrapping on a fund's own residuals.

which illustrates such shrinkage is the Fidelity Magellan Fund which over 1963-2000 has an OLS alpha of 10.4%**pa. Is this plausible? Using, *'learning across funds'* and estimates, $(\mu_\alpha, \sigma_\alpha)$ = (1.5, 1.5) the posterior alpha for Magellan is 4.8% p.a. - a substantial 'shrinkage' on the standard alpha[49]. So pooling information from other funds seems to reduce the variability in (posterior) estimates of extreme fund alphas (for any set of priors) – this should help to avoid errors when choosing 'extreme performers'.

Overall, what these latest studies demonstrate is the importance of looking at performance of funds in the tails of the distribution (rather than the average fund) and then making appropriate adjustments for idiosyncratic risk across all funds before making inferences. The clear message from recent UK and US results is that there are a few 'top funds' who demonstrate genuine skill but the majority have either no skill and do well because of luck or, perform worse than bad luck and essentially waste investors time and money.

4.6 MARKET TIMING

Without asset holdings data we can use the Treynor-Mazuy,TM (1966) or Henriksson – Merton, HM (1981) regressions using conditioning variables. For the US, using monthly data Treynor-Mazuy (1966) find only 1 fund out of 57 has significant timing ability while Henriksson (1984) finds only 3 funds out of 116 show significant market timing ability. Similarly, the non-parametric (frequency) test used by Jiang (2003) on US monthly returns finds no evidence of successful market timing (over the period 1980-1999, using 1827 surviving and 110 non-surviving funds).

In a recent study Bollen and Busse (2001) argue that market timing may occur at a frequencies higher than monthly and therefore use daily data (1984-1995) on 230 domestic equity funds (and compare these results with those using monthly data). Their methodology involves generating simulated returns and then testing the size and power of conventional t-tests on the timing coefficient in the TM and HM models.[50] When using both daily and monthly data they find the tests have the correct size but the daily data has more power. For example, for moderate

---

[49] The Bayesian approach incorporates uncertainty in estimates of $(\mu_\alpha, \sigma_\alpha)$ by using Gibbs sampling techniques. Also, portfolio allocation (based power utility from next periods terminal wealth) is far less sensitive to priors when 'learning across funds' is used since without the latter, portfolio allocations are very sensitive to 'high', 'some' and 'no' skepticism in the priors – another example of 'shrinkage'.

[50] To evaluate the size of the test, set $\gamma_{i,m}$ = 0 in equation (6), simulate returns, perform a standard t-test and count the frequency of rejections of the null. Power is assessed by using various values for $\gamma_{i,m}$ > 0, generating simulated data and then testing the frequency of rejections of the null of no market timing.

market timing[51] (e.g. $\gamma_{i,m} = 5$) the simulated daily data gives significantly significant positive coefficients for $\gamma_{i,m}$ in the HM and TM models around 85% of the time, whereas for the monthly data this drops to only 30%. Using *actual* daily (monthly) returns Bollen and Busse find that about 40% (34%) of the funds generate a significantly positive $\gamma_{i,m}$ and 28% (5%) a significantly negative $\gamma_{i,m}$, for the TM model. So, daily data indicates more significant market timing effects but with roughly equal numbers of significantly 'good' and 'bad' market timers - and 32% who have no significant market timing skills at all [52].

Market timing using portfolio data and individual stock holdings data has yielded broadly similar results. Analyzing the relationship between cash and equity holdings to see if the manager increases (decreases) her exposure to the market return just before a rise (fall) in the market index, Graham and Harvey (1996) and Becker et al (1999) find no evidence of successful market timing. Using the characteristic timing measure CT, studies on US data find no evidence of successful market timing over a horizon of one year (Wermers 2000) – funds therefore cannot successfully time the characteristic benchmarks (e.g. accurately forecasting when returns on small stocks will be higher than those on large stocks). However, directly regressing fund betas (calculated using fund holdings and estimated stock betas) on the HM and TM market return variables, Jiang et al (2005) find stronger evidence for positive market timing over a 3-month horizon, particularly for aggressive growth and growth funds, in the top 25th percentile of funds (over the period 1980-2003).

There has been relatively little research carried out on the market timing skills of UK funds. Fletcher (1995) evaluates the market timing of 101 mutual funds between 1980 and 1989 using the HM and TM models and finds that the funds have no market timing skill. In fact, the results suggest that funds, on average, reduced their market exposures when market returns were high and *vice-versa*. Leger (1997) using UK equity investment trusts between 1974 and 1993 also finds negative and statistically significant timing (using the TM model).

---

[51] For $\gamma_{i,m}$ =5, if the market is expected to rise by 5% over the month, then beta will increase by 0.25.

[52] Using the same daily data set Busse (1999) examines the allied concept of volatility timing. Do funds reduce their market betas when conditional volatility is higher than average and hence enhance risk adjusted returns? Although the Sharpe ratio is higher for those funds which reduce market exposure in times of abnormally high volatility, there is no relation between the strength of volatility timing in the previous six months and fund performance in the next six months (in terms of either the 4F-alpha or the Sharpe ratio) – so volatility timing is not an exploitable strategy for investors (see also, Busse 2001).

Overall, the evidence seems to be quite conclusive that market timing and volatility timing are not likely to provide profitable strategies on a net return, risk adjusted basis.

5. PICKING WINNERS: PERSISTENCE

Ex-post alphas (and CS measures) indicate average abnormal performance over some past data period. However it is also important to assess whether there are ex-ante rules which can be used to choose funds which subsequently earn statistically and economically significant abnormal returns, either for the fund or for the investor (i.e. after deduction of all costs) – in short whether there is persistence in fund performance. This section evaluates alternative methods used to measure persistence from a 'statistical' and economic viewpoint and report results for US and UK mutual funds. For investors as a whole to benefit from persistence we also have to establish that they are 'smart' and allocate additional funds to 'winners' and withdraw funds from 'losers' – this is discussed in section 7.2. Of course, if we cannot establish persistence then investors may be wasting resources in chasing potential 'winner funds'[53].

STATISTICAL PREDICTABILITY

Predictability and persistence are slightly different concepts. Persistence implies funds ranked as 'past winners' (losers), tend to stay winners (losers) in the future, so there is a positive correlation between past and future performance as funds maintain their *relative* positions. Predictability allows for the latter but also for past winners to become future losers (i.e. reversals, implying negative correlation). Tests for persistence/predictability fall into two broad categories. One can test for 'statistical predictability' or 'economically significant' predictability or both. Statistical measures of persistence rank funds over some past horizon and measure the association between past performance and future performance – where the performance metrics may be different in the ranking and post-ranking periods (e.g. ranking into decile portfolios based on past raw returns but the post-rank metric being future alpha performance). The statistical approach measures the average association between the *relative orderings* of funds in the pre- and post-sort periods – using correlation or regression procedures. However, such tests do not directly tell us whether the future abnormal return of past winners or losers, is positive or negative. For example, (rank) correlations or a regression of pre- and post-sort CS or alphas can be used to establish predictability - although there may be a high correlation between the alphas of past decile ranked funds and their subsequent alphas, nevertheless all of the post-sort decile-

[53] As we have seen Berk and Green (2004) demonstrate that if investors chase potential winners, the market may be efficient even though investors do not earn an *equilibrium* return above that on a passive fund. Picking winners usually involves a changing set of funds at each rebalancing period, so the finding of *short-run* persistence is not necessarily inconsistent with Berk and Green's equilibrium result.

alphas may be negative, indicating (relative) predictability but poor future abnormal performance for all decile portfolios.

Similarly, tests based on contingency tables (e.g. log-odds ratio, Wilcox test) – see Teo and Woo (2001) and Tonks (2006) - involve a 'frequency count' of fractiles of repeat winners WW and losers LL (relative to the number of WL and LW's) in two different periods. But any measured 'persistence' only involves relative frequencies, so we cannot directly assess the economic significance of the results (e.g. in terms of the risk adjusted returns to the persistent winner/loser portfolios) and it is often not clear how such results may be exploited by investors. Also, in the contingency table and regression/correlation approaches, measured persistence may be due mainly to repeat losers rather than repeat winners. In addition (Spearman) rank correlations treat each point in the ranking equally and lack power against the hypothesis that predictability in performance is concentrated in the tails of fund performance – an issue we take up in section 6. While the above approaches can be used to establish statistical predictability, investors are presumably more interested in the future absolute performance of both winners and losers (taken separately) – an issue to which we now turn.

RECURSIVE PORTFOLIOS

A popular method of testing for both the statistical and economic importance of persistence amongst winners and losers is the recursive portfolio method. First a sorting rule is established. For example, using monthly data we might classify funds into deciles at time-t, based on *any* fund attribute (e.g. its turnover, size, previous one year return, past CS measure or its alpha estimated over the previous 36 months). At the portfolio formation date t, (equally weighted or value weighted) decile portfolios are formed. The portfolio holding period (t, t+h) is then established (e.g. h=12 months) and the monthly returns noted, after which rebalancing takes place and new decile portfolios are formed.[54] This gives rise to a sequence of monthly ex-ante *'forward looking'* (or 'post-sort') returns $R_i^f(t,T)$ - where $t = t+1, t+2, \ldots T$ [55]. If the ranking criterion is based on a fund's alpha then the recursive portfolio method allows (past) factor loadings to change over time. If the complete (concatenated) monthly time series $R_i^f(t,T)$ is used to estimate the 'forward looking' post-sort α_i^f then we assume the 'forward looking' factor

[54] When a fund dies sometime over the forward looking horizon, it is usually included in the portfolio until it dies and the portfolio is then rebalanced amongst the remaining 'live' funds, until the next rebalancing period. If we do not implement this procedure then 'lookahead bias' ensues.

[55] In some studies there is a gap between the rebalancing dates and the measurement of post-sort returns. For example, at each rebalancing date we might track future returns only over horizons from t+3 to t+15 rather than from t to t+12 – this allows a test of longer horizon persistence, without confounding the results with short-horizon persistence (e.g. see Teo and Woo 2001).

loadings (and alpha) over (t, T) are constant - but this should be tested[56]. One method to mitigate the latter problem is to estimate the factor model using a moving window with a minimum of (say) successive 36 monthly values of $R_t^f(t,T)$ - here we assume α_i^f (and factor loadings) for fractile portfolios are constant over each 36 month period. Clearly, testing for 'short horizon persistence' *and* time-varying factor loadings, can only be achieved by using relatively high frequency data (e.g. daily). Note that time varying *fund* factor loadings may arise either from time varying factor loadings for individual stocks or, funds changing the proportions they hold in particular stocks (e.g. due to short-term market timing, or strategic asset allocation, or accommodating inflows/outflows due to liquidity trades).

When constructing the CS measure from the $R_i^f(t,T)$, one potential advantage over the parametric alpha method is that persistence on a risk adjusted basis can be measured over any horizon which is an integer multiple of the frequency of stock holdings or trade data – the latter minimum period is usually one quarter. This is because 'risk adjustment' using characteristics of the stocks in the portfolio, only requires data for the benchmark returns (and stock holdings). The CS measure is semi-parametric and in measuring persistence, we can apply an event study approach. Each time the portfolio is rebalanced at t, we track $CS^f(t,T)$ = { $CS_{t,t+1}$, $CS_{t+1,t+2}$, $CS_{t+2,t+3}$, …, $CS_{T-1,T}$ } to the end of the data set. Aggregation over each rebalancing period {t, t+1,… ,T-1} gives mean values $\overline{CS}_{t,t+1}, \overline{CS}_{t+1,t+2}$ … and we can test each *separate* event horizon for persistence. Aggregation over time provides post-sort returns over longer horizons and averaging these allows tests of the null $\overline{CS}(t,t+k)=0$ for alternative values of k (using standard errors corrected for overlapping data problems). Thus there are no problems in testing whether persistence is valid over short (quarterly) or longer horizons[57].

It is worth noting that sorting may involve any rule that is thought to separate funds into future 'winners' and 'losers' (e.g. the number of letters in the name of the fund) but the performance metric will usually be a risk adjusted measure. Single or multiple sorts (e.g. by past return and by fund size) are possible but clearly there are data limitations on how far one can undertake multiple sorts. Whatever ranking criteria is chosen, then within any fractile, funds may be quite heterogeneous (e.g. the top past return decile may contain funds with quite a large variation in returns, fund size, styles, turnover, etc.). Sorting into finer fractiles (e.g. top 1%) may

---

[56] If we use *recursive* past alphas to rank the funds but then use *all* the post-ranking data $R_i^f(t,T)$ to estimate ex-ante performance we are being somewhat inconsistent since we are implicitly assuming that past *fund* factor loadings and alphas maybe time varying but future fractile parameters are constant.

[57] Although note that testing the null on $\overline{CS}(t,t+2)$ is not independent of a joint test on $\overline{CS}_{t,t+1}, \overline{CS}_{t+1,t+2}$.

involve a trade off between power (due to high and non-normal idiosyncratic risk) versus the heterogeneity of funds within the fractile. Here bootstrapping across either individual funds or a cross-section bootstrap may be desirable – see section 5.4.

The (post formation) forward looking returns will usually consist of a changing portfolio of different mutual funds, although not all funds will necessarily be 'switched' at each rebalancing date – the implications for transactions costs of this 'fund-of-funds' rebalancing strategy is discussed below. There is an added danger when the model used to assess ex-ante performance is the same as that used in the (ex-post) ranking criterion, since any bias in the performance measure will apply to both periods, which may lead to an erroneous inference of persistence (Brown, Goetzmann, Ibbotson and Ross, 1992).

DATA SNOOPING BIAS

All tests are of course subject to data snooping bias. This arises when a *given* data set is used more than once for inference or model selection. In any finite set of data, a search over a large number of models (or profitable trading strategies) will unearth some successful 'outcomes' (e.g. positive abnormal returns) purely by chance – and these are the ones that may be reported. Indeed with enough permutations we can find a successful mechanical trading rule on a set of random numbers - provided that we can test our rule on the *same* set of random numbers which we use to discover the rule. Add to this the likelihood of survivor bias in rules, (i.e. rules that do not work on *new* data are discarded) then the probability of finding at least one successful rule in even a long time series of data, may be quite high (Sullivan et al 1999, 2001).

Data snooping bias arises because tests are usually only conducted on a subset of surviving rules and not on all the other trading rules which did not survive. The dangers of data snooping bias in tests of persistence of mutual fund performance are high because of the large number of permutations across the different rules for portfolio formation, the rebalancing period chosen and the horizon over which future returns are evaluated. Guarding against these biases requires that successful 'rules' remain successful when confronted by genuinely new data.

6. PERSISTENCE: EVIDENCE

Having examined empirical results on ex-post performance in section 4 we now turn to evidence of possible ex-ante strategies whereby investors can 'beat the market' - in short, an analysis of the EMH applied to mutual funds. In section 6.1 we present evidence on 'predictability' (i.e. statistical association between past and future performance) and in section 6.2

we consider the economic significance of persistence (i.e. the size as well as the statistical significance in abnormal performance). In each section we begin with studies on the US mutual fund industry before commenting on the less numerous studies for UK funds. We define short (long) horizon predictability/persistence as involving statistically significant effects over a period of less (greater) than one year.

6.1 PREDICTABILITY: STATISTICAL MEASURES

US STUDIES

In a key early study of long run predictability using 279 US equity mutual funds, Grinblatt and Titman (1992) split their 10 year sample period (December 1974 to December 1984) into two 5 year sub-periods, estimate α_i for each fund and perform a cross-section regression of $\alpha_{i,t}$ on $\alpha_{i,t-5}$.[58] They find that both past winner and past loser deciles exhibit predictability, with stronger evidence for predictability among past losers. Hendricks, Patel and Zeckhauser (1993) using a contingency table approach (on raw returns and Jensen's alpha) with 165 equity funds (1974-84) find evidence of predictability over successive 1-year periods – the so-called 'hot-hands' investment strategy. Similarly, Goetzmann and Ibbotson (1994) in a study of about 700 mutual funds (1976 – 1988) using a variety of sorting rules and statistical metrics,[59] find evidence of 'statistical predictability' over 1-month, 1-year and 2-year horizons. In general these early studies tended to use databases that included only surviving funds but Brown et al (1992) pointed out that survivorship bias can result in a level of persistence as found in the above studies, even though non exists. Brown and Goetzmann (1995) and Elton, Gruber and Blake (1996b) are two studies which use a database free of survivorship bias and over the 1970's to 1990's find evidence of statistical predictability for rebalancing periods of 1 and 3-years, when using a number of performance ranking criteria (past 1-year and 3-year alphas and t-alphas, from a three factor model plus a bond return).

The influential paper by Carhart (1997) uses a comprehensive database of over 1,800 US equity mutual funds (1963 – 1993), sorts funds into deciles based on past one-year net returns or past 3-year, 4F-alphas and finds some evidence of one-year persistence for the top and bottom decile ranked funds using a contingency table approach. He then tracks each decile fund's gross returns over the subsequent 1-5 years and finds that persistence of up to 3 years

---

[58] The authors calculate abnormal returns from a risk model incorporating various size, dividend yield and momentum characteristics. They also check for possible bias by repeating the regressions with a randomly chosen stratified control sample of 109 funds with similar characteristics.

[59] Sorting rules include for example, previous years raw returns and past 2-year alphas and the statistical metrics include contingency tables, regressions and rank correlations.

occurs for the lowest decile ranked fund but for all other deciles there is little or no evidence of persistence[60]. More recently Teo and Woo (2001) use the contingency table approach to test predictability in raw returns and style adjusted returns. Style adjusted returns for fund-i are defined as $SAR_{i,t} = R_{i,t}(i \in S) - R_{s,t}$, where $R_{s,t}$ is the (equally weighted) net return on all funds with style-S and they classify funds into the nine styles used by Morningstar.[61] Sorting by either past one year SARs or raw returns (using a median split) they find more 'winner-winner' WW and 'loser-loser' LL portfolios than one would expect by chance over the next year and for SAR's (but not raw returns) contingency tables also indicate persistence for up to 3 years. Hence there is evidence of statistical predictability over long horizons for funds sorted by SAR's but not by raw returns.

In a further assessment of persistence Blake and Morey (2000) study the Morningstar 5-star rating service as a predictor of US domestic equity mutual fund performance and find that the top (5-star) rating funds did not outperform the 3-star rated funds. However, using a sample of around 1,800 funds Morey et al (2006) find that the change in Morningstar ratings methodology that took place in 2002 does predict future relative risk adjusted returns over the next 3 years - so 5-star funds do better than lower ranked funds - but the future 4F-alpha of the 5-star funds is negative[62].

UK STUDIES

Leger (1997) evaluates performance persistence by simply *counting* the number of funds with positive or negative Jensen's alphas in non-overlapping 5 year sub-periods (1984-93). Of the 72 funds in the study less than 4 funds record a positive abnormal performance in any of either two, three or four consecutive periods. Using a contingency table approach Lunde et al (1999) sort UK equity mutual funds into quartiles based on 4F-alphas with rebalancing every 3 years, over the 1972-95 period (using 1,400 surviving and 900 non-surviving funds) and finds evidence of WW and LL persistence. Similar results for 131 investment trusts over the period 1989 – 1995 are found by Allen and Tan (1999) with one-year rebalancing and they reject the null

---

[60] Also, Gorman (2003) over the 1986-2000 period, ranks 35 US *small-cap* funds on one, two, or three-year past conditional alphas. Using a regression of future alphas on past alphas (only), he finds evidence of positive persistence over the next year and reversals in years two and three.

[61] Morningstar 'styles' are not available pre-1993 so some funds have to be matched with CRSP returns data in a more eclectic manner.

[62] Results are based on a regression of the 3 year forward looking 4F-alpha (over 2002-2005) on (0, 1) dummy variables for the Morningstar ratings (1 to 4) given in 2002. The intercept is statistically significant with a value of about -0.15, implying that the average 5-start rated fund has a forward looking alpha of about -1.8% p.a. Other performance metrics, such as the Sharpe ratio, Jensen's alpha and a conditional beta model (measured both with and without returns adjusted for loads) are also used and give qualitatively similar results.

of no persistence for raw returns and Jensen's alpha measures.[63] Finally, Fletcher and Forbes (2002) use 724 funds over 1982-96 and find statistically significant persistence over one year but only for the 1989-90, 1993-94 and 1995-96 periods[64]. When ranking on (excess) fund returns, persistence in the above 3 periods is due equally to repeat winners and repeat losers but when returns in excess of the *market return* are used, persistence is clearly driven by repeat losers.

Overall, studies of predictability on US data using statistical measures, find evidence that poor performance persists for up to 3 years while there is mixed evidence that winners repeat over periods in excess of one year. Results on UK data are somewhat sparse but indicate there may be some short-run persistence while evidence for long-run persistence is rather weak. But as we have noted such statistical measures of persistence do not necessarily imply an exploitable trading strategy – an issue we take up below.

6.2 RECURSIVE PORTFOLIO METHOD

LONG HORIZON

The Elton, Gruber and Blake (1996b) study noted above was one of the first to examine the economic significance of persistence in a survivorship bias free sample of US funds (188 funds, 1977-1993) using a 3F model plus a bond return. For the top decile funds (ranked on past 1-year or 3-year alphas or past t-alphas) and either 1-year or 3-year rebalancing, they find evidence of a small positive forward looking alpha of around 0.5% p.a. and for the bottom decile a forward looking alpha of between -2.4% and -5.4% p.a. (depending on the ranking and rebalancing periods used) – non of the top and bottom decile forward looking alphas are statistically significant taken individually, but their difference is statistically significant at the 5% significance level or above. Similar result are reported by Brown and Goetzmann (1995).

There was something of a watershed when Carhart (1997) using all US equity funds (1963 – 1993) suggested that persistence found in earlier studies using the CAPM or 3F-alpha, may be a manifestation of the momentum effect in stocks which are 'accidentally held' by funds, rather than funds actively choosing stocks with a high loading on the momentum factor. He applies the recursive portfolio formation methodology sorting into winner and loser deciles based on lagged one-year (net) returns, rebalanced annually. A concatenated returns series $R_t^f(t,T)$ is then used to estimate $\alpha_{4F}^{f,net}$ which is negative for all decile portfolios (significantly so for the

[63] Similar results are obtained using a regression of last periods ranked (abnormal and raw) returns on next periods returns or using the Spearman rank correlation coefficient.

[64] Winners and losers are classified using either annual excess returns (over the risk free rate) or returns in excess of the market return, with the median as the 'break point'. Tests use the log-odds ratio.

bottom 3 deciles) – so poor performance persists but good performance does not[65]. Carhart also sorts funds by their loading on the momentum factor but funds with high past $\beta_{MOM,i}$ are found to underperform in the post-ranking period, using $\alpha_{4F}^{f,net}$. Thus funds which have high momentum factor *loadings* are not actively pursuing a momentum strategy but instead "accidentally" happen to hold last year's winning stocks and by so doing, enjoy a high average (raw) return over the next year but a negative risk adjusted return.

Kosowski et al (2005) examine persistence over 1978-2002 using recursive (equally weighted) portfolios, sorting on the past 12 or 36 month (unconditional) 4F-alpha, *with one year* rebalancing periods. The estimated net return alpha's for various fractiles are then computed over the period 1978-2002 and compared with the bootstrap distributions for each fractile (which impose α_i = 0 across all funds). Using the whole universe of funds (ranking on past 36 month alpha) they find the top decile exhibits persistence over one year with an alpha of 1% p.a. (p=0.05), with growth funds largely responsible for persistence in 'winner' funds. At the bottom of the performance distribution (for all funds), deciles 6-10 have significantly negative abnormal performance (of about -1% p.a. for deciles 6-9 and -3.5% for decile-10).[66]

Many studies of persistence rank funds using past returns. Teo and Woo (2001) show that past winners contain different styles from year to year, so that ranking on past returns may not pick up managerial skill within a style but merely differential style performance over time. For example, past 'winners' may be mainly growth funds in one year, while in the next year value funds predominate. But this does not fit with our prior notion that past winners should contain a roughly uniform mix of fund styles – namely the best managers within each style. They therefore suggest that an ex-ante portfolio formation rule based on *style adjusted returns* may be more appropriate when trying to pick future winners.

Sorting funds into deciles (1984-99) on past style adjusted returns $SAR's$, forming recursive portfolios and concatenating future *style adjusted (gross)* returns $SAR_i^f(t,T)$, Teo and Woo use the *spread* between the past winner minus past loser decile returns

---

[65] The top decile has a 4F-alpha of -1.4% p.a. (t=1.6) and for the bottom three deciles alpha is -1.6% (t=2.5), -2.4 (t=3.1) and -4.8 (t=4.3), respectively. Similar qualitative results ensue when ranking into deciles based on past 4F-alphas (estimated recursively, using the previous three years of data and rebalancing annually).

[66] Koswoski et al when ranking on *past 1-year,* 4F-alphas and using a 1-year rebalancing period find the top two winner deciles have significant persistence with 4F-alphas of 1.4% p.a (p<0.01) for the top decile and 0.84% p.a. (p<0.01) for the second best decile fund, and again deciles 6-10 have statistically negative average performance of around -1.8% p.a. Note that Carhart (1997) finds no persistence for the top decile funds and the difference between the two studies may be attributed to the different sample period and more importantly to the non-normality in the specific risk of the top decile funds, which is taken into account in the bootstrap but not necessarily with Carhart's parametric p-values.

$SPRD^{f}_{W-L} = SAR^{f}_{W}(t,T) - SAR^{f}_{L}(t,T)$ as the dependent variable in the 4-factor model. They find statistically significant 4F-alphas of around 0.3%*** p.m. (3.6% p.a.), for portfolios formed on either past 1-3, 2-4, 3-5, or 4-6 years $SAR's$. But there are no results reported for the separate post-formation alphas of past winners and losers, only for the difference between the winner and loser deciles. But if you cannot short-sell MF's (or successfully use ETFs or replicate the portfolio composition of the loser funds and short sell their constituent shares) then it is difficult to see how in practice one might exploit the estimated "spread alphas". However, the *cumulative* SAR's over 70 months (i.e. about 6 years) for past winner and loser portfolios (sorted on past 1-3 year SAR's) are 2% and minus 16% respectively, so the economic impact of 'persistence' seems to lie with the past losers rather than the past winners[67].

Overall it appears that the evidence for economically significant *long-run* persistence for US past loser funds is well established, whereas evidence for persistence of past winners is less well established. We now turn to evidence of short-run persistence based on monthly or quarterly rebalancing.

SHORT HORIZON

Mamaysky, Spiegel and Zhang (2004) concentrate on short horizon persistence for US funds and question both the use of unconditional constant factor models and sorting on badly mis-measured variables such as past alphas. They argue that it is unlikely that any fixed parameter model of performance can adequately capture the myriad of portfolio risks arising from the diverse range of trading strategies pursued by the universe of mutual funds. Therefore, sorting on estimated alphas might result in the top and bottom deciles containing both genuine winners and losers but also a substantial number of funds with the highest estimation errors – particularly if less than 60 monthly observations are used. To overcome this problem they suggest a back-testing technique in which the statistical model must exhibit some past predictive success before a fund may be included in a given portfolio.

They apply a variety of different models (e.g. CAPM and 4F models, with recursive OLS or Kalman filter time varying parameters[68]) to each US mutual fund and suggest two sequential filter rules to identify potential winner funds. For example, a successful strategy for including a

[67] Teo and Woo (2001) also find that persistence (based on 4F alphas) for top minus bottom deciles also holds when ranking portfolios by *raw* returns for formation periods of 1-3 years, 2-4 and up to 3-5 years – note that this squares with Carhart's (1997) results since the persistence found in the top minus bottom decile excess returns in Teo and Woo appears to be predominantly due to the bottom decile.

[68] Recursive estimation using the Kalman filter can be interpreted as an unobservable variables approach to the estimation of conditional models, which seeks to mimic dynamic trading strategies. Kosowski (2006) documents different alphas in recessions and expansions and the Mamaysky et al (2004) technique may be exploiting this source of predictability.

fund at time t is first to estimate the factor model with 60 observations prior to time t-2 and then (i) only accept a fund if the forecast made at t-2 of the fund's alpha for t-1 has the same *sign as its excess return over the market* $R_{i,t-1} - R_{m,t-1}$ and (*ii)* re-estimate the factor model using 60 months of data up to t-1 and only accept a fund if its estimated alpha and market beta are in the ranges $\pm 2\% \, p.m.$ and 0 to 2, respectively. This group of 'accepted' funds are then sorted into deciles based on their estimated alphas. When the rebalancing period is every 12 months there is little evidence of persistence in abnormal (alpha) performance but for *monthly* rebalancing the top decile portfolio has a statistically significant abnormal (net return) alpha of 2.5% (t=4.0) to 4.5%p.a. (t=3.2) depending on the model used[69]. There is also significant evidence of negative abnormal performance for the bottom five deciles.

In a recent study Kacperczyk, Sialm and Zheng (2006) investigate whether a fund's actions *between* portfolio disclosure dates provides incremental information which could be used by investors to 'pick winners'. Broadly speaking, they measure the 'unobserved actions' of funds by the 'return gap' – the difference between the observed quarterly net return and the quarterly buy and hold return (using previously reported portfolio weights). Thus the 'return gap' measures any benefits of interim trades less any hidden costs (e.g. transactions costs), over each quarter. Sorting funds into deciles on the past return gap either over the previous 4 or 15 months gives a statistically significant $\alpha_{4F}^{f,net}$ with monthly rebalancing, for the loser but not the past winner decile. However, adding a 'predictive filter' results in statistically and economically significant persistence. The 'predictive filter' used is similar in spirit to that of Mamaysky, Spiegel and Zhang (2004) and is to only include funds where the *sign* of the average excess return (over the market return) equals the *sign* of the return gap – this gives the past winner-decile portfolio an $\alpha_{4F}^{f,net}$ = 2.52**% p.a. and the past loser portfolio an $\alpha_{4F}^{f,net}$ = -4.0*** % p.a..

Further evidence of an incremental effect of the return gap on short-term persistence is confirmed in panel regressions of $\alpha_{4F}^{f,net}(i,t)$ on the lagged return gap (and other fund characteristics, including the lagged excess fund return). The source of the persistence is due in part to higher IPO allocations to high return gap funds. So part of the 'unobserved actions' of funds is due to their acquisition of favourable IPOs between portfolio reporting dates (Reuter 2005).

---

[69] Similar result apply for the top-5, top-10 and top-20 funds.

Instead of using a predictive filter to identify 'winner funds' Bollen and Busse (2004) seek to refine the choice of 'winners' by allowing factor loadings to change each quarter. They investigate performance persistence using *daily* data (from January 1984 to December 1995, using 230 mutual funds). They form decile recursive portfolios *each quarter* ranking on 4F-alpha (augmented by the HM or TM market timing variables). They find the top decile portfolio has a statistically significant abnormal net return alpha (bootstrap standard errors) in the post-ranking quarter of 25-39*** bp (an average of about 1.3% p.a.)[70] while deciles 6-10 have statistically significant negative abnormal performance between minus 20 and 80*** bp per quarter (-0.8% to -3.2% p.a.). Because the post-ranking recursive regressions are estimated quarterly they effectively mimic (non-parametrically) a conditional or time varying parameter model (e.g. Mamaysky, Spiegel and Zhang 2004), since the alpha and factor loadings can change each quarter.

EXTRANEOUS AND PRIOR INFORMATION

In section 5.5 we noted that the use of extraneous information may improve inference about ex-post performance of funds in the extreme tails. Can this methodology improve sorting rules to identify ex-ante future winners and losers? Pastor and Stambaugh (2002) and Busse and Irvine (2006) show that additional precision is achieved in estimating *individual* fund alphas when incorporating extraneous data on returns on seemingly unrelated passive assets[71]. The procedure is based on the idea in Stambaugh (1997) that truncating a set of returns so all return series are of equal length is inefficient – but this is exactly what occurs in standard estimation of fund alphas. The intuition behind this approach is that returns on passive non-benchmark assets may be correlated with fund holdings, so incorporating a *long* data series for non-benchmark returns increases precision for alpha. First, Pastor and Stambaugh (2002) show that incorporating a long time series of passive asset returns can substantially alter the estimate of alpha and its precision (compared with the standard approach) and also that differences in alpha estimates across models (e.g. CAPM, 3F-model) are attenuated when using passive asset returns. Busse and Irvine (2006) apply the methodology to tests of persistence, using daily data (1985-95) on 230 equity fund returns and data on non-benchmark assets (from 1968 onwards). Using a variety of alternative factor models they rank funds into deciles based on both standard alpha, a frequentist alpha (which incorporates non-benchmark returns) and a Bayesian alpha. With quarterly rebalancing they demonstrate increased predictability (based on Spearman rank

[70] These result are not inconsistent with Carhart (1997) who finds no positive persistence. This is because Carhart ranks funds on raw returns, uses longer pre- and post-sort horizons and concatenates the post-ranking returns when estimating alphas. Bollen and Buse simulate the latter effects with their daily data and consistent with Carhart, also find no positive persistence over horizons longer than one quarter.

[71] Passive assets are those that do not appear in the factor model for fund returns. So, for example the SMB factor would constitute passive assets for the CAPM factor model, while industry returns would constitute passive assets for the 4F-model.

correlations of pre- and post-alphas), for both the 'non-benchmark' frequentist-alphas and the Bayesian alpha rankings, compared with the standard frequentist alpha rankings. So extraneous information may help identify portfolios which yield performance persistence.

The idea that not all information about a fund's future performance is encapsulated in a fund's past alpha is also taken up by Cohen, Coval and Pastor (2005) but their 'additional information' uses the *holdings of other successful funds*, to help predict a particular fund's performance. They draw on the 'home-bias' mutual fund literature which notes that physical proximity may facilitate relevant information transmission, which results in a concentration of fund assets in geographically 'nearby companies' or fund managers in the same city having similar portfolios (Coval and Moskowitz 1999, Hong, Kubik and Stein 2005).

Specifically, Cohen, Coval and Pastor (2005) provide an alternative ranking metric for fund performance based on 'commonality', that is how far a particular fund's stock holdings currently mimic the stock holdings of funds which have performed well (based on all their past alphas)[72]. Essentially fund-i's 'skill' (= $sk_{i,t}$) is a weighted average of other funds' alphas, with the weights depending on the covariances between fund-i's portfolio weights and the current weights of the other managers. If fund-i holds only stocks that are held by no other manager then $sk_{i,t}$ collapses to α_i, otherwise $sk_{i,t}$ is high if it has portfolio weights which are similar to portfolio weights of other funds with high alphas. The idea is that additional precision about fund-i's performance is improved if information across all funds is pooled. They use data from April 1982-September 2002 (with a maximum of 1,502 funds) and rebalance quarterly. Sorting separately on either past α_i or $sk_{i,t}$ they find the respective forward looking top decile alphas are $\alpha_{4F}^{f,g}$ = 2.48% p.a. (t=2.6) and $\alpha_{4F}^{f,g}$ = 4.3% p.a. (t=3.5)[73] – so sorting on past $sk_{i,t}$ gives a higher abnormal return for the winner decile. Using a double sort into 25 quintiles, they also find that sorting on $sk_{i,t}$ has an *incremental* effect on performance over and above that from first sorting

[72] The analogy they use is that after observing a group of basketball players you note that the average score is 8/10 for the two-handers, but only 4/10 for the one-handers. Then if two players, one one-hander and one two-hander are observed, each currently with a 4/5 score, then you would bet that the two-hander is more likely to have a higher score out of 10 - the track records of the other two-handers are better than the one-handers, so you assume the current two-hander has a better technique and the one-hander is more likely to have been lucky with his first 5 shots.

[73] These results are based on ranked alphas estimated from the previous 12 months of data – results are similar when past alphas are estimated using all past data on the fund.

on α_i [74] - the "top-alpha, top-sk" sorted portfolio has $\alpha_{4F}^{f,g}(5,5) =$ 4.61% p.a. and the bottom sorted portfolio has $\alpha_{4F}^{f,g}(1,1) =$ - 3.79% p.a. with the long-short portfolio yielding 8.4% p.a. (t= 4.0)[75]. Once again there is evidence for the incremental predictive power of extraneous information but note that reported alphas are gross of TERs and therefore represent abnormal returns to the fund rather than to investors – as rebalancing is quarterly, transactions costs (bid-ask spreads, load fees) may also be high.

Using 855 funds (1972-1995) for five sectors[76] Blake and Timmermann (1998) form (equally weighted) recursive quartile portfolios based on alphas estimated over the previous 24 months and hold these portfolios for only *one month*. They find statistically significant positive and negative persistence in alpha-performance[77] particularly for smaller company funds but results vary depending on the risk adjustment model used and to exploit this persistence may require significant transactions costs due to monthly rebalancing.

In summary the above studies, particularly for the US, provide prima facie evidence for short-run positive persistence among past winners. But with *quarterly* rebalancing it is unlikely that the positive abnormal net return of 1.3% p.a. found for US funds by Bollen and Busse (2004) is sufficient to cover both load fees (which average about 1.2% p.a.) and quarterly rebalancing costs (brokerage fees and bid-ask spreads and price impact). However, Cohen, Coval and Pastor's (2005) 'best' winner portfolio has a substantial *gross* abnormal return to the fund of $\alpha_{4F}^{f,g}(5,5) =$ 4.61% p.a. - also with quarterly rebalancing. For *monthly* rebalancing Mamaysky, Spiegel and Zhang (2004) find the abnormal net return is between 2.5% to 4.5% p.a. - but again this could be outweighed by transactions costs of frequent rebalancing. So, as far as the use of filter rules, extraneous information and Bayesian approaches are concerned they do appear to provide incremental predictive power for future fund performance – so, for picking ex-ante winners these approaches are promising but as yet are far from definitive.

---

[74] A double sort of funds into 5x5 quintiles is undertaken based first on past α_i quintiles and then on past $sk_{i,t}$ quintiles, with rebalancing quarterly (equally weighted) – there is also a quarterly delay in calculating $sk_{i,t}$ to allow for observation of portfolio weights, which can be reported with a delay.

[75] No t-statistics are given for the (5,5) and (1,1) portfolios so we cannot infer if the strategy gives statistically significant abnormal returns solely to the 'winner' portfolio, which would remove the problem of short-selling mutual funds. Also, the figures reported above are the largest found for the various sorts, across a variety of models.

[76] The sectors are equity growth, equity income, general equity, smaller companies and a balanced sector.

[77] The unconditional factor model has the excess returns on the stock market index, r_{mt} the excess return on small cap stock $R_{s,t}$ over the market index $R_{m,t}$ and the excess returns on a five-year UK government bond $(R_{b,t} - r_t)$ as the independent variables: $r_{i,t} = \alpha_i + \beta_{mi} r_{m,t} + \beta_{si}(R_{s,t} - r_t) + \beta_{bi}(R_{b,t} - r_t) + \varepsilon_{it}$.

6.3 STOCK HOLDINGS AND TRADE DATA

We now turn to the question of whether the use of stock holdings and trade data provide more definitive results for persistence in performance amongst US equity funds. As we see below, what makes results based on stock holdings and trade data rather difficult to summarize are the number of variants on the output metrics considered (e.g. the CS measure based on stock holdings, stock purchases or stock sales or "buy-sells", each over different horizons) - coupled with the usual variants on possible sorting rules. For example, we might find persistence in CS(buys/sells) but not for the CS(holdings) measure. Therefore given the above and the relative novelty of studies using stock holdings and trade based returns, we present some illustrative results in tables – the reader can then better judge whether our account is accurate and balanced.

Wermers (1997) using stock holdings data 1975-94 finds that past (decile) winners earned positive gross returns over the next year (in excess of the average fund) of 3%. But using a momentum variable based on an *active* rather than a passive momentum strategy, he inferred that the presumption of an active momentum strategy *(i.e. increasing* holdings in high past return stocks) was just as tenable as a passive momentum strategy. Chen, Jegadeesh and Wermers (2000) followed up this important question of whether 'winning' funds actively follow momentum strategies or whether as described by Carhart (1997), winning funds *accidentally* hold the previous periods winning stocks and hence benefit from the well documented momentum effect in stocks. They use *stock* holdings and buy and sell trades for all (equity) mutual funds between 1975-1995, using quarterly rebalancing and where 'winners' and 'losers' are the top and bottom quintile of funds ranked on past one year *net* returns. Performance measures in the post-ranking period are *gross* returns (table 1, Panel A) and the risk adjusted CS measure (table 1, Panel B).

[Table 1 here]

Average *gross* returns on 'all holdings' of stocks of winning funds in the two quarters prior to the ranking period (Qtr-1, Qtr-2) are significantly higher than the returns of losing funds by around 5% p.q., so that past winners *hold* proportionately more momentum stocks than past losers (table 1, Panel A). In addition, the past returns of the 'buys' of winning funds are significantly higher than the past returns of the 'buys' of losing funds by around 2% p.q. - this seems to point to active momentum investing by winning funds relative to losing funds. However, the 'all holdings' and 'buys' of winning funds outperform the gross returns of losing funds only in the portfolio ranking (Qtr-0) and subsequent quarter (Qtr+1) - so momentum in gross returns is very short lived (table1, Panel A) - which is consistent with Wermers (1997) earlier results. A very

similar picture emerges when the post-ranking performance is based on risk adjusted (CS) returns (Table 2). The winner funds' CS(holdings) data indicates persistence in the formation (0.83%*** p.q.) and subsequent quarter (0.37%** p.q.) but CS(buys) is only significant in the formation quarter (1.21%**) – so persistence when measured by CS is also short-lived (table 1, Panel B).

[Table 2 here]

An interesting question addressed by Wermers (2003b) is whether persistence in gross returns is due to persistence in CS, CT or AS and whether persistence carries over to net return alphas. Each year funds are ranked into past winner and loser (deciles) based on their net returns and rebalancing is annual. For all equity funds, the top minus bottom decile (W-L) average gross return one-year ahead is $\overline{R}^{g}_{stk}$ (all holdings) = 5.3%** p.a., which is mostly due to the statistically significant AS measure of 4.1%* p.a. – this indicates that past winner funds accidentally held more stocks (one year earlier) which *currently* outperform those held by past losers. The contribution of stock picking skill to the overall gross return of 5.3% is relatively small since CS_{W-L} (all holdings) = 0.83%** p.a.[78] Although the CS measure indicates stock selection skill over the post-ranking year by both past winner (CS=1.75%* p.a.) and loser funds (CS=0.92%* p.a.), this does not carry over to α^{net}_{3F} and α^{net}_{4F} measures, which are negative (but usually not statistically significant). So although one year persistence in risk adjusted gross returns (CS) for stocks is evident, indicating skill, this does not carry over to abnormal net returns for investors.

FUND STYLES: GROWTH FUNDS

There is evidence that herding can affect stock prices (Lakonishok, Shleifer and Vishny 1992) and that fund purchases themselves may impact on prices particularly at reporting dates (Carhart 2002a) or due to the allocation of 'hot' IPOs to particular funds within a fund family (Gaspar, Massa and Matos, 2006). Wermers (1999) follows up these ideas for different fund styles. Using stock holdings data and quintile sorts based on several measures of herding he finds that herding influences abnormal fund returns over the following 6 months[79]. This effect is more pronounced among small stocks and growth orientated funds and the effect is strongest in

[78] The characteristic timing CT is positive for past winners and losers and for W-L, but CT is not statistically significant, indicating that past winning funds do not increase their holdings of benchmark stocks with high future returns – they cannot *market time* the stock characteristics.

[79] Here, abnormal returns are returns on stocks in excess of their (equal weighted) average size quintile return. Grinblatt, Titman and Wermers (1997) also find evidence of herding by funds and Carhart et al (2002a), Musto (1997, 1999) find evidence of herding around reporting dates.

the first 10 year period 1975-1984 rather than the last 10 years 1985-94, where only herding in small stocks effects subsequent returns. So herding by mutual funds particularly in small stocks may be a source of the momentum effect in stock returns – but given the large bid-ask spreads it is an open question whether such effects are exploitable, even at the fund level.

In a later paper Wermers (2003) follows up the idea that return persistence may be stronger in particular fund styles (rather than in averages across all fund styles). This is done using quarterly rebalancing (into deciles based on past 1 year net returns) and for holding periods from one to four years (Table 2). Only for *growth funds* does he find strong statistical evidence for persistence in winner minus loser (W-L) decile (raw) returns of around 4.3%*** p.a. in the first post formation year, falling to 2.5%*** p.a. over 4 years (table 2, Panel A). However, using the CS(holdings) gross returns index, persistence is found over 4 years for the past winner decile only – of around 2% p.a. (for a cohort of 26 growth funds). Wermers finds that this arises mainly from 'same stock' purchases rather than purchases of new stocks. Thus any short run herding by past 'growth fund' winners is into stocks already held, this tends to quickly push up their prices which remain high after 4 years reinforcing the underlying momentum effect.

Overall, our conclusions from studies based on stock holdings and trades is that past US winner funds do ('accidentally') *hold* more momentum stocks than past losers and persistence is relatively short lived - even before trading costs and expenses. There is stronger evidence that long run persistence applies to growth funds due to herding into existing momentum stocks.

The above studies have examined persistence based on ranking funds on some measure of past performance but some studies have also investigated other 'sorting rules' based on variables such as turnover, industrial concentration of fund holdings, deviations of stock holdings from their benchmark index and the utility derived from the chosen portfolio. We briefly discuss these below.

TURNOVER

It is possible that funds with well informed traders have high turnover (='winners'), which then results in higher future abnormal returns. Conversely, there may be many uninformed traders who 'churn' stocks in order to look skilled but in fact are not. Wermers (2000) tests these competing views over the 1975 – 1993 period using gross returns on stock holdings and finds the W-L turnover ranked decile has $\overline{R}^{g}_{stk}$ =4.3%*p.a. (one-year ahead).[80] So high turnover funds tend

---

[80] Funds are ranked into deciles (also repeated for quintiles) by their levels of turnover during the previous year and portfolios are rebalanced annually. Most of the W-L gross return of 4.5%p.a. is due to AS_{W-L} = 2.2%***p.a., rather than

to earn higher gross returns on their stock holdings (than do low turnover funds), which more than covers their additional transactions costs (2.37%***pa) and expense ratios (0.28%*** pa). However, evidence for high turnover funds having a positive CS measure is rather weak but the evidence that *all* (decile turnover) funds have an $\alpha_{4F}^{f,net}$ of about minus 1% p.a. is much stronger - so ranking by prior turnover does not provide a trading rule which gives positive abnormal net returns to the investor.

INDUSTRIAL CONCENTRATION

It has been argued that funds that have relatively high levels of industrial concentration ICI (relative to the market index) may have higher future performance, because these 'winner' funds obtain an informational advantage by carefully analyzing just a few industries.[81] Using a recursive portfolio approach (quarterly rebalancing, equally weighted) and decile sorts based on ICI , Kacperczyk et al (2005) find that for the decile 'winner' portfolio, $\alpha_{4F}^{f,g}$ = 0.53* % p.q. (2.1% p.a.) but $\alpha_{4F}^{f,net}$ and the CS (holdings) performance measures are both not statistically significant. A similar approach has been adopted by Cremers and Petajisto (2006) who find stronger support for persistence after sorting funds into (5x5) quintiles based on an index of active management, (which they call Active Share, AS[82]) and the previous years' average benchmark-adjusted return ($\overline{R}_i - \overline{R}_b$). Rebalancing annually they find the 'past winner' quintile portfolio (i.e. high AS-high past return) has $(\overline{R}_i - \overline{R}_b)^{net}$ = 3.69% p.a. (t=2.31) and $\alpha_{4F}^{f,net}$ = 2.29% p.a. (t=1.87), with 'winner minus loser' portfolio having $\alpha_{4F}^{f,net}$ = 3.09% p.a. (t=2.54).[83]

---

due to stock picking or market timing skills since $CS_{W\text{-}L}$ = 1.22% p.a. and $CT_{W\text{-}L}$ =0.23% p.a. which are not statistically significant (Wermers 2000, table VI).

[81] Each share held by the mutual fund is allocated to one of 10 industrial classifications and the industrial concentration index; $ICI_{i,t} = \sum_{k=1}^{10}(w_{k,t} - \overline{w}_{k,t})^2$ where $\overline{w}_{k,t}$ = market (value) weight of industry-k in the market index and $w_{k,t}$ is the weight of fund-i's holdings of stocks in industry sector k.

[82] The Active Share index is similar to that for ICI, namely $AS_{i,t} = \sum_{k=1}^{N}\left|w_{k,t} - \overline{w}_{k,t}\right|$ where $\overline{w}_{k,t}$ = (value) weight of stock-k in the fund's benchmark index and $w_{k,t}$ is the weight of fund-i's holdings of stocks. Note that for AS the absolute value of the difference in weights is used and the index is the 'style index' of the fund, whereas for ICI the square of the weights is used and only a single market index is used - see Cremers and Petajisto (2006) for a discussion of the relative merits of these two measures.

[83] It is really this two-way sort on past returns and past Active Share that gives relatively large alpha estimates for the past 'winner' portfolio. If you merely sort on past Active Share, then the highest quintile has $\alpha_{4F}^{f,net}$ = 1.36 (t=1.43). The persistence of 'past winners' is even stronger if there is a three-way sort by size, AS and past returns with the smallest size quintile (but with highest AM and past return) giving $\alpha_{4F}^{f,net}$ = 5.63% pa (t=3.66). In a broadly similar vein Fang and Kosowski (2006) report that *stocks* sorted on the basis of mimicking *top* analysts stock recommendations give portfolios with higher post-sort 4F-alphas than sorts based on mimicking *average* analyst's recommendations – demonstrating the superior information content of "semi-private" information over public information.

Overall, there is some statistical evidence that sorting funds on the basis of their Active Share, together with past return or fund size, gives positive future abnormal return performance to investors, but these effects are not particularly large and the question remains as to whether such abnormal performance is exploitable at a (risk adjusted) net returns level after transactions costs of rebalancing.

UTILITY

While all of the above studies of persistence involve portfolios of funds, the portfolios of past winners and losers are formed on fairly ad-hoc grounds. In contrast, Avramov and Wermers (2005) chose portfolio weights at each rebalancing date in order to maximize next periods utility. Does this result in a better ex-post abnormal performance than say, decile sorts on the basis of past raw returns ('hot hands') or past 4F-alphas, which tend to give at best zero or small positive forward looking values for $\alpha_{4F}^{f,net}$. Avramov and Wermers (2005) allow investors to choose optimal (positive) weights to hold in 1,301 (no-load domestic equity) mutual funds[84] in order to maximize a quadratic utility function in next period's (dollar) excess return and variance, with rebalancing *each month* (over the period December 1979 to November 2002)[85]. This results in a portfolio which yields an $\alpha_{4F}^{f,net}$ that is large, positive and statistically significant, ranging between 9-12% p.a. – *providing* investors utilize the predictability in the factors which are driven by macroeconomic variables such as the dividend yield, default and term spreads and the interest rate.[86] In the model investors are Bayesians and take account of estimation risk in choosing the optimal weights (Baks et al 2005, Barberis 2000). So far this result is something of a 'black box', but Avramov and Wermers (2005) show that the optimal weights for investors who take account of predictability, results in these investors successfully timing industrial sectors such as energy, utilities and metals over the business cycle (Moskowitz 2003) and being overweight in those funds which hold these outperforming sectors. Clearly, the ex-post abnormal return to this 'recursive optimal portfolio strategy' is substantial at between 9 and 12% p.a. but it appears that investors would have to hold up to 1,300 funds and rebalance every month – an issue that needs further investigation particularly with respect to transactions costs.

---

[84] These include actively managed funds, index funds, sector funds and exchange traded funds (ETFs)

[85] In this mean-variance framework there are no hedging demands – the importance of the latter seems to be model and data specific – see for example, Ait-Sahalia and Brandt (2001), Campbell et al (1999, 2003), Viceira (2001) and Cuthbertson and Nitzsche (2004).

[86] The model of returns is the 4F conditional alpha-beta model, where the 4-factors then each depend on the lagged macroeconomic predictor variables z_{t-1} and the latter are themselves forecast using a VAR model – a set up similar to Barberis (2000).

UK STUDIES

There are far fewer studies of persistence using UK data. The study by Quigley and Sinquefield (2000) takes 752 funds (1978-97) and forms (equally weighted) decile portfolios each year, ranked by raw returns over the previous twelve months. The gross (raw) return for the W-L deciles is 3.54% p.a.. While this initially seems to point to an easy 'beat the market' strategy, in fact pursuing this strategy involves an annual turnover of 80% in the composition of the top portfolio and with a bid/offer spread of 5%, abnormal returns would be eliminated. In addition, the 3F-alphas from the top two (post-formation) portfolios while positive (averaging 0.6% p.a.) are not statistically significant and yet the bottom 4 portfolios have statistically significant negative 3F-alphas (averaging minus 1.8% p.a.).[87] Similarly, Fletcher (1997) using a CAPM model, ranking by past alpha and forming (equally weighted) quintile portfolios on five-year, two-year and one-year ranking and evaluation periods, finds very little evidence of positive persistence. Fletcher and Forbes (2002) using a variety of factor models (over January 1983 to December 1996) tend to find that past winners suffer reversals, with their alphas in the post-formation period being negative and statistically significant – while past losers have negative future alphas.

SUMMARY

The evidence on persistence is voluminous, yet attempting to give a brief yet balanced summary of performance persistence across US and UK studies is difficult. Different studies examine different sample periods, some use purely 'statistical measures' (e.g. correlation, rank correlation, contingency tables) and others use economic measures of performance such as raw returns (gross or net) or adopt different measures of risk adjusted abnormal return (e.g. different factor models, or different benchmark returns). Furthermore, different studies sort funds on different fractiles and use different rebalancing and holding periods. Finally, not all studies control for survivorship and look ahead bias or cover the whole universe of funds and with so many sorting rules used, data snooping bias is an issue.

Even after noting the above caveats, it seems that there is some persistence amongst the top decile of all US funds sorted on several characteristics including past raw returns, 4-factor alphas or CS measures. Using a risk adjusted *gross returns* (CS) measure persistence may last up to four years for a small number of growth funds and for up to one year when the top decile is formed from 'all funds'. Persistence amongst past winners does not seem to last longer than 1-year when we use an abnormal *net returns* metric such as the 4-factor alpha. For example, in Kosowski et al (2005), sorting on past 3-year, 4F-alpha and rebalancing annually gives a top

[87] Qualitatively similar results are found when they rank funds each year into deciles based on their past 3F-*alphas* (using the previous three years of data), while ranking funds into deciles based on their past β_{SML} or β_{HML} factor weightings does not produce any statistically significant forward looking 3F-alphas.

decile $\alpha_{4F}^{f,net}$=1% p.a. Unless investors can mimic, with a small number of funds, the performance of the top decile portfolio (which may currently contain over 180 funds) and avoid load fees to minimize rebalancing costs, it is doubtful that a significant exploitable 'persistence anomaly' exists. In contrast, there is strong evidence that poor performance persists fairly uniformly across deciles 5-9 with $\alpha_{4F}^{f,net}$ around -1%p.a. and the bottom decile has $\alpha_{4F}^{f,net}$ = -3.6% p.a.. Broadly similar results apply to the relatively few comprehensive UK studies on persistence.

Recent work using extraneous information (Pastor and Stambaugh 2002, Cohen, Coval and Pastor 2005) or incorporating the predictability of factors in an optimal portfolio (Avramov and Wermers 2005) have demonstrated a prima facie case for successful ex-ante strategies portfolios – but such strategies are only be feasible for rather sophisticated investors and any predictability would need to be shown to be robust over time and outweigh any transactions costs of frequent rebalancing. Overall our analysis of persistence provides useful insights for investors about which funds to avoid - but offers much less certainty about which funds to purchase.

7. FUND FLOWS AND PERFORMANCE

In section 7.1 we examine characteristics which might influence a fund's (or portfolio of fund's) relative risk adjusted performance, concentrating particularly on relative costs. In section 7.2, we ask the question "Is Money Smart?" If there is persistence in performance of past winner or loser funds and money is smart, then we expect to observe investors switching from loser to winner funds, with the former either ceasing to exist (or changing to a successful strategy) and the latter giving rise to positive abnormal returns – at least over the short-run. If performance persists at the fund manager level, rather than at the fund level, then the 'smart money' should follow successful managers – this is examined in section 7.3.

7.1 PERFORMANCE AND FUND CHARACTERISTICS

In this section we analyze the possible sources of ex-post and ex-ante abnormal performance. Do large funds perform better than small funds, or do high turnover or high cost funds provide a better net return than low cost or low turnover funds? In short, what type of fund characteristics influence performance.

Since each fund's abnormal return and the relationship between the abnormal return and fund characteristics may vary over time a Fama-MacBeth rolling regression is often adopted. A cross-section regression of (quarterly) CS_i on fund characteristics is straightforward on a rolling

basis. For abnormal performance measured by $\alpha_{i,t}$ the following is calculated (say) every month, $\alpha_{i,t} = r_{i,t} - \beta'_{i,t-m} F_t$, where the $\beta'_{i,t-m}$ are estimated using a moving window of t-1 to t-m observations (often m=36 months). The following cross-section regression is then run each month:

[12] $$\alpha_{i,t} = \theta_t + \delta'_t X_{i,t-k}$$

where $\delta'_t = \{\delta_{1t}, \delta_{2t,} \ldots \ , \delta_{mt}\}$ and X_{t-k} is the m-vector of fund characteristics at time t-k[88]. The advantage of this approach is that the factor loadings, the abnormal performance and the impact of the firm characteristics X_{t-k} on α_i are allowed to vary over time. Note that one of the regressors in X_{t-k} could be the fund's previous performance, in which case we are measuring persistence, after accounting for other fund characteristics - see section 6. Instead of using the Fama-MacBeth rolling regression, equation [12] can be estimated using a suitable estimator for an unbalanced panel (usually with time fixed effects) but this method is less popular, in part because it allows fewer parameter estimates to be time varying (Petersen 2005 provides a comparison of the two methods).

Carhart (1997) tries to get a handle on what determines abnormal performance using a month rolling Fama-McBeth regression of 4F-net alphas of *individual funds (*estimated using the previous 36 months of data) on fund characteristics X_{it} which include size ($\ln TNA$), total expense ratio (TER), turnover ($TURN$) and maximum load fees ($LOAD$)[89]. Carhart finds expenses, turnover and load fees all have a negative impact on (abnormal net return) performance - of particular interest is the coefficient on the expense ratio: for every 100 basis point increase in TER, the net return alpha falls by 1.54%. Carhart also notes that the negative coefficient on load fees contradicts the oft-cited claim that such managers are more skilled than those of no load funds (see also, Chen et al 2004).

---

[88] For testing any individual $\delta_t = 0$ we use, $t = \bar{\delta} / se(\delta_t)$ where $\bar{\delta} = \sum_{t=1}^{T} \delta_t / T$ and $se(\delta_t) = \sqrt{\sum_{i=1}^{T} (\delta_i - \bar{\delta})^2 / T}$.

[89] TURN and TER are monthly averages of annual figures and are contemporaneous while LOAD and TNA are lagged one year. TURN is reported turnover plus 0.5 times the percentage change in TNA (adjusted for investment returns and mergers). LOAD is the sum of maximum front-end, back-end and deferred sales charges. The rolling cross-section regression is from July 1966 to December 1993.

Similar results are found by Chalmers, Edelen and Kadlec (1999) who rank funds over the 1984-1991 period into quintiles (rebalanced quarterly) using either total costs, trading costs, expense ratios or turnover. For all three cost ranking criteria, there is a strong negative and statistically significant relationship between costs and future performance, while turnover and performance are also negatively related but not statistically significant (see also Warther 1995)[90].

FUND RESTRICTIONS, INCENTIVES AND PERFORMANCE

There is a wide dispersion in the legal restrictions applied across funds – for example in their use of derivatives, margin purchases, short-selling, borrowing and categories of restricted (usually illiquid) stock. However, such restrictions do not appear to influence abnormal returns (Almazan, Brown, Carlson and Chapman 2004).[91] Funds which employ incentive fees (based on a benchmark such as the S&P500) might be expected to motivate existing managers or attract better managers. In 1999, only 108 US funds (1.7% of all funds covering 10.5% of total fund assets) used incentive fees and these had an average multi-index alpha (over the 1990-1999 period) of about 0.6% p.a., which exceeded that on non-incentive fee funds who had an average alpha of -0.4% pa – but it appears as if this differential performance is due to the lower expense ratio of the incentive-fee funds (Elton, Gruber and Blake 2003).[92]

INDUSTRIAL CONCENTRATION

As noted above Kacperczyk et al (2005) by sorting on a fund's industrial concentration (in stock holdings) show that this helps predicts future performance. This is also found to be the case when quarterly decile sorts on three alternative forward looking performance measures $PERF^{f}(t,T) = \{\alpha_{4F}^{f,net},\ \alpha_{4F}^{f,g}, CS^{f,g}\}$ are regressed on fund characteristics (e.g. *TER, TURN, ln(AGE), ln(TNA), NCF*) as well as $ICI_{i,t-1}$[93]. They find ICI is statistically significant but its economic impact is not large. For example, one standard deviation increase (=5%) in ICI

[90] Given the difficulty in obtaining relevant UK data, relatively little work has been conducted on the relationship between winner and loser funds and fund characteristics. However, Fletcher and Forbes (2002) examine whether annual charges, load charges and fund size are correlated with quartile (raw return) ranked portfolios of funds (rebalanced annually). The authors report very little cross-sectional variation in these characteristics and hence the authors suggest that such characteristics do not explain the "winner minus loser" abnormal returns.

[91] The restrictions listed are combined into an index and the analysis then proceeds in the usual manner either by sorting based on the index (above and below average) and then forming long-short portfolios that are re-balanced annually or using a Fama-McBeth cross-section regression of the rolling value of $\alpha_{3F,i}$ on the score index (and other fund characteristics). They use 324, US domestic equity funds from January 1997 to December 1999.

[92] The performance model is the 3F model with additional factors for bond returns and an international index. The non-incentive fee funds were paired with the incentive fee funds by size and investment objective.

[93] They use (time) fixed effects, unbalanced panel estimation (quarterly, 1984 -1999). AGE is the age of the fund and NCF is the previous quarters net cash flow. All variables are lagged one quarter except for TER and TURN which are lagged one year due to data availability. Variables ICI, lnAGE and TER are statistically significant in the net return 4F-alpha regressions over the whole period and sub-periods 1987-93 and 1994-1999.

increases $\alpha_{4F}^{f,net}$ by 13bp per quarter (0.5% p.a.) whereas a 1% increase in the TER leads to a fall in $\alpha_{4F}^{f,net}$ by 40bp per quarter (1.6% p.a.).

LOAD AND NO-LOAD FUNDS

The presence of load fees has been attributed to fund managers trying to separate investors with different liquidity needs (Chordia 1996, Nanda, Narayanan and Warther 2000). But what is the relative performance of load and no-load funds? With the exception of Ippolito (1989), who find that load funds earn rates of return that plausibly offset the load charge, most studies generally find that there is no significant difference between the performance of load and no-load funds (even before the former are adjusted for load charges) – see *inter alia*, Elton et al (1993), Grinblatt and Titman (1994), Droms and Walker (1994), Gruber (1996), Carhart (1997), Fortin and Michelson (1995) and Morey (2003). In addition, Morey (2003) examines relative performance *within l*oad funds using a number of risk adjusted measures (e.g. one and 4F alphas and finds there is little significant difference in abnormal returns between high load funds and low load funds[94]. The above empirical results contradict the predictions of the theoretical model of Nanda, Narayanan and Warther (2000), where investor returns in load funds exceed those in no-load funds and investors in funds with high-load fees earn a higher return than those in low-load funds.

To summarize, for funds as a whole their performance seems to be strongly and negatively related to costs such as TER, load fees and trading costs, with high cost funds resulting in negative abnormal returns. As far as load\no-load funds are concerned the evidence strongly suggests that abnormal returns on load funds do not cover the additional load fees charged. The latter may account for the decline in both the number of funds and cash under management in load funds (in the US) and is also consistent with (no-load) funds trying to recoup charges in higher 12-1b fees (Mahoney 2004). Here the evidence is clear, investors should choose no-load funds and funds with low expenses, if they wish to increase the probability of higher net abnormal returns. There also appears to be no strong evidence of an economically significant difference in performance based on differences in 'industrial concentration', the number of restrictions applied to funds and the use of incentive fees.

7.2 INVESTMENT FLOWS AND PERFORMANCE: IS MONEY SMART?

In a competitive market we might expect active investors to re-allocate cash away from past poor performers and towards past winners, in the expectation that this will increase future

[94] Houge and Wellman (2006) also note that load funds have higher expense ratios than non-load funds (50 bp difference over 2000-2004).

returns[95]. Key areas for investigation are first, the relationship between past fund performance and subsequent fund flows and second, whether fund flows provide an investment signal which can be used to give economically significant future returns – that is, whether money is smart. These propositions can be tested in an event study framework by sorting funds into appropriate portfolios (e.g. on past flows or performance) and following their subsequent returns. Alternatively one can use a (Fama-MacBeth) cross-section regression approach with either performance or flow metrics as the dependent variable, with their lagged values as independent variables - plus other fund characteristics such as TER, LOAD, TURN, AGE, Ln(TNA) as control variables.

PAST RETURNS AND FUTURE FUND FLOWS

A strong and significant positive relationship between past 'performance' and subsequent cash flows (after allowing for the influence of other fund characteristics), for a number of alternative performance measures (such as α_{4F}^{net}) is reported in Gruber (1996)[96]. There is also evidence that funds with larger stocks of unrealized capital gains (i.e. the 'tax overhang') experience lower inflows and outflows (so investors avoid realizing the gains) and that inflows are more sensitive to after-tax returns than pre-tax returns (Barclay, Pearson and Weisbach 1998, Bergstresser and Poterba 2002) – so potential tax burdens appear to influence fund flows.

More recently, using Fama-McBeth (quarterly and monthly) rolling cross-regressions approach, Barber, Odean and Zheng (2004) re-examine the impact of past returns and different types of fees on future fund flows. They find that (quarterly) flows into individual funds are positively related to past (excess) fund returns, returns squared and 12b-1 (advertising) fees but are negatively related to front-end loads - however, funds with high operating expenses do not lead to reduced inflows[97]. They argue that this demonstrates greater sensitivity of flows to advertising and load fees which are 'visible', rather than to operating expenses which are less 'visible' (see also Wilcox 2003 and Del Guercio and Tkac 2002). In a similar vein, Sirri and Tufano (1998) and Jain and Wu (2000) find that funds which spend more in 12b-1 fees to advertise their recent good performance, experience higher inflows (relative to good funds who

---

[95] As we have seen this scenario is the basis of the equilibrium model of Berk and Green (2004), while the possibility of 'strategy switching' by poorly performing funds in the Lynch and Musto (2003) model predicts relatively low outflows from poorly performing funds.

[96] Percentage net cash flows are defined as $NCF_t = \{NAV_t - (1+R)NAV_{t-1}\}/NAV_{t-1}$. Data on gross cash inflows and outflows are usually not available. Gruber weights each fund's performance metric (e.g. alpha) using net cash inflow weights ('new money') to obtain the portfolio abnormal return. He uses 227 funds over the period January 1985-December 1994.

[97] The quarterly cross-section regressions begin in 1970Q1 and end in 1999Q3. Control variables include ln(NAV), AGE and fund return volatility. In some regressions there is a positive perverse influence of higher operating expenses leading to higher fund inflows.

do not advertise).[98] The latter could be interpreted as funds minimizing the search costs for investors, about past good performers which is useful providing future performance is also high.

So studies clearly show that inflows into actively managed funds respond to good past performance, high advertising expenditure and low costs in a rational way. What about flows to index funds? Elton, Gruber and Busse (2004) using 52 US (S&P500) *index* funds (Jan 1996-Dec 2001) also find that flows respond to past relative performance (measured either as the CAPM alpha or excess fund returns over the S&P500 index).[99]

Fund flows amongst poor performers are examined by Lynch and Musto (2003) who find that flows are less sensitive to past performance when past performance is relatively poor[100]. This they attribute to low past returns having less information content for future returns because poor performing funds are more likely to change strategy. They find some fairly weak evidence that future performance of past losers is higher when the fund manager is replaced (but not when two other 'strategy change' variables based on changing factor loadings are used). Other key US studies also find evidence that cash flow is less sensitive to poor performance – for a variety of performance measures (e.g. raw returns, style adjusted returns, returns in excess of the market, Jensen's alpha – see Sirri and Tufano (1993), Del Guercio and Tkac (2002). Once again for the UK, data deficiencies imply that the relationship between fund performance and fund flow is comparatively unexplored[101].

The *relative* lack of fund outflow from past poor performers is partly explained by Lynch and Musto's (2003) "change of strategy" and subsequent higher returns of really poorly performing funds. But the lack of a significant outflow from poorly performing funds as a whole

---

98 Similar results are reported for flows into the fund family (Gallaher, Kaniel and Starks 2006) and in addition the convex performance-flow relationship also applies to advertising expenditure and fund flows, but this convex relationship only applies to the highest performing fund families (see also, Cronqvist 2005).

99 'Cash flow' is actually the unexpected cash flow and is the residual from a regression of actual cash flow on fund size. The cross section regression has annual 'year' intercept dummies and two control variables: a load dummy and the number of funds in the same family, which both have positive and statistically significant effects on future fund flows. The former they interpret as stronger selling incentives to brokers after receiving the load fee which outweighs the direct higher costs to investors and the latter they interpret as the convenience of easy inter-fund transfers and record keeping for the investor. These results on fund flows are of added importance because the correction for risk of an index fund is relatively uncontentious. Risk measures such as $(\beta_i - 1)$ or the tracking error (measured by the CAPM, R-squared of the fund relative to the average R-squared across all funds) are found to be statistically insignificant determinants of fund flows.

100 Chevalier and Ellison (1999b) also report that poorly performing funds who change managers, suffer less outflow than those that retain their managers.

101 However, for the UK Fletcher and Forbes (2002) examine whether performance is linked to fund flows. Ranking funds recursively into quartiles annually on past year excess returns reveals that the highest performing quartile experiences the largest cash inflow during the year. The worst quartile experience the least cash inflow, but do not suffer an absolute cash outflow, which suggests little penalty for their relatively poor performance. This is corroborated by Keswani and Stolin 2005, who have monthly data separately for inflows and outflows and where past performance is measured using 4F-alpha (estimated over the previous 36 months.

(many who remain subsequent poor performers) is worrying for the equilibrium model of Berk and Green (2004).

FLOWS AND FUTURE RETURNS

So, the evidence that investors chase past winners is clear, but we now examine if this results in higher future returns. Edelen (1999) addresses this question by differentiating between 'liquidity trades' and 'discretionary trades' (due to skill in market timing or stock selection) and finds that the latter have a positive effect and the former a negative effect on subsequent performance. This is because liquidity trades unexpectedly alter the funds relative cash/equity holdings, move the fund from its target portfolio and cause managers to undertake non-discretionary trades which are likely to lose money. This view is reinforced by Alexander, Cici and Gibson (2006) who take 324 US equity funds (January 1997-December 1999) and use a double sort into 25 quintiles based on net flows and the dollar value of trades. 'Valuation motivated trades' are defined as large dollar-buys (sells) which take place when there are heavy net outflows (inflows) on the other hand 'liquidity motivated trades' are funds in quintiles where small dollar buys (sales) are accompanied by large inflows (outflows). They find that valuation motivated trades earn substantially more on a risk adjusted gross CS basis over the subsequent year than do liquidity motivated trades - even after trades resulting from possible tax loss selling or mandated reporting months (which might involve window dressing) are excluded. So there is some evidence of skill for 'value trades'.

However, investors are presumably concerned about the overall return on the cash that is moved from past poor performers to past winners. Gruber(1996) finds that the average net return alpha (from a regression on market, size, value and a bond index) on 'new cash' is 29 bp per annum and the average investor saves 22 bp per annum, by removing their capital from poorly performing funds[102]. In a broadly similar study but using about 1,800 funds Zheng (1999),[103] over January 1970-December 1993, finds that high positive inflow fund's subsequently outperform low inflow funds (using conditional and unconditional $\alpha_{3F}^{f,net}$) but neither alpha is individually statistically significant and therefore the strategy requires shorting low inflow funds, which may not be feasible. In any case, the outperformance is relatively short lived (i.e. one quarter) and 'new money funds' statistically, do not beat the market as a whole. There is stronger evidence

[102] Cash inflows into a fund in quarter *t* are multiplied by the risk adjusted return of the fund in (a number of) subsequent periods. Returns are then aggregated over all funds and all time periods.

[103] For example funds are sorted each quarter into two portfolios based on $NCF_i>0$ and $NCF_i<0$ or alternatively sorts are undertaken using median cash flow as the 'break point' – in all, Zheng (1999) uses 6 'new money' sorting rules.

that smaller funds with large cash inflows subsequently earn abnormal positive risk adjusted returns, with the most favourable 'new-cash-*inflow* rule' giving $\alpha_{3F}^{f,net}$ = 2.2%p.a. (t=4.8).

Gruber (1996) and Zheng (1999) do not use a momentum factor so their documented flow-future performance link may be due to passive short-term momentum effect from existing stocks in the fund (Carhart 1997) or from new purchases of existing momentum stocks (Wermers 2000). Sapp and Tiwari (2004) test this proposition by repeating Zheng's portfolio sorts on new money flows and find that abnormal performance using $\alpha_{3F}^{f,net}$ does not carry over when using $\alpha_{4F}^{f,net}$.[104] They also undertake a *fund-by-fund (*Fama-MacBeth) cross-section regression to ascertain whether investors' net cash flows are determined by an active momentum strategy rather than investors' blindly chasing past returns. A regression of $NCF_{i,t}$ is repeated quarterly and they find that $(\beta_{MOM,i})_{t-1}$ is not statistically significant - but past returns and previous NCF are significant and positive.[105] This suggests that money is not 'smart', because it does not chase funds with active momentum styles but merely chases past (raw) return winners and the latter strategy does not earn positive future abnormal returns based on $\alpha_{4F}^{f,net}$.

Due to paucity of data, Keswani and Stolin (2005) is the only UK study which links new cash inflows and outflows to future performance (measured by 4F-alpha) over the period 1992-2000, using around 500 funds. With monthly portfolio rebalancing they find that 'new money' flows earn a higher abnormal return than 'old money' - but in each case the abnormal 4F return is negative.

From the above it can be seen that early studies suggest that money is 'smart', in the *limited* sense that most cash inflows are into past winner funds who subsequently experience higher future returns than past losers – note that this implies a *relatively* better outcome but does not imply that investors can earn positive abnormal returns. We would rather retain the word 'smart' for ex-ante investment strategies that earn positive abnormal returns – taking action to improve your *relative* position, but still earning negative abnormal returns, might be better described as 'less dumb than average' strategy. Indeed, later studies find that investors' cash

---

[104] Agarwal, Daniel and Naik (2004) find that for hedge funds, past performance leads to higher future inflows but greater inflows are associated with poorer future performance – consistent with the Berk and Green (2004) model.

[105] Control variables are $\{\ln TNA_i, TURN_i, TER_i, LOAD_i\}$. They also test this proposition by sorting funds each quarter into decile *portfolios* based on $(\beta_{MOM,i})_{t-1}$ and find that there is little difference in future cash inflows across the these deciles in any of the next 4 post-ranking quarters, whereas future cash flows do blindly follow past raw return winner decile funds.

blindly follows past raw return winner-funds (rather than funds with an active momentum strategy) and such funds do not have positive future abnormal returns after correcting for the momentum effect – so on our definition, money is not 'smart'.

The above conclusion is reinforced by Cooper, Gulen and Rau (2005) who examine the flow-performance relationship (1994-2001) for 332 funds which changed their names to reflect a current 'hot style'. Suppose it is the case that when a fund changes its *name* to a hot style (e.g. from 'growth' to 'value', or 'small' to 'large') but does not actually change its style[106], this results (ceteris paribus) in a large cash inflow[107], yet the subsequent abnormal performance of this fund is poor. We would not then infer that "money is smart" - this is exactly the conclusion reached by Cooper et al (2005). They find that funds which change their names to 'hot styles' are either funds which have been doing badly (based on 3F-alpha), are established funds or are funds with low advertising and low recent cash inflows[108]. The subsequent extra cash inflow attributable simply to the (cosmetic) name change is a substantial 25% after one year (in excess of flows to matched funds with no name change). But the subsequent returns and 3F alpha performance of the cosmetic name change funds is worse than their pre-name change performance and worse than funds with no-name changes. For example, pre and post name change $\alpha_{3F}^{f,net}$ are minus 0.11% p.a. (t = -1.92) and minus 0.23% p.a. (t = -3.63) respectively and average raw returns 1.42% to 0.33%, respectively.

Of course, some fund *managers* are 'smart', since simply by undertaking a cosmetic name change they can attract additional funds from investors (which are enhanced by additional advertising of the name change). But these investors pay around 3.75% transactions cost on average (for loads, expenses and fees), yet they subsequently earn no extra return. At a minimum this implies that more disclosure of fund holdings may be required so that investors can make more informed decisions. But somewhat pessimistically, it may also imply that investors do not use what knowledge is available in a sensible manner.

---

[106] Cosmetic name changes are those which do not result in a change in style as measured by the change in factor loadings on SMB and HML. If the fund's style factor loading in the 3F-model (measured over the 2 years after the name change) does not exceed that of the quintile BMV and size sorted control portfolio 'break points', then the name change is 'cosmetic'.

[107] Flows (3 months after the name change) also respond positively to changes in past advertising (12b-1 fees), performance (e.g. 3F alpha, net returns decile rankings relative to all equity funds) and negatively to changes in expenses and load fees (not significant) and to fund size (lnTNA) – as found in earlier studies. The 'hot style' is defined by a (0, 1) dummy, taking the value 1 when the corresponding style premium (e.g. R_{HML}) is 'up' and zero otherwise.

[108] 'Name change' is a (0, 1) dependent variable in a logit regression on fund characteristics.

Index funds are one of the simplest investment products, so Elton, Gruber and Busse **(**2004) investigate whether investors can use simple rules to move into index funds with relatively high future returns. First they show that past performance (using the differential return over the market or CAPM-alpha) or TERs over either 1 or 3 years, has high predictive power for future one and 3-year performance[109] – in contrast to studies of statistical prediction for actively managed funds. They then measure the actual performance of index funds over one and three year horizons[110] and compare this with returns to two types of 'alternative' index portfolios. The first are naïve portfolios (i.e. equally weighted or weights proportional to TNA) and the second are 'smart' investor portfolios (i.e. top deciles of index funds based on highest past returns or, highest past CAPM-alphas or lowest past total expense ratios, TERs).

The average return earned from actual net inflows into index funds are generally worse than any of the above alternative portfolios. For example, actual investors in index funds *as a whole,* do 15 bp per annum worse than if they had purchased the top 10% of funds based on past returns (for 1-year holding periods). True, this is not a particularly large differential to 'active search' but the future differential return between past decile winners and loser index funds is 92 bp per year – which is economically (and statistically) significant.

Elton et al (2004) suggest that the lower returns from actual flows into index funds (compared with the above alternatives) is due to higher marketing costs - as funds actually held have higher loads and 12b-1 fees and also have higher expenses. Of course, higher expenses could lead to more advice from brokers but what is clear is that such compensation seems incompatible with the interests of investors, since it leads to inferior performance compared with simple mechanical rules for investing in index funds. So, financial advisers and brokers benefit from relatively high fees from the funds recommended but there is no extra return for investors relative to the index itself, or to simple strategies based on past performance of index funds. It is also the case that poorly performing index funds also receive substantial cash inflows even thought their subsequent performance is relatively poor. Once again we have evidence of inertia or ignorance on the part of many investors who are investing in the simplest mutual fund product, while advisers have an economic incentive to sell inferior products. The counter-argument is that differential fees provide differential information which is genuinely valued by investors – even though they could do better with simple rules (see Hortacsu and Syverson 2004).

---

[109] The R-squared in various alternative regressions of '3 year performance' on 'past 3 year performance' are in the range 0.77 to 0.88 and the coefficients are close to 1 for differential returns on past differential returns, alpha on past alpha and are close to -1 for returns or alpha on past TERs. Similar results apply for 1 year horizons.

[110] Actual (cash flow weighted) returns are $R^{A}_{t,t+j} = \sum_{i=1}^{N} w_i R_{i,t,t+j}$ *where* $w_{i,t} = CF_{i,t} / \overline{CF_t}$ and $\overline{CF_t}$ = mean cash inflow over all N index funds (If the cash inflow is negative, the fund is excluded for this month). Funds are rebalanced monthly.

7.3 MUTUAL FUND MANAGERS

Skill and persistence in performance may reside at the fund manager level rather than at the fund level. If so, investors should 'chase' past top performing managers not necessarily top funds. Using a cross-section Fama-McBeth approach, Chevalier and Ellison (1999b) evaluate whether mutual fund performance (after controlling for other fund risk characteristics) is related to fund manager 'skill' as measured by age, the average SAT score of the manager's undergraduate institution and whether or not the manager held an MBA. Using a sample of 492 mutual fund managers who had sole responsibility for a fund for at least some part of the 1988 – 1994 period, the most robust performance differential identified is that managers with higher undergraduate SAT scores obtain higher risk adjusted returns. They attribute this outperformance to better natural ability, education and professional networks associated with having attended a higher SAT score undergraduate institution. Although Evans (2003) shows that alpha has an influence on promotions and demotions of managers nevertheless Baks (2002) who tracks managers as they move between funds, concludes that the fund typically has a greater influence on performance than the manager.

Do we get similar results on manager skill and performance when using holdings and trade data on stocks held by funds? Ding and Wermers (2004) sort funds (each quarter into deciles) based on either cumulative years of experience of the fund manager or the cumulative value of the manager's CS index and then track the returns and CS index over the next year. Top managers, based on their past cumulative CS index (but not on cumulative years experience) gives a statistically significant $\overline{CS}_{t,t+4Q}$ = 2.2**% p.a. (for data period December 1985-December 1999) – but the latter is statistically insignificant when the data period is extended to December 2002 (Ding and Wermers 2005). Although net cash inflows are mainly into past winner managers' funds (28% p.a. increase) rather than past losers (7.8% inflow), the past winners do not earn a statistically significantly higher net return or conditional or unconditional (Ferson-Schadt 1996) values for $\alpha_{4F}^{f,net}$.[111]

---

[111] Similar results are reported for the performance of pension fund mandates over 1993-2003, with a total of 9,581 decisions over 3,715 plan sponsors examined (Goyal and Wahal 2005). Fund managers that underperform their benchmarks over the last 3 years are fired and new mandates given to past winners. However, in the 3 years before hiring these managers earn excess returns (over the benchmark) of 14.1% (s.e.=2.3) but in the post-hiring 3 years they earn only 1.6% (s.e.=1.6) (p3). For mangers fired for bad performance (as opposed to say a change of mandate) the pre-firing 3-year excess return is -5.3% (s.e.=1.2%) but in the subsequent 3 years is 6.3% (se=2.9). (p.4). Consistent with these figures, using a matched sample of round trip firing and hiring decisions, there is a net loss in terms of return to the change in mandate –as well as the transition costs of around 2%. Put another way, plan sponsors cannot successfully market time changes in mandates - and the only benefit appears to be maintaining incentives among incumbent funds, due to the threat of the loss of the mandate.

Because of the limitations of any sort procedure[112] Ding and Wermers (2004) undertake a Fama-McBeth cross section regression (repeated annually, 1986-2000) of $CS_{i,t+1}$ on fund characteristics such as years of experience ("Years"), the average cumulative value of CS (*CumCS*). A representative result for funds with a *growth-oriented* style is[113]

[13] $$CS_{i,t+1} = 0.86 + 0.03\ Years_{i,t} + 0.09\ CumCS_{i,t}$$
$$(1.38)\ (1.92) \qquad (3.04)$$

1986-2000, Average number of funds=676, Average $\overline{R}^2 = 0.01$
Source: Ding and Wermers (2004), table VII.

Hence, past cumulative performance of a fund manager ($CumCS$) influences future risk adjusted performance (CS) and years of experience is marginally significant. In a later study, with data to end 2002, Ding and Wermers (2005) continue to find that past cumulative performance is statistically significant for future CS returns for *growth* oriented (but not income oriented) funds, while years of experience is statistically significant but only for managers of large funds.

Overall, it appears that choosing funds on the basis of the past skill characteristics of their *managers* has some statistical predictive power for future abnormal returns. However, the link from differential skill to future abnormal CS return on stock holdings is not particularly large, applies mainly to growth funds and does not carry over to higher net return alphas for all funds. Thus any skill that resides at fund manager level seems to accrue to those running the fund and not to investors.

So, "Is Money Smart?" Our overall conclusion from the above studies must be that most (but not all) money is pretty dumb. Investors blindly chase past winners (both active funds and index funds), chase funds with cosmetic name changes, respond strongly to fund advertising and they do not chase funds with a high loading on the momentum factor. Past winner funds do earn positive risk adjusted gross but not net returns, in subsequent periods – fund managers therefore

---

[112] These include the assumption of homogeneity within deciles, the possible impact of other fund characteristics being correlated with the characteristic used to sort, and when using factor models the assumption of constant factor loadings over time.

[113] The statistics reported are the time-series averages of the cross section regression parameters. Other control variables tried in the regression such as the standard deviation of market returns over the life of the manager (=risk taking), the manager career turnover ratio (=aggressiveness) and manager replacement in the previous year (= a 0, 1 dummy) are not statistically significant. When *all funds* or just *income funds* are used, the cumulative CS manager performance is just statistically significant at a 10% significance level in the former but not in the latter – so for income funds manager stock picking skills do not predict future risk adjusted performance.

expropriate the returns to their skills, as in the equilibrium model of Berk and Green (2004). However, what is also clear is that at the other end of the performance distribution, the absence of large cash outflows from poorly performing funds probably inhibits the competitive process, since most of these funds continue to earn persistently poor abnormal returns (Kosowski et al 2005, Barras et al 2005). So, some money may be smart but it is at the lower end of the performance distribution that money is really dumb – in contrast to the predictions in the Berk and Green (2004) model.

8. CONCLUSIONS

We have surveyed the recent literature on mutual fund ex-post performance, together with fund characteristics that contribute to that performance and also addressed the question "Is money smart?". What does this literature reveal about investment decisions ?

First consider the simplest investment product, index funds -the average risk adjusted net return is about minus 0.4% p.a. (matching the average TER of 0.41% p.a.). Also, the future differential return between past decile winner and loser index funds is 92 bp per year which is economically and statistically significant and there is a substantial 2% p.a. differential between the best and worse performing US index funds. Investors therefore appear to ignore simple ex-ante strategies for increasing performance and clearly with expense ratios as high as 1.35%, many investors do not avoid high cost index funds.

Turning now to active funds, work based on *stock* holdings and buy/sell data (based on the gross CS measure) has demonstrated that some stock picking skill exists among 'all' active equity funds and especially for the average aggressive growth fund - but this outperformance barely covers management fees and transactions costs. In any case investors are primarily concerned with the performance of the overall fund, not just the stocks held or traded by the fund.

When looking at US ex-post *average* performance (over 1974-94) for active equity funds, the difference between gross returns on *stock* holdings and net returns on *all* asset holdings is about 2.3% p.a. and this difference is accounted for in equal measure by lower returns on non-stock holdings, the TER and trading costs (Wermers 2000). On a risk adjusted basis the 4F-net return alpha for the average (equally weighted) fund over 1975-2002 is around minus 0.5% p.a. (but not statistically significant – Kosowski et al 2005, Barras et al 2005) but the average fund

does earn a positive risk adjusted return in recession periods (Kosowski 2006)[114]. UK studies find that the average fund underperforms and over the 1975-2002 period α_{3F}^{net} = -1.7***% p.a. (Cuthbertson et al 2006).

The large dispersion in the cross-section of fund abnormal returns in both the US and UK suggests it is worth examining funds in the tails of the performance distribution and assessing the role of luck versus skill. In terms of ex-post performance recent US and UK studies find around 2-10% of funds in the extreme right tail have positive net return alphas and at least 20% of funds spread throughout the right tail have genuinely poor performance (Kosowski et al 2005, Barras et al 2005, Cuthbertson et al 2006)[115]. For example, the net return alpha for the US fund (1975-2002) at the 95th percentile is 4.8% p.a. and in the left tail at the 5th percentile is minus 6% p.a. – both of which are statistically significant. US data reveal that the top performers are in growth and aggressive growth styles, while the top 'growth and income' and 'balanced or income' funds do not beat their 4-factor benchmarks. In contrast, in the UK, skilled funds tend to be in the income style rather than in growth or small cap funds.

What causes the differential performance across funds? UK evidence is very sparse in this area, but US studies show that the main influences on the cross-section of active fund's abnormal net return alphas are the strong negative impact of fees (e.g. 12b-1 fees, TER's, loads) and to a lesser extent, high turnover.

What about ex-ante rules for picking winners? Statistical measures of persistence such as contingency tables and cross-section regressions of current on past abnormal returns indicate relatively weak short-term persistence among the past winners (which is strongest over horizons of one year or less) and longer horizon persistence (up to 3-5 years) for past losers – for both US and UK studies. However, what is important to investors is the economic significance of any persistence.

Using the recursive portfolio approach, US studies show statistically significant persistence in risk adjusted (CS) *gross* returns to stock holdings and trades, in the current and next quarter for the top decile of 'all funds' of around 1% p.a. and persistence up to 4-years ahead of around 2% p.a. for *growth* funds (using data 1975-93, Wermers 2003b). Short-term

---

[114] Post 2000, some US mutual fund companies have introduced funds that emulate hedge funds and Agarwal, Boyson and Naik (2006) find that on average these 'hedged mutual funds' outperform traditional mutual funds by at least 3% p.a. (after controlling for expenses, past performance, risk and fund characteristics).

[115] For the US there is evidence of more winners in the 1975-1989 period relative to the 1990-2002 period but widespread poor performance is prevalent in both periods – Kosowski et al 2005.

persistence among past winners is somewhat weaker when using abnormal *net* returns to the fund as a whole. Based on decile sorts (on past 4F-alphas or raw returns over 1978-2002) the US evidence shows some persistence over one year for the 'winner' decile ($\alpha_{4F}^{f,net}$ = 1%p.a., p =0.05)[116] but what is absolutely clear is that it is worth avoiding funds in the bottom 4 'loser' deciles since these have persistent negative net return alphas (e.g. for the bottom decile $\alpha_{4F}^{f,net}$ = -3.5% p.a., p <0.01) - Kosowski et al (2005).

Results assessing manager performance rather than fund performance find some statistical relationship between various measures of management skill and future performance. But the effect is not very large and is unlikely to yield sorting rules which give substantive abnormal net returns. The same goes for sorting on variables such as turnover, industrial concentration and while extraneous fund information may help in picking winners the evidence here is relatively new and involves a relatively sophisticated sort procedure (Cohen, Coval and Pastor 2005).

The above figures all ignore load fees, rebalancing costs and investor's taxes - so is chasing a net return alpha of around 1% p.a. (p=0.05) for the top decile winner portfolio worth it? (Kosowski et al 2005). Even ignoring estimation and model error, many practical difficulties need to be addressed. Most investors would have to mimic the performance of the winner decile (of around 180 funds at the end of 2002) with a smaller number of funds - and when rebalancing (at least annually) would need to avoid any funds with load fees, take account of bid-ask spreads and finally, consider any tax implications of fund purchases/sales. (The average load fee is around 3.6% and applies to about 56% of funds – Mahoney 2004). To ascertain the true impact of such costs we have to know the number and dollar value of funds purchased/sold at each rebalancing date but it seems likely that such costs would outweigh the relatively small abnormal net return alpha of around 1% p.a. Picking winners from equity mutual funds appears to be a difficult task.

Because of space constraints our survey does not cover the performance of bond and hedge funds in detail. But, it may be worth noting at this point that broadly speaking, persistence in US bond fund performance is similar to that found for domestic equity funds. Blake, Elton and Gruber (1993) using around 300 bond funds find little evidence of persistence and expenses are the key determinant of ex-post relative performance. However, more recently Huij and Derwall (2006) using a survivorship-bias free sample of over 3,500 US bond funds (1993-2003) find that there is some statistical persistence when using correlation and contingency tables. They also

[116] Bollen and Busse (2004) find a similar figure with quarterly rebalancing, over 1984-94.

examine economic persistence by sorting into deciles on past 3-year alphas (based on several alternative multifactor models) and find that for equally weighted decile portfolios (rebalanced monthly) there is statistically significant negative persistence in forward looking alphas (of between -1% and -3%) but no positive persistence. However when the top decile weights are based on 'modern portfolio theory'[117] they find forward looking alphas of 1.27%, 1.13% and 0.53% p.a. for monthly, quarterly and annual rebalancing, respectively and these figures are around 0.5% higher for no-load bond funds. So there is evidence that before rebalancing costs, the top decile-alpha bond funds persist and stronger evidence that losers persist.

What about hedge fund performance? This poses additional data problems (Fung, Hsieh and Ramadorai 2005) many of which are mitigated by the comprehensive study of Kosowski, Naik and Teo (2006). Using a cross-section bootstrap, they find that in a sample of over 2,700 funds (January 1994-December 2002), all of the individual funds in the top 10% of hedge funds ranked by t-alpha have highly statistically significant large positive (seven-factor) ex-post alphas,[118] while all of the funds in the bottom 10% have negative alphas but these are due to bad luck rather than 'bad skill'. Using the recursive portfolio approach with annual rebalancing and sorting on past t-alpha, they find that *all* decile portfolios exhibit statistically significant forward looking alphas of between 4% and 6% p.a. (p-values<0.03)[119]. Clearly, 'winner persistence' is much stronger statistically and economically in hedge funds compared with (US or UK) equity mutual funds (see also Agarwal and Naik, 2002).

.

Is money smart? It turns out that for investment in US equity mutual funds, a substantial amount of cash, flows into past winner funds (usually defined in terms of either raw return, alpha or CS) and the winner decile portfolio earns positive gross abnormal returns - yet this only translates into future abnormal net returns (before switching costs) of around 1% p.a. at best. Thus the Berk and Green (2004) equilibrium model is broadly correct for winners – fund managers expropriate any rents from their skill and the return to smart investors after all transactions cost is probably rather small. However, the strong persistence in poorly performing

---

[117] The weights are $w_i = \{\alpha_i / \sigma_{\varepsilon_i}\} / \sum_{i=1}^{N} \alpha_i / \sigma_{\varepsilon_i}$ for funds in the top decile-sort which also must have $\alpha_i > 0$ to be included.

[118] For example, the 90th percentile fund has an ex-post alpha of 15% p.a. (p<0.00) =0. The improvement in ex-post performance of hedge funds relative to mutual funds can also be gleaned from the fact that to end 2002, the (equally weighted) average (seven factor) alpha for US hedge funds is 5.0% p.a. while that for US mutual funds is considerably lower at -0.5% pa.

[119] Short-term persistence (at three month horizons) in hedge fund raw returns has been found in previous studies (e.g. Brown, Goetzmann and Ibbotson 1999, Agarwal and Niak 2002, Liang 2000), but not at longer horizons. (Getmansky, Lo and Makarov 2004 argue that short-term persistence may be due to illiquidity in stock returns). Kosowski, Naik and Teo (2006) also find that a sort based on Pastor and Stambaugh's (2002) Bayesian alphas or t-alphas (i.e. utilizing returns on seemingly unrelated assets) provides stronger statistical evidence of persistence (up to a 4-year horizon), than sorting on frequentist alpha or t-alpha.

funds together with low positive inflows (rather than large outflows) suggest that the Berk and Green model does not apply at the negative end of the performance distribution. Although really poor funds may change their strategy and improve their performance in the future (Lynch and Musto 2003), the prevalence of a large number of persistent poor performers indicates that a lot of money is dumb - and any move towards a competitive equilibrium among loser funds, appears to be relatively slow.

Why any mutual fund, particularly a long-lived fund, which truly underperforms would be permitted to survive in a competitive market is puzzling. In part it may be that for some investors, interpreting performance measurement statistics is a difficult task and for precision requires a long fund life-span. Various rational explanations for the continued existence of poorly performing funds include investors being 'locked in' (e.g. pension plans) or having accrued capital gains (Gruber 1996). Other reasons include biases in investor information sets, the influence of advertisements and blindly following brokers recommendations, inertia, ignorance, actual or psychological costs (e.g. disappointment aversion, disposition effect) – in short, an element of irrationality, if your baseline model is one of informed decisions in relatively frictionless and low information cost markets (see for example, Elton, Gruber and Busse 2004, Sapp and Tiwari 2004, Cooper, Gulen and Rau 2005).[120]

In this survey we have noted the wide range of innovative studies used in measuring the performance of mutual funds and these methodologies have also been applied to pension funds and more recently to hedge funds and even to venture capital funds (Cochrane 2005). Work on mutual funds has revealed a great deal about their performance and the reasons behind this performance. Indeed, recent studies have concentrated on the role of luck in performance statistics together with an emphasis on individual fund performance or the performance of small 'fund-of-fund' portfolios – this is important since it brings academic work closer to the interests of investors and industry practitioners.

There is a large empirical literature on agency problems of mutual funds which we have not been able to cover in this survey. One argument is that funds should be viewed as businesses owned by their shareholders and managed by fund sponsors, and there is a presumption that agency problems are best dealt with via detailed external regulation of funds. An alternative is that shareholders view themselves as customers rather than owners and hence their weapon for combating acute agency problems by specific funds is to shift their assets

[120] Instances of 'non-rational' behavior for other financial decisions is widely documented in the behavioural finance literature - an excellent recent survey is Barberis and Thaler (2003). Note that poor ex-post performance for even the worst ranked individual *hedge funds* is not statistically significant and is therefore due to 'bad luck' – nevertheless, investors should switch into those hedge funds with genuine skill (Kosowski et al 2006).

elsewhere – the presumption here is that this view should entail less external regulation (Tkac 2004). On a public policy level, the academic literature cited above suggests there is certainly scope for more transparency about individual fund costs, trades and performance measures and also a strong argument for a more impartial educative role directed towards fund investors[121].

Overall, academic work demonstrates that there are relatively few mutual funds which have genuinely positive alphas and picking ex-ante winners is very difficult when one considers potential data snooping bias, model/estimation error and transactions costs due to rebalancing. In contrast, persistence among past loser funds is well established. The evidence on fund performance suggests that any 'national' savings schemes (e.g. 401K schemes, part privatization of US social security and Turner's(2006) 'BritSaver' scheme) should warn against trying to 'pick winners' and seek to provide impartial, independent information on fund performance[122]. Sensible advice for most investors would be to hold low cost index funds and avoid holding past 'active' loser funds. Recent work with multiple sorts or using Bayesian approaches have demonstrated the possibility of 'picking winners' from the mutual fund universe but overall, current evidence suggests that only very sophisticated investors should pursue an active investment strategy of trying to pick winners - and then with much caution.

---

[121] For example, see Mahoney (2004) on market timing scandals in the US and the literature on choices in 401K pension schemes (e.g. Bernatzi and Thaler 2001, Huberman and Jiang 2006). Sandler (2002) and Turner (2004, 2005) analyze private long-term savings in the UK (including the design of compulsory versus voluntary schemes) and OECD (2005) provides a survey of nascent programes in developing financial education in member states. Wermers (2001) analyses the issues associated with more frequent disclosures of asset holdings.

[122] Given the lack of consensus on the provision of long-term savings in the UK, it has been suggested in evidence to the Turner Commission that in the UK one should set up an independent Pensions Policy Committee PPC along the lines of the Bank of England's Monetary Policy Committee (Cuthbertson et al 2005b). The PPC would act as an independent organization focused on education concerning generic principles and 'good practice' in assessing the performance of long term savings vehicles including mutual, (private) pension and hedge funds. This suggestion is all the more important if a voluntary 'opt-out' national savings scheme or later maybe a compulsory savings scheme is adopted in the UK (Turner 2005).

References

Abrevaya, Jason and Wei Jiang, 2005, A Nonparametric Approach to Measuring and Testing Curvature, *Journal of Business and Economic Statistics*, 23 (1), 1-19.

Admati, A.R., S. Bhattacharya, Stephen A. Ross, and P. Pfleiderer, 1986, On Timing and Selectivity, *Journal of Finance,* 41, 715-730.

Agarwal, Vikas and Narayan Y. Naik, 2002, 'Multi-period Performance Persistence Analysis of Hedge Funds' *Journal of Financial and Quantitative Analysis*, 35, 327-342

Agarwal, Vikas, Naveen D. Daniel and Narayan Y. Naik, 2004, Flows, Performance and Managerial Incentives in Hedge Funds, London Business School, Working Paper, July.

Agarwal, Vikas, Nicole M. Boyson and Narayan Y. Naik, 2006, Hedge Funds for the Rest of us? An Examination of Hedged Mutual Funds, London Business School, Working Paper, August.

Ait-Sahalia, Y., and M. Brandt, 2001, Variable Selection for Portfolio Choice, *Journal of Finance,* 56, 1297-1351.

Alexander, Gordon, J., Gjergji Cici and Scott Gibson, 2006, Does Motivation Matter When Assessing Trade Performance?, *Review of Financial Studies*, forthcoming.

Allen, D.E., and M.L. Tan, 1999, A Test of the Persistence in the Performance of UK Managed Funds, *Journal of Business Finance and Accounting,* 25, 559-593.

Almazan, Andres, Keith C. Brown, Murray Carlson and David A. Chapman, 2004, Why Constrain Your Mutual Fund Manager, *Journal of Financial Economics,* 73, 289-321.

Ang, Andrew, and Geert Bekaert, 2006, Stock Return Predictability – Is it There?, *Review of Financial Studies*, forthcoming.

Avramov, Doron and Russ Wermers, 2005, Investing in Mutual Funds when Returns are Predictable, *Journal of Financial Economics, forthcoming.*

Baker, Malcolm, Lubimor Litov, Jessica A. Wachter and Jeffrey Wurgler, 2005, Can Mutual Fund Managers Pick Stocks? Evidence From Their Trades Prior to Earnings Announcements, Harvard Business School, Working Paper.

Baks, Klaas P., 2002, On the Performance of Mutual Fund Managers, Working Paper, Emory University

Baks, Klaas P., A. Metrick, and J. Wachter, 2005, Should Investors Avoid All Actively Managed Funds? A Study in Bayesian Performance Evaluation, *Journal of Finance,* 56, 45-86.

Ball, R., and S.P. Kothari, 1989, Nonstationary Expected Returns: Implications for Tests of Market Efficiency and Serial Correlations in Returns, *Journal of Financial Economics,* 25, 51-74.

Barber, Brad M., Terrance Odean and Lu Zheng, 2004, Out of Sight, Out of Mind: The Effects of Expenses on Mutual Fund Flows, *Journal of Business*, 78, 2095–2120.

Barberis, Nicholas, 2000, Investing for the Long Run When Returns are Predictable, *Journal of Finance,* 55, 225-264.

Barberis, Nicholas, and R.H. Thaler, 2003, A Survey of Behavioral Finance, in G.M. Constantinidis, M. Harris and R. Stulz (eds), *Handbook of the Economics of Finance,* Elsevier Science B.V.

Barclay, Michael, N. Pearson and Michael Weisbach, 1998, Open Ended Mutual funds and Capital Gains Taxes, *Journal of Financial Economics*, 49, 3-43.

Barras, Laurent, Olivier Scaillet, and Russ Wermers, 2005, False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas, FAME Research Paper No.163, University of Geneva, October.

Becker, C., W. Ferson, D.H. Myers and M.J. Schill, 1999, Conditional Market Timing with Benchmark Investors, *Journal of Financial Economics,* 52, 47-78.

Benartzi, S., and R.H. Thaler, 2001, Naïve Diversification Strategies in Defined Contribution Savings Plans, *American Economic Review,* 91, 79-88.

Bergstresser, Daniel and James Poterba, 2002, Do After-Tax Returns Affect Mutual Fund Inflows, *Journal of Financial Economics*, 63, 3, 381-414.

Berk, Jonathan B., and Richard C. Green, 2004, Mutual Fund Flows and Performance in Rational Markets, *Journal of Political Economy,* 112, 1269-95.

Bildersee, J.S., 1975, The Association Between a Market Determined Measure of Risk and Other Measures of Risk, *Accounting Review,* 50, 81-98.

Blake, C.R., E. J. Elton and M.J. Gruber, 1993, The Performance of Bond Mutual Funds, *Journal of Business,* 66, 371-403.

Blake, Christopher A., and M. Morey, 2000, Morningstar Ratings and Mutual Fund Performance, *Journal of Financial and Quantitative Analysis,* 35, 451-483.

Blake, David, and Allan Timmermann, 1998, Mutual Fund Performance: Evidence from the UK, *European Finance Review,* 2, 57-77.

Blake, David, B. Lehmann and Allan Timmermann, 1999, Asset Allocation Dynamics and Pension Fund Performance, *Journal of Business,* 72, 429-461.

Bogle, J., 1999, *Common Sense on Mutual Funds,* J. Wiley

Bollen, Nicolas P.B., and Jeffrey A. Busse, 2001, On the Timing Ability of Mutual Fund Managers, *Journal of Finance*, LVI(3), 1075-1094.

Bollen, Nicolas P.B., and Jeffrey A. Busse, 2004, Short-Term Persistence in Mutual Fund Performance, *Review of Financial Studies,* 18(2),569-597

Breen, W., Jagannathan, R. and Ofer, A. (1986). Correcting for Heteroscedasticity in Tests for Market Timing Ability, *Journal of Business*, 59, pp 585-598.

Bris, Arturo, Huseyin Gulen, Padma Kediyala and P. Raghavendra Rau, 2005, Good Stewards, Cheap Talkers or Family Men? The Impact of Mutual Fund Closures on Fund Managers, Flows, Fees and Performance, Yale School of Management, Working Paper.

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

Brown, Stephen J., and William N. Goetzmann, 1995, Performance Persistence, *Journal of Finance* 50, 679-698.

Brown, Stephen J., William N. Goetzmann and James Park, 2001, Careers and Survival: Competition and Risk in the Hedge Fund and CTA Industry, *Journal of Finance*, 56, 1869-1886.

Brown, Stephen J., William N. Goetzmann and Roger. G. Ibbotson, 1999, Offshore Hedge Funds Survival and Performance 1989-1995, *Journal of Business*, 72, 91-118.

Brown, Stephen J., William N. Goetzmann, Roger G. Ibbotson, and Stephen A. Ross, 1992, Survivorship Bias in Performance Studies, *Review of Financial Studies* 5, 553-580.

Busse, Jeffrey A. and Paul J Irvine, 2006, Bayesian Alphas and Mutual Fund Persistence, *Journal of Finance*, 61, 5, 2251-2288.

Busse, Jeffrey A., 1999, Volatility Timing in Mutual Funds: Evidence from Daily Returns, *Review of Financial Studies*, 12(5), 1009-1041.

Busse, Jeffrey A., 2001, Another Look at Mutual Fund Tournaments, *Journal of Financial and Quantitative Analysis,* 36(1), 53-73.

Campbell, John Y., and Luis M. Viceira, 1999, Consumption and Portfolio Decisions when Expected Returns are Time Varying, *Quarterly Journal of Economics,* 114, 433-496

Campbell, John Y., Y.L. Chan, and Luis M. Viceira, 2003, A Multivariate Model of Strategic Asset Allocation, *Journal of Financial Economics,* 67, 41-80

Capocci, D. and G. Hubner, 2004, Analysis of Hedge Fund Performance, *Journal of Empirical Finance*, 11, 55-89.

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

Carhart, Mark M, J. Carpenter, A. Lynch, and D. Musto, 2002b, Mutual Fund Survivorship, *Review of Financial Studies* 15, 1439-1463.

Carhart, Mark M, Ron Kaniel, David K. Musto and Adam V. Reed, 2002a, Leaning for the Tape: Evidence of Gaming Behavior in Equity Mutual Funds, *Journal of Finance,* LVII(2), 661-693.

Carpenter, J. and Lynch, A. (1999). Survivorship bias and attrition effects in measures of performance persistence, *Journal of Financial Economics*, vol. 54, pp 337-374.

Chalmers, J., R. Edelen and G. Kadlec, 1999, Transaction Cost Expenditures and the Relative Performance of Mutual Funds, Working Paper, University of Oregon.

Chan, Louis K.C, J. Karceski, and Josef A. Lakonishok, 2000, New Paradigm or Same Old Hype in Equity Investing?, *Financial Analysis Journal* 56, 23-36

Chan, Louis K.C., 1988, On the Contrarian Investment Strategy, *Journal of Business* 61, 147-164.

Chan, Louis K.C., and Josef A. Lakonishok, 2004, Value and Growth Investing: Review and Update, *Financial Analysts Journal,* 60, 71-86.

Chan, Louis K.C., Narasimhan Jegadeesh, and Josef A. Lakonishok, 1996, Momentum Strategies, *Journal of Finance* 51, 1681-1713.

Chen, Hsiu-Lang, Narasimhan Jegadeesh, and Russ Wermers, 2000, The Value of Active Mutual Fund Management: An Examination of the Stockholdings and Trades of Fund Managers, *Journal of Financial and Quantitative Analysis* 35, 343-368.

Chen, Joseph, Harrison Hong, Ming Huang and Jeffrey Kubik, 2004, Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organisation, *American Economic Review*, 94, 1276-1302.

Chen, L., N. Jegadeesh and J. Lakonishok, 1996, Momentum strategies, *Journal of Finance*, 51, 1681-1713.

Chevalier, J., and G. Ellison, 1999a, Risk Taking By Mutual Fund Managers as a Response to Incentives, *Journal of Political Economy*, 105, 1167-1200.

Chevalier, J., and G. Ellison, 1999b, Are Some Mutual Fund Managers Better than Others? Cross-Sectional Patterns in Behavior and Performance, *Journal of Finance,* 54, 875-899*.*

Chordia, Tarun, 1996, The Structure of Mutual Fund Charges, *Journal of Financial Economics*, 41, 3-39.

Christopherson, Jon A., Wayne E. Ferson, and Debra A. Glassman, 1998, Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance, *Review of Financial Studies,* 11, 111-142

Cochrane, John, 2005, The Risk and Return of Venture Capital, *Journal of Financial Economics*, 75, 3-52.

Cohen, Randolph B., Joshua D. Coval and Lubos Pastor, 2005, Judging Fund Managers by the Company They Keep, *Journal of Finance,* LX(3), 1057-96.

Coles, Jeffrey, L., Naveen D Daniel and Frederico Nardari, 2006, Does the Choice of Model or Benchmark Affect Inference in Measuring Mutual Fund Performance?, Working Paper, Arizona State University, January.

Connor, G. and R. Korajczyk, 1986, Performance Measurement with the Arbitrage Pricing Theory, *Journal of Financial Economics*, 15, 373-394.

Cooper, Michael J., Huseyin Gulen and P. Raghavendra Rau, 2005, Changing Names With Style: Mutual Fund Name Changes and Their Effects on Fund Flows, *Journal of Finance,* 60, 2825-2858.

Coval, Joshua D., and Tobias J.Moskowitz, 1999, Home Bias at Home: Local Equity Preference in Domestic Portfolios, *Journal of Finance,* 54, 2045-2074.

Cremers , Martijn and Antti Petajisto, 2006, How Active in Your Fund Manager? A New Measure that Predicts Performance, Yale School of Management, Working Paper.

Cronqvist, Henrik, 2005, Advertising and Portfolio Choice, *Journal of Financial Economics,* forthcoming.

Cuthbertson, Keith and Dirk Nitzsche, 2004, *Quantitative Financial Economics: Stocks, Bond and Foreign Exchange,* J. Wiley, Chichester.

Cuthbertson, Keith, Dirk Nitzsche and Niall O'Sullivan, 2005, Live Now, Pay Later or Pay Now Live Later, evidence to the Turner Commission on UK Pension Reform, Pensions Commission, London.

Cuthbertson, Keith, Dirk Nitzsche and Niall O'Sullivan, 2006, Mutual Fund Performance: Skill or Luck?, *Cass Business School, London, SSRN* Working Paper.

Daniel, Kent M., Mark Grinblatt, Sheridan Titman and Russ Wermers, 1997, Measuring Mutual Fund Performance With Characteristic Based Benchmarks, *Journal of Finance* 52, 1035-1058.

Davis, Gerald, F. and E. Han Kim, 2005, How Do Business Ties Influence Proxy Voting By Mutual Funds, *Journal of Financial Economics,* forthcoming.

Del Guercio, D. and P.A. Tkac, 2002, The Determinants of Flow of Funds of Managed Portfolios: Mutual Funds versus Pension Funds, *Journal of Financial and Quantitative Analysis,* 37(4), 523-57.

Ding, Bill and Russell Wermers, 2004, Mutual Fund "Stars": The Performance and Behavior of US Fund Managers, Robert H. Smith School of Business, University of Maryland, Working Paper.

Ding, Bill and Russell Wermers, 2005, Mutual Fund Performance and Governance Structure: The Role of Portfolio Managers and Boards of Directors, Robert H. Smith School of Business, University of Maryland, Working Paper.

Droms, W. and D. Walker, 1994, Investment Performance of International Mutual Funds, *Journal of Financial Research*, 17(1), Spring, 1-11.

Edelen, R.M., 1999, Investor Flows and the Assessed Performance of Open-End Mutual Funds, *Journal of Financial Economics*, 53, 439-466.

Efron, B., and R.J. Tibshirani, 1993. *An Introduction to the Bootstrap*, Monographs on Statistics and Applied Probability (Chapman and Hall, New York).

Elton, Edwin J., Martin J Gruber and Christopher Blake, 2001, A First Look at the Accuracy of the CRSP Mutual Fund Database and a Comparison of the CRSP and Morningstar Mutual Fund Databases, *Journal of Finance*, 56, 2415-2430.

Elton, Edwin J., Martin J Gruber and Christopher Blake, 2003, Incentive Fees and Mutual Funds, *Journal of Finance,* 58(2), 779-804.

Elton, Edwin J., Martin J Gruber and Jeffrey A. Busse, 2004, Are Investors Rational? Choices Among Index Funds, *Journal of Finance,* 59, 261-288.

Elton, Edwin J., Martin J Gruber, Yoel Krasny and Sadi Ozelge, 2006, The Effect of the Frequency of Holding Data pm Conclusions About Mutual Fund Management Behavior, Working Paper, NYU, July.

Elton, Edwin J., Martin J. Gruber and Christopher Blake, 1996a, Survivorship Bias and Mutual Fund Performance, *Review of Financial Studies*, 9(4), 1097-1120.

Elton, Edwin J., Martin J. Gruber and Christopher Blake, 1996b, The Persistence of Risk Adjusted Mutual Fund Performance, *Journal of Business*, 69(2), 133-157.

Elton, Edwin J., Martin J. Gruber and Christopher, R. Blake, 1995, Fundamental Economic Variables, Expected Returns and Bond Fund Performance, *Journal of Finance,* 40, 1229-1256.

Elton, Edwin J., Martin J. Gruber, Das, S. and Hlavka, M. 1993, Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios, *Review of Financial Studies*, 6, 1-21.

Evans, Richard, B., 2003, Does Alpha Really Matter? Evidence from Mutual Fund Incubation, Termination and Manager Change, Wharton, Working Paper.

Fama, Eugene F and Kenneth R. French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance*, 47 427-465.

Fama, Eugene F. and J. MacBeth, 1973, Risk, Return and Equilibrium : Empirical Tests, *Journal of Political Economy*, 81, 607-636.

Fama, Eugene F. and Kenneth R. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics,* 33, 3-56.

Fang, Lily and Robert Kosowski, 2006, Comparing Stars – Trading on Star Mutual Funds' Holdings and Star Analysts' Recommendations, Tanaka Business School, Working Paper.

Ferson, Wayne E. and Khang, K 2002, Conditional Performance Measurement Using Portfolio Weights: Evidence for Pension Funds, *Journal of Financial Economics,* 65, 249-282.

Ferson, Wayne E. and Rudi W. Schadt, 1996, Measuring Fund Strategy and Performance in Changing Economic Conditions, *Journal of Finance,* 51, 425-62.

Fletcher, Jonathan and David Forbes, 2002, An Exploration of the Persistence of UK Unit Trusts Performance, *Journal of Empirical Finance,* 9, 475-493.

Fletcher, Jonathan, 1995, An Examination of the Selectivity and Market Timing Performance of UK Unit Trusts, *Journal of Business Finance and Accounting* 22, 143-156.

Fletcher, Jonathan, 1997, An Examination of UK Unit Trust Performance Within the Arbitrage Pricing Framework, *Review of Quantitative Finance and Accounting*, 8, 91-107.

Fortin, R. and S. E. Michelson, 1995, Are Load Mutual Funds Worth the Price, *Journal of*

Investing, 4(3), 89-94.

Friend, I., Blume, M. and Crockett, J. 1970, *Mutual Funds and Other Institutional Investors*, McGraw Hill, New York.

Fung, W. and D. A. Hsieh, 1997, Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds, *Review of Financial Studies*, 10, 275-302.

Fung, W., Hsieh, D., Naik, N. and Ramadorai, T., 2005 Lessons from a Decade of Hedge Fund Performance: Is the Party Over or the Beginning of a New Paradigm? Working Paper, London Business School.

Gallagher, Steven, Ron Kaniel and Laura Starks, 2006, Madison Avenue Meets Wall Street: Mutual Fund Families, Competition and Advertising, University of Texas, Working Paper, January.

Gaspar, Jose-Migual, Massimo Massa and Pedro Matos, 2006, Favoritism in Mutual Fund Families? Evidence on Strategic Cross-Fund Subsidization, *Journal of Finance*, LXI(1), 73-104.

Getmansky, M., Lo, A. and Makarov, I., 2004, An Econometric Model of Serial Correlation and Illiquidity of Hedge Fund Returns, *Journal of Financial Economics,* 74, 529-610.

Goetzmann, W. and R. Ibbotson, 1994, Do Winners Repeat? Patterns in Mutual Fund Performance, *Journal of Portfolio Management*, 20, 9-18.

Goetzmann, W., Ingersoll Jr., J., and Ivkovich, Z. (2000). Monthly Measurement of Daily Timers, *Journal of Financial and Quantitative Analysis*, 35, pp 257-290.

Gorman, Larry, 2003, Conditional Performance, Portfolio Rebalancing and Momentum of Small-Cap Mutual Funds, *Review of Financial Economics,* 12, 287-300

Goyal, Amit and Sunil Wahal, 2004, The Selection and Termination of Investment Managers Plan Sponsors, Goizueta Business School, Emory University, Working Paper.

Graham, John R. and Campbell R. Harvey, 1996, Market Timing Ability and Volatility Implied in Investment Newsletter' Asset Allocation Recommendations, *Journal of Financial Economics,* 42, 397-421.

Grinblatt, Mark and Sheridan Titman, 1989, Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings, *Journal of Business*, 62, 393-416.

Grinblatt, Mark and Sheridan Titman, 1992, The Persistence of Mutual Fund Performance, *Journal of Finance,* 47, 1977-1984.

Grinblatt, Mark and Sheridan Titman, 1994,. A Study of Monthly Mutual Fund Returns and Performance Evaluation Techniques, *Journal of Financial and Quantitative Analysis*, 29 (3), 419-444.

Grinblatt, Mark and Sheridan Titman, 1995, *Performance Evaluation*, in R. Jarrow, V. Maksimovic and W. Ziemba, eds: Handbook in Operations Research and Management Science, Vol 9, Elsevier Science, North Holland.

Grinblatt, Mark, Sheridan Titman, and Russ Wermers, 1997, Momentum Investment

Strategies, Portfolio Performance and Herding: A Study of Mutual Fund Behavior, *American Economic Review* 85, 1088-1105.

Grossman, S.J. and Stiglitz, J.E., 1980, The Impossibility of Informationally Efficient Markets, *American Economic Review*, 66, 246-253.

Gruber, M., 1996, Another Puzzle: The Growth in Actively Managed Mutual Funds, *Journal of Finance*, 51(3), 783-810.

Hall, P., 1986, On the Bootstrap and Confidence Intervals, *Annals of Statistics*, 14, 1431-1452.

Hall, P., 1992. *The Bootstrap and Edgeworth Expansion,* Springer Verlag, Hamburg.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser, 1993, Hot Hands in Mutual Funds: Short Run Persistence of Performance, 1974-88, *Journal of Finance* 48, 93-130.

Henriksson, R. and Robert C. Merton, 1981, On Market Timing and Investment Performance : Statistical Procedures for Evaluating Forecasting Skills, *Journal of Business,* 54, 513-533.

Henriksson, R., 1984, Market timing and mutual fund performance: an empirical investigation, *Journal of Business*, 57, 73-96.

Hochman, S., 1983, The Beta Coefficient: An Instrumental Variable Approach, *Research in Finance* 4, 392-407.

Hon, Mark, and Ian Tonks, 2003, Momentum in the UK Stock Market, *Journal of Multinational Financial Management,* 13, 43-70.

Hong, Harrison, Jeffrey D. Kubik and Jeremy Stein, 2005, Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Trades of Money Managers, *Journal of Finance,* LX(6), 2801-24.

Hortacsu, Ali and Chad Syverson, 2004, Product Differentiation, Search Costs and Competition in the Mutual Fund Industry: A Case Study of S&P500 Index Funds, *Quarterly Journal of Economics,*, 19(2), 403-56.

Houge, Todd and Jay Wellman, 2006 The Use and Abuse of Mutual Fund Expenses, University of Iowa, Working Paper, January.

Huberman, Gur and Wei Jiang, 2006, Offering versus Choice in 401(K) Plans: Equity Exposure and Number of Funds, *Journal of Finance,* LXI(2), 763-801.

Huij, Joop and Jeroen Derwall, 2006, "Hot Hands" in Bond Funds or Persistence in Bond Performance, Erasmus University, Working Paper, February.

Huij, Joop and Marno Verbeek, 2006, On the Use of Multi-Factor Models to Evaluate Mutual Fund Performance, Erasmus University, Working Paper, June.

Investment Company Institute, 2005, Mutual Fund Fact Book, Washington D.C., Investment Company Institute.

Jain, Prem, C. and Joanna S. Wu, 2000, Truth in Mutual Fund Advertising: Evidence on Future Performance and Fund Flows, *Journal of Finance*, 55, 937-958.

Jegadeesh, Narasimhan, and Sheridan Titman, 2001, Profitability of Momentum Strategies : Evaluation of Alternative Explanations, *Journal of Finance,* 56, 699-720.

Jensen, Michael C., 1968, The Performance of Mutual Funds in the Period 1945-1964, *Journal of Finance* 23, 389-416.

Jiang, George J., Tong Yao and Tong Yu, 2005, Do Mutual Funds Time the Market? Evidence from Portfolio Holdings, University of Arizona, Working Paper.

Jiang, Wei, 2003, A Non Parametric Test of Market Timing, *Journal of Empirical Finance* 10, 399-425.

Jones, Christopher S. and Jay Shanken, 2005, "Mutual Fund Performance with Learning Across Funds", *Journal of Financial Economics*, 78, 507-552.

Kacperczyk, Marcin, Clemens Sialm and Lu Zheng, 2005, On the Industry Concentration of Actively Managed Mutual Funds, *Journal of Finance,* LX(4), 1983-2011.

Kacperczyk, Marcin, Clemens Sialm and Lu Zheng, 2006, Unobserved Actions of Mutual Funds, Sauder Business School, University of British Columbia, Working Paper.

Keim, Donald B. and Ananth Madhavan, 1997, Transactions Costs and Investment Style: An Inter-Exchange Analysis of Institutional Equity Trades, *Journal of Financial Economics,* 46, 265-292.

Keswani, Aneel and David Stolin, 2005, Which Money Is Smart? Mutual Fund Buys and Sells of Individual and Institutional Investors, Cass Business School, London, Working Paper.

Khorana, Ajay, Henri Servaes and Peter Tufano, 2006, Explaining the Size of the Mutual Fund Industry Around the World, forthcoming *Journal of Financial Economics.*

Koski, Jennifer and Jeffrey Pontiff, 1999, How Are Derivatives Used? Evidence From the Mutual Fund Industry, *Journal of Finance,* 54(2), 791-816.

Kosowski, Robert, 2006, Do Mutual Funds Perform When It Matters to Investors? US Mutual Fund Performance and Risk in Recessions and Expansions, Working Paper, INSEAD, August.

Kosowski, Robert, Allan Timmermann, Hal White, and Russ Wermers, 2005, Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis, forthcoming *Journal of Finance*.

Kosowski, Robert, N. Y. Naik and M. Teo, 2006, Do Hedge Funds Deliver Alpha? A Bayesian and Bootstrap Analysis, forthcoming, *Journal of Financial Economics.*

Kothari, S.P and Jerold B. Warner, 2001, Evaluating Mutual Fund Performance, *Journal of Finance,* LVI (5), 1985-2010

Lakonishok, Josef A., Andrei Shleifer, and Robert W. Vishny, 1992, The Structure and Performance of the Money Management Industry, *Brookings Papers on Economic Activity,* 339-391.

LaPorta, Rafael, Josef A. Lakonishok, Andrei Shleifer, and Robert W. Vishny, 1997, Good News for Value Stocks: Further Evidence on Market Efficiency, *Journal of*

Finance, 52, 859-874

Leger, L., 1997, UK Investment Trusts : Performance, Timing and Selectivity, *Applied Economics Letters,* 4, 207-210.

Lehmann, Bruce, N. and David M. Modest, 1987, Mutual Fund Performance Evaluation: A Comparison of Benchmarks and a Benchmark of Comparisons, *Journal of Finance*, 42(2), 233-265.

Liang, B., 2000, Hedge Funds: The Living and the Dead, *Journal of Financial and Quantitative Analysis,* 35, 309-327.

Lindbeck, Assar and Mats Persson, 2003, The Gains from Pension Reform, *Journal of Economic Literature,* XLI, March, 74-112.

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

Lynch, Anthony W. and David K. Musto, 2003, How Investors Interpret Past Fund Returns, *Journal of Finance,* LVIII(5), 2033-2058.

Mahoney, Paul G. 2004, Manager-Investor Conflicts in Mutual Funds, *Journal of Economic Perspectives*, 18(2), Spring, 161-182.

Malkiel, G., 1995, Returns from Investing in Equity Mutual Funds 1971 to 1991, *Journal of Finance*, 50, 549-572.

Mamaysky, Harry, Matthew Spiegel, and Hong Zhang, 2004, Improved Forecasting of Mutual Fund Alphas and Betas, Yale School of Management, ICF Working Paper 04-23.

Mandelker, G.N., and S.G. Rhee, 1984, The Impact of the Degrees of Operating and Financial Leverage on Systematic Risk of Common Stock, *Journal of Financial and Quantitative Analysis,* 19, 45-57.

Marcus, A. J., 1990, The Magellan Fund and Market Efficiency, *Journal of Portfolio Management*, 17, 85-88.

Meier, Iwan and Ernst Schaumberg 2004, “Do Funds Window Dress? Evidence for U.S. Domestic Equity Mutual Funds, HEC Montreal, Working Paper.

Morey, Matthew, R. and Aron Gottesman, 2006, Morningstar Mutual Fund Ratings Redux, Pace University NY, Working Paper.

Morey, Matthew, R., 2003, Should You Carry the Load? A Comprehensive Analysis of Load and No-load Mutual Fund Out-of-Sample Performance, *Journal of Banking and Finance*, 27, 1245-1271.

Moskowitz, T., 2003, An Analysis of Covariance Risk and Pricing Anomalies, *Review of Financial Studies,* 16, 417-457.

Musto, David, 1997, Portfolio Disclosures and Year-End Price Shifts, *Journal of Finance*, 52, 1563-1588.

Musto, David, 1999, Investment Decisions Depend on Portfolio Disclosures, *Journal of*

Finance, 54, 935-952.

Myners, P., 2001, Institutional Investment in the United Kingdom: A Review, *Report prepared for the Chancellor of the Exchequer*, H.M Treasury, London.

Nanda, Vikram, M. P. Narayanan and Vincent Warther, 2000, Liquidity, Investment Ability and Mutual Fund Structure, J*ournal of Financial Economics,* 57, 417-443.

Nanda, Vikram, Zhi Wang, and Lu Zheng, 2004, Family Values and the Star Phenomenon, *Review of Financial Studies*, 17(3), 667-698.

Newey, Whitney D., and Kenneth D. West, 1987, A Simple, Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica,* 55, 703-708.

OECD, 2003, Monitoring the Future Social Implication of Today's Pension Policies, *OECD*, Paris, unpublished.

OECD, 2005, *Improving Financial Literacy: Analysis of Issues and Policies*, November, OECD, Paris.

Pastor, Lubos, and Robert F. Stambaugh, 2002, Mutual Fund Performance and Seemingly Unrelated Assets, *Journal of Financial Economics,* 63, 315-350.

Pesaran, Hashem M., and Allan Timmermann, 1994, Forecasting Stock Returns: An Examination of Stock Market Trading in the Presence of Transaction Costs, *Journal of Forecasting,* 13, 335-367

Pesaran, Hashem M., and Allan Timmermann, 1995, Predictability of Stock Returns : Robustness and Economic Significance, *Journal of Finance,* 50, 1201-1228.

Pesaran, Hashem M., and Allan Timmermann, 2000, A Recursive Modelling Approach to Predicting UK Stock Returns, *Economic Journal,* 110, 159-191.

Petersen, Mitchell, A., 2005, Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches, Working Paper, Northwestern University.

Politis, D.N., and J.P. Romano, 1994, The Stationary Bootstrap, *Journal of the American Statistical Association,* 89, 1303-1313.

Pozen, Robert, 1998, The Mutual Fund Business, MIT Press, Cambridge, MA.

Presidential Commission on Social Security Reform, 2001, Strengthening Social Security and Creating Personal Wealth for All Americans, *Report of the President's Commission,* Washington D.C. (available at www.csss.gov).

Quigley, Garrett, and Rex A. Sinquefield, 2000, Performance of UK Equity Unit Trusts, *Journal of Asset Management,* 1, 72-92

Reuter, Jonathan, 2005, Are IPO Allocations for Sale? Evidence from Mutual Funds, *Journal of Finance,* forthcoming.

Sandler R., 2002, Sandler Review: Medium and Long-Term Retail Savings in the UK, *Report prepared for the Chancellor of the Exchequer*, H.M Treasury, London.

Sapp, Travis and Ashish Tiwari, 2004, Does Stock Return Momentum Explain the "Smart

Money" Effect?, *Journal of Finance,* LIX(6), 2605-2622.

Serfling, R., 1980, *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Sharpe, W.F., 1966, Mutual Fund Performance, *Journal of Business*, 39, 119-38.

Shukla, Ravi, 2004, The Value of Active Portfolio Management, *Journal of Economics and Business,* 56, 331-346.

Sirri Erik R. and Peter Tufano, 1993, *Competition and Change in the Mutual Fund Industry*, chapter 7, 181-214, HBS Press, Boston, Mass.

Sirri Erik R., and Peter Tufano, 1998, Costly Search and Mutual Fund Flows, *Journal of Finance,* 53(5), 1589-1622.

Stambaugh, Robert F., 1997, Analyzing Investments Whose Histories Differ In Length, *Journal of Financial Economics,* 45, 285-331.

Sullivan, R., A. Timmermann and H. White, 1999, Data Snooping, Technical Trading Rule Performance and the Bootstrap, *Journal of Finance,* 65(5), 1647.

Sullivan, R., A. Timmermann and H. White, 2001, Dangers of Data Mining: The Case of Calendar Effects in Stock Returns, *Journal of Econometrics*, 105(1), 249-286.

Teo, M., and Sung-Jun Woo, 2001, Persistence in Style-Adjusted Mutual Fund Returns, Manuscript, Harvard University. Available at SSRN.

Thomas, A. and I. Tonks, (2001). Equity Performance of Segregated Pension Funds in the UK, *Journal of Asset Management*, 1 (4), 321-343.

Tkac, Paula, 2004, Mutual Funds: Temporary Problem or Permanent Morass?, *Economic Review,* Federal Reserve Bank of Atlanta, 4th Quarter, 1-21.

Tonks, Ian, 2005, Performance Persistence of Pension Fund Managers, *Journal of Business*, 78, 1917-1942.

Treynor, Jack, and K. Mazuy, 1966, Can Mutual Funds Outguess the Market, *Harvard Business Review,* 44, 66-86.

Turner, Adair, 2004, Pensions: Challenges and Choices : The First Report of the Pensions Commission, The Pensions Commission, *The Stationary Office*, London.

Turner, Adair, 2005, A New Pensions Settlement for the Twenty-First Century: The Second Report of the Pensions Commission, *The Stationary Office*, London.

Turner, Adair, 2006, Implementing an Integrated Package of Pension Reforms: The Final Report of the Pensions Commission, *The Stationary Office*, London.

Viceira, Luis M., 2001, Optimal Portfolio Choice for Long-horizon Investors with Nontradable Labor Income, *Journal of Finance,* 56, 433-470.

Volkman, D. and Wohar, M. (1995). Determinants of Persistence in Relative Performance of Mutual Funds, *Journal of Financial Research*, 18(4), Winter, 415-430.

Warther, Vincent A., 1995, Aggregate Mutual Fund Flows and Security Returns, *Journal of Financial Economics*, 39, 209-235.

Wermers, Russ, 1997, Momentum Investment Strategies of Mutual Funds, Performance Persistence and Survivorship Bias, University of Maryland, Working Paper.

Wermers, Russ, 1999, Mutual Fund Herding and the Impact on Stock Prices, *Journal of Finance,* 54, 581-622.

Wermers, Russ, 2000, Mutual Fund Performance : An Empirical Decomposition into Stock Picking Talent, Style, Transactions Costs, and Expenses, *Journal of Finance,* 55, 1655-1703.

Wermers, Russ, 2001, The Potential Effects of More Frequent Portfolio Disclosure on Mutual Fund Performance, *Perspective,* Investment Company Institute, 7(3),1-11.

Wermers, Russ, 2003a, Is Money Really "Smart"? New Evidence on the Relation Between Mutual Fund Flows and Performance Persistence, University of Maryland, Working Paper.

Wermers, Russ, 2003b, Are Mutual Fund Shareholders Compensated for Active Management "Bets"?, University of Maryland, Working Paper.

Wilcox, Ronald T., 2003, Bargain Hunting or Star Gazing? How Consumers Choose Mutual Funds, *Journal of Business.*

Zheng, L., 1999, Is Money Smart? A study of Mutual Fund Investors' Fund Selection Ability, *Journal of Finance*, 54(3), 901 – 933.

Table 1: Performance Persistence (All US Mutual Funds)

Between January 1st 1976 and January 1st 1995 funds are sorted into quintiles at the end of each quarter, based on the return on their stock portfolio over the previous year. Three separate portfolios are formed based on stocks held (*All Holdings*), bought (*Buys*) and sold (*Sells*) by all funds in the highest and lowest past return quintiles at the end of (or during, for buys and sells) the portfolio formation quarter Qtr-0. Buy-and-hold returns on *All Holdings* portfolios are based on the aggregate shareholdings of each stock at the end of each quarter. The 'buy-and-hold' returns for Buys/Sells are based on the *changes* in shareholdings each quarter. Panel A reports the 'raw' gross portfolio returns, while Panel B gives the characteristic selectivity risk adjusted measure (CS).

The portfolio returns are averaged (across all event dates) during event quarters Qtr-1, Qtr-2 and for various holding periods following Qtr-0, with portfolio weights based on the end of quarter-0 holdings (or quarter-0 shares bought/sold). Returns are in percent per quarter. Time-series t-statistics adjusted for overlapping observations where appropriate, provide tests of statistical significance at the 10%, 5% and 1% levels indicated by *, ** and ***, respectively.

Panel A : Gross Returns (% p.a.)							
Performance Ranking	**Qtr-2**	**Qtr-1**	**Qtr-0**	**Qrt+1**	**Qrt+1 through Qtr+2**	**Qtr+1 through Qtr+3**	**Qtr+1 through Qtr+4**
Top Quintile							
All holdings	6.66	7.26	5.25	4.40	8.66	12.55	16.23
Bottom Quintile							
All holdings	1.68	1.77	3.54	3.13	6.52	10.33	14.10
Top minus bottom quintile							
All holdings	4.98***	5.49***	1.71***	1.27**	2.14	2.22	2.13
Buys (trades>0)	1.86***	2.36***	1.78***	0.36**	1.10	1.38	1.26
Sells (Trades<0)	5.07***	5.59***	0.60	1.22**	1.52	1.52	1.50
Buys-sells	-3.21***	-3.23***	1.18	-0.87**	-0.42	-0.14	-0.24
Panel B : Characteristic Selectivity CS (% p.a.)							
Performance Ranking	**Qtr-2**	**Qtr-1**	**Qtr-0**	**Qrt+1**	**Qrt+1 through Qtr+2**	**Qtr+1 through Qtr+3**	**Qtr+1 through Qtr+4**
Top Quintile							
All holdings	1.70***	1.99***	0.83***	0.37**	0.57	0.43	0.31
Bottom Quintile							
All holdings	-0.85***	-0.93***	0.15	-0.13	-0.13	0.0004	0.03
Top minus bottom quintile							
All holdings	2.55***	2.92***	0.68***	0.51**	0.70	0.43	0.28
Buys (trades>0)	0.63**	1.02**	1.21**	-0.02	0.18	0.13	-0.01
Sells (Trades<0)	2.81***	3.09***	-0.40	0.56**	0.81**	0.58	0.76
Buys-sells	-2.18***	-2.08***	1.62	-0.59	-0.63	-0.45	-0.77

Source: Chen et al (2000), Table 8.

Table 2: Performance Persistence of US Growth Funds

Between December 31st 1975 and December 31st 1993 funds that are classified as 'growth' or 'aggressive growth' are sorted into fractile portfolios at the end of each quarter, based on their average monthly net return over the previous year. Quarterly buy and hold returns are calculated each quarter for each fractile portfolio and portfolios are rebalanced quarterly. Annual returns are computed using average total net asset-weighted returns for each quarter which are then compounded.

Panel A shows the time series average of the number of funds and the total net asset-weighted returns average annual (raw) net returns. Panel B shows the characteristic selectivity measure CS. The table shows the returns in each post-formation year up to year 4. Portfolio returns are averaged (across all event dates) and the time-series t-statistics (adjusted for overlapping observations where appropriate), provide tests of statistical significance at the 10%, 5% and 1% levels indicated by *, ** and ***, respectively.

Source: Wermers(2000b) Table V, Panel A and Panel C.

Panel A : Net Returns (all holdings)						
	No. of funds	**Yr+0 (% p.a.)**	**Yr+1 (% p.a.)**	**Yr+2 (% p.a.)**	**Yr+3 (% p.a.)**	**Yr+4 (% p.a.)**
Top 10%	26	33.0	17.1	17.3	16.2	16.4
Bottom 10%	26	-0.3	12.7	14.4	13.8	13.9
Top-Bottom	26	33.3	4.3**	2.9**	2.5*	2.5***
Panel B : Characteristic Selectivity (all holdings)						
		Year+0 % p.a.	**Year+1 % p.a.**	**Year+2 % p.a**	**Year+3 % p.a.**	**Year+4 % p.a.**
Top 10%	26	10.0	2.1**	2.0***	1.8***	1.5***
Bottom 10%	26	-5.9	0.8	1.4	0.2	0.8
Top-Bottom	26	15.9	1.3	0.6	1.6**	0.7